



Universiteit
Leiden
The Netherlands

Analyzability and semantic associations in referring expressions : a study in comparative lexicology

Urban, M.

Citation

Urban, M. (2012, October 10). *Analyzability and semantic associations in referring expressions : a study in comparative lexicology*. Retrieved from <https://hdl.handle.net/1887/19940>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/19940>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/19940> holds various files of this Leiden University dissertation.

Author: Urban, Matthias

Title: Analyzability and semantic associations in referring expressions : a study in comparative lexicology

Date: 2012-10-10

Chapter 6

Results II:

Semantic Associations and Their

Cross-linguistic Distribution

6.1. INTRODUCTION

Chapter 5 was concerned mostly with quantitative, statistics-based analysis of the behavior of individual languages with regard to the structure of their lexicon. This chapter, in contrast, is concerned with the semantic side of things, and investigates the patterns of associations found in motivated lexical items. § 6.2. systematizes the results for a selection of particularly noteworthy semantic fields and the ties between the meanings within them. It shows, inspired by Hjelmslev's (1963) structuralist analysis of the organization with respect to the meanings 'tree,' 'wood' and 'forest' (compare also Haspelmath 2003: 237), the differences across languages in how they "carve up" the relevant semantic space, as well as some common metaphorical extensions (from 'eye,' 'mouth,' and 'faeces') to other entities not contiguously related to them. § 6.3. then asks whether there is non-random variation in lexico-semantic associations depending on the climatic and geographical environment languages are spoken in. The extensive discussion in § 6.4. is concerned with yet other possible sources of non-random variation, namely the possibility of the spread of particular patterns within languages families by genealogical inheritance, as well as spread due to language contact and resulting areality with regard to semantic associations. Another concern of this section lies in globally recurrent and frequent patterns of lexico-semantic associations. This is the closest the present work comes to the *locus classicus* of linguistic typology: the hunt for universals. Furthermore, the section asks what, if anything, we can learn from these about cognition.

6.2.1. INTRODUCTION

Figure 1 is an adjacency network of lexico-semantic associations based on the entire database. It is based on an adjacency matrix of the lexical associations in the sample data, and plotted by using a visualization technique kindly computed by Michael Cysouw. In the network, the closer the meanings are associated with each other, that is, the more frequent the respective association is found in the database, the closer the meanings are to each other, and the shorter the branches connecting them. To make sure that the network remains readable, only associations found in more than nine languages of the sample are displayed. This is a constraint imposed by problems with readability: if all associations are displayed simultaneously, the diagram becomes unreadable. There is no reason inherent in the data why not the full network should be plotted.

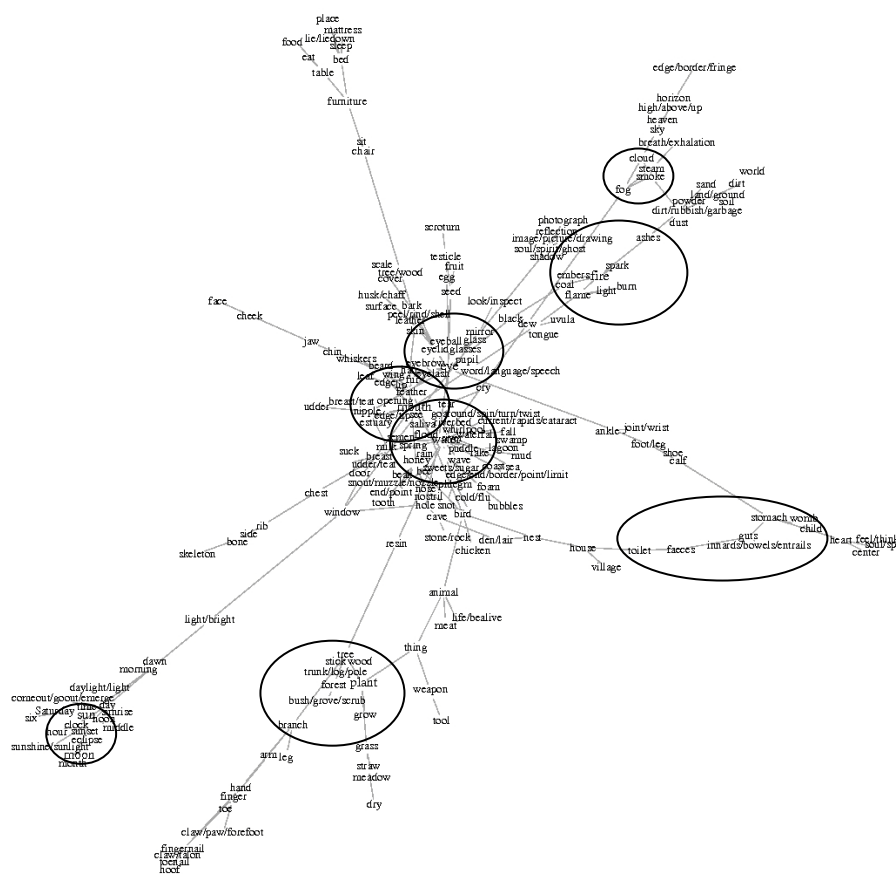


fig. 1.: an adjacency network of lexico-semantic associations

There are a number of meanings which occupy central positions in the network, while others are peripheral. This is not accidental, because the former are associated with a large number of concepts. Specifically, this is true of the heavenly bodies, water- and fire-related meanings, aerosols, and plants and their parts. The following paragraphs successively pick out one of these clusters and discuss in more detail how the meanings occupying central positions relate to the semantic fields surrounding them. At the same time, they point out differences in the lexical distinctions different languages make in a given semantic domain. In addition, further sections are devoted to internal organs of the body and body fluids; many of them do not show up in the diagram because the ties between them are relatively weak, but they showcase interesting interrelations.

Furthermore, there are four meanings that figure prominently in the network not because they are related to a large number of meanings contiguously, but because they are frequent source concepts in metaphor-driven denominations for a large number of meanings from a wide variety of semantic domains. These are ‘eye,’ ‘mouth,’ ‘faeces,’ and kinship terms. Given that the network only shows associations occurring nine or more times in the languages of the sample, these associations are only hinted at there: note, for instance, the proximity of ‘eye’ to ‘spring’ and of ‘mouth’ to ‘estuary.’ More thorough discussion in the relevant paragraphs underscores that metaphorical transfer of the aforementioned meanings is also found to other referents.

Dixon (2010: 256) maintains that “[f]or a study to qualify as lexical typology it should involve comparison of a tightly knit set of terms, the meaning of each being with respect to the meanings of the other terms in the set (just as in a grammatical system),” and the discussion to follow goes in that direction.

6.2.2. SEMANTIC FIELDS AND THEIR ORGANIZATION ACROSS LANGUAGES

6.2.2.1. *The heavenly bodies*

Figure 3 provides a diagrammatic representation of lexical associations between words for the heavenly bodies, that is, the ‘sun,’ the ‘moon,’ and the ‘stars.’

The policy for this and all diagrams to follow is, in order to keep the discussion manageable, that only meanings figuring on the list of meanings on the original list are displayed (of course it would be possible to also include other associated meanings within the respective semantic domains as described in the discussion in Appendix E.). The thickness of the lines represents the strength of the association in the languages of the sample (the thicker the line between the boxes, the more languages exhibit an association between the meanings within the boxes). In addition, arrows indicate the direction of the mapping as revealed by morphologically complex terms: if a line has only an arrow in one direction, as that between ‘sun’ and ‘moon’ in figure 2, it means that the meaning ‘moon’ may be expressed by morphologically complex terms with one constituent being ‘sun,’ but not the other way around. Size of the arrows gives a rough idea of the prevalence of the mapping directions. Thus, a large arrow on one side of the line and a smaller arrow on the other indicates that the mapping is in both directions, but more frequent in one than the other. A thick line with small arrows on one or either side would indicate that the association is mostly by colexification, with some cases of realization by morphological complex-

ity, a line with no arrows at all means that the association is exclusively by colexification, etc.

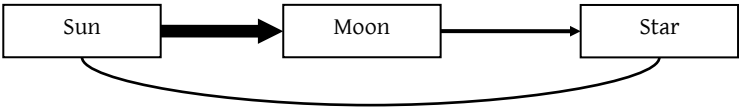


fig 2.: the heavenly bodies and lexical ties between their expressions

There is a common lexical relationship between ‘sun’ and ‘moon’ in languages of the Circum-Pacific language area as defined by Bickel and Nichols (in press), cf. (Urban 2009). Even in these languages, however, words for ‘star’ are virtually always distinct in the sense that they are either completely unrelated lexically and do not share morphological material, or that they are motivated by a complex expression based on the common word for both ‘sun’ and ‘moon.’ There is no language in the sample in which the former situation, colexification of all three referents, is found, and only one potential case of the latter situation: in Hupda, *wædhəm’æh* ‘star’ might consist of *wædhə* ‘sun, moon’ and *mæh* ‘small,’ but this is considered unsure in the consulted source. Aside from languages of the Circum-Pacific area, it is the normal situation to have distinct words for ‘sun’ and ‘moon,’ and mostly also for ‘star,’ although in some cases the latter meaning may be expressed by complex terms deriving from either ‘sun’ or ‘moon,’ as is the case for instance in Guaraní and Wayampi. There is, however, one case of a language that expresses the meanings ‘sun’ and ‘star’ with the same word, namely the Australian language Burarra and also one case of a language with one term for ‘star’ and ‘moon,’ Abipón. Moreover, Bislama *sta* is glossed as “any heavenly body (e.g. moon, star, meteorite).” Table 1 provides examples of the lexical differentiation of the field in different languages (an asterisk after terms indicates that the language also features semantically more specific terms for one or more of the meanings colexified).

	Macaguán	Burarra	Abipón	Wayampi	Kosarek Yale
‘moon’	<i>jomét, -omét*</i>	<i>anjirderda, ran.gu</i>	<i>eergRaik</i>	<i>yai</i>	<i>wal</i>
‘sun’		<i>marrnga</i>	<i>mpaeRa, grahaolai</i>	<i>kwalai</i>	<i>heng</i>
‘star’	<i>jarwát</i>		<i>eergRaik</i>	<i>yai-tata</i>	<i>douang, imbidea</i>

table 1: lexical differentiation for the heavenly bodies cross-linguistically

At any rate, the Burarra, Abipón, and Bislama cases appear to be extreme typological rarities judging from the evidence of the sample (it would be interesting to know if the situation that is encountered in Burarra has parallels in other Australian languages, although it does not appear to be too widespread).

6.2.2.2. *Aerosols*

Figure 3 represents cross-linguistic associations between terms for aerosols, that is, ‘smoke,’ ‘steam,’ ‘cloud,’ and ‘fog’ diagrammatically. As can be inferred from the figure, ties are quite strong and, with the exception of the pair ‘smoke’ – ‘steam,’ asymmetric when it comes to analyzable terms: complex terms for ‘cloud’ and ‘fog’ on the basis of ‘smoke’ are attested, as are complex terms for ‘fog’ based on ‘cloud,’ but not the other way around.

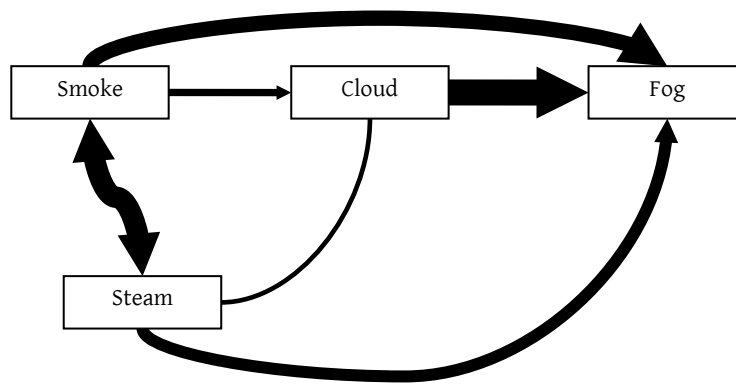


fig 3.: the semantic space of aerosols and lexical ties between its elements

There is just one candidate among the languages in the sample for employing a single term to cover the entire space of the semantic map of aerosols: the Zaparoan language Arabela, where *najaca* is used for all four meanings, though the complex expression *cohuaja najaca* (*cohuaja* means ‘white’) is (also) in use for ‘cloud’ and ‘fog.’ At the very least, it seems safe to say that all four referents contain the *najaca* element. Further, there are no morphologically unrelated synonyms listed for any of the four meanings in the source, which suggests that indeed *najaca* is the only conventional lexical expression associated with the meanings. Candidates for this type are also the Barbacoan language Cayapa and Tsafiki, where the lexemes *ñivijcha* and *poyó* respectively cover the meanings ‘cloud,’ ‘smoke,’ and ‘steam.’ They are candidates only because both sources do not cite the respective word for ‘fog,’ thus leaving open the possibility that this meaning is in fact expressed by a distinct lexical item. Tehuelche may be another case of a language with an at least likely diachronic relationship between terms for all four items in the semantic space: *p'aʔwn ~ p'awn ~ p'eʔwn ~ paʔwn* is synchronically ‘cloud, fog,’ while the phonologically very similar *p'aʔn ~ pa:n* is used for ‘smoke’ and ‘steam’ (there is also the possibility that the -w-consonantism in the forms for ‘cloud, fog’ might be due to fossilized derivation by infixation diachronically).

Otherwise, in languages that cover three of the meanings with one single term, but employ a different one for the fourth, ‘cloud’ appears to be the one that is most commonly lexically distinguished. Examples are found in languages of Australia, more specifically Burarra and Yir Yoront. In Yir Yoront, *thorrqn* covers ‘smoke,’ ‘steam,’ and ‘fog’ (alongside ‘haze’ and ‘spray from waves;’ for ‘smoke,’ there is also a compound with *thum*

‘fire’), while *yirrp* is used for ‘cloud.’ *Yirrp*, however, also means ‘rain,’ an instance of Australian “actual/potential-polysemy,” and it is intriguing to speculate if this common Australian pattern contributed to the organization of this lexical field in Yir Yoront. Similarly, in Burarra, *jolnga* is used to refer to ‘smoke,’ ‘haze,’ ‘vapor,’ and ‘fog,’ while ‘cloud’ is *nguparr*, though note that there are competing unrelated synonyms or near-synonyms for some of the meanings. In contrast, in Anggor, the semantic range of *mburingai* is ‘fog, mist, vapor, cloud,’ while the lexically unrelated *hasahemi* is used to convey the meaning ‘smoke.’ There are no totally clear-cut instances of languages which treat ‘smoke,’ ‘cloud,’ and ‘fog’ lexically similarly, but ‘steam’ differently: in Maxakalí, *gōy* covers the three aforementioned meanings, but the source does not indicate how ‘steam’ is expressed (and in addition, there are compounds on the basis of *gōy* with *hām*, reduced from *hahām* ‘land,’ and *tex*, reduced from *tehex* ‘rain,’ for ‘fog’). Furthermore, the Nez Perce lexical affix *?ipé-* is glossed as ‘pertaining to smoke, cloud, fog.’ As the diagram in figure 4 also underscores, cross-linguistically, the ties between the meanings ‘smoke’ and ‘steam’ are more tightly knit than those with the other two meanings in the lexical field.

It is illuminating to move on to investigate whether there are languages which have two terms each of which cover two of the four meanings in the domain, because, surprisingly, such languages are quite rare. Next to the Tehuelche case already mentioned, Kwoma is an example of such a language; here, *hejagwayap* is used for ‘cloud’ and ‘fog’ and *hirika* for ‘smoke’ and ‘steam.’

In contrast, languages which express two of the meanings in the semantic space by one term and the other two by unrelated terms are of course amply attested, though not in all possible configurations. In line with the stronger ties between the meanings ‘smoke’ and ‘steam’ on the one hand and ‘cloud’ and ‘fog’ on the other, one finds languages that use a single term for ‘smoke’ and ‘steam’ and unrelated ones for ‘cloud’ and ‘fog’ (Buli is an example) as well as the reverse situation expected from the general strength between the lexico-semantic connections, that is, languages that use the same word for ‘cloud’ and ‘fog’ but different ones for ‘steam’ and ‘smoke’ (Baruya is an example). There appears to be an ontology-based motivation for this situation: while both smoke and steam can be observed to emanate and rise up from elemental natural phenomena, fire and water respectively, neither fog nor clouds do, whilst there is an element of perceptual similarity to the latter meanings in that low clouds may appear similar to fog, and indeed the boundary, both meteorologically and perceptually, between the two is fluid to a certain degree. Examples of languages where the lexical organization of the domain in question cross-cuts this rather general division are less easy to find, but do exist: in Gurindji, for instance, *kaparru* means ‘fog’ and ‘smoke-haze,’ while *ngapurung* is used for ‘steam’ and ‘fragrance from cooking’ and *maarn* for ‘cloud,’ while in Kyaka, *popo* covers ‘steam’ and ‘fog,’ while there are unrelated items for ‘smoke’ and ‘cloud,’ and in Sedang, *kia hia* covers ‘cloud,’ ‘smoke,’ and ‘air,’ and there are different terms for ‘steam’ and ‘fog.’ Other combinations are not unambiguously attested. Table 2 summarizes the discussion and provides examples for each of the configurations mentioned.

	Ara- bela	Yir Yoront	Anggor	Buin	Kwoma	Buli	Baruya	Sedang	Kyaka	Gurindji
'smoke'	(coh- uaja) najaca	thorrqn	hasa- hemi	iito* ¹	hirika	nyuik	jita	kia hia*	(isare) suk- wua	kapa- rru
'steam'			mbur- ingai*	numa			mud- inya	xoh, riðh	popo*	nga- purung
'fog'				iito* iito*	hejag- wayap	koal-uk	yir-aaya	idrik, inoa		kappa-rru
'cloud'		yirrp				ching- mari		kia hia*	kopa	maarn

table 2: the semantic space of aerosols and different lexical configurations

Although the lexical associations recur on a global scale, there is nevertheless an areal hotspot in South America around the eastern slopes of the Andes.

6.2.2.3. Internal Organs of the Trunk

Lexico-semantic ties between the internal organs of the trunk, as visualized by the thin arrows between the boxes in figure 4, are on average relatively weak cross-linguistically.

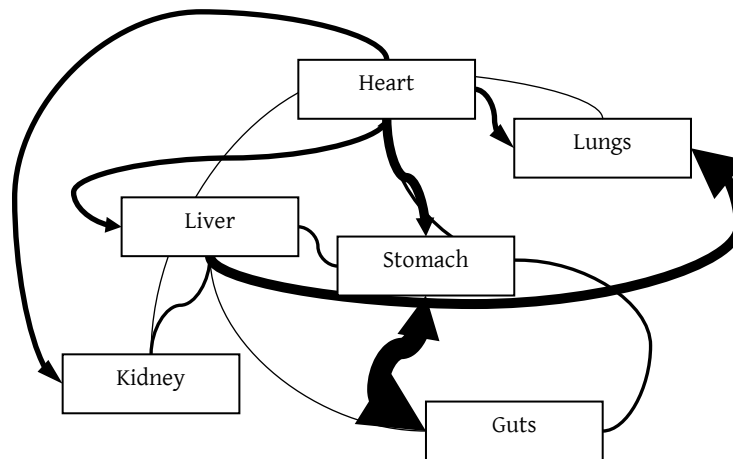


fig. 4.: internal organs of the body and ties between their lexical expressions

Perhaps the most noteworthy fact is that where such terms occur, they may connect a wide variety of the internal organs with each other (though not without restrictions, as will become clear later), and this fact is mirrored in diachrony in that terms for internal parts of the trunk are frequently subject to semantic change in which a term for one internal organ shifts to another. For instance, in Indo-Aryan, Vedic *vṛkká-* 'kidney' under-

¹ 'white cloud,' 'white smoke' more specifically and, according to the English-Buin finderlist, also 'fog.'

went semantic (and phonological) change to *bukkā*- ‘heart’ in a later stage of development. Still later, cognates came to mean ‘belly’ in Sinhalese, ‘lungs’ (among other meanings) in some Romani dialects, and came back full circle semantically to ‘kidney’ for instance in Tōrwāli (Turner 1966); see also Matisoff (1978) for some data from Tibeto-Burman.

The strongest connection is that between ‘stomach’ and ‘guts,’ which is unsurprising given the close spatial and functional proximity. Languages featuring a single term for ‘stomach’ and ‘guts’ are common, but complex terms also occur. Here, complex terms for ‘guts’ on the basis of ‘stomach’ (e.g. Toaripi *ére horou* ‘belly rope’) clearly outnumber complex terms for ‘stomach’ on the basis of ‘guts’ (e.g. Ngambay *kéy bò sìn* ‘house big guts’). Relatively strong ties are also found between the meanings ‘lungs’ and ‘liver’ (compare Blust 2005 for Austronesian specifically). Colexification is attested for instance in Laz, but more often it is the case that the ‘lungs’ are expressed by a morphologically complex term based on the word for ‘liver,’ in which case the most frequent structures highlight the lesser weight of the lungs (e.g. Hawaiian *ake-māmā* ‘liver-light’) or their lighter color (e.g. Bislama *waet-lewa* ‘white/bright-liver’). While the ‘liver’ is thus often the source concept for the ‘lungs,’ so is the ‘heart.’ This is nicely illustrated by the terminological system for internal organs of the body found in the sampled varieties of Quechua, Ancash Quechua and Imbabura Quechua, shown in table 3.

	Ancash Quechua	Imbabura Quechua
‘heart’	<i>shunqu</i> - <i>shonqu</i>	<i>shungu</i>
‘liver’	<i>ñatin</i> , <i>yana ñatin</i> ‘black liver’	<i>yana shungu</i> ‘black heart’
‘kidney’	<i>uru-n</i> ‘egg/testicle-3sg’	
‘lungs’	<i>yuraq ñatin</i> ‘white liver,’ <i>yuraq shunqu</i> ‘white heart’	<i>yurak shungu</i> ‘white heart’

table 3.: terminology for internal organs of the body in varieties of Quechua

Both languages have an apparently cognate unanalyzable word for ‘heart.’ In Imbabura Quechua, this term serves to conceptualize other internal organs of the body: both ‘liver’ and ‘kidney’ are *yana shungu* ‘black heart,’ while the lungs are in contrast *yurak shungu* ‘white heart.’ The same structure for the latter meaning is found in Ancash Quechua, too. In fact, Imbabura Quechua is the language in the sample with the strongest lexical relationships between terms for internal organs of the body. In contrast, Ancash Quechua also uses the ‘liver’ as a conceptualization source: the lungs can also be called *yuraq ñatin* ‘white liver,’ and the meaning of the simplex *ñatin* can be reinforced and contrasted to the word for ‘lungs’ by *yana ñatin* ‘black liver.’

Interestingly, the same situation -one meaning being the conceptualization target of both ‘liver’ and ‘heart’- also pertains to the ‘stomach,’ although the ties are weaker in this case, and the data are not entirely straightforward. In Yanomámi, *amo* means ‘piece of liver’ (as well as ‘central part of a plant’ inter alia). The meaning ‘liver’ itself is rendered by *amoko* or *amoki* (-*ko* is a plural suffix and *ki* a quantal classifier, see § 4.4.1.1.), with *amoko* also being capable to refer to the ‘stomach.’ The situation with regard to the meanings ‘heart’ and ‘stomach’ is more straightforward: many languages colexify the meanings (for

instance Yuki, in which both meanings can be expressed by *tu*'), and in two languages, Kiowa and Malagasy, complex terms are found for 'stomach' (the Kiowa term, for instance, is *t'eqin-t'ou* 'heart-water'; when questioning consultants, the lexicographer received the reply that it is so called "because it is the place that the vomit comes from"). Note that both 'mouth' and 'stomach' are part of the digestive system and hence also contiguously associated (compare the diachronic connection of Greek *stómachos* 'stomach' with *stóma* 'mouth').² However, with regard to the source concepts 'heart' and 'liver,' there is no discernible directional pattern evidenced by analyzable terms. In Mbum, the 'heart' is *làù wârké* 'liver male,' while, as already seen, in Imbabura Quechua the 'liver' is *yana shungu* 'black heart.'

There is one internal organ of the body that stands out in that lexical ties with other internal organs are comparably weak cross-linguistically. These are the 'kidneys.' Colexification is found in three sampled languages with 'heart,' and also in three languages, complex terms for the 'kidneys' on the basis of 'heart' are found, one of them Imbabura Quechua. Colexification with 'liver' is found in Badaga (although the relevant term also conflates the meanings 'larynx' and 'lungs'), and there is one language, Kiowa, in which the same root *tād* is used for 'kidney' and 'liver' exclusively. However, a number of morphologically complex structures exist to disambiguate, among them *tād*-*syh* *n* 'liver/kidney-small' = 'kidney' and *tād*-*eid* *l* 'liver/kidney-be.large' = 'liver.' The fact that cross-linguistically more complex structures exist for 'kidneys' than for 'liver' is suggestive that it is the latter meaning which is dominant and lexically more entrenched. Notably, there are no particular lexical ties of 'kidney,' 'lungs,' or 'stomach' in the sample, apart from cases in which one term has broad reference over a wide range of internal organs to be discussed now.

Sometimes, languages cover more than two points of the semantic space regarding the internal organs with one single term, and there appear to be few if any restrictions as to which organs can and cannot be so treated. For instance, in Ngambay, *wùr* may refer to 'liver,' 'belly' and 'heart' (and has a figurative meaning 'patience'). In Kwoma, the most salient meaning of *wopu* is 'liver,' judging from the microstructure of the dictionary entry, but a secondary reading is "vital organs generally (e.g. liver, heart, lungs)." An even more extreme case is presented by Khoekhoe *!nāb*, which means 'belly, stomach' in a narrow sense but also "innards, offal (i.e. lungs, heart, liver, kidneys)" in a more general sense. This is, next to the situation in Badaga already discussed above, the only case in which a term with such broad reference also includes the 'kidneys,' and it seems to be the only possible generalization that inclusion of this meaning in colexification is rare.

At any rate, it is conceivable that this case of synchronic colexification is a snapshot of an ongoing process of semantic generalization (as noted above, semantics of terms for internal organs of the body seem to be quite unstable). Interestingly, the apparent dominant reading is different in each case: 'liver' in Kwoma and 'belly/stomach' in Khoekhoe, which can be read as evidence that the starting point of the generalization is not

² In earlier language *stómachos* denoted 'throat,' 'gullet' and 'mouth (of the bladder, uterus)' and assumed the meaning 'upper orifice of the stomach' later (Beekes 2010: 1408).

necessarily one particularly salient internal organ, but can be constituted by several different ones. Corroborating this, the synchronic cases of colexification involving a narrow and a more general reading have a diachronic correlate for instance in Indo-Aryan: Classical Sanskrit *phupphusa-* ‘lungs’ (that is, again a different internal organ) is continued in Sindhī as *papuvā* with the same meaning, but in the plural form *papu*, it refers to ‘heart and liver and lungs, breast, bosom’ (Turner 1966). Table 4 gives examples of cross-linguistic patterns of colexification in this semantic field (again, an asterisk indicates presence of more specific alternative terms), with Greek illustrating a language with maximal differentiation.

	Khoekhoe	Ngambay	Badaga	Laz	...	Greek
‘heart’	<i>lnāb*</i>	<i>wùr*</i>	<i>karu*</i>	<i>guri</i>		<i>kardiá</i>
‘stomach’		<i>wùr*</i>				<i>stómachi</i>
‘lungs’		<i>pùpú</i>	<i>cuttage ~ suttage*</i>	<i>cigeri</i>		<i>pnéumōn</i>
‘liver’		<i>wùr*</i>				<i>sykóti, žpar</i>
‘kidney’		<i>mùnjù</i>		<i>n/a</i>		<i>nefró, nefrós</i>

table 4: internal organs of the trunk and cross-linguistic patterns of colexification

An obvious question that arises is whether the patterns in the linguistic treatment of the internal organs of the body have any physiological grounding, that is, whether they can be explained by the perceptual properties of the organs. This is most clearly the case for the lexical connections between ‘lungs’ and ‘liver’: they are situated in close spatial proximity in the human body; they are both big organs, but differ in color (the liver is reddish brown while the lungs are pink) and in weight (the liver is the heaviest internal organ of humans, which nicely explains the conceptualizations mentioned above). In general, lexico-semantic ties are strongest for the four organs positioned roughly in the center of the trunk: the ‘heart,’ the ‘liver,’ the ‘lungs,’ and the ‘stomach’ (although the latter has for obvious reasons also pronounced connections with ‘guts’ cross-linguistically). Thus position within the trunk seems to be one explanatory dimension. Together with a second dimension, that of size, an even more complete picture emerges. Given that there is a hierarchy between the organs with respect to size (liver > lungs > heart > stomach > kidney), one can explain the strong ties between ‘liver’ and ‘lungs’ on the one hand, as well as the relatively strong ties between ‘heart’ and ‘stomach,’ in particular by colexification, on the other. Note that organs on the endpoints of the hierarchy tend to show few connections. This is true of ‘liver’ and ‘stomach’ (in spite of anatomical proximity), but it is particularly conspicuous with respect to the linguistic recognition of the ‘kidneys’: their peripheral position as well as their small size explain the paucity of lexico-semantic ties with other organs (in addition, they have a notably pronounced shape, and thus motivated terms in many languages make reference to that rather than to other internal organs of the body, see Appendix E, 129 for full discussion).

6.2.2.4. *Body fluids*

As can be seen from the diagram in figure 5, there are comparably weak lexical ties between the terms for body fluids cross-linguistically. However, these ties exist interestingly between the majority of the individual body fluids and are not, as one might suspect, restricted to just a few of them while others are completely unconnected in all sampled languages. These are diagnosed as being metaphorical in the present framework (compare Rice's to appear metaphor EFFLUVIA ARE OTHER EFFLUVIA to account for such semantic associations in Dene Sųliné).

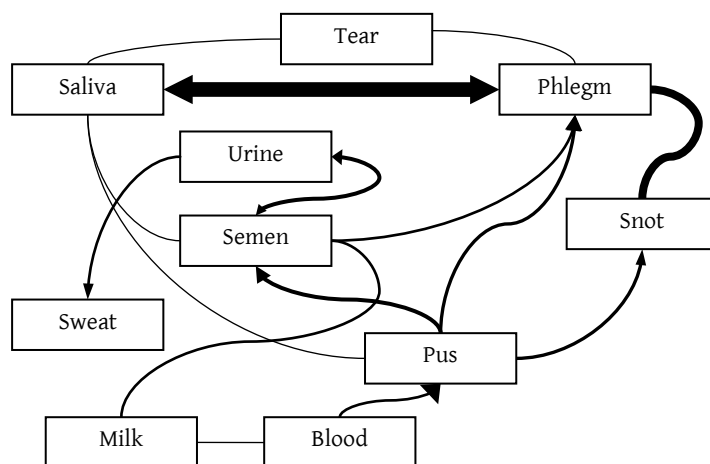


fig. 5.: body fluids and ties between their lexical expressions

Among the strongest ties in the above diagram are those connecting the meanings ‘saliva,’ ‘phlegm,’ and ‘snot,’ that is, those body fluids that have their origin in parts of the respiratory system. Perhaps the most surprising finding is the relatively central role that the concept ‘pus’ plays: complex terms for ‘semen’ and ‘snot’ are in fact found at times on the basis of ‘pus,’ for instance, Abzakh Adyghe has *pe-šan* ‘nose-pus’ for ‘viscous snot’ and Nez Perce *simqéheqs* /*simqé-heqes*/ ‘penis-pus’ for ‘semen,’ but the other logically possible direction is unattested in the sample. However, terms for ‘pus’ that are secondary to those for ‘blood,’ like Tetun *raan-kroek* ‘blood-rotten’ and *raan-mutin* ‘blood-white’ are attested.

Otherwise, the distribution of the associations allow for little systematization, which is not the least due to the fact that most of them are only attested in one language (‘milk’ – ‘blood’ in Kwoma, ‘tears’ – ‘mucus’ – ‘spittle’ in Khalkha, which also has distinct words for the two latter meanings, ‘spit’ – ‘semen’ in Rotokas, ‘saliva, spittle’ – ‘pus’ in Sedang, as well as perhaps the expression of the meaning ‘sweat’ on the basis of ‘urine’ in Guarani). The association between ‘milk’ and ‘semen’ occurs in two languages, but here perceptual similarity in color is available as a hypothesis for the motivation of the association, and that with ‘urine’ occurs in two languages: by colexification in Tuscarora and by an archaic derived term for ‘semen’ from a verb meaning ‘to urinate’ in Khoekhoe.

The overall lesson to learn from the sample data is that lexical connections between body fluids are relatively rare (that is, they are referents most of the time expressed by unrelated lexical items), but they do occur, and there appears to be no general constraint as to what names for body fluids are particularly prone to be lexically associated. Such an apparently relatively unconstrained situation may be the outcome of taboos or, in a less strong form, by euphemistic meaning extension of terms for more “harmless” body fluids to more delicate ones. For diachrony, then, the upshot is that it is at least not impossible that terms for body fluids shift in meaning to other body fluids without any apparent semantically-based explanation for the shift other than that both referents belong to the same semantic domain.

6.2.2.5. Configurations of water

As is obvious from the diagrammatic representation in figure 6, ‘water’ plays a central role in this semantic field.

However, there are many more observations to be made about the organization of the field in individual languages. There are languages in the sample with no clear areal distribution in which ‘water’ and at least one major type of body of water, that is ‘river’ or ‘lake,’ and are not lexically distinguished at all (such a system is described in detail in Burenhult 2008b for Jahai and was alluded to earlier). The most extreme and unique case in the sample is Jarawara, which uses a single lexical item, *faha*, not only for ‘water,’ ‘river,’ and ‘lake,’ but also for ‘rain’ (there is also the word *isi/iso* for ‘rain,’ which also means ‘leg,’ ‘handle’ and ‘stalk;’ the only lexical alternative available for ‘lake’ is *rako*, which is a loanword from Portuguese).

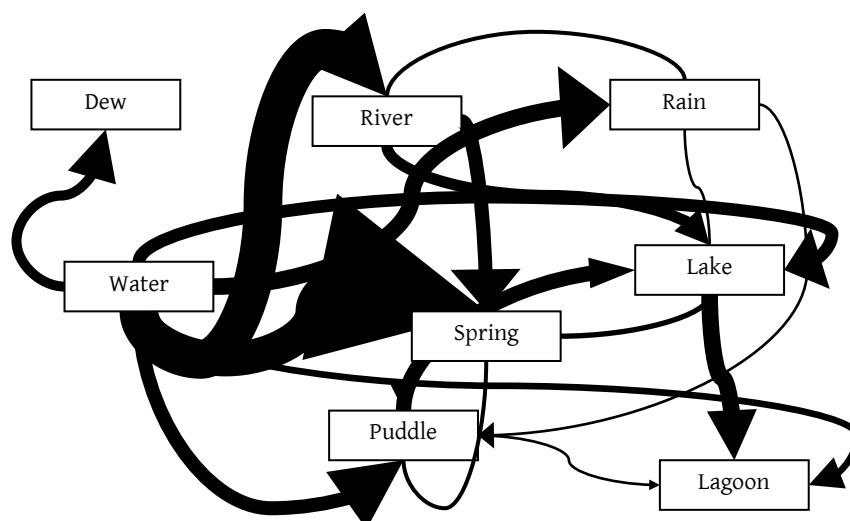


fig. 6.: terms for configurations of water and lexical ties between them

While typically languages with a system similar to Jahai employ different and unrelated terms for ‘rain,’ and, conversely, languages which use the same word for ‘water’ and ‘rain’ usually have unrelated or at least morphologically complex expressions for the different bodies of water, some languages with an overlap exist. The meaning of Waris *po*, for instance, ranges over the referents ‘water,’ ‘river,’ and ‘rain,’ while ‘lake’ is treated differently linguistically (the term might be a compound of *po* and the Waris word for ‘son,’ a similar system is found in Bakueri, where ‘lake’ is literally ‘sea child’). The same situation is found in Kosarek Yale, also spoken in New Guinea. In contrast, in Berik, *fo* ranges semantically over ‘water,’ ‘river,’ and ‘lake,’ while ‘rain’ is expressed by the lexically unrelated *aro*, and in Itzaj *ja'* means ‘water,’ ‘rain,’ and ‘lake,’ but not ‘river,’ for which a number of other lexical labels exist (among them *ok ja'*, literally ‘foot/leg water,’ there are also formally redundant complex terms on the basis of *ja'* for ‘lake’). Systems like these, in which two different configurations of water are denoted by a single lexical item which is also the designation for the substance ‘water’ are relatively rare cross-linguistically. The semantic field is somewhat more differentiated in Hupda: here one encounters the same word denoting both ‘water’ and ‘rain,’ but different lexical expression for ‘river’ and ‘lake’ (one of the Hupda terms for ‘river’ is *deh-mí* ‘water-waterway’). This system is fairly common cross-linguistically, but is particularly frequent in the Americas. The mirror-image of Hupda is Quileute, in which an even more common system is found. In languages of this type, the same monomorphemic lexical item is used to denote the substance ‘water’ and ‘river,’ while ‘rain’ and ‘lake’ are lexically differentiated. A conceivable situation is also one in which ‘water’ and ‘lake’ are colexified, but ‘river’ and ‘rain’ are treated differently linguistically. Comanche is the closest in the sample to that: *umahpaa?* means ‘rainwater’ and ‘pond, lake,’ while the substance ‘water’ is *paa*, ‘rain’ is *umap̣* and ‘(small) river’ is *okwèeṭ*. Not in all languages which lack lexical differentiation for different bodies of water is it necessarily the case that the term covering them is always at the same time expressing the meaning ‘water.’ For instance, Khalkha has a single term, *møren*, which may refer to both ‘river’ and ‘lake,’ and Bakueri has a single term, *m̥or̥ô*, for both ‘river’ and ‘spring’ while the substance ‘water’ is designated by the unrelated *m̥aliwá*. The different systems are summarized in table 5, with Kildin Saami illustrating full lexical differentiation.

	Jarawara	Waris	Berik	Itzaj	Hupda	Quileute	Khalkha	Kildin Saami
‘water’	<i>faha*</i>	<i>po</i>	<i>fo</i>	<i>ja'</i>	<i>děh</i>	<i>kʷáya</i>	<i>usun</i>	<i>čāz'</i>
‘rain’			<i>aro</i>	<i>ja'</i>		<i>libókʷ</i>	<i>boruɣa(n),</i> <i>xura</i>	<i>ābb'r</i>
‘river’			<i>fo</i>	<i>b'ekan,</i> <i>riiyoj,</i> <i>ok ja'</i>	<i>dehmí,</i> <i>má</i>	<i>kʷáya*</i>	<i>møren*</i>	<i>jōgk</i>
‘lake’		<i>polomb</i>		<i>ja'</i> <i>(noj-ja')</i>	<i>móh</i>	<i>tlókʷot</i>		<i>jāvv'r</i>

table 5.: configurations of water and cross-linguistic patterns of colexification

The concept ‘dew’ is only connected to the semantic field by analyzable terms, and there is only one language in the sample, Nez Perce, which colexifies ‘dew’ and ‘water’ directly.

6.2.2.6. *Fire and associated meanings*

Similar to the semantic field of configurations of water discussed above, where terms for ‘water’ occupy a central position, the domain of fire-related concepts is obviously organized around ‘fire,’ as seen in figure 7.

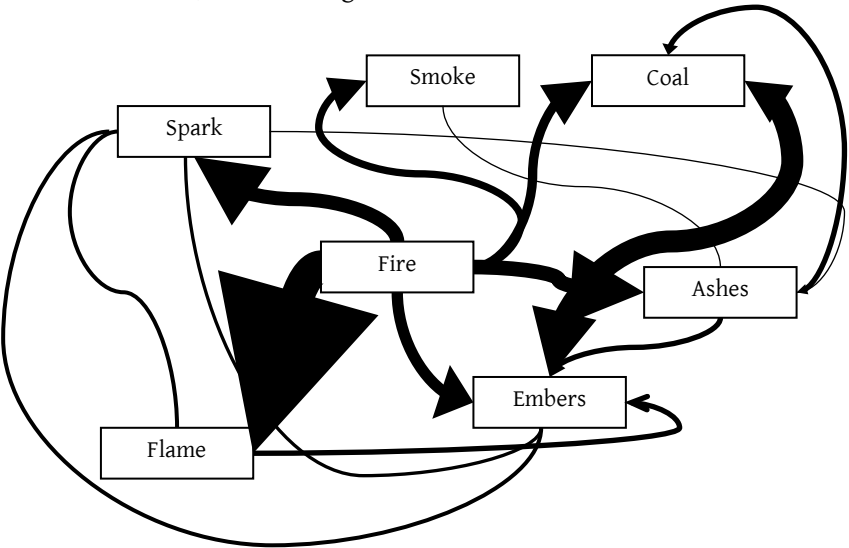


fig. 7.: terms for fire-related meanings and lexical ties between them

A language in which fire-related meanings are consistently contiguity-anchored is Toaripi. The majority of the relevant terms, as seen in table 6, however, are not fully analyzable on the basis of the consulted source.

Meaning	Toaripi equivalent
‘fire’	<i>a</i>
‘flame’	<i>a-uri</i> ‘fire-tongue’
‘spark’	<i>a-e</i> ‘fire-faeces’
‘ashes’	<i>a-futae</i> ‘fire-??’
‘embers’	<i>a-koela</i> ‘fire-??’
‘smoke’	<i>a-ikaera, a-ikoela, a-ivuka, a ikohela</i> ‘fire-??’
‘coal’	<i>a-ro</i> ‘fire-??’

table 6.: Toaripi terms for fire-related meanings.

Within this semantic field, however, colexification is relatively rare, which in all likelihood has something to do with the disparate perceptual properties of the referents. The most common pattern of colexification is that of ‘flame’ with ‘fire’ itself, followed by that with ‘embers’ and ‘spark,’ which are both relatively weak, however. Stronger associations

by colexification are found again between the meanings ‘embers’ and ‘ashes,’ as well as between ‘embers’ and ‘coal’ and ‘ashes’ and ‘coal.’ This is most likely due to the obvious fact that ‘coal’ and ‘ashes’ have a common semantic denominator: they are remnants of a burning fire. Rather than asking which meanings may be colexified, perhaps a more interesting question in this particular case is to ask which patterns of colexification are actually not attested. As already implied in the above discussion, the meanings in the above diagram essentially form two clusters for which there is an ontological basis: colexification is attested for meanings having to do with an actually burning fire (‘flame,’ ‘spark’) and those that are remnants of a once burning fire (‘ashes,’ ‘coal’), with ‘embers’ occupying an intermediate position and forming a link between the two clusters that mirrors its intermediate position in the process of a fire burning down.

6.2.2.7. *Plants and their parts*

Strongest ties in the semantic field of plants and their parts, as can be inferred from figure 8, are found for meronyms of ‘tree,’ and this is true for both colexification and morphologically complex terms.

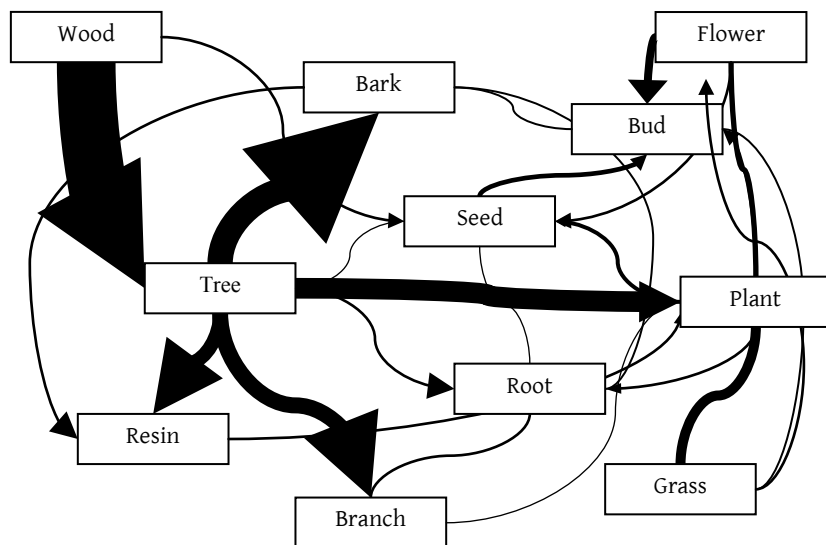


fig. 8.: terms for plants and their parts and lexical ties between them

In particular, colexification of ‘tree’ and ‘wood’ is frequent. In Witkowski et al.’s (1981) sample, colexification occurs in two thirds of sampled languages, and the figure obtained in the present study draws close to this value. As also noted by Witkowski et al. (1981: 5), there is a tendency for morphologically complex terms for ‘tree’ to be based on ‘wood,’ although in the data underlying this study, this pattern is not as strong as discussion in Witkowski et al. (1981) suggests. In fact, this situation is only found unambiguously in one sampled language, Manange (which has *2fiŋ-3tuŋ* ‘wood-copse/trunk’), though interest-

ingly Lesser Antillean Creole French has a similar pattern (*pié-bwa* ‘wood-stem’). In addition, the Cubeo term for ‘tree’ is *jocɥ-cɥ*, which is the term for ‘wood,’ *jocɥ*, suffixed with the classifier *-cɥ* for tree-like objects and furthermore, Upper Chehalis *ʔiʃ-aʔʃ* ‘a clump of trees’ is a reduplication of the term for ‘wood.’ Whether the reverse relationship also occurs and if so, how frequent it is, cannot presently be assessed since ‘wood’ itself is not among the items under investigation. Colexification of ‘tree,’ with ‘trunk,’ ‘pole,’ or ‘log’ is found in eighteen sampled languages.

As a comparison with those languages in which the colexification is with ‘tree’ makes clear, there is some overlap between the groups. Ngambay and Khalkha have single terms for both ‘tree,’ ‘wood,’ and ‘forest’ (and in the case of Ngambay, even ‘branch’) due to the extremely frequent colexification of ‘tree’ and ‘wood’ themselves (see Appendix E, 65), but there is also a relatively large number of languages which do not have this pattern of colexification, instead either having distinct terms for ‘wood’ and ‘tree’ and colexifying ‘forest’ with one of them. Table 7 provides examples of different patterns.

	Ngambay	Khalkha	Waris	Yaqui	Imbabura Quechua
‘wood’	<i>kake*</i>	<i>modu(n)*</i>	<i>ti</i>	<i>kuta</i>	<i>kaspi</i>
‘tree’				<i>juya</i>	<i>yura</i>
‘forest’			<i>sungeit, ekla</i>		<i>sacha</i>
‘branch’		<i>gesigyn</i>	~ <i>klal, tikla</i>	<i>buja</i>	<i>malki</i>
		<i>gesigyy*</i>			

table 7. terms for parts of plants and cross-linguistic patterns of colexification

In some sampled languages, ‘forest’ is expressed by morphologically complex terms on the basis of ‘tree,’ as in Baruya *yɪ’darya*, literally ‘tree area’ and Ancash Quechua *sacha marka* ‘tree/plant area’ (see Appendix E, 26 for more details). An interesting variation of complex terms of the Baruya and Ancash Quechua type is found in three languages of South America, Jarawara, Lengua, and Yanomámi. Here, the general meaning ‘place’ is colexified with ‘forest’ (on semantically general terms of this kind in South America and Jarawara specifically, see § 6.4.3.15.). This is also interesting in the light of the impact of environmental factors on the colexification of particular meanings, since at least Jarawara and Yanomámi are spoken in the tropical rainforest of the Amazon basin.

Moreover, meronyms of ‘tree’ are commonly expressed by morphologically complex terms on the basis of a term with just that meaning. A language in which ‘bark,’ ‘resin,’ and ‘branch’ are all expressed by morphologically complex terms is San Mateo del Mar Huave; in fact it is the only sampled language in which all three meanings are expressed by analyzable terms of which ‘tree’ is one constituent (see table 8), but the semantic relations between these are not very frequent cross-linguistically (see discussion of individual concepts in the relevant sections Appendix E).

Meaning	Sam Mateo del Mar Huave Equivalent
'bark'	<i>mipang xiül</i> 'shell tree'
'resin'	<i>aonts xiül</i> 'excrete tree'
'branch'	<i>omal xiül</i> 'point tree'

table 8: San Mateo del Mar Huave terms for meronyms of 'tree'

Of these, 'resin' may also be based on 'bark,' as in Piro *mta-ha* 'bark-water.'

Apart from meronyms of 'tree,' the lexico-semantic ties are cross-linguistically relatively weak, but one less tight cluster is discernible which consists of meanings having to do with the reproductive system of plants: 'bud,' 'flower,' 'seed' (and 'fruit'). Obviously, the 'flower,' 'bud,' and 'fruit' stand in a relationship of temporal contiguity with each other, and, in addition, 'seed' stands in a meronymic relationship with 'fruit.' Within this field, ties between 'bud' and 'flower' and 'seed' and 'fruit' respectively are particularly strong (relatively speaking). For instance, in Wayampi, 'bud' is *pəti-yaʔi* 'flower-child' (which is also an interesting denomination because of the metaphorical transfer of 'child,' having to do with reproduction in humans or animates more generally, to the fauna). Very strong are the ties (found in seventeen languages) between the meanings 'seed' and 'fruit,'³ but in spite of the obvious contiguous relationship between 'flower,' 'bud' and 'fruit,' colexification of 'bud' with 'fruit' or 'flower' is comparatively rare. In Kaluli, the meaning colexified with 'flower' is more precisely 'inedible tree fruit,' and it is this fact which points to a possible explanation of the observed frequencies. 'Flowers' and 'buds' are of no or quite limited use for humans, while 'fruits' are in that some of them are edible, and thus their quality to potentially serve as foodstuff may be an important component of the lexical semantics of terms for 'fruit' cross-linguistically. Thus, it is not surprising that colexification of the kind mentioned above is relatively rare, in that 'fruits,' in terms of Gibson (1979), are likely to be conceptualized under the perspective of human affordance, whereas 'flower' and 'bud' are not. Table 9 provides an overview over the elaborateness of lexical differentiation for the meanings just discussed.

	Efik	Sahu	Lesser Antillean Creole French	Baruya	Kiliwa
'fruit'	<i>m'fri</i> <i>mfuri</i> *	~ <i>palingasa</i> *	<i>n/a</i>	<i>n/a</i>	<i>tkwma?</i> , <i>-pay</i>
'flower'			<i>flé</i>	<i>purirya</i>	<i>tpyawp</i>
'seed'		<i>moi'i</i>	<i>jem</i> *	<i>wia</i>	<i>tyit</i>
'bud'		<i>boro</i>		<i>purirya</i>	<i>chiilp</i>

table 9.: terms for some parts of plants and cross-linguistic patterns of colexification

³ The situation in Toaripi is also discussed in Brown's (1972: 171) comparative semantic analysis of Toaripi and the related Oroko. He states that "while the term *fare/hae* covers both the meanings 'seed' and 'fruit', with fleshy types of fruit it has reference rather to the seed or nut, and not to the fruit as a whole."

Among the other minor patterns not made explicit in figure 9 since not all of the respective meanings figure on the original meaning list are colexification of ‘leaf’ and ‘branch’ in Gurindji and Nuuchahnulth, which can be explained by the spatial contiguity between the two meanings, colexification of ‘flower’ and ‘pod’ in Rao, of ‘leaf’ and ‘flower’ in Cheyenne, and that of ‘bud’ and ‘young leaf’ in Efik, Sko, Jarawara and Lesser Antillean Freole French.

6.2.3. BODY-PART METAPHORS

6.2.3.1. *The Eye*

The ‘eye,’ as the most salient feature of the human face (Shepherd et al. 1981), is an extremely common conceptualization source for a wide range of meanings, including both other, presumably less salient body parts, and many meanings in other semantic domains. There is also literature on this for Austronesian languages specifically (Barnes 1977, Chowning 1996), but this type of transfer is common across the globe. The diagram in figure 9 provides an overview of lexico-semantic associations with ‘eye’ (dashed lines indicate that the association is only present by semianalyzable terms in this and further diagrams in the following two sections).

In the upper left corner, there are contiguity-based conceptualizations on the basis of ‘eye’ for body-parts that are immediately adjacent to the eye, or are more properly put parts of it, as well as for ‘tear,’ a body-fluid that is in contiguity with the ‘eye’ since this is where it originates. From a conceptual point of view, these are fairly uninteresting. What is noteworthy, though, is the large number of languages, as indicated by the thick black arrows, in which ‘eyebrow,’ ‘eyelash,’ ‘eyelid,’ ‘pupil,’ and ‘tear’ are transparent complex expressions based on the respective terms for ‘eye.’

A more interesting question, however, is whether there is an all-or nothing situation, that is, whether languages either favor having complex lexemes for all of the body-parts in contiguous association with the eye or to have unanalyzable terms for the entire set of meanings. The answer is that there is little evidence for such a principled linguistic treatment. There are languages in which terms for the entire set of contiguously related meanings are analyzable (as summarized in table 10 for Kashaya, which is such a language), but a more frequent situation is that languages fall somewhere in between, with some analyzable terms and some unanalyzable ones.

Lexical Item	Underlying Representation and Gloss	Meaning
<i>huʔuʔ sime</i>	/huʔuy sime/ 'eye fur'	'eyebrow'
<i>huʔuʔ pitemʔ⁴</i>	/huʔuy pitemʔ/ 'eye droop.of.eyes'	'eyelash'
<i>huʔuy šiʔda</i>	/huʔuy šiʔda/ 'eye skin'	'eyelid'
<i>huʔuʔ qʰaʔbe</i>	/huʔuy qʰaʔbe/ 'eye rock'	'eyeball'
<i>ʔuʔqʰa⁵</i>	/huʔuy ahqʰa/ 'eye water'	'tear'
<i>huʔuʔ šihta</i>	/huʔuy šihta/ 'eye bird'	'pupil'

table 10.: analyzable items for meanings related to 'eye' in Kashaya

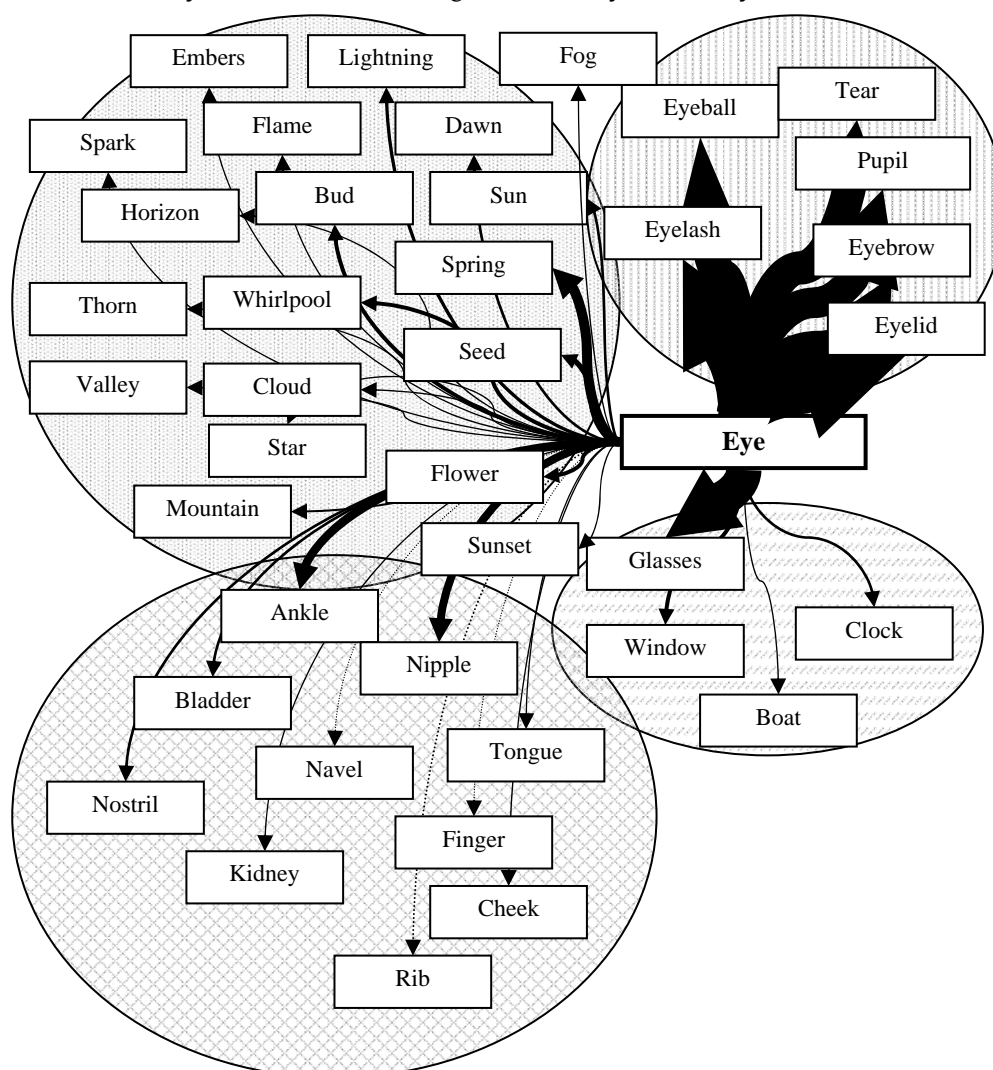


fig. 9.: lexico-semantic associations for the 'eye'

⁴ *huʔuʔ sime* may be used as well.⁵ Glossed as 'eye water' in the consulted source, but possibly lexicalized.

In the lower right corner of the diagram in figure 9, there are a variety of artifact terms which cross-linguistically are sometimes conceptualized via ‘eye.’ In the case of ‘eyeglasses,’ this is obviously due to contiguity, while in the case of ‘clock,’ it is due to the similarity in shape between the two referents, perhaps aided by the additional similarity in shape with the ‘sun.’ While the presence of the word for ‘eye’ in the Yoruba word for ‘boat,’ *òkò-oju-emi* ‘vehicle-eye-water’ remains unclear, similarity in shape is also available as an explanation for the presence of the words for ‘eye’ in terms for ‘needle’ and ‘window.’ Similarity in roundish shape is also a likely factor explaining the conceptualization of ‘windows’ via ‘eye’ in many languages (and note the etymology of English *window*, which, according to the Oxford English Dictionary, goes back to Old Norse *vindauga*, a compound of *vindr* ‘wind’ and *auga* ‘eye,’ as well as the precise parallel noted in Yoruba in table 11).

In fact, discussion of the case of ‘window’ provides a good transition to the many meanings in the domains of nature-related and body-part terms that may be expressed cross-linguistically by analyzable terms on the basis of ‘eye.’ The similarity in roundish shape explains most of the associations that occur with body-part terms. The ‘ankle,’ the ‘nipples,’ the ‘navel,’ the ‘bladder,’ the ‘kidney,’ and ‘the nostrils’ quite obviously are round in shape, and many of them are also roughly comparable in size to the ‘eye.’ More remotely roundish are the ‘tongue,’ the ‘cheeks,’ the ‘ribs,’ and the ‘finger.’ In these cases, it can be conjectured that there is another semantic feature in which these entities may be perceived as being similar to the ‘eye,’ namely three-dimensionality.

Similarity in round shape explains also the most frequent metaphor-driven transfers to terms for topological features, most prominently ‘spring’ and ‘whirlpool’ as well as those to ‘seed,’ ‘bud,’ and, from there on, ‘flower.’ (the associations with ‘fog’ in Kiliwa *yuw=hi?* ‘eye-cover’ and with ‘horizon’ in Khoekhoe *mūs lkhâu-s* ‘eye radiate-3SG.FEM’ appear to be contiguity-based). For two of the associations with meanings related to ‘fire,’ namely ‘embers’ and ‘spark,’ roundish shape may also be adduced as underlying the metaphorical transfer.⁶ However, an additional component of brightness may well play a role, which would then also account for the associations with ‘lightning’ and ‘dawn’ and, most importantly, ‘sun,’ an association which is particularly common in languages of Southeast Asia and Oceania (Urban 2010). The association with ‘dawn,’ however, may also be explained with reference to shape-based similarity alone, given that the very first light of the day at dawn emerging from the horizon in fact is remotely roundish in shape.

The obviousness with which shape-based similarity is detectable is gradient. Thus, the associations with nature-related features such as ‘mountain,’ ‘valley,’ ‘thorn,’ and ‘cloud’ intuitively appears to be conceptually more remote than that with, say, ‘seed.’ However, all may be conceived on some level of abstraction as being roundish entities. Two additional remarks are in order: first, as mentioned above, the referents in question are not only round, but also three-dimensional objects, and this appears to be in some cases a secondary motivating factor. Second, the bolder conceptual transfers cluster in a

⁶ The association with ‘flame,’ occurring in Kyaka, is a little less clear. However, *lenge*, which is the relevant term, has many meanings in Kyaka, among them ‘node or knuckle,’ ‘stratum or narrow (vertical) panel in man’s skirt-net,’ ‘woven body of a bag,’ ‘eye of boil or carbuncle’ and last but not least ‘eye of the head.’

relatively small number of languages, and it may well be the case that the presence of a number of the more obvious metaphorical patterns is a prerequisite for the development of conceptually less nearby transfers, with concomitant semantic bleaching and generalization of the term for ‘eye.’ Among the languages which have a variety of the latter is Yoruba, as seen in table 11.

Lexical Item	Gloss	Meaning
<i>ojú òrun</i>	‘eye heaven’	‘cloud’
<i>ojú-sanmà</i>	‘eye-sky’	‘cloud’
<i>okò-oju-omi</i>	‘vehicle-eye-water’	‘boat’
<i>ojúaféfé</i>	‘eye-wind’	‘window’

table 11: metaphor-driven complex lexical items in Yoruba involving *oju* ‘eye’

Another language in which a variety of terms for natural kinds and artifacts are expressed using ‘eye’ as a source concept is Rama (Chibchan):

Lexical Item	Gloss	Meaning
<i>king-úp, kung-úp</i> ⁷	‘head/top-eye’	‘mountain’
<i>kat up</i>	‘tree eye’	‘fruit, peanut’
<i>kú up</i>	‘bird’s wing eye’	‘thorn, prickle’
<i>ngústi úp</i>	‘pissing eye’	‘bladder’
<i>isúl-uk up</i>	‘??-skin eye’	‘finger’
<i>píns-up ~ pínsh-uk</i>	‘?? eye’	‘navel’

table 12: metaphor-driven complex lexical items in Rama involving *up* ‘eye’

Valuably, Rigby and Schneider (1989) have at times included comments by their Rama consultant for lexical items. For *píns-up ~ pínsh-uk* ‘navel,’ the consultant provided the literal meaning ‘belly eye,’ while in the entry for ‘finger,’ a comment on *up* as occurring in this term is ‘something round.’ This may have something to do with the term for ‘finger’ being only semianalyzable, but it is interesting for a general discussion of the semantic extension of ‘eye’ to round-shaped objects that here the semantic content ‘eye’ seems to be bleached to ‘something round’ when occurring in some complex constructions. This is in line both with the suggestion that *up* acts as a device for nominal classification (see § 4.4.1. for discussion) as well as with the observations to be made in § 6.2.3.3. in discussing extensions of ‘faeces’ or ‘excretion’: in some languages, there appear to be semantic templates on the basis of certain meanings to conceptualize a wide variety of referents on the basis of a bodily notion (see Levinson 1994 for discussion of body-part extensions in Tzeltal). These observations have obviously parallels to both grammaticalization and lexicalization: they are similar to grammaticalization in that a certain amount of semantic bleaching and generalization is involved, and to lexicalization in that the complex structures of this type may well not be transparent anymore to language users precisely be-

⁷ There are more semianalyzable terms for ‘mountain’ where one of the constituents is *up* ‘eye.’

cause of the semantic bleaching that seems to be a concomitant effect of an increasing use of the same source concept to name a large number of entities.

Summarizing, one can note three subtypes of metaphorical transfer of terms for 'eye' to other concepts, which, at least in the domain of body-parts, may receive additional support if the target concept is similar to the 'eye' in size.

- (i) Shape-based similarity alone: roundness
- (ii) Shape-based similarity with the putative additional motivating component of brightness
- (iii) Shape based similarity with the putative additional component of three-dimensionality.

6.2.3.2. The Mouth

Figure 10 represents semantic associations for 'mouth' diagrammatically.

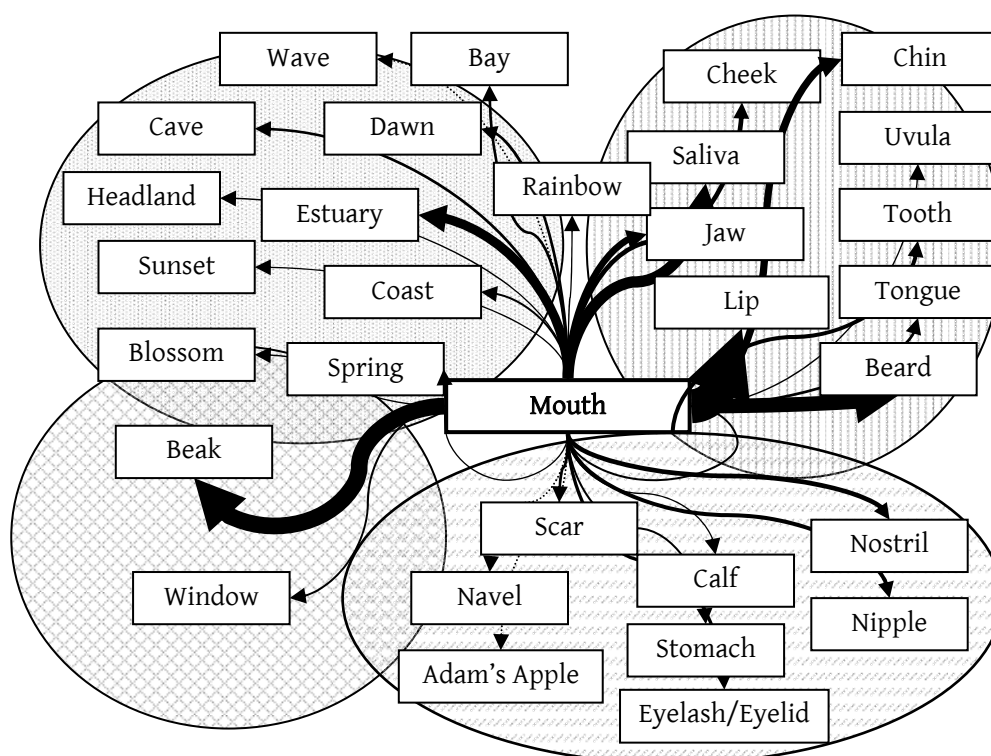


fig. 10.: semantic associations for 'mouth'

As in the above diagrams, concepts – facial features in this case – which are contiguously related to the 'mouth' are found in the top right corner. The situation here is comparable with that for 'eye', namely that the lexico-semantic ties in this area are strongest: terms for 'lips' frequently consist of the respective terms for 'mouth' and 'skin'

crosslinguistically, as in Cayapa *fi'pa'-quica* 'mouth-skin.' Terms for 'beard' which may be "literally" translated as 'mouth-hair', such as Wintu *qol-čekey*, are common as well. 'Saliva' is frequently expressed by morphologically complex terms consisting of terms for 'mouth' and 'water' or 'liquid' more generally, such as Wappo *na-méy* (Appendix E, 133), and 'jaw'-terms are sometimes analyzable as complex structures involving 'mouth' and 'bone' (Appendix E, 118), as in Baruya *maanaginya*, which contains *maanga* 'mouth' and *yaginya* 'bone' and also means 'chin.' In one language, Aymara, 'tooth' is *laka ch'akha* 'mouth bone' (see Appendix E, 144 for more thorough discussion).

Sometimes the associations with topological features may exist because the word for 'mouth' also has a semantically bleached more general reading in the languages mentioned explicitly in the consulted sources. Often this is 'opening.' This is for instance the case in Bororo, one of the languages in which the respective term is present in the word for 'cave' (*ia-ri, ri* 'stone'), in Toaripi, where it is present in the word for 'nostril' (*ever-ape*, *ever* 'nose'). In Kiliwa, *ha?* in *miy=ha?* 'calf' also means 'face' alongside 'mouth,' and in Welsh *pen* in *pen-rhyn* 'headland' also means 'end, head, top.' Thus, one can basically observe the same pattern as with 'eye,' namely a certain amount of semantic generality and bleaching of the respective terms, in particular when occurring in morphologically complex expressions.

In spite of this, there is relatively little evidence for clustering of either contiguity-based analyzable terms for facial features or similarity-based extension to topological features of the environment in particular languages. A language where terms for facial features including the word for 'mouth' are frequent is Abzakh Adyghe, but here terms are often only semianalyzable (table 13):

Lexical Item	Gloss	Meaning
<i>zac'e</i>	/ze-č'e/ 'mouth-end'	'beard'
<i>ze-pq'</i>	/ze-pq'(ə)/ 'mouth-skeleton'	'lower jaw, chin'
<i>l'ə-pš'e</i>	'oral.cavity/opening -??'	'lip'
<i>l'ə-ps</i>	/l'ə-psə/ 'oral.cavity/opening-water'	'saliva'

table 13: contiguity-driven terms for facial features in Abzakh Adyghe

Two metaphor-based conceptualizations for the facial features 'beard' and 'chin, jaw,' with the term for 'mouth' acting as a contiguity anchor as defined in chapter 3, are found in Takia, as well as one for 'nostril' in which the 'mouth' serves as the source concepts' (table 14):

Lexical Item	Gloss	Meaning
<i>awa-n dabi-n</i>	'mouth-3SG root-3SG'	'beard'
<i>awa-n to-n</i>	'mouth-3SG arm-3SG'	'chin, jaw'
<i>ŋdu-n awa-n</i>	'nose-3SG mouth-3SG'	'nostril'

table 14: metaphor-driven terms for facial features in Takia

In contrast, pervasive conceptualization of topological features or other body-parts do not notably cluster in a particular language (at least not for the concepts presently looked at). Extensions of ‘mouth’ to nature-related and topological concepts are, comparable to the situation for ‘eye,’ often found when the target concept has roundish shape. This is true of virtually all of the meanings in this domain for which an association with ‘mouth’ is found, although, again, for the temporal concepts ‘dawn’ and ‘sunset’ the level of abstraction is somewhat higher than with the topological concepts. Unlike the patterns observed for ‘eye,’ however, objects which either have an opening, such as a ‘cave,’ or which involve the end or starting point of the passage of a substance or object (a figure) along some trajectory (here it seems useful to adopt terminology borrowed from Gestalt psychology into Cognitive Linguistics) appear to be a particularly amenable to being conceptualized via ‘mouth.’ This is most obviously the case for the associations between ‘estuary’ and ‘mouth’: just like the mouth is located at one end of the esophagus and respiratory tract, so the estuary constitutes one of the end points of a river.⁸ But the patterning is also noticeable when it comes to the meanings ‘sunset’ and ‘dawn’ which relate to the beginning or end-point of the movement of the sun in the sky. The case of the association with ‘beak’ is due to functional similarity, since the beak is the corresponding body part of birds to the mouth in that its functions include ingestion.

In the domain of human body-parts it is likewise roundness that appears to be the most prominent feature that triggers the fact that ‘mouth’ is used as a source concept, again obviously aided in the cases of the ‘nostrils’ and the ‘stomach’ by the fact that these body-parts have an opening (although the association with ‘stomach’ may be additionally supported by the contiguity between the two concepts as they both participate in the process of ingestion).

There are a number of concepts which can, judging from the evidence of the sample, cross-linguistically both be named with reference to ‘eye’ as well as to ‘mouth’: these include the ‘nostril,’ the ‘nipple,’ the ‘window,’ and the ‘dawn’ (and there are semianalyzable terms for ‘navel’ on the basis of ‘eye’ and ‘mouth’ in one language each). But notably, languages in which the word for ‘nipple’ is based on ‘eye’ are more frequent than those in which it is based on ‘mouth.’ Conversely, ‘mouth’ is a more frequent source concept for ‘nostrils’ than ‘eye’ is, and the same is true of ‘dawn,’ where associations with ‘mouth’ occur in two languages each but only in one with ‘eye.’ Thus, while their round shape makes them in principle amenable to being conceptualized both via ‘eye’ and ‘mouth,’ there appear to be some cross-linguistic preferences that have to do with whether the referents in question have openings or may be viewed as constituting parts of the trajectory of some objects. Size of the respective body-part term might play a role here, too: a ‘nipple’ is relatively small, comparable in size better to the ‘eye’ than to the ‘mouth;’ also note that ‘eye’ as a source concept for the slightly bigger ‘calf of the leg’ is not attested. For the case of the ‘nostrils,’ which are also comparatively small, the salience of their function as a trajectory (note also terms such as Baruya *siduta* /sinna-tuta/ ‘path-

⁸ The reverse naming pattern for ‘estuary’ is found in Jahai where *kit tom* ‘river-mouth’ is literally ‘water-bum’ (Burenhult 2008b: 186). Body-part metaphors for hydrological features are pervasive in this language.

nose/nostril') may override the factor size. Thus, tentatively: THE 'EYE' WILL BE A MORE FREQUENT SOURCE CONCEPTS FOR ROUND (THREE-DIMENSIONAL) OBJECTS, WHEREAS 'MOUTH' WILL BE MORE OFTEN UTILIZED AS A SOURCE CONCEPT WHEN THE TARGET CONCEPT IS ROUND, BUT WITH THE ADDITIONAL COMPONENT OF IT HAVING EITHER AN OPENING OR FORMING THE BEGINNING OR FINAL POINT OF THE TRAJECTORY OF SOME ENTITY.

There is one exception to the above generalization, namely 'window,' which is more frequently conceptualized via 'eye' than via 'mouth,' although a salient feature of windows is obviously that they are openings in the walls of houses. This is against what one would expect under the tentative generalization just made. However, there are other important factors that come into play here, namely that windows are functionally associated with seeing, as is the 'eye,' and that terms for 'window' may be embedded into a broader conceptual transfer pattern that likens human faces to houses, as in the examples in (1.) from Ma'di (Central Sudanic):

- (1.) a. /dʒó tī/ 'house mouth' = 'door'
 b. /dʒó mī/ 'house eye' = 'window'
 c. /dʒó dri/ 'house head' = 'roof'

(Blackings and Fabb 2003: 51, surface forms omitted)

Metaphorical denominations for 'door,' which are based on 'mouth' seem to be particularly frequent (see Zamponi 2009: 539 on Arawakan languages, Monod Bequelin 2006: 220 on Tzeltal), although the precise extent of this phenomenon cannot be assessed presently, since 'door' is not among the concepts on the wordlist under investigation here. And the passing-through function of doors is much more salient than that of windows (in addition to the fact that the latter have something to do with sight, which is another reason why 'eye' might be preferred as a source concept for 'window' cross-linguistically). This is probably why among the meronyms of 'house,' the 'door' rather than the 'window' are named using 'mouth' as the source concept, and the observations made with respect to 'door' and 'window' would then be readily accountable under the assumption of metaphors of 'mouth' for entities that are both roundish in shape and where some sort of spatial transition takes place. Summarizing the putative metaphoric transfer patterns, there are:

- (i) Shape-based similarity alone: roundness
- (ii) Shape-based similarity with the putative additional motivating component of target sources having an opening
- (iii) Shape based similarity with the putative additional component of the target concepts forming the starting or ending point of a trajectory.

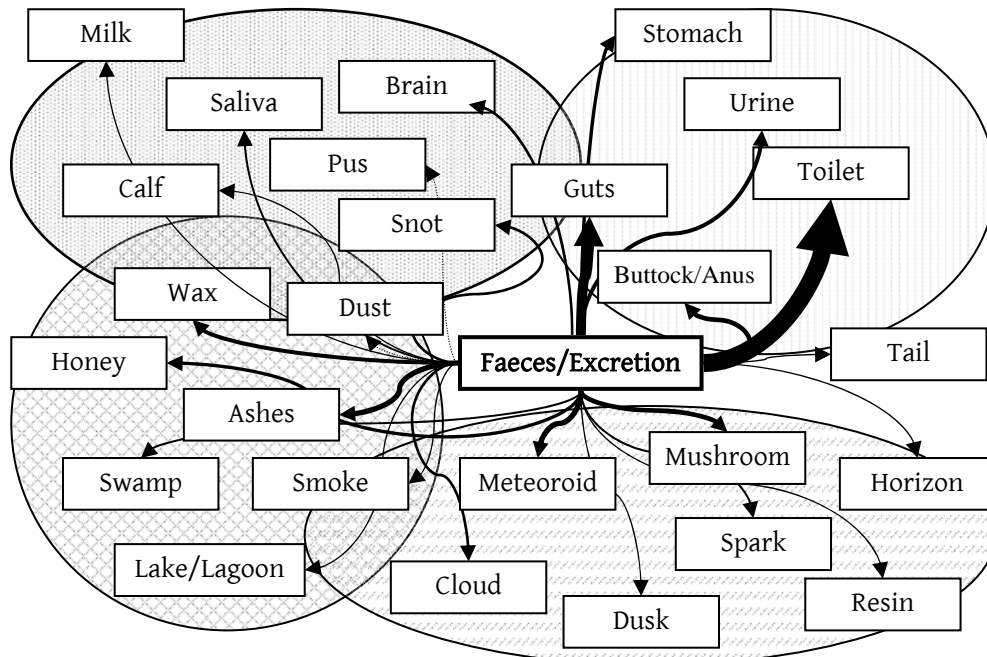
6.2.3.3. *Faeces/Excretion*

fig. 11: concepts with lexico-semantic associations to 'faeces' or 'excretion'

As represented diagrammatically in figure 11, lexico-semantic associations with 'faeces' may be grouped into four at times overlapping categories (not shown in the diagram is Bororo *baigabe* 'lightning,' which may be spurious: possible constituents are *baiga* 'type of Bororo bow' and *be* 'faeces'): on the one hand, in the top right corner are relatively obvious contiguity-based conceptualizations. These are noteworthy, but do not require extensive further discussion. In the top left corner other body-parts and body secretions are found, which are in some languages based on a semantic extension of 'faeces.' While there is a relatively obvious analogy between 'snot' and 'faeces' (this pattern is for instance found in Kashaya, where *?ilahp^ha* 'dry snot' consists of *?ila* 'nose' and *ahp^ha* 'excrement'), that between 'faeces' and 'brain' is somewhat less clear. Interestingly, this association occurs in the sample only in the Barbacoan languages Cayapa (*mishpe* /mishu-pe/ 'head-excrement') and Tsafiki (*fu-pe* 'hair-excrement'). The Tsafiki word for 'wax' is also based on 'faeces.' What some of the body fluids in this group have in common is their undesirability.

The circle in the lower left corner groups together a number of referents where the association with 'faeces' is still fairly well motivatable in most cases, and which are clearly metaphorical in nature. 'Honey' and 'wax' are in a sense the most clearly perceivable correlates to excretions when it comes to bees, while 'ashes' are the (probably mostly useless, which can be construed as the *tertium comparationis* with 'faeces') remnants of a fire, and similarly, 'smoke' is emitted by a burning fire but is not readily utilizable in terms

of human affordance. The case of the association with ‘lake’ is fairly unclear. This comes from Cubeo, where *macajitabũ* appears to consist of *maca* ‘faeces’ and *jitabũ* ‘puddle.’

The semantic associations in the lower right corner are the most interesting ones, because they tend to be highly abstract metaphors mostly. Table 15 lists languages with complex terms for ‘mushroom’ on the basis of ‘faeces’ (furthermore, Cashinahua colexifies these meanings directly).

Language	Term	Gloss
Rendille	<i>u'dú-yeyyah</i>	‘moon-faeces’
Toba	<i>huaqajñi l-'atec</i>	‘star 3SG.POSS-excrement’
Hawaiian	<i>kūkae-lío</i>	‘excrement/dung-horse’
San Mateo del Mar Huave	<i>aonts potwit</i>	‘excrete black.vulture’

table 15: Languages with a metaphorical term for ‘mushroom’ on the basis of ‘faeces’

A natural reaction to this data is to question whether the terms given in the consulted sources are really the generic terms for mushrooms as opposed to the name of a particular kind of mushrooms. However, this does not appear to be the case, since for instance the San Mateo del Mar Huave term is explicitly glossed as “los hongos” suggesting a generic function, and in the other sources there is no indication either that a specific kind of mushroom is designated by the above terms.⁹

Moreover, there are four instances in which celestial phenomena are expressed by terms on the basis of ‘faeces’: ‘meteoroid,’ ‘cloud,’ ‘horizon,’ and ‘dusk.’ While at least the connection with ‘meteoroid’ is motivatable, the other associations are fairly unclear, at least at first glance. It is probably an areal phenomenon of New Guinea to have complex words for ‘cloud’ consisting of terms for ‘wind’ and ‘faeces.’ The other two remaining patterns are found in Austronesian languages. In Tetun, ‘dusk’ is *loro-teen* ‘sun-excrement’ (the term also denotes a species of moss), while in Lenakel, ‘horizon’ is alternatively *noua-nisii-tehe* ‘fruit-excrement-sea’ or *noua-nisii-neai* ‘fruit-excrement-sky.’ The Tetun data appear to be explainable by assuming that the last light of day seen at dusk is something the sun has left behind -excreted- before disappearing, and similarly, one could speculate

⁹ In addition, there is some more, albeit not very compelling, evidence for a global prevalence of the association. Hladký (1986: 11) mentions that “the Czech word *houby* ‘mushrooms’ functions as an euphemism for *hovno* ‘shit,’” although this fact is not necessarily sufficient evidence for a connection between the two meanings, since replacement of swearwords can also occur with phonologically similar words not standing in any semantic relation with them, as in Engl. *shoot* replacing *shit*. English may be another case in point, since *stool* as found in *toadstool* can, as also suggested by Hladký (1986: 15), mean inter alia “a discharge of faecal matter of a specified colour, consistency, etc.; the matter discharged” (Oxford English Dictionary), in other words, ‘faeces.’ However, the earliest attestation of *toadstool* in the Oxford English Dictionary (as *tadstoles*) dates to 1398, whereas the earliest attestation of *stool* in the sense cited above is from 1597 and occurs in a medical context, so that it is possible that when *toadstool* was coined, the sense of *stool* understood was indeed that of ‘seat’ rather than ‘faeces.’ At any rate, the underlying metaphor seems generally to be that mushrooms are the residua of the respective entities that form the second member of the compounds; also noteworthy in this context is the association with ‘fart’ in Kiliwa and perhaps in Yay (see Appendix E, 40).

that in Lenakel the horizon, as the perceived end of sky and sea, “is” its excrement. Generalizing, one can note the following metaphorical transfer patterns:

- (i) From ‘faeces’ to other body secretions with a mostly negative connotation
- (ii) From ‘faeces’ to secretions of animals (‘wax,’ ‘honey’)
- (iii) From ‘faeces’ as the result of digestion to other natural processes which leave manifest remnants (‘ash,’ ‘smoke,’ etc.)
- (iv) From ‘faeces’ to natural phenomena which, loosely, may be conceived as the excretion of some entity

In many cases in which a physical object in the broadest sense is expressed using this extension, it tends to be rather smallish in size and to have rather well-defined boundaries. This is true of the concepts ‘calf of leg,’ ‘cloud,’ and ‘mushroom.’

Another noteworthy point is that often more than one instance of the same conceptualization strategy is found in the same language. This points to the possibility that the distribution of lexical patterns such as this one is not fortuitous. More generally speaking, it points to the existence of abstract underlying schemes of semantic processes in word-formation present in one language, but not in another (see also § 6.2.3.1. on ‘eye’). This is in principle reconcilable with cognitive accounts of word-formation as outlined e.g. by Tuggy (1987, 2005), albeit in a quite different manner. While such cognitive approaches to word-formation highlight the abstract nature of schemes on the formal side, here there appears to be a case of an abstract semantic pattern that can be realized in a number of ways on the semantic side.

Lexical Item	Gloss	Meaning
<i>aonts xiül</i>	‘excrete tree’	‘resin’
<i>aonts mijiw-aran</i>	‘excrete breast’	‘milk’
<i>aonts potwit</i>	‘excrete black.vulture’	‘mushroom’
<i>mi-xiüt aonts</i>	‘AL.POSS-line excrete’	‘guts’
<i>aonts najloc</i>	‘excrete wound’	‘pus’
<i>aonts ombeayaran</i>	‘excrete mouth’	‘saliva’
<i>aonts oxingueran</i>	‘excrete nose’	‘snot’

table 16: complex lexical items in San Mateo del Mar Huave involving *aonts* ‘excrete’

To illustrate what is meant, it is instructive to look at a number of complex lexemes in San Mateo del Mar Huave (table 16), which all contain *aonts* ‘excrete’ (one of the San Mateo del Mar Huave terms for ‘faeces,’ *aonts-aran*, is based on this root, *-aran* being a suffix indicating (probably inalienable) possession, compare Stairs Kreger and de Stairs 1981: 291). While there is frequent parallelism in form (but note the differing structure of the word for ‘guts’), even more striking is that the same pattern of semantic transfer is employed to conceptualize a wide variety of disparate referents.

6.2.3.4. Kinship semantics and their extensions

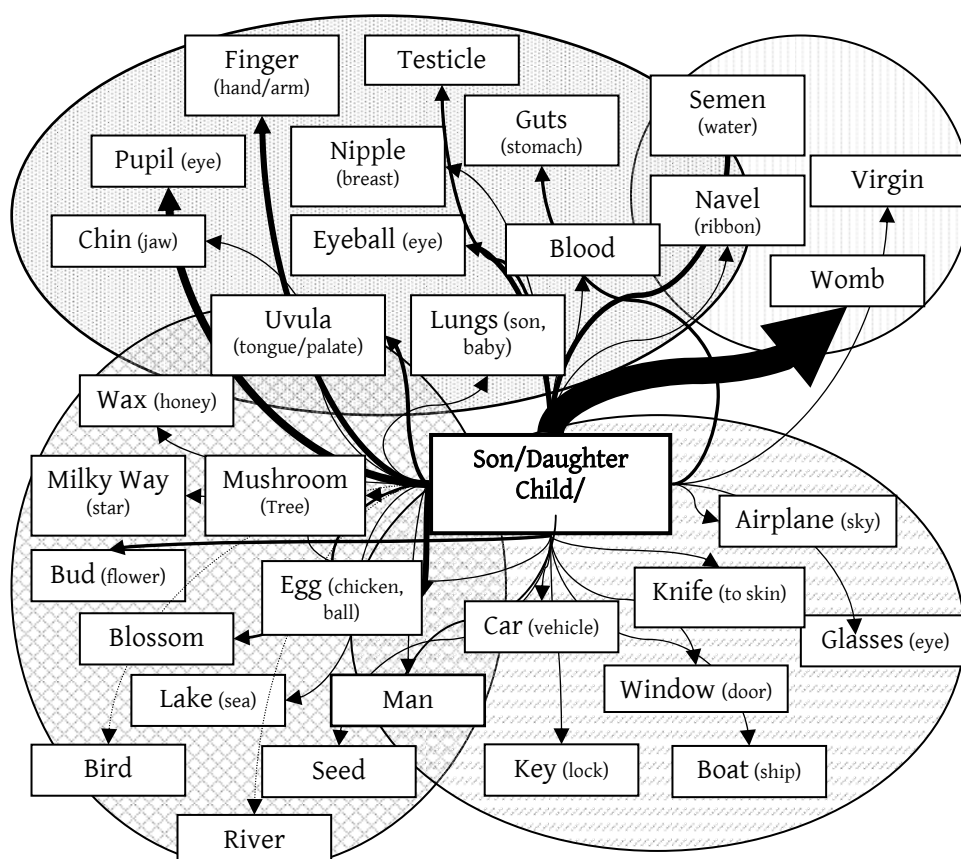


fig. 12.: lexico-semantic associations for 'son,' 'daughter,' or 'child'

The diagram in figure 12 provides an overview of colexifying and morphologically complex terms in which either of the meanings 'son,' 'daughter,' or 'child' generally is expressed by one constituent, with the meaning of (one of the) other term(s) in complex terms given in smaller size in parentheses.

As in the diagrams in §§ 6.2.3.1. – 6.2.3.3., contiguity-based associations are in the upper right corner. In this case, there are such associations with four of the meanings on the meaning list: some terms for 'virgin' contain an element meaning 'girl,' and frequently, terms for the 'womb' are complex featuring elements with either of the meanings 'son,' 'daughter,' or 'child' (see Appendix E, 148). Furthermore, in Miskito, *won klua* 'navel' is analyzable as 'child ribbon' (and probably also denotes the 'umbilical cord,' although it is not explicitly glossed so), and in Mbum, 'semen' is *mbii gûn* 'water child' (Kiowa colexifies 'semen' with 'child,' with the optionally complex term *'ih-ṭḥẹ* 'child/semen/egg-white' for 'semen'). The rest of the cross-linguistic associations can be classified as being metaphorical in nature. For the majority of associations, the general underlying transfer pat-

tern can be, inspired by Jurafsky's (1996) analysis of semantic sources for diminutives, summarized as CHILDREN ARE SMALL THINGS. Thus, one finds denominations for smallish body parts such as the 'finger' like Katcha *bibala ma nizo* 'child GEN hand,' for the 'uvula' like Tetun *nanarak-oan* 'palate-son,' and so on (the association with 'lungs' is only found in Dadibi, where *ogwa wai* 'spirit of man, lung' appears to be analyzable as 'son baby'). The associations with 'pupil' and 'eyeball,' while arguably also metaphorical in nature, are likely based on a slightly different perceptual property of the pupil, namely to reflect a small image of oneself in one's interlocutor's eye (Tagliavini 1949, Brown and Witkowski 1981, Urban forthcoming) that yields terms for the 'pupil' on the basis of meanings such as 'child,' 'small person,' 'doll,' etc. As far as the association with 'child' specifically is concerned, however, it can be reconciled with the general association of 'child' ('son,' 'daughter') with small things. As with body-parts, CHILDREN ARE SMALL THINGS arguably also underlies complex terms for nature-related terms on the basis of 'child,' like those for 'mushroom,' 'bud,' 'egg,' 'lake,' and 'Milky Way.' It is interesting to note that, with the exception of the 'Milky Way,' in effect an agglomeration of distant and hence small stars, these referents are roundish in nature, and this seems to be nonaccidental. Matisoff (1992: 304), in discussing complex terms in Thai based on *lûuk* 'child' (including for instance *lûuk-faj* 'child-fire' = 'spark'), maintains that the semantic development of *lûuk* was from 'child' first to 'fruit' (with the additional conceptual similarity that fruits serve the reproduction of plants, just as children do in the case of humans) to 'small thing' in general.

More difficult to analyze is the association with 'wax' (Kanuri *kàmàgàn-mí* 'honey-son.of'). Perhaps this is because 'honey' is more desirable than 'wax'? Likewise, in the domain of artifacts, 'keys' are smaller than the 'locks' they are used to open and close, a 'window' is smaller (but similar) to a 'door,' a 'boat' is smaller than a 'ship' (but used for the same purpose), a 'car' may be conceived of as a small 'vehicle' when the standard of comparison is, say, 'trucks,' and an 'airplane' can be seen as a small point on the sky (Ngaanyatjarra *yilkaringkatja* contains *yilkari* 'air' and *katja* 'son'). The same language is responsible for the association with 'glasses' by the term *kurungkatja* (*kuru* 'eye,' *katja* 'son'). The association with 'knife' is due to an advanced stage of conventionalization in Kiliwa, where *na(y)* in fact colexifies 'child' and 'small,' the relevant term is analyzable as *na(y)-c-ruuw* 'child/small-INST/MOUTH-to.skin' and the "literal" translation offered in the consulted source is "small skinner". In a few cases, however, the direction of the mapping is apparently not from 'child' (or 'son, daughter') to the other meaning, but rather the other way around: for instance, in Samoan, *gā'au* 'guts, intestines' is also a jocular designation for 'son.'

A diagram showing the associations found in the sample for 'mother,' 'father,' or 'parents' generally is in figure 13.

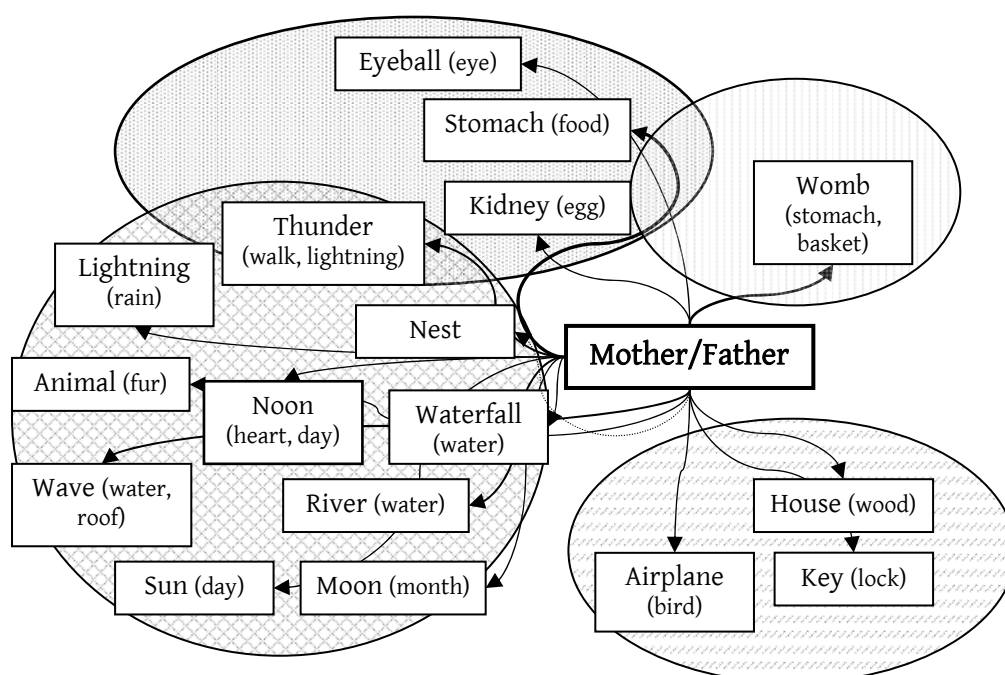


fig. 13.: lexico-semantic association for 'mother' and 'father'

Contiguity-based associations realized by complex terms are restricted to precisely one meaning, which is, as in the case for 'child,' the 'womb': Ket has *ām-d huīj* 'mother-POSS stomach' and Bislama *basket blong mama* 'basket GEN mother/pregnant.'

Jurafsky (1996: 546) proposes, among others, the metaphorical transfer pattern BIG THINGS ARE MOTHERS, which would account for the complex terms for 'eyeball,' 'river' (note that frequently, terms for 'river' contain morphemes meaning 'big' and 'water,' see Appendix E, 47), and 'house,' which is a large structure made of 'wood,' and presumably also 'wave' and 'waterfall.' Another pattern, perhaps related to Jurafsky's (1996: 547) GROUPS ARE FAMILIES, in which the opposition between 'mother' = 'big' and 'child' = 'small' plays as central role, is the extension of 'mother' to parts of an object situated in the center of that object, which would account for the associations with 'eyeball' and 'noon.' The association with 'lightning' is due to Mbum *māā-mbām* 'mother-rain,' and those with 'thunder,' 'sun,' and 'month' are found in one and the same language, San Mateo del Mar Huave: *qjüy teat monteoc* 'walk father thunderbolt' is the term for 'thunder,' *müm caaw* 'mother month' that for 'moon' and *teat nüt* 'father day' that for 'sun.'

It must be noted that there are a number of cases where presence of 'mother' or 'father'-terms remains conceptually somewhat unclear. Koyraboro Senni has *duma-ñaa-guuri* 'kidney-mother-egg' for 'kidney,' and Anggor has *ninihondi* 'animal, game animal,'

presumably analyzable as /nine-hondi/ ‘fur/feather-mother.’¹⁰ Also puzzling is Manange *Itantsa-lama* ‘lock-mother’ for ‘key’¹¹ in the light of Matisoff’s (1992: 300, 306) examples from languages of Southeast Asia and Oceania where the mother-child opposition is exploited to express the meanings ‘lock’ and ‘key,’ such as Thai *mêe-kuncɛɛ* ‘lock, padlock’ (*mêe* ‘mother’) – *lûuk-kuncɛɛ* ‘key’ (*lûuk* ‘child’) and Malay *ibu kuntji* ‘master/skeleton key, lock’ (*ibu* ‘mother’) – *anak kuntji* ‘key’ (*anak* ‘child’). In fact, these languages are not the only ones in which such an opposition is found, and the existence of such patterns has not gone unnoticed by both Matisoff (1992) and Jurafsky (1996). ‘Thumb’ is ‘mother of hand’ and the ‘finger’ ‘child of hand’ “throughout the Mayan family.” For instance, Tzeltal has *smeʔak’ab* ‘your thumb’ (literally “its mother your hand”) and *yalak’ab* ‘your finger’ (“its child your hand,” Matisoff 1992: 346fn97, quoting Terrence Kaufman p.c.). This is confirmed by the sample data: In Itzaj, *al k’ab* ‘finger’ is analyzable as ‘child arm/hand’ (‘thumb’ is not on the meaning list). Again in Malay, *ibu panah* is ‘bow’ and *anak panah* ‘arrow’ (Matisoff 1992: 301), and in Meyah, *otkonú efesá* ‘stomach child’ is ‘intestine,’ and *otkonú mosú* ‘stomach mother’ is ‘large intestine.’

Curiously, it is much more frequently the female parent, the ‘mother,’ rather than the ‘father,’ whose designant is used in complex terms to convey size or importance. This difference also turns up in a different context, namely that evidence from several languages suggests that if the opposition is not ‘mother’ – ‘child,’ but ‘female’ – ‘male’ (these competing patterns have also been noted by Matisoff 1992 and Jurafsky 1996), it is still the female semantics associated with big size and the male semantics with normal or small size.¹² In Mali, the masculine and feminine noun classes are employed in a similar fashion, with the feminine class conveying big size and the masculine normal size: thus *churet-ki* (-*ki* is the feminine noun class marker) is ‘a large flame’ and the masculine *churet-ka* ‘an average sized flame’ (Stebbins 2005: 101); for particularly small objects the diminutive class suffix is -*ini* is chosen in Mali (that indeed gender is a factor in the semantics of the noun classes is shown by examples such as *asingal-ka* ‘male forest spirit’ and *asingal-ki* ‘female forest spirit,’ Stebbins 2005: 103, where size difference does not seem to play a role). In Bora, Manguaré drums used for communication come in a set of a big and a small one, the big one being called ‘female’ and the small one ‘male’ (Seifart and Meyer 2010: 4). In Yeli Dnye, ‘hisfingers’ are *kóó pyââ dmi* ‘hand/arm woman bundle’ (where ‘bundle’ is a classifier) and ‘his thumb’ *kóó k:aa pyââ* ‘arm taro woman’ (Levinson 2006b). In Koyraboro

¹⁰ Note also Samoan *tamatama’ilima* ‘finger,’ containing *lima* ‘hand’ and *tama’i* ‘small thing.’ *Tamā* is ‘father,’ but the initial two syllables of the complex term seem more likely due to partial reduplication of *tama’i*.

¹¹ For ‘lock’ Manange also has a redundant compound *Itantsa-2tsatsa* ‘lock-young’ which again suggests the opposite direction from the apparently more widespread Southeast Asian pattern by virtue of featuring *2tsatsa* ‘young’ (*Itantsa* itself is perhaps borrowed from Nepali *taalca* ‘lock’).

¹² In Khoekhoe, different nominal designants indicating gender and number may be suffixed to one and the same root to yield different meanings. Here, the evidence from the present data is somewhat inconclusive, and gender assignment seems to be to a large part arbitrary, but there is slight evidence that it is indeed the feminine gender associated with smallness or lesser “fierceness:” *llnâ-b* (-*b* being the nominal designant for the third person singular masculine) is ‘horn of an animal,’ *llnâ-s* (-*s* indicating third person singular feminine) is ‘trumpet, brass instrument,’ *lnanu-s* is ‘rain, raincloud,’ *lnanu-b* ‘rain, thunderstorm,’ *khara-s* is ‘testicle,’ *khara-b* ‘scrotum.’

Senni, the ‘mother’ – ‘child’ and the ‘man’ – ‘woman’ opposition coexist. Compounds with *-ñaa* ‘mother’ denote “a complete object that constitutes the source of the entity denoted by the compound initial” (Heath 1999: 108, compare also Matisoff 1992 on the extensions of ‘mother’ to ‘source, origin’). Thus, *dugu* is ‘incense’ and *dugu-ñaa* is the ‘incense plant;’ plants and their products are the most common application of the system. In contrast, compounds with *-ize* ‘child’ have diminutive semantics (*tasa-yze* ‘small bowl’) or fulfill the ‘unit-excerpting’ function of the diminutive noted by Jurafsky (1996) in the realm of botany, as in *hayni-ize* ‘grain of millet,’ but also in body-part terms, as in *himbiri-ize* ‘single hair’ versus *himbiri* ‘hair’ collectively and *kabe-ize* ‘finger’ as opposed to *kabe* ‘hand.’ Here, corresponding pairs as in other languages as discussed above are found: *daarey-yze* is the ‘jujube fruit,’ and *daarey-ñaa* the ‘jujube tree’ (Heath 1999: 108). There are also compounds with *aru* ‘man’ and *woy* ‘woman,’ which are used to specify gender of an animal (*čirow* ‘bird,’ *čirow-aru* ‘male bird,’ *čirow-woy* ‘female bird’), but also to distinguish size: *hoŋko* ‘water lily fruit,’ *hoŋko-aru* ‘large water lily fruit,’ *hoŋko-woy* ‘small water lily fruit’ (Heath 1999: 109). Thus, in Koyraboro Senni, it is the compound with ‘man’ that conveys bigness and that with ‘woman’ conveys smallness, in contrast to the evidence from languages discussed above. This is the paradoxical situation noted by Jurafsky (1996: 545) that ‘woman’ or more generally ‘female’ may be associated cross-linguistically with big as well as small size. The evidence for this presented by Jurafsky comes mainly from gender alternation, such as Hindi *ghantā* ‘bell’ (masculine) and *ghantī* ‘small bell’ (feminine) rather than from morphologically complex terms. Such instances are not found in the present sample, and the evidence from the sample suggests that the extension of ‘woman’ to big size is more common than that with small size, especially as far as complex terms of the lexical type are concerned,

There is also one instance where a term for ‘grandmother’ rather than ‘mother’ is used to convey big size: in Mali, ‘flood’ is *milat-ka av-uouk* ‘coconut.shell-M.SG SG.POSS-grandmother;’ furthermore, the Rama terms *dama árkalí* ‘lightning’ and *dama yatangi* ‘thunder’ contain *dama* ‘grandfather.’

6.3. ENVIRONMENTAL FACTORS

Of course, especially in the conceptualization of the natural surroundings by a speech community, the properties of these may play a role in the development of certain patterns of colexification. One example is the colexification of ‘forest’ and ‘mountain’ which is attested in a number of languages in the sample (unsurprisingly, semantic shift in diachrony is also attested, for instance in Uralic, Redéi 1988: 571). There is a straightforward explanation for this pattern available: in mountainous terrain, dense vegetation with larger plants is found on mountain slopes, whereas in the valleys, vegetation is more sparse, typically grassland (compare the Kiliwa term *?+mat=xu?sawý* ‘DN+earth/land=clean/clear’ for ‘valley’). Obviously, such a landscape is likely to trigger the colexification of ‘forest’ and ‘mountain’ (see also Fränkel 1938 for semantic shift between terms for environmental features within Indo-European). It is then a fine example of spatial contiguity semantically – where there are mountains there is forest and vice versa.

Another lexical feature that may be directly due to environmental features is the colexification of ‘cloud’ and ‘fog.’ Laycock (1970: 1138) states that in New Guinea, this pattern occurs typically in high mountain areas. It is empirically testable whether there is indeed a correlation between this pattern of colexification and altitude on a global scale. Data for altitude were gathered for this purpose from GTOPO30, a digital elevation model of the world available at <http://www.gpsvisualizer.com/elevation.html> (latitude and longitude data from the World Atlas of Languages Structures were used for the altitude query).¹³ Under the interpretation that there is a scale of lexical differentiation between the two referents, from fully differentiated when lexically unrelated terms are present via semi-differentiated when a complex term for ‘fog’ is present on the basis of ‘cloud’ (as described in Appendix E, 25) to lack of lexical differentiation in the case of colexification, there is a statistically significant correlation between altitude and the degree of differentiation in the languages of the statistics sample ($S = 33422.35$, Spearman’s $\rho = .3894346$, $p = .0009417$), but even when one removes analyzable terms from the calculation due to their ambiguous status, there still is a strong difference, significant at $p = .00108$ ($S = 26237.3$, Spearman’s $\rho = .3993293$), to the effect that languages spoken at higher altitudes are more likely to feature this pattern. A visualization of the differences is provided in figure 14.

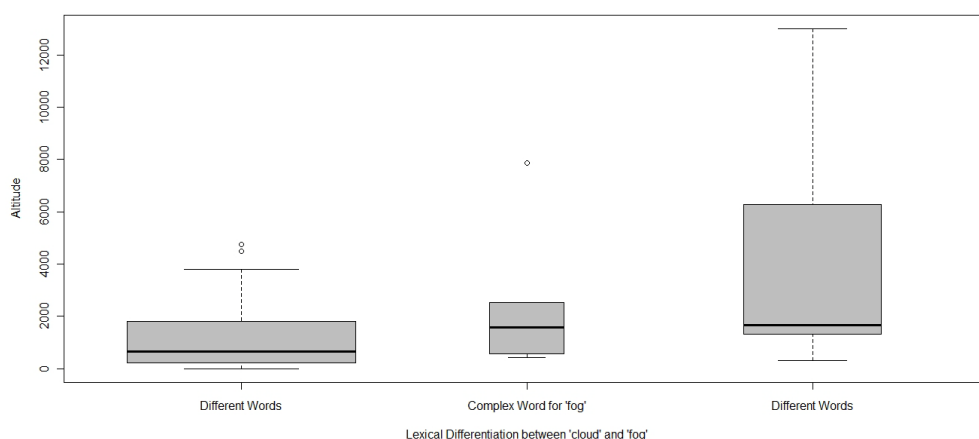


fig. 14: colexification of ‘cloud’ and ‘fog’ depending on altitude.

However, there are also a lot of languages spoken at very high altitudes which use different words for the two referents, and the statistics is not sensitive to alternative synonyms or near-synonyms that may be present in languages with colexifying terms.¹⁴

¹³ Unfortunately, the query for unknown reasons yielded errors for the coordinates for two of the languages, Embera and Bislama, which is why they are excluded from calculation.

¹⁴ Another factor was tested, but with negative results: One can also speculate whether there are extra-linguistic factors that might predict a language’s behavior with respect to the carving up of the lexical domain of bodies of

6.4. GENEALOGICAL AND AREAL LEXICO-SEMANTIC PATTERNS

6.4.1. INTRODUCTION

This section addresses another major topic concerning the lexicon from a cross-linguistic point of view. Alongside asking questions about the distribution of quantitative aspects of lexical motivation that were discussed at length in chapter five, it is at least equally interesting to ask about the distribution of individual lexico-semantic patterns, both from a genealogical and areal perspective.

6.4.2. INTRA-FAMILY COMPARISONS FOR SEMANTIC ASSOCIATIONS

From the point of view of genealogical linguistics, an interesting question that can be asked is whether there are lexico-semantic patterns that are peculiar to a particular language family. Departing from this basic question, it is possible to extend the discussion into more theoretical matters, that is, to elucidate whether patterns in semantics and in the structure of morphologically complex lexical items can be diachronically stable within language families so as to make them useful additional features that can, alongside e.g. regular correspondences in phonology, be of use for genealogical classification and the establishment of language families. Laycock (1975: 228), for instance, summarizes a number of areal patterns of New Guinea reported in Laycock (1970), and notes that “[i]t seems that some of these distributions may be highly regional, and may prove, if charted on the language map of the New Guinea area, to be useful in the establishment of linguistic supgroupings” and goes on to say that “[s]uch a use of semantic domains for linguistic taxonomy is a new approach which is as yet untried, but which shows signs of promise for the future.” Matisoff (1978: 231) suggests that “[t]he shifting patterns of semantic association within a language or a language-family are at least as interesting as phonological changes through time, and may prove to be equally criterial for establishing degrees of genetic relationship” (see also François 2010 on the reconstruction of semantic patterns and the issue of disentangling them from areal diffusion).

Within-family stability of lexico-semantic associations were investigated selectively for two well-established and uncontroversial families, namely Tupian (actually,

water, in particular with respect to the colexification of ‘water’ and ‘river.’ It is for instance possible to speculate that in languages with comparably many speakers which are thus presumably spoken in a relatively larger territory when compared with languages with a small speech community, people will be familiar with several different rivers in the territory, whereas in a smaller language in terms of speaker size, confined to a small territory, there may well only be one natural watercourse people encounter on a day-to-day basis, serving as the source of fresh water. To assess whether empirical data support this hypothesis, figures for the size of the respective speech communities were gathered from Lewis (2009), see appendix D for data. However, statistical analysis does not yield a significant difference between the populations under the interpretation, as in the analysis for ‘fog’ and ‘cloud’ above, as a continuum of decreasing lexical differentiation ($S = 54230.14$, $\rho = .05117413$, $p = .674$, Wilcoxon rank sum test). Probably the initial hypothesis is too simplistic anyway, given that small speech communities need not be settled in a particular territory, but instead may be highly mobile (Johnson and Earle 2000).

Tupi-Guaraní more narrowly) and Uto-Aztecan. For this task, a quite simply methodology was employed: when for a given meaning equivalents are available for all sampled languages within the language family, and where at least one has a morphologically complex expression or particular pattern of colexification, the other language(s) was/were checked for presence or absence of the relevant pattern.

The Tupi-Guaraní languages Guaraní and Wayampi have a total of 24 patterns in common, most of which are found in the domain of nature-related and topological terms. They include the following pairs of colexification: ‘nose’ – ‘beak’ – ‘prow of canoe,’ ‘bark’ – ‘skin,’ ‘coal’ – ‘embers,’ ‘foam’ – ‘bubbles,’ ‘grass’ – ‘plant,’ ‘river’ – ‘water,’ ‘smoke’ – ‘steam,’ ‘tree’ – ‘wood,’ ‘boat’ – ‘canoe,’ ‘paper’ – ‘book,’ ‘nipple’ – ‘breast, teat,’ ‘fingernail’ – ‘claw’ and extension of ‘skin’ to ‘surface, cover’ more generally. As for complex terms with common structure, there are complex terms for ‘cave’ (‘stone-hole’), ‘Milky Way’ (‘tapir-way’), ‘star’ (‘moon-fire’), a term for ‘belt’ containing the word for ‘waist,’ complex terms for ‘tear’ (‘eye-water’), words for ‘whirlpool’ containing verbs meaning ‘to twist’ and ‘to turn around,’ and complex terms for ‘milk’ with constituents ‘breast’ and ‘water, liquid, juice.’ Moreover, there is an association between ‘horn’ and ‘point’ by colexification in Guaraní and by a complex term in Wayampi, and of ‘semen’ with ‘child’ by colexification in Wayampi and by a derived term in Guaraní.

Many of the shared patterns are so frequent cross-linguistically that their value as a characteristic of a genealogical grouping is strongly diminished as they might easily also have come into being independently (see relevant sections in Appendix E). However, in Tupi-Guaraní there is also a genealogical signal consisting of rare or even absent structures outside of this family, in particular the terms for ‘Milky Way’ and ‘star.’ Based on glottochronological calculations, Rodrigues (1964) dates the split of Tupi-Guaraní to approximately 2,500 BP, and Silva Noelli (2008: 663) informs that radiocarbon dating of artifacts from archaeological sites even suggests a “much earlier” date. Another unrelated piece of evidence for an at least relative stability of certain lexico-semantic associations comes from Malagasy. The Malagasy word for ‘sun’ is morphologically complex and is of the type ‘eye of day,’ which is typical for languages of the Austronesian family and languages of Southeast Asia (Urban 2010). However, Malagasy has replaced its inherited word for ‘eye’ by a loanword from Bantu, *màso*, but retained notably enough the complex term for ‘sun’ of the Austronesian-Southeast Asian type employing the Bantu loanword. This shows that lexico-semantic structures at least have some potential for diachronic stability over the not inconsiderable time-depth of probably more than one millennium in the case of Malagasy ‘sun’ (Adelaar 1989: 35 tentatively posits a migration of Austronesian speakers to Madagascar in or after the 7th century AD, while noting that there is also evidence for prolonged contacts with South Sumatra after the migration event) and at least two and a half millennia in the case of Tupi-Guaraní.

However, there is also a lot of negative evidence. The Uto-Aztecan language family has considerable greater internal diversity than Tupi-Guaraní, and its common ancestor is thought to have been spoken earlier than that of Tupí-Guaraní. Holman et al. (2011) estimate a breakup date around 4,000 BP. The evidence from this family is sobering, in spite of the high similarity in the abstract values (percentage of analyzable terms and of

metaphor-driven terms) for this particular family: there is not a single pattern that is common to all of the four languages in the sample, and the four cases of a correspondence of three languages, colexification of 'skin' and 'bark' in Comanche, Pipil, and Yaqui, 'sun' and 'day' in Cahuilla, Comanche, and Yaqui, 'tree' and 'wood' in Cahuilla, Comanche, and Pipil, and of 'moon' and 'month' in Cahuilla, Comanche, and Yaqui, are all among the most common semantic associations in the world's languages as a whole. Even correspondences between two languages are not very common: Patterns of colexification shared by two of the four sampled Uto-Aztecan languages include: 'gold' – 'money' (Comanche, Pipil), 'water' – 'river' (Cahuilla, Pipil), 'pit' or 'seed' – 'eye, face' (Cahuilla, Pipil, in which latter the meaning 'face' is restricted to compounds), 'land' – 'earth' (Pipil, Yaqui), 'boat' – 'canoe' (Comanche, Yaqui), 'fingernail' – 'claw' (Pipil, Yaqui), 'skin' – 'leather' (Yaqui, Pipil), and 'skin' – 'shell' (Pipil, Comanche). As for complex terms, Cahuilla and Pipil have a complex term 'mouth-hair' for 'beard,' and there are complex terms for 'forest' with a constituent meaning 'tree' in Comanche and Pipil, which are however structurally quite different otherwise. Also, Comanche and Pipil both have terms for 'sky' based on the notions 'high' or 'above,' and there are complex terms for 'scissors' containing a verb meaning 'to cut' in Comanche and Yaqui. Further more heterogeneous commonalities include an association between 'honey' and 'candy, sweets' by colexification in Yaqui and by a complex term in Pipil, and a complex term for 'lake' involving a constituent meaning 'water' in Cahuilla, and semianalyzable terms of that kind in Comanche and Yaqui. Mirroring the colexification of 'water' and 'river' in Cahuilla and Pipil, Yaqui has the complex term *batwe /ba'a-bwe'u/* 'water-big.' There are complex terms for 'swamp' with one constituent meaning 'water' in Pipil and Yaqui, but with otherwise different structure (and there is a semianalyzable term in Comanche). Comanche colexifies 'mouth' and 'lip,' and a complex term for 'lip' with a constituent 'mouth' is featured in Yaqui. Finally, there are complex terms for 'nostrils' with a constituent 'nose' in Pipil and Yaqui, and complex terms for 'eyebrow,' 'eyelash,' and 'eyelid' with one constituent meaning 'eye' in many of the languages but with varying other constituents. Very many of the abovementioned associations are common cross-linguistically and well attested outside Uto-Aztecan (see relevant sections in Appendix E), another one, the association of 'sky' with meaning like 'high, above,' is common in North America as a whole, and still others, such as colexification of 'gold' with 'money' and the terms for the 'scissors' cannot be interpreted reasonably as indicating a deep historical signal because they almost certainly postdate the time of European contact.

As a preliminary conclusion from the small set of investigated families, the genealogical signal of lexico-semantic structures is weak, and thus appears to be of rather limited use for traditional historical linguistics concerned with single language families alone (albeit not of no use at all, as the case of Tupi-Guaraní shows, though in general this family "is noted for a high degree of lexical and morphological similarity among its member languages in spite of their extensive geographical separation," Jensen 1999: 128). However, this does not mean that lexico-semantic patterns are not amenable at all to historical interpretation (note also the areal skewing with regard to lexical differentiation of 'sun' and 'moon' which is amenable to a historical interpretation). This is one of the topics to be

discussed in the following section, which deals with areal factors that are at times responsible for the distribution of individual patterns.

6.4.3. AREAL PATTERNS IN SEMANTIC ASSOCIATIONS

6.4.3.1. *Introduction*

Alongside intra-family comparison, on the other hand, one can also take the opposite perspective and ask whether certain patterns in the lexicon are not genealogically but still areally restricted. If such patterns are found, this would be an indication that the lexicon is susceptible to significant influence from neighboring languages in language contact situations. Evans (1990: 137), for instance, maintains “that the whole continent of Australia is characterisable as a linguistic area from the point of view of certain types of polysemy and semantic change that are common right across the continent, but rare or unreported elsewhere.” As the quote already makes clear, areal influences in the lexicon pertain both to morphologically complex lexical items as well as colexification, although the latter has not received ample discussion: when it comes to influences languages in contact can exert on each other with respect to the lexicon, what immediately comes to mind is lexical borrowing, that is the transfer of a word, in its phonological form and typically also semantic content, from one language to another. What probably comes to mind next are calques or loan-translations, and indeed, there are a number of studies devoted to areal calquing exclusively or as part of larger discussions of linguistic areas (to be mentioned in a minute). In addition, there is an ugly duckling that has not received the same amount of attention: this is the transfer of semantic structure from one language to one or more other neighboring languages and the convergence of neighboring languages with respect to the internal semantic structure of their vocabulary items. In contrast to lexical borrowing, this process does not involve transfer of linguistic material, but rather of semantic structure alone, which is then superimposed onto native lexical items. Saying that this phenomenon has not received wide attention does not mean that the phenomenon is not known in theory, but discussions in the literature are typically not longer than half a page. In the German literature, “Lehnbedeutung” is typically used (e.g. Blank 1997: 349); this term probably goes back to Betz (1949). Haugen (1950: 219) uses the term “loan synonym” for this process “which only adds a new shade of meaning to the native morpheme;” Curnow (2001: 427), in a brief discussion, adopts Haugen’s terminology, and defines loan synonymy as the process “where the meaning of a word is extended to fit the pattern of lexical extensions of a word in another language with a similar basic meaning.” Geeraerts (2010: 29) speaks of “semantic borrowing,” “the process by means of which a word *x* in language *A* that translates the primary meaning of word *y* in language *B* copies a secondary meaning of *y*,” noting that the process is also known as a “semantic calque.” One example adduced by Geeraerts (2010: 29–30) is Greek *angelos* which acquired, under the influence of (translation of religious texts from) Hebrew *ml’k*, the additional meaning ‘heavenly messenger,’ i.e. ‘angel’ alongside its basic and original meaning ‘human messenger, envoy.’ Smith-Stark (1994: 17) uses “loan shift” to refer to “calquing the internal structure of a lexical item.” Another major relevant line of research is that concerning relexification in the process of creole genesis, one of the facets of which is that the internal lexical seman-

tic structure of lexical items from the substrate language is mapped onto the phonological shape of words of the lexifier language (for an overview see e.g. in Lefebvre 2001 and Lefebvre 1998 for a more detailed case study). Relexification is primarily adduced to account for properties of creole languages.

Also related but slightly different because concerned with polysemy of grammatical morphemes in grammaticalization processes is Heine and Kuteva's (2005: 100) notion of "polysemy copying," as well as Mous's (2003) "lexical manipulation" in language mixing.

The terminological multiplicity reflects the paucity of detailed influential research on this phenomenon. Empirically oriented discussions of semantic transfer are mostly equally short as their theoretical counterparts, with the exception of Enfield (2003), who offers a detailed case study of semantic and grammatical convergence in Mainland Southeast Asia. Moreover, Ross (1996) mentions the remodeling of the lexical semantics of Takia vocabulary under influence of Waskia in the process of the "metatypy" of Takia; this includes loan translation, but Ross also explicitly mentions semantic transfer. Similarly, Aikhenvald (1999: 406) offers a brief discussion of lexical semantic influence of East Tucano languages on the Arawakan language Tariana, and recognition of the phenomenon is implicit in Campbell et al. (1986) and Matisoff (1978, 2004).

The following discussion of areal lexico-semantic patterns is concerned both with loan translations as well as with semantic transfer, in short, any lexico-semantic patterns comprised under the definition of lexical motivation as described in § 3.5. Matisoff (2004) uses "areal semantics" to refer to both sorts of patterns; however, strictly speaking this is a misnomer, since loan translations in fact are not actually concerned with semantics proper (compare Marty's 1908 critique of the equation of semantic patterns within complex expressions with semantics itself mentioned in § 2.4.), which is the reason for speaking about areal lexico-semantic patterns rather than areal semantics as a cover term. Where lexical semantics belongs in the general theory of language contact is likewise unclear. Thomason and Kaufman (1988: 121), unsurprisingly given the sparseness of dedicated literature on the topic, report not having found many examples of contact-induced change in lexical semantics, but state that they would expect it to pattern with changes in phonology and syntax, which they show to occur even if contact is not very intense. However, the motivation for convergence in the lexicon seems to be the same as for contact phenomena in general: it is an adaptation serving to increase intertranslatability of the languages in the repertoire of bi- or multilingual speakers (Gumperz and Wilson 1971), and, as Sasse (1985: 84-85) points out, lexical and grammatical calquing and syntactic convergence share this same ultimate cause.

While this is not the first study on this subject, it appears to be the first one with global scope. Past research on areal patterns in the lexicon has at times suffered from a lack of informedness on the cross-linguistic distribution of the phenomena in question. As Evans (1990: 152) also notes, claims concerning areality such as the one made by him for Australia as a whole "can only be fully substantiated when proposed paths of semantic change and grammaticalisation in Australia can be shown to differ significantly from those found in other language families and areas." For instance, Campbell (1979) as quoted in Smith-Stark (1994: 18-19), mentions 'skin-tree' for 'bark' and lack of a separate word for

'river' as areal patterns of Mesoamerica. The data of this study show, quite to the contrary, that these phenomena are widespread globally (Campbell also mentions other patterns that may truly be characteristic of Mesoamerica in particular). Campbell et. al (1986: 553, table 1) provide a list of 55 lexico-semantic patterns that are candidates of being diagnostic of the Mesoamerican linguistic area, of which twelve survive closer scrutiny (that is, examining whether a given pattern occurs in control languages immediately to the north and the south that are not part of the putative area) and are accepted by Campbell et al. as areal features of Mesoamerica. Among them are 'vein' = 'road (of blood),' 'edge' = 'mouth,' and 'lime' = 'ash,' 'stone-ash.' As Campbell et al. (1986) themselves note, it can be imagined that the former two patterns might also be coined independently and remark in footnotes that they in fact are also found in South America. The data of the present study confirm this, and reveals further cases elsewhere. Similarly, Matisoff (2004), while correctly pointing out that certain associations (such as 'fire' + 'tongue' = 'flame') are common cross-linguistically because they "are so 'natural' to human thought processes" (2004: 351), reports some lexico-semantic structures considered by him to be peculiar to Southeast Asia in the absence of known data from other areas. Among them are terms for 'meteor' on the basis of 'star' and 'faeces' in Hmong and Lahu (2004: 367). The present sample identifies this pattern in Sedang in Southeast Asia, but also shows that the association is also, and apparently more frequently, found in the Americas, namely in Central Yup'ik, Haida, Highland Chontal, and Toba as well as in the Austronesian language Tetun (see Appendix E, 36). Matisoff (2004: 367) also mentions a metaphorical transfer from 'eye' to 'anklebone' under the heading "Southeast Asian lexico-semantic areal features," but this association is even more frequent cross-linguistically than that between 'star' and 'faeces,' and occurs in very many areas of the world (see Appendix E, 99).

Here, the question is when one is willing to accept a particular feature, be it grammatical or lexical, as areal: is it enough if it is absent in the immediate vicinity of the area that is to be demonstrated, while it is acceptable if it shows up further South or North on the same continent or with some frequency in completely other areas of the world? Arguably, the above evidence does not mean that the respective patterns suggested by Matisoff (2004) should not be considered areal features of Southeast Asia, but they raise the question where one should draw the boundary at which one cannot speak of areal features anymore.

The assorted examples mentioned above point to two related vexing problems: the first concerns the need to test putative areal patterns against solid cross-linguistic evidence, or at the very least against a control sample of neighboring languages. The present data make it possible to assess areal lexico-semantic patterns against solid cross-linguistic evidence; and they also make it possible to offer for the first time a more extensive study of semantic transfers without concomitant transfer of linguistic material. The second issue already hinted to above is the problem of defining linguistic areas in the first place (for relevant discussion, see e.g. Bisang 2006, Muysken 2008, Campbell 2009). One big issue in that task is the problem of circularity: linguistic areas, by definition, are areas in which a number of languages have come to share structural features that are not due to genealogical inheritance, and the commonalities between these same languages is then

explained by their participation in the linguistic area. Furthermore, mere eyeballing of maps showing the distribution of a linguistic feature for potentially interesting areal patterns is methodologically dubious (Cysouw 2005, *inter alia*).

Because of the intricate difficulties in delimitating linguistic areas in a methodologically principled way, Bickel and Nichols (2009, *in press*) have taken a relatively radical measure: instead of relying on linguistic evidence in setting up areas, they first define areas without actually recurring to linguistic evidence. In this approach, geographically and sociohistorically delimited areas are first seen as hypothetical linguistic areas as well, and statistical procedures are then employed to test whether there is a linguistic correlate between geographically contiguous regions of the world which are known to be sociohistorically linked to each other.¹⁵

Here, a geography-based approach is adopted as well, as described in particular in Bickel and Nichols (2009: 487). Initially, a breakdown of the world in eleven areas modeled on that of Nichols (1992: 25-26) was used. These areas are: Africa (including Malagasy), Europe, Eurasia (excluding Europe, but unlike in Nichols 1992: 25-26, not the Caucasus), South and Southeast Asia (conventionally delimited, as in Bickel and Nichols 2009, by the Wallace Line), New Guinea, Oceania, Australia (including Tasmania), Western North America (including Kiliwa but excluding the Eskimo-Aleut language Central Yup'ik as in Nichols 1992: 25-26), Eastern North America (including Lesser Antillean Creole French and delimited from Western North America by the Rocky Mountains), Mesoamerica, and South America. This entails that the areas for which patterns can be detected are necessarily rather large, but since the density of the present sample is low, identification of comparably small sprachbund-sized areas is unfortunately not possible on the basis of the sample in the first place. For the same reason, the assessment of areality necessarily departs from classical definitions of linguistic areas such as that of Emeneau (1956: 15fn28) as quoted in Masica (2001: 209): "an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families," which is not a feasible definition for present purposes since in many cases only one language per family was sampled in the first place. As is known today, linguistic areas come in all sizes, from very small to very large (as has emerged from studies such as Dryer 1989, 1992 and Nichols 1992), and it is of course possible, and even likely, that a given lexico-semantic association detected in the sample for only one language somewhere in the world in fact participates in a smaller linguistic area and might be a diagnostic feature of it. One instance would be the Vaupés area as described by Epps (2007), who also notes similarities in lexical structure. Another example of such a smaller linguistic area is the Clear Lake area in California, where speakers of Lake Miwok, Patwin, Wappo, and Pomoan languages lived in close proximity. Lake Miwok and Patwin belong to different subgroups of Penutian, while Wappo and Pomoan are unrelated or unrelatable by traditional methods both to one another as well as to Penutian. There were frequent con-

¹⁵ A similar approach is sketched by Masica (2001: 219): "an alternative discovery procedure, ... and perhaps more 'objective', would be to start, not with languages or an area of interest, but with a few selected features of interest drawn from universal typology (i.e. rather than 'all' features), determine their distribution, and see what patterns emerge, particularly convergent patterns involving more than one feature" (emphasis removed).

tacts between the groups, intermarriage and widespread multilingualism (Callaghan 1964: 47fn3), the perfect medium for the emergence of linguistic areality. The Clear Lake area happens to be represented by Lake Miwok and Wappo in the present sample, and the Pomoan languages are represented by Kashaya, which is not immediately spoken around Clear Lake, but not far from it either. And indeed, there is evidence for convergence with respect to the lexicon (apart from a number of loanwords from neighboring languages in Lake Miwok and massive phonological convergence, Callaghan 1964). Some of the commonalities summarized in table 17 must be of fairly recent origin.

Concept	Lake Miwok	Wappo	Kashaya
'train'	<i>wikî karéeta</i> 'fire wagon'	n/a	<i>hokare'ta</i> /ʔoho-kare'ta/ 'fire-wagon'
'eyeball'	n/a	<i>huči-lél</i> 'eye-stone'	<i>huʔu: qʰaʔbe</i> /huʔuy qʰaʔbe/ 'eye rock'
'toilet'	n/a	<i>čéyu čhùya</i> 'feces house'	<i>ʔahpʰahca</i> /ʔahpʰa-ahca/ 'shit-house'

table 17.: Some lexical commonalities in languages of the broader Clear Lake area

Returning to methodological questions, statistical evaluation of the evidence is not performed, the reason being that when the global statistics sample is broken down into smaller areas, the coverage of each area is very sparse, and the results one would arrive at in statistical analysis are not amenable to meaningful interpretation since the power of potential statistical tests would be extremely low. Furthermore, under these particular circumstances, statistical analysis would likely filter out a considerable number of genuine cases of areality as a direct consequence of the low sample density for each individual area. Instead, the following interim measure to assess areality is used: lexico-semantic associations are considered potentially areal if (i) the number of unrelated languages with the pattern in one of the areas outnumbers the number of unrelated languages in the entire rest of the world more than twice and (ii) at least 15 per cent of genealogically unrelated languages within the area in question feature the association (for the case of Australia, which in Dryer's 2005 classification hosts only one very large language family, Australian, only the second criterion is applied, since otherwise no areality could emerge for this continent by definition. Note also that the unity of Pama-Nyungan and non-Pama-Nyungan, and even the genealogical unity of Pama-Nyungan is not universally accepted, see Dixon 2002 and Evans 2005 as representatives of the different opinions and further references therein). The assessment of areality is carried out on the basis of the EXT-2 sample, since this task requires the densest possible coverage of the world's languages. As a measure of genealogical control, however, datapoints obtained from genealogically related languages are collapsed to 1. That is, genealogical inheritance, which apparently occurs at times (compare the Tupi-Guaraní words for 'star' and 'Milky Way'), is ruled out – an important requirement when areal rather than genealogical factors are at stake! This criterion is for consistency also applied to terms denoting artifacts, which are often unlikely to be of great age due to the recent introduction of their referents. The second criterion is a control for the size of the area in terms of distinct language families it hosts. Given that South America, for instance, has a much higher genealogical diversity than, say, Europe, it would be much easier for a particular association to be diagnosed as potentially

areal for that area than for ones with less linguistic diversity, and hence lower sample density. Without additional control there would be the danger that the number of diagnosed patterns is merely a function of genealogical diversity, which is not the case when the present criterion is applied.

The criteria have a certain probability of leading to a number of false positives, i.e. patterns that are found to be areal in spite of existing counter-examples reported in the literature, but not in the sample. For instance, one of the lexico-semantic patterns is the metaphorical extension of terms for ‘beak’ to ‘prow of canoe’ in languages of South America, next to one case in Hawaiian. As a comparison with the data for the association ‘beak’ – ‘nose’ quickly shows, usually the very same terms also mean ‘nose.’ Now, a lexical connection between ‘nose’ and ‘prow of canoe’ is also reported to be common in Austronesian languages (Blust 2009: 314), as suggested by the presence in Hawaiian in the present sample, and occurs also in Australia (Schebeck 1978: 175). The reason why this pattern is not detected in the sample is likely that the respective words in Austronesian languages do not simultaneously also cover ‘beak’ (‘nose’ itself is not among the meanings presently surveyed). Thus this case is not damaging to the validity of the sample, but it shows that still some caution is required before the findings are interpreted as being incontrovertible facts. In spite of this, a criterion that is in danger of being too lax rather than too stringent is opted for because rather than sorting out too many potentially areal cases and to report only a smallish list, it seems more desirable at this stage of research to provide a more comprehensive overview of preliminary candidates for areal lexico-semantic patterns in the lexicon as a resource for further research, even if some of them might turn out to be spurious. In spite of the shortcoming of not being able to offer statistical backup of the patterns, this method still has one big advantage over previous studies on areality in the lexicon, namely that it allows assessing areal patterns against solid cross-linguistic evidence from the entire world. In addition, it satisfies, in spite of its departure both from traditional accounts and from Bickel and Nichols’s statistics-based approach, Masica’s (2001: 207) programmatic statement as to the goals of areal linguistics: “[p]rimarily, areal linguistics should mean the study of significantly non-random distribution of linguistic features in space – first of all the facts and if possible the reasons behind them” (emphases removed).

In addition, the method used also by definition excludes the possibility that one and the same lexico-semantic pattern is diagnosed as being characteristic of more than one area at the same time. For the moment, the discussion is concerned with positive evidence, in spite of Masica’s (2001: 215–216) valid reminder that linguistic areas can and should also be defined negatively, i.e. by absence of a certain feature (see § 6.4.3.14. for some casual notes on Eurasia in this respect).

The following tables show the lexico-semantic patterns that are diagnosed as being characteristic for the areas defined above, starting with Africa. Throughout, the first column states the lexico-semantic association, the second identifies the number of languages in the area with the pattern before the dash, and the number of languages outside the area (if any) after the dash. In the third column, the languages within the area participating in the pattern are named, and after them, the language(s) featuring the association

outside the area is/are stated in parentheses. The fourth column identifies whether the pattern is one of recurrent constituents in morphologically complex expressions (C), whether it is due to colexification (CL), or whether the pattern is mixed (M), i.e. the same lexico-semantic association is realized in at least one language as a morphologically complex expression and in at least one other by colexification. Such a situation should not be surprising, given the close ties between morphologically complex terms and colexification that have been emphasized throughout this study. The rightmost column is an important one: here references to the literature are stated. As already alluded to above, it is by no means the case areal patterns in the lexicon have entirely escaped attention of scholars so far. However, they are often mentioned en passant in other contexts (for instance, for the New Guinea area typically in discussions of difficulties in the applicability of lexicostatistics), and this column is intended to bundle previous statements in the literature that correspond to the detected patterns on the basis of the sample, and, importantly, to provide, where applicable, additional information on the historical emergence of the pattern and the cultural underpinnings that are likely responsible for them. Furthermore, the column also identifies statements in the literature where the lexico-semantic pattern is mentioned outside of the area for which it is presently detected, that is, where such information is available, it serves to flag patterns that may be spurious. Since the present method of assessing areality cannot be anything more than preliminary, patterns that may be areal, but on a smaller scale, are mentioned in prose casually after the tables, in particular if suggestions to this effect can be found in the literature, but relevant patterns fail to be diagnostic for the larger scale areas under investigation. As a final note, the analysis does not take into account the cross-linguistic differences in motivated terms between languages: it is conceivable that a language in which they are frequent has a greater chance of participating in areal patterns simply by virtue of having more motivated terms than another.

6.4.3.2. Areal Patterns of Africa

	Lexico-Semantic tations	Associa-	Number of Lan-	Languages	Type	Comments
1.	'seed' – 'kind'		3-0	Efik, Hausa, Koyraboro Senni	CL	
2.	'sunrise' – 'sprout'		3-1	Hausa, Khoekhoe, Noni Khalkha	- C	

table 18: areal lexico-semantic patterns of Africa

There are a number of other associations only found in the present sample in two unrelated languages of Africa. For instance in Hausa and Khoekhoe, there is a metaphorical transfer from 'branch' to 'descendants,' and in Hausa, as well as the Niger-Congo languages Mbum and Yoruba, there are complex terms for the 'eclipse' featuring constituents

meaning ‘to catch.’ In Khoekhoe, the ‘Adam’s Apple’ is called *dom-!khom-s* ‘throat-bundle-3SG,’ and similarly, in Yoruba it is *kókó-òfun* ‘lump-throat.’ Hausa and Yoruba colexify ‘buttocks’ with ‘root,’ and Hausa and Buli colexify ‘guts’ with ‘inner tube of tire.’ In Hausa and Katcha, perhaps by functional contiguity, words for ‘ball’ also denote the ‘nut of the dumpalm.’ There is an association between ‘cheek’ and ‘bag’ or ‘sack’ in Mbum and Ngambay, and in Hausa and Swahili ‘dawn’ is colexified with ‘morning prayer.’

6.4.3.3. Areal patterns of Europe

	Lexico-Semantic Associations	Number of Languages	of Languages	Type	Comments
1.	‘bay’ – ‘breast’	2-0	Basque, Greek	CL	also present e.g. in German, where it is a loan translation of Lat. <i>sinus maritimus</i> and attested from the 17 th century onward (Kluge 2002)
2.	‘estuary’ – ‘sea’	2-0	Basque, Welsh	C	
3.	‘house’ – ‘company, firm’	2-0	Basque, Greek	CL	

table 19: areal lexico-semantic patterns of Europe

Another putative European association is that between ‘Milky Way’ and ‘milk,’ attested in the present sample in Basque, and by a semianalyzable term in Greek. At least in German, the term is a 17th century calque from Latin *via lactea* (Kluge 2002); Latin may itself have been influenced by Ancient Greek. Likewise, a pattern having probably originated in Europe (though it may have Semitic origins) is the association between ‘Adam’s Apple’ and ‘apple,’ which is attested in the present sample in Basque, Greek, and Welsh, but also in Itzaj, where it was apparently calqued from Spanish.

6.4.3.4. Areal patterns of South and Southeast Asia

	Lexico-Semantic Associations	Asso-	Number of Languages	Lan-	Languages	Type	Comments
1.	‘waterfall’ – ‘cliff’		3-1		Bwe Karen, White Hmong, Yay – M Copainalá Zoque		

table 20: areal lexico-semantic patterns of Mainland Southeast Asia

Surprisingly given the amount of literature mentioning Southeast Asia as a host for areality, also in the lexicon, there is only one pattern emerging as a candidate for an areal association by employing the present methodology. This has something to do with the fact that presently not only Mainland Southeast Asia is considered, and it may be that the Southeast Asian mainland, if considered on its own and/or with inclusion of Japanese data, would yield a considerably higher number of areal associations. For instance, there are

terms for ‘animal’ containing elements meaning to ‘move’ and ‘thing’ in Mandarin, Vietnamese, and Japanese, probably due to widespread borrowing from Chinese. Moreover, in Manange and White Hmong, but also in Haida, there is a term for ‘lightning’ involving an element meaning ‘to blink eyes’ (Bauer 1992 argues “that the three meanings ‘wink’, ‘lightning’, and ‘wave’ comprise a Southeast Asian word family,” but also mentions that this association might extend to Austronesian languages). Also with regard to ‘lightning,’ there are associations with spirits in White Hmong and Yay, and there are terms for this meaning involving a constituent meaning ‘sword’ in Sedang and Yay. Also, a characteristic of this area are terms for the ‘eclipse’ involving a constituent meaning ‘to eat’ and terms for animals; however, ‘eat’ as a constituent of relevant terms is also attested in other languages outside Southeast Asia (compare Appendix E, 35 and Matisoff 2004).

6.4.3.5. Areal patterns of Eurasia

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	‘flower’ – ‘picture’	2-0	Khalkha, Yukaghir	CL	
2.	‘forest’ – ‘taiga’	2-0	Khalkha, Yukaghir	CL	
3.	‘lightning’ – ‘arrow’	2-0	Khalkha, Kildin Saami	C	
4.	‘rainbow’ – ‘thunder’	3-0	Ket, Nivkh, Kildin Saami	C	See Räsänen (1947)
5.	‘river bed’ – ‘deep’	2-0	Ket, Kolyma Yukaghir	C	
6.	‘guts’ – ‘end’	2-0	Abzakh, Adyghe, Kolyma Yukaghir	C	

table 21: areal lexico-semantic patterns of Eurasia

6.4.3.6. Areal patterns of Australia

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	‘moon’ – ‘snail’	2-0	Burarra, Gurindji	CL	
2.	‘nest’ – ‘raft’	2-0	Nunggubuyu, Tasmanian	CL	
3.	‘boat’ – ‘collamon’	2-0	Gurindji, Yir Yoront	CL	
4.	‘pen’ – ‘poke’	2-0	Gurindji, Yir Yoront	C	
5.	‘breast/milk’ – ‘Burton’s legless lizard’	2-0	Burarra, Nunggubuyu	CL	
6.	‘night’ – ‘sleep’	2-0	Ngaanyatjarra, Nunggubuyu	CL	a fine example of actual-potential polysemy

table 22: areal lexico-semantic patterns of Australia

6.4.3.7. *Areal patterns of New Guinea*

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'animal' – 'pig dog'	3-0	Kyaka, Sentani, Takia	C	See Wälchli (2005), Farr (2001: 124) on Korafe.
2.	'egg' – 'nut'	4-0	Dadibi, Kaluli, Kyaka, Takia	CL	See Brown (1977: 299) on Eleman.
3.	'table' – 'floor'	3-1	Baruya, Dadibi, Toaripi, (Hawai-ian)	M	
4.	'bone' – 'strong/strength'	– 3-1	Baruya, Kwoma, Waris, (Ngambay)	C	See Aikhenvald (2007: 549) on Manambu, which has been in contact with Kwoma

table 23: areal lexico-semantic patterns of New Guinea

Another New Guinea pattern mentioned in the literature for a particular language (Farr 2001: 126 on Arop-Lokep) is that of having terms for 'cloud' that can be literally translated as 'faeces of the wind.' Alongside Arop-Lokep, this pattern is found in another Austronesian language, Takia, that is spoken in close proximity. However, it is also found in Toaripi, spoken near the Southern Coast of New Guinea, and thus quite far removed from the other languages where it is known to occur. According to Brown (1977: 299), here the complex term replaces an older inherited Eastern Eleman word. Given the high idiosyncrasy of the pattern, it seems quite unlikely that the cases are independent of each other, and thus one would expect the association to be found in other New Guinea languages as well. Colexification of 'seed' and 'egg' is, as noted by Laycock (1970: 1141), indeed very common in the New Guinea area, but it also occurs in some other languages. Furthermore, there is an apparently widespread cultural association between the 'womb' and 'netbags' in New Guinea (see for instance Stewart and Strathern 1997). This is confirmed by the sample on the lexical level for two New Guinea languages, but the association is also present in Buin and Burarra.

6.4.3.8. *Areal patterns of Oceania*

There is just one association common enough to be diagnosed as areal according to the present methodology, namely that between 'star' and 'starfish,' which occurs by colexification in Buin, Lavukaleve, Rotokas, and Bislama.

6.4.3.9. Areal Patterns of Eastern North America

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'animal' – 'move on earth'	2-0	Blackfoot, Lakota	C	
2.	'straw' – 'hat'	2-0	Pawnee, Cheyenne	M	
3.	'house' – 'teepee'	2-0	Comanche, Kiowa	CL	
4.	'train' – 'run'	2-0	Biloxi, Chickasaw	C	
5.	'eyeball' – 'round object'	3-0	Blackfoot, Chickasaw, Oneida	C	
6.	'Saturday' – 'younger sibling'	2-0	Biloxi, Chickasaw	C	See Brown (1999: 26)

table 24: areal lexico-semantic patterns of Eastern North America

There are terms for 'Saturday' involving constituents meaning 'Sunday' and 'small, little' in Cheyenne and Kiowa, but taken by themselves, both 'Sunday' and 'small, little' recur as constituents in terms for 'Saturday' in other configurations (see Appendix E, 158).

6.4.3.10. Areal Patterns of Western North America

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'beak' – 'peck'	3-0	Upper Ineseño, Chehalis, Chumash, Nuuchahnulth	C	perhaps present in Hani as well
2.	'branch' – 'knot in tree/knot in wood'	4-1	Carrier, Upper Chehalis, Lake Miwok, Nez Perce, (Central Yup'ik)	CL	
3.	'dew' – 'wet/moist'	3-0	Upper Ineseño, Chehalis, Chumash, Kiliwa	M	
4.	'thorn' – 'sticker'	3-1	Cahuilla, Lake Miwok, Wintu, (Pawnee)	C	
5.	'car' – 'move self'	3-1	Carrier, Nez Perce, Kashaya, (Lakota)	C	see some remarks in Appendix E, 77

6.	'car' – 'machine'	4-1	Ineseño Chumash, CL Kashaya, Lake Miwok, Wappo, (Kildin Saami)	all terms in Western North America bor- rowed from Spanish
----	-------------------	-----	---	--

table 25: areal lexico-semantic patterns of Western North America

6.4.3.11. *Areal Patterns of Mesoamerica*

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'resin' – 'birdlime'	2-0	Itzaj, San Lucas Quiaviní Zapotec	CL	
2.	'rope' – 'lasso'	4-1	Itzaj, San Mateo del Mar Huave, Xicotepec de Juárez Totonac, Copainalá Zoque, (Cu- beo)	CL	

table 26: areal lexico-semantic patterns of Mesoamerica

The paucity of the results is surprising, since there is, as noted already above, a series of publications that demonstrate that Mesoamerica forms a linguistic area, also when it comes to the lexicon. The present sample confirms the association between 'ashes' and 'lime' noted by Smith-Stark (1994), which occurs here in Itzaj and Copainalá Zoque, but also in Tetun by the analyzable term *ahu-metan* 'lime-black.' As already noted earlier, some of the features mentioned in Campbell et al. (1996) and Smith-Stark (1994) do not emerge as being characteristic of Mesoamerica under the present method of assessing areality. The fact that a region of the world which has been studied in great detail from the point of view of areal phenomena in the lexicon yields a very small number of results when compared with the other areas suggests that there is a lot more interesting work to be done for other areas of the world in areal linguistics.

6.4.3.12. *Areal Patterns of South America*

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'beak' – 'prow of canoe'	5 (6)-1	Aguaruna, Bora, Guaraní, Huambisa, Piro, Wayampi, (Hawaiian)	M	Also common in Austronesian according to Blust (2009: 314), as suggested by the presence in Hawaiian
2.	'cloud' – 'smoke'	– 7 (8)-3	Arabela, Bora, Cashinahua, Cavineña, Cayapa, Hupda, Maxakalí, Tsafiki, (Buin, Sedang, Nez Perce)	M	
3.	'dust' – 'smoke'	5-1	Carib, Guaraní, Hupda, An-	M	

			cash Quechua, Tsafiki, (Fijian)
4.	'knife' – 'ma- chete'	7-3	Bora, Cavineña, Chayahuita, M Cubeco, Hupda, Jarawara, Yanomámi – Meyah, Sko, Basque
5.	'neck' – 'nape'	9 (10)-1	Abipón, Aguaruna, Arabela, M Aymara, Cubeco, Piro, Ancash Quechua, Tehuelche, Toba, Wayampi, (Hani)

table 27: areal lexico-semantic patterns of South America

There are at least two further conspicuous patterns that occur on a smaller scale in South America: one is to have a semantic extension of the word for 'beard' to also mean 'antennae of an insect.' This is found in the sample in three languages spoken on the eastern slopes of the Andes: Arabela, Cavineña, and Piro (as well as in Mesoamerica in Xicotepec de Juárez Totonac). Furthermore, South America is a hotspot for complex terms for 'mouth,' either as 'teeth hole,' 'speak hole' or 'language hole.' Such terms occur in Jarawara, Maxakalí and Tsafiki, in other words, without clear confinement on a smaller scale within South America. There are also a number of languages in South America with complex terms for 'dew,' either on the basis of 'urine' or 'saliva' (for the latter in the Vaupés area compare Epps 2007: 285, see also Zamponi 2009: 590 for data from Maipure), but ones on the basis of 'urine' also occur in Australia. Also common is a lexical association between 'stomach' and 'faeces,' either by polysemy, as for instance in Bororo, or by way of morphological complexity, as in Piro, where *hitška-mapa* contains *hitška* 'faeces' and *mapa* 'bag, bladder.'

6.4.3.13. Larger linguistic areas?

6.4.3.13.1. *Introduction.* As stated in the general introduction to the discussion of areality in semantic associations, on the basis of the present sample it is only possible to deal with quite large areas due to the insufficient coverage for small geographically restricted regions of the world (for instance, there is only one language, Greek, which participates in the Balkan Sprachbund). However, there is some evidence from evaluation of the areal clustering of semantic associations that even larger linguistic areas exist than those discussed so far, namely in those cases when there is a notable areal bias to the distribution of a given feature, which however fails to be diagnosed as potentially areal for any of the areas looked at so far, not because the pattern occurs with some frequency at random scattered elsewhere in the world, but because its representation is relatively strong also in areas geographically adjacent to that in question. This is the case, first, for Eurasia as a whole, including Europe, second, for Mainland Southeast Asia and Oceania broadly (including New Guinea), third, for the Old World, that is, Eurasia and Africa, and finally, for North America as a whole. The following discussion presents the relevant evidence that emerges when the methodology introduced in § 6.4.3.1. is applied to these macro-areas.

6.4.3.13.2. *Putative semantic associations characteristic of Eurasia as a whole*

	Lexico-Semantic Associations	Associa- tions	Number of	Lan- guages	Languages	Type	Comments
1.	'fog' – 'darkness'		2-0		Welsh, Khalkha	CL	
2.	'Saturday' – 'Saturn'		3-1		Badaga, Khalkha, Welsh, (Japa- nese)	M	
3.	'Saturday' – 'unique'		2-0		Basque, Abzakh Adyghe	C	

table 28: putative areal patterns of Eurasia as a whole

6.4.3.13.3. *Putative semantic associations characteristic of Mainland Southeast Asia and Oceania.*

	Lexico-Semantic Associations	Number of	Languages	Type	Comments
1.	'foam' – 'saliva'	9 (10) - 3	Baruya, Buin, Kwoma, Lavukaleve, Muna, Nga- anyatjarra, Sko, Tasmanian, Bislama, Lenakel, (Badaga, Lesser Antillean Creole French, Tsafiki)	M	
2.	'beard' – 'feather'	7 (8) - 3	Berik, Kwoma, Meyah, Toaripi, Bwe Karen, Hawaiian, Lenakel, (Efik, Chickasaw, Toba)	M	

table 29: putative areal patterns of Mainland Southeast Asia and Oceania

A common pattern in both Southeast Asia and Oceania is to have complex terms for 'sun' on the basis of 'eye' (Urban 2010b). Furthermore, many languages of this area, namely Angkor, Muna, Sahu, Sedang, and Bislama colexify 'bark' with 'husk' (this association is also heard of in Niger-Congo languages), and Kwoma, Toaripi, Hawaiian, and Samoan have a lexical association between 'kidney' and 'fruit,' while there is a semianalyzable term in Bwe Karen. Also common in the broader area of New Guinea and Oceania, but rare though not unheard of elsewhere, are associations between 'egg' and 'fruit' and 'egg' and 'seed' (for New Guinea, compare Laycock 1970: 1141, and Brown 1977: 299 for Eleman more specifically). Holmer (1966) draws attention to associations between 'place' and 'time' in Oceania more generally; two languages in the sample behave accordingly, Kyaka and Yir Yoront, but this is not enough to satisfy criterion (ii).

6.4.3.13.4. *Putative semantic associations characteristic of the Old World.*

	Lexico-Semantic Associations	Associa-	Number of Languages	Lan-	Languages	Type
1.	'house' – 'lineage'		5-1		Buli, Rendille, Abzakh Adyghe, Badaga, Basque, (Wintu)	CL
2.	'pen' – 'feather'		5-1		Efik, Khoekhoe, Basque, Nivkh, Kildin Saami, (Hawaiian)	CL

table 30: putative areal patterns of the Old World

Also common, though not common enough to satisfy criterion (ii) of the present methodology are the following associations: colexification of 'animal' with 'livestock,' 'sky' with 'god,' 'star' with 'blaze,' 'tail' with 'buttocks,' 'ball' with 'bullet,' 'clock' with 'bell,' 'street' with 'journey,' 'heart' with 'boldness, courage,' and that of 'neck' with 'neck of vessel;' see relevant sections in Appendix E.

6.4.3.13.5. *Putative semantic associations characteristic of North America*

	Lexico-Semantic Associations	Asso-	Number of Languages	Languages	Type
1.	'clearing' – 'field/meadow/lawn'	–	5-2	Carrier, Kashaya, Lake Miwok, Pawnee, Tuscarora, (Central Yup'ik, Badaga)	CL
2.	'mirror' – 'window'		6-1	Upper Chehalis, Kashaya, Kiowa, Pawnee, Tuscarora, Wintu, (Fijian)	CL
3.	'bed' – 'lie/lie down'		9 – 4 (8)	Carrier, Upper Chehalis, Cheyenne, Ineseño Chumash, Haida, Kiliwa, Lakota, Wintu, Yuki, (Efik, Kwoma, Jarawara, Fijian, Hawaiian, Malagasy, Rotuman, Tetun)	M

table 31: putative areal patterns of the North America

In some languages of North America, namely Carrier, Cheyenne, and Lakota, there is an association that likens the 'rainbow' to a 'snare' (see Hall 1997: 56). Notable are also the associations between thunder and a mythological thunderbird (compare Eells 1889: 335). This pattern is apparently most widespread in the Northwest, occurring in Carrier and Upper Chehalis (see also Sapir 1916/1949), but clearly extends further west and south (e.g. Chamberlain 1890 on Algonquian). Interestingly, the association is also found in Miskito, and, outside of the Americas, in Waris, spoken in New Guinea. Other perhaps notable patterns in North America are complex terms for 'sky' involving a constituent meaning 'blue' or colexification of these meanings, which is the case in Biloxi, Upper Chehalis, Cheyenne, Oneida, Tuscarora, but also in Hawaiian. Also, in three languages of North America, Lakota, Nuuchahnulth, and Pawnee, there is an association between 'paper' and 'cloth.'

6.4.3.13.6. *On the possibilities of the emergence of macro-areality in semantic associations.* How could these patterns, if they are real and can be substantiated, have arisen? Is there any plausible scenario emerging from the history of the relevant regions that could account for these distributions, however abstract it may be?

There is at least one of the associations identified as being characteristic of a large continent-sized area, Eurasia, for which the history is documented: that of 'Saturday' with 'Saturn,' and it is illuminating to discuss the emergence of this association as a case study of how large-scale areality in the lexicon can emerge. Eurasia is the host of a number of commonalities in the lexicon, and it is here that one can probably best trace at least some of its developments. The continent is a spread zone in terms of Nichols (1992), which is to a great extent due to its geography. Eurasia does not feature major geographical boundaries to west-east travelling and the geography thus provides a fostering environment for cultural and linguistic contact. One instance of these is the long-established west-east trade along the Silk Road (Beckwith 2009). Nichols and Peterson (2005) and Bickel and Nichols (2009) have pointed to linguistic outcomes of this situation. Consonantism patterns in pronouns involving a bilabial nasal in the first person pronoun root and an alveolar stop or affricate in the second person pronoun root, for instance, are hypothesized to be the result of precisely the aforementioned cultural continuities that link the whole continent since ancient times. At a later point of time, after the initial establishment of the east-west trading networks the origins of which go back to the Bronze Age, the conquest of Persia by Alexander the Great and his advance further east brought about a fusion of the Ancient Greek culture with that of the Middle East and Asia, renewing pan-Eurasian cultural ties. This is known as the Hellenistic period, commencing in the third century BC, and it is here that at least one of the lexico-semantic patterns characteristic of Eurasia has its origin, namely that of the association between 'Saturday' and the planet 'Saturn.' The practices of naming the days of the week after the seven planets (including sun and moon) visible with the bare eye is of Mesopotamian origin, and towards the end of the Hellenistic period, the system was adopted in the Hellenistic world, and names were given to the days of the week on the basis of the Greek names of the planets (the precise line of development is a little more complicated, see Cumont 1935 for details). Subsequently, the system established itself in Greek-controlled Egypt and in the Roman World by the first century BC (Sarton 1959), from where it in turn made its way to the vernaculars of Europe due to Roman dominance. It spread to the Indian subcontinent along with Hellenistic astronomy, and coexisted there with the indigenous Indian calendar based on lunar cycles (Markel 1995). Its presence in Japanese is of relatively recent origin, it being a loan translation from Chinese in the context of Sinicization that is first attested in Japanese in 1444; Mandarin as spoken today does not use the planetary-based model term anymore (Schmidt 2009).

Some space was devoted to this particular case in spite of the danger to drift into a *Wörter und Sachen*-style hybridization of linguistic and encyclopaedic cultural facts, because it allows one to catch a glimpse at the development of and the historical facts responsible for a particular feature common to the lexicon of languages spoken in a very large, continent-sized area. No similar historical account is available for the emergence of

other patterns, such as the apparently widespread Eurasian association between ‘rainbow’ and ‘thunder,’ and its origins therefore remain obscure, but their existence nevertheless demonstrate the long history of cultural and linguistic continuities over Eurasia that result in some similarities in linguistic structures over the entire continent.

As for potential still larger areas such as the Old World, discussion, unfortunately, must remain more speculative. Notably, though, Africa, in particular Northern Africa, has had historical ties with Eurasia throughout the historical period: first by way of the Ancient Egyptian empire, later, by Hellenistic influence in North Africa; still later, the spread of Arabic culture in medieval times to both Africa and Europe as well as colonization of the African continent by European powers may all have played a role in the shaping of individual commonalities in lexico-semantic associations. As for Southeast Asia and Oceania, a candidate for bringing about areality that immediately comes to mind is the Austronesian spread from Southeast Asia to Oceania, which is known to have resulted in prolonged language contact with notable effects on linguistic structure in New Guinea in particular. All this, including the putative large-scale associations in particular, requires empirical substantiation, but at least, from the point of view of history, the possibility of the emergence of such large-scale patterns does not seem to be ruled out entirely.

6.4.3.14. *A short note on negative evidence, concerning Eurasia*

While in general the investigation is concerned with positive rather than negative evidence for assessing areality, there is nevertheless evidence that Eurasia also forms a large linguistic area that is constituted jointly by the striking absence or rarity of some lexico-semantic patterns that are so common otherwise that their presence can be considered the norm rather than the exception. These include:

- (i) rarity of the association between ‘milk’ and ‘breast’ exceptions: Ket *mam-ul* ‘breast-water’ and Kolyma Yukaghir *ibiši* ‘milk, breast, nipple’
- (ii) absence of complex terms for ‘lip’ of the type ‘mouth-skin’

With respect to the association between ‘skin’ and ‘bark’ a similar west-east cline is observable: Eurasian languages which have the association tend to be spoken in the east rather than in the west. The westernmost representative in the sample is Abzakh Adyghe, followed by Ket, Kolyma Yukaghir and Japanese (in Chukchi, the current form *itqilyən* is lexicalized from **ut(tə)-qulyə(n)* ‘tree-skin’ and non-transparent synchronically).

6.4.3.15. *‘Things’*

This section takes up a topic alluded to in various places in the discussion so far. The phenomenon at stake is that instead of derived terms, in some languages there is a high number of analyzable terms of the lexical type, in particular in the domain of artifacts, involving a constituent simply meaning ‘thing.’ Prominent among these is Cheyenne, and two examples from this language are in (2.).

- (2). a. *šééšestôtse* /šééše-hestôtse/ 'lie-thing' = 'bed'
 b. *he'enénestôtse* /he'e-nén-hestôtse/
 'female-nurse-thing' = 'nipple'

Table 32 provides data on the occurrence of fully analyzable terms of the lexical and derived type involving a constituent glossed as 'thing' within the meanings under investigation, separately for the domain of artifacts, since this is where most such terms accumulate and terms in other domains. Figure 15 is a plot of the distribution of such terms in the entire (EXT-2) sample.

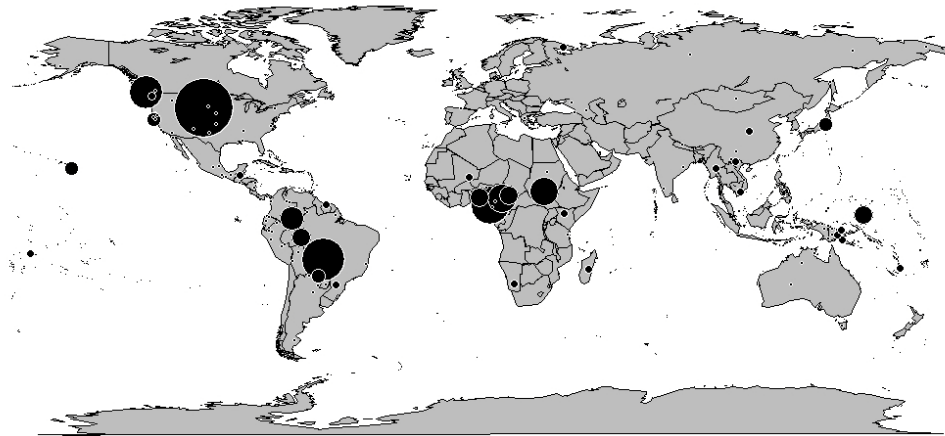


fig. 15: terms involving a constituent 'thing' in the full (EXT-2) sample

Language	'Thing'-terms in Artifacts	'Thing'-terms elsewhere
Efik	2	5
Katcha	2	3
Khoekhoe	0	1
Mbum	5	0
Ngambay	3	0
Rendille	0	1
Yoruba	1	2
Dadibi	1	0
Toaripi	1	0
Japanese	0	2
Kildin Saami	1	0
Cheyenne	9	2
Itzaj	0	1
Kashaya	2	0
Nuuchahnulth	2	4
Quileute	1	0
Bororo	4	4

Guaraní	1	0
Hupda	4	0
Jarawara	0	3
Kaingang	1	0
Lengua	2	0
Hawaiian	1	1
Kapingamarangi	1	2
Bwe Karen	0	1
Lenakel	0	1
Malagasy	0	1
Mandarin	0	1
Samoan	1	0
Takia	1	0
Vietnamese	0	1
Yay	0	1

table 32: terms involving a constituent meaning ‘thing’ in the sampled languages

Two words of caution are in order: the first pertains to the question as to how accurately the sampled meanings represent the situation in the entire lexicon. For instance, although not figuring prominently in the selection of vocabulary items on the wordlist and being sometimes redundant from a purely semantic point of view, compounds with *mar* ‘thing’ are also very frequent in Meyah, for instance *márféb* /*már eféb*/ ‘thing string’ = ‘string,’ *márfók* /*már ofók*/ ‘thing bud’ = ‘flower,’ *már éij* ‘thing throw’ = ‘trash’ (Gravelle 2004: 53; see also § 4.5.1.4.1). Second, table 32 does not take into account the number of available terms per language for the investigated meanings, but merely counts terms with the structure of interest.

Still, bearing in mind these points, the distribution does not appear to be entirely fortitious: notable is an areal bias in Africa and the Americas that is significant at $p = .01203$ when performing a Kruskal test using the standard six-way areal breakdown of the world ($\chi^2 = 14.6382$, $df = 5$, only using languages of the statistics sample to control for inheritance).

However, there is more to be said: ‘thing’-terms, both for artifacts and non-artifacts, are found with a high frequency in languages with a rather modest apparatus of nominal morphology. This is true of most African languages in the sample, but it is equally true of Macro-Gê languages in South America, Vietnamese in Southeast Asia, and of Oceanic, in particular Polynesian, languages of Oceania. One might be inclined to conjecture that in such languages forming complex terms on the basis of ‘thing’-words replaces a missing nominalizing morpheme (and note Moser’s 2004: 133 discussion of *né* ‘thing’ in Ngambay as a deverbal nominalizer as well as the dictionary gloss for Nuuchahnulth *-mis* ‘thing, used as a nominalizer’ and the same situation which obtains in Bwe Karen; terms involving these elements were taken into account in table 32), in other words, that ‘thing’-terms are circumlocutory transcategorial operations in languages that lack morphological means to do so (indeed, a grammaticalization cline ‘thing’ > ‘nominalizer’ is perfectly con-

ceivable, although not mentioned by Heine and Kuteva 2002). Notably, though, languages in which ‘thing’-compounds accumulate do not necessarily lack noun-deriving morphology, in particular a general nominalizer, altogether, so this cannot be the whole story. Some languages with many ‘thing’-terms typically lack dedicated derivational morphemes for instrument and locative nominalization (or employ these so rarely that they do not figure in the database), but this generalization is far from going the whole way as well: for instance, Kapingamarangi and Malagasy feature prefixes *k-* and *faN-* respectively for forming instrument nouns, and there are other languages for which this explanation is not available either.

However, there is some evidence, although limited, that one is dealing, at least in non-Andean South America, with an areal phenomenon. In this regard, Jarawara *jama* ‘thing’ deserves special mention. It can and apparently very frequently does refer to any object when the context allows identifying the referent unambiguously. Dixon (2004: 540) states that

[*jama* is the generic term par excellence. The normal gloss offered is ‘thing.’ It can be used for any new object or foodstuff, for which a name is not known, or simply as a vague term instead of employing a more specific name. *Jama* can be used for ‘season’ or ‘time’ ... *Jama* can be used to refer to the forest (a more specific name is *jama.kabani*; no etymology is known for *kabani*) or to game in the forest, or to the spirits of the forest. *Jama* can also be used to refer to fishes in the river. A PN [possessed noun] is generally used with a free noun; *jama* is often the ‘dummy head’ with a PN. For instance, free noun X plus PN *abe/ebene* is ‘living being associated with X’, e.g. the people inhabiting a place called X. ... Insects in general can be described as *jama abe*, literally ‘creatures associated with a thing’, using *jama* in its most general sense. ... The wide range of meaning and use of *jama* is exhibited in one sentence from a text which has *jama* as its A argument and also *jama* as its O argument (but with different reference): *jama jama firi kasa* ‘the thing (here, lightning flash) fully illuminates the thing (here, a dead body)’ ...

Alongside being capable of referring to the weather, especially salient seems to be the use of *jama* to refer to a specific time or place:

- (3.) a. *jama hiwa-bote ama-ke*¹⁶
 thing(f) be.hot-VERY be-DECf
 ‘the weather (lit. thing) is very hot’ (adapted from Dixon 2004: 337)

¹⁶ Glosses: 1EXC ‘1st person non-singular exclusive (excluding addressee),’ A ‘transitive subject,’ AUX ‘auxiliary,’ CINT ‘content interrogative,’ DEC ‘declarative mood,’ f ‘feminine,’ IMMED ‘immediate mood,’ NOM ‘nominalization,’ PERI ‘peripheral,’ s ‘intransitive subject,’ sg ‘singular.’

- b. *himata jama jaa ti-ka-ma-ri-be?*
 what thing(f) PERI 2sgS-in.motion-BACK-CINTF-IMMEDf
 ‘when (lit: at what thing) will you return?’
 (adapted from Dixon 2004: 409)
- c. *jama jabo-ke faha otaa kii ni kaaro*
 thing(f) be.far-DECf waterf 1exCA search.in AUX+NOM PERIf
 ‘the place where we fish is far off (lit. the thing (place) is far off, we
 search the water at it’
 (adapted from Dixon 2004: 500)

Jama is also very frequently employed for word-formation tasks, and the lexemes formed with it appear to have varying degree of conventionalization. Instrument nouns are frequently formed by combining *jama* with a reduplicated form of a verb to denote items of acculturation, such as *jama ho-howe* ‘thing RED-clean.out’ = ‘rake’ (adapted from Dixon 2004: 534); none of them is found in the database, though. In this function, *jama*, given its extremely vague semantic content, approaches a nominalizer in function. *Jama* is equally readily combined with another noun for the purpose of word formation, as seen e.g. in *jama soki* ‘thing be.dark’ = ‘night’ and, conversely, *jama wehe* ‘thing light’ = ‘day.’

Interestingly, in Bororo, a similar, although not identical situation is encountered. Bororo *boe* is prominently used as the autonym with which the Bororo refer to themselves:

- (4.) *Boe e-tu-re*
 Bororo 3PL-go-NEUTRAL
 ‘The Bororos left.’
 (Crowell 1979: 227)

However, as Crowell (1979: 226) remarks, the term, like Jarawara *jama*, “occurs with great frequency, along a scale of specificity.” *Boe* can also be used to refer to other Indians or people, or it can (or must) be translated by ‘thing.’ Parallel to the range of use of *jama* in Jarawara, *boe* is also used when talking about the weather or about time:

- (5.) a. *Boe uru-re*
 thing hot-NEUTRAL
 ‘It’s hot’
- b. *Boe xo-re*
 thing black-NEUTRAL
 ‘It’s dark (or night).’
 (Crowell 1979: 226, glosses adapted)

While there may be an areal factor in play (note also the parallelism with respect to the structure of the expressions for ‘night’ in Bororo and Jarawara), there also is an alternative explanation: high frequency of a semantically underspecified noun, presupposing a high amount of implicitly shared cultural and real-world knowledge, may be a symptom of languages used primarily for intra-group communication (Thurston 1989, Wray and Grace

2005). As Dixon (2010: 301) notes, “[i]n contrast, there are languages which lack anything resembling a generic noun ‘thing’.” In fact, the behavior of languages with respect to this seems to be not unconstrained, but might rather be accountable for by sociolinguistic factors.

6.4.3.16. *General Discussion.*

A clear division between facts and historical explanation is important for the context of the present study. The question that arises now after surveying the patterns is whether all of them necessarily must have a historical explanation. In spite of the fact that with a broad geography-based approach to areal linguistics, one should not necessarily expect isoglosses as in geographically more restricted areas or even dialect geography (Bickel and Nichols in press), some of the patterns found in continent or subcontinent-sized areas, for instance North America, have a very discontinuous geographical distribution while the overall frequency is not necessarily very high (though, by definition, higher than 15%). While it is not impossible that they indeed are indicative of historically grown areality that extends across the continent, this need not necessarily be so. This is partly due to the sparseness of historical data for some areas of the world that would allow coming up with more specific contact scenarios and due to the lack of research of areality in the lexicon that is concerned specifically with large areas. However, this statement should not distract from the fact that some of the diagnosed patterns may simply be spurious. In connection with this, it is notable that there are also a number of terms for items of acculturation, in particular in the Americas, that are diagnosed as being areal. Rather than being an indication for continent-wide diffusion, these are by far more likely to be relatively uniform responses to items of acculturation that were previously unknown (however, some diffusion of semantic associations in historical times seems to have happened. For instance, Siouan and Algonquian languages feature terms for ‘distilled spirits’ literally translatable as ‘fire-water’ according to Rankin 2003: 193, a pattern they share with Dene Słłłłł as reported by Rice to appear. Similarly, languages of the Southeastern United States feature terms translatable as ‘bitter-water’ for ‘whiskey’ according to Brown 1999: 146, table 11.1.). Obviously, then, these are of little to no value for an areal linguistics that seeks to identify diffused structural traits in neighboring languages due to language contact. However, they are still amenable to a somewhat less interesting, because obvious, historical explanation in that the artifacts they designate, such as ‘pens’ or ‘tables,’ were previously unknown in the areas where the morphologically complex terms for them occur.

A general observation is that there is some variation between areas whether the found patterns are more frequently due to colexification or due to parallelism in morphological structure. While, for instance, the patterns found in the Old World are characteristically due to colexification (this is not an absolute statement but a generalization only: the areal patterns found in Europe, to the contrary, are entirely due to calquing), those in the Americas tend to be found in parallelisms in morphologically complex lexical items – a finding that is in line with the general distribution of morphological complexity on a macro-areal scale: comparably low degree of analyzability in the Old World, comparably high degree in the New World. All in all, there are 21 patterns of colexification and 17

patterns in morphologically complex terms that are diagnosed as being associated with one of the tested areas under the present method. In other words, if the diagnosed patterns are genuine, the results show that areal influence and convergence in the semantics of lexical items alone, without replication of morphological structure, is roughly equally frequent as calquing in the traditional sense, and thus seems to be of value for research in linguistic areality and language contact generally.

6.5. CROSS-LINGUISTICALLY COMMON LEXICO-SEMANTIC ASSOCIATIONS

One value of the discussion of individual concepts in Appendix E noted by Blank (2003) is that such a list of cross-linguistic associations allows one to predict to some extent that, should a neologism be coined in some language of the world, speakers are likely to choose one of the conceptualization strategies listed there. However, another value in the opposite direction that has been noted early on in the literature as discussed in chapter 2 is that, should terms be etymologizable, they are also likely to exhibit one of the lexico-semantic associations found synchronically in other languages, and in this sense, the issue is intricately linked with questions of diachrony.

Having discussed structures that are probably areal in their distribution, one is left with two types of distributions in lexico-semantic associations: those that are very rare cross-linguistically, occurring for instance in only two or three languages in the sample without any appreciable hotspot anywhere, and those that are so common that they are encountered in very many parts of the world in very many languages. The purpose of this section is to present the latter. The data for the entire (EXT-2) sample are used for this purpose, on the one hand to allow for a maximally inclusive dataset for this explorative investigation, on the other since, as discussed in § 6.4.2., there is little (though not no) evidence for inheritance and genealogical stability of most patterns anyway, a fact which justifies this decision.

However, before it is possible to do that, one needs to decide how to sort out the areally unrestricted rare patterns from the areally unrestricted common patterns. This may seem like an odd question to discuss explicitly, but the two types of data are not separated from one another by some sort of intrinsic criterion, but form a continuum. As with the very first typology in § 4.2. pertaining to the relative prevalence of analyzable terms of the lexical and derived type, the relative frequency of each association (both by morphological analyzability and colexification) is calculated, and then divided by the representation score for the meaning it pertains to in order to account for the variable recoverability of relevant terms from the consulted sources. All in all, there are 1,892 lexico-semantic associations (few of them are doublets since both associated meanings are on the 160-items list used). Figure 16 is a histogram showing the frequency distribution.

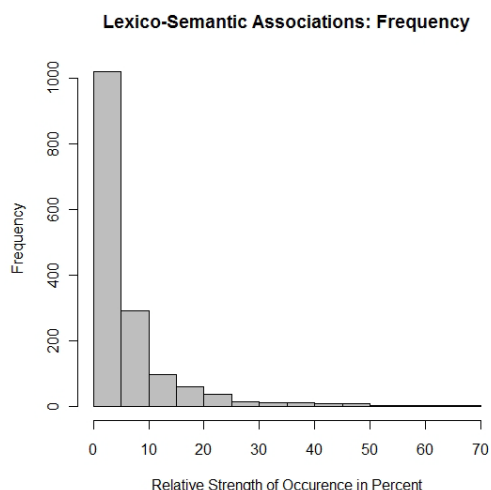


fig. 16: relative strength of occurrence in percent for recurrent lexico-semantic associations

The distribution is approximately negatively exponential: the vast majority of lexico-semantic associations is relatively rare, and there are only very few extremely frequent associations. The most common is that between ‘eyeball’ and ‘eye,’ found in 94.34 per cent of sampled languages for which data are available, mostly because languages have analyzable terms for ‘eyeball’ featuring a wide variety of mostly metaphor-based denominations for the concept which is contiguity-anchored by ‘eye’ (see Appendix E, 129 for details). One of the least frequent association is that between ‘moon’ and certain types of snails, found only in two languages of Australia, a pattern that may in fact be an areal association peculiar to Australia (there are very many associations present in only two languages; this one comes out as one of the least frequent since lexical equivalents for ‘moon’ are available for almost all sampled languages, and hence its relative strength is lower than if a pattern is attested twice, but for a meaning for which few terms could be extracted; note also that an occurrence in two languages is the lowest possible figure here, since all associations occurring only once are discarded according to the methodology described in § 3.7.3.2.).

Also parallel to the discussion of analyzable terms of the lexical and derived type in § 4.2., a division of the associations into four quartiles according to their frequency may be carried out, with everything in the fourth quartile defined (arbitrarily, but at least with a boundary that is generated out of the frequency counts themselves) as being a common association. This comprises all associations with a strength ranging from 7.14 per cent to the maximum (the abovementioned association between ‘eyeball’ and ‘eye’). Of course, this means that there is still huge variation within this group, but there also is the advantage that for an exploratory investigation, the likelihood of missing a pattern that may in fact be more common than suggested by the present sample is greatly reduced. In absolute numbers, 485 associations are found within this range.

In the present sample, the most common associations are contiguity-based. Alongside the already mentioned association between ‘eyeball’ and ‘eye,’ it is also very common cross-linguistically to have analyzable terms for ‘nostril’ having a constituent ‘nose’ acting as the contiguity anchor (with little variation as to the meaning of the second constituent, which is most commonly ‘hole, opening,’ see Appendix E, 129). Complex terms for ‘eyelid,’ ‘eyelash’ and ‘pupil’ featuring, like those for ‘eyeball,’ a constituent meaning ‘eye’ are also very common. As for colexification, very commonly (each figuring among the 20 most common associations), ‘moon’ and ‘month,’ ‘tree’ and ‘wood,’ ‘milk’ and ‘breast,’ ‘soil’ and ‘land, ground’ generally, ‘paper’ and ‘letter’ or ‘book,’ and ‘rope’ and ‘thread, string, twine,’ or ‘fibre’ are colexified. Rather than listing each association here separately, Appendix D features a column in which common associations for each meaning as just defined (as well as putative areal patterns), should there be any, are listed as a resource for further research.

In fact, of the 476 common associations for which no areal hotspot was discernible according to the methodology used in § 6.4.3, roughly two thirds are mostly contiguity-based, and only about one third mostly metaphor-based, while for some of them several analyses are possible. From these figures alone, however, one cannot yet infer any reliable generalizations about the way languages conceptualize referents as revealed by motivated terms. This on the one hand due to the fact that many of the contiguity-based associations that are frequent are in fact only contiguity anchors for varying metaphorical transfers, each of which taken by itself is usually too infrequent to end up in the fourth quartile (terms for the ‘eyeball’ are an exception). On the other hand, even without taking into account this type of association, contiguity-based associations are more numerous among those that recur frequently in different languages (note that metonymy has come to be regarded by some Cognitive Linguists as the more important cognitive process when compared with metaphor, e.g. Taylor 2003: 126). Given this result, it is instructive to look in more detail at the frequent metaphor-based associations to see whether one can derive any generalizations as to the kinds of meanings for which they are common.

The following is a rough division for the sake of bringing order into the wealth of different associations (it is intended to only give a rough classification, and assignment of individual patterns may be debatable; also, not listed are associations for which several analyses are possible, for instance that between ‘heart’ and ‘center, middle’):

- (i) Analogies (biologically speaking, often homologies) in humans and plants or animals: ‘bark’ – ‘skin’ (59.42%), ‘fingernail’ – ‘claw/talon’ (48.94%), ‘bark’ – ‘peel/rind/shell’ (42.75%), ‘feather’ – ‘hair’ (35%), ‘beak’ – ‘nose’ (31.19%), ‘breast’ – ‘udder/teat’ (30.34%), ‘feather’ – ‘fur/wool’ (29.29%), ‘bark’ – ‘hide/leather’ (23.91%), ‘skin’ – ‘rind/peel’ (23.61%), ‘breast’ – ‘nipple/teat’ (23.45%), ‘beak’ – ‘mouth’ (21.1%), ‘branch’ – ‘arm’ (20.71%), ‘nest’ – ‘house/home’ (19.08%), ‘skin’ – ‘shell’ (18.75%), ‘eyelid’ – ‘peel/rind/shell’ (17.14%), ‘eyelid’ – ‘bark’ (15.24%), ‘beak’ – ‘snout/muzzle’ (14.68%), ‘fingernail’ – ‘hoof’ (14.18%), ‘beard’ – ‘whiskers’ (13.97%), ‘eyelash’ – ‘fur’ (13.18%), ‘eyelash’ – ‘feather’ (12.4%), ‘eyebrow’ – ‘fur’ (11.03%), ‘nipple’ – ‘udder/teat’

- (10.75%), 'branch' – 'hand' (10.71%), 'eyelid' – 'leather/hide' (10.48%), 'beak' – 'tooth' (9.17%), 'tendon' – 'root' (9.01%), 'branch' – 'leg' (7.86%), 'feather' – 'leaf' (7.86%), 'skin' – 'husk/chaff' (7.64%), 'bark' – 'scale' (7.25%)
- (ii) Body parts and body liquids: 'tendon' – 'vein/artery' (49.55%), 'fingernail' – 'toenail' (27.66%), 'finger' – 'toe' (26.06%), 'phlegm' – 'saliva/spittle' (24.32%), 'vein' – 'nerve' (20.17%), 'snot' – 'phlegm/sputum' (15.57%), 'brain' – 'marrow' (13.53%), 'phlegm' – 'snot' (13.51%), 'nipple' – 'eye' (10.53%), 'uvula' – 'child/son' (10%), 'uvula' – 'vagina/clitoris' (10%), 'nipple' – 'head' (9.47%), 'calf' – 'belly/stomach' (9.4%), 'ankle' – 'eye' (8.59%), 'eyebrow' – 'feather' (8.09%), 'semen' – 'pus' (7.58%), 'uvula' – 'tonsil' (7.5%)
- (iii) Aerosols: 'steam' – 'smoke' (31.73%), 'fog' – 'cloud' (29.77%), 'steam' – 'fog/mist' (15.38%), 'fog' – 'smoke' (14.5%), 'cloud' – 'smoke' (7.86%)
- (iv) Artifacts: 'ladder' – 'stairs/staircase' (26.09%), 'window' – 'door' (20.95%), 'glasses' – 'mirror' (18.6%), 'airplane' – 'boat/canoe' (15.63%)
- (vi) Mythology-/culture-based associations: 'sky' – 'heaven' (30.77%), 'shadow' – 'soul/spirit/ghost' (27.74%), 'shadow' – 'reflection/mirror/image' (24.82%), 'heart' – 'feel/think' (18.06%), 'shadow' – 'image/picture/drawing' (20.44%), 'thunder' – 'god/spirit' (9.63%), 'heart' – 'soul/spirit' (8.33%)
- (vii) Generalizing/narrowing: 'rope' – 'thread/string/cord/twine' (49.61%), 'mountain' – 'hill' (42.54%), 'coast' – 'edge/end/border' (35.63%), 'train' – 'wagon/vehicle' (35%), 'buttocks' – 'bottom/base' (23.39%), 'resin' – 'water/liquid/juice' (21.65%), 'mouth' – 'opening' (20.83%), 'ashes' – 'dust' (19.72%), 'dust' – 'dirt/rubbish/garbage' (19.12%), 'semen' – 'water/juice' (18.18%), 'chair' – 'furniture' (17.6%), 'valley' – 'gully/furrow/ditch/gorge/channel' (17.54%), 'lake' – 'lagoon' (17.36%), 'skin' – 'surface/cover' (15.97%), 'lip' – 'edge' (15.83%), 'bed' – 'furniture' (15.38%), 'table' – 'furniture' (14.91%), 'estuary' – 'opening' (12.77%), 'window' – 'hole/opening' (12.38%), 'beak' – 'end/point' (11.93%), 'mouth' – 'edge/tip' (10.42%), 'car' – 'cart/carriage' (10.2%), 'bark' – 'cover' (9.42%), 'tail' – 'end' (8.51%)
- (v) "Bold" metaphors: 'Milky Way' – 'Trail/Road/Street' (37.5%), 'eyeball' – 'seed' (28.3%), 'estuary' – 'mouth' (25.53%), 'horizon' – 'edge/border/fringe' (24.39%), 'flame' – 'tongue' (22.02%), 'eyeball' – 'egg' (20.75%), 'sunrise' – 'come out/go out/merge' (15.58%), 'headland' – 'nose' (15.38%), 'eclipse' – 'die/kill' (15%), 'egg' – 'testicle' (13.29%), 'bay' – 'corner' (12.77%), 'pupil' – 'child/son/daughter' (12.68%), 'semen' – 'seed' (12.12%), 'eyeball' – 'child' (11.32%), 'pupil' – 'seed/grain' (11.27%), 'testicle' – 'seed' (10.48%), 'headland'

- ‘head/forehead’ (10.26%), ‘spring’ – ‘eye’ (10.08%), ‘eclipse’ – ‘eat’ (10%), ‘uvula’ – ‘child/son’ (10%), ‘horizon’ – ‘end/finish’ (9.76%), ‘sunset’ – ‘sink’ (9.76%), ‘sunset’ – ‘enter’ (9.76%), ‘sunset’ – ‘fall/drop’ (9.76%), ‘womb’ – ‘house’ (9.57%), ‘eyeball’ – ‘fruit’ (9.43%), ‘vein’ – ‘way/street’ (9.24%), ‘sunset’ – ‘descend/go down/lower’ (8.54%), ‘meteoroid’ – ‘faeces’ (8.45%), ‘meteoroid’ – ‘tail’ (8.45%), ‘meteoroid’ – ‘fall’ (8.45%), ‘meteoroid’ – ‘fly’ (8.45%), ‘street’ – ‘manner/method/system’ (8.4%), ‘airplane’ – ‘bird’ (8.33%), ‘rainbow’ – ‘snake’ (8.27%), ‘sunrise’ – ‘appear’ (7.79%), ‘mouth’ – ‘door/entrance’ (7.64%), ‘eyelid’ – ‘lid/cover’ (7.62%), ‘eyeball’ – ‘grain’ (7.55%), ‘rainbow’ – ‘bow/arc/bend’ (7.55%), ‘spark’ – ‘lightning’ (7.5%), ‘whirlpool’ – ‘whirlwind’ (7.41%), ‘horizon’ – ‘meet/meeting’ (7.32%), ‘horizon’ – ‘basis’ (7.32%), ‘Milky Way’ – ‘river’ (7.14%)
- (ix) Other: ‘lagoon’ – ‘lake/pond’ (44.64%), ‘moon’ – ‘sun’ (13.51%), ‘shadow’ – ‘photograph’ (10.95%), ‘puddle’ – ‘swamp’ (10.29%), ‘mirror’ – ‘shadow’ (8.77%), ‘wrinkle’ – ‘crease/fold/pleat’ (8.45%), ‘flood’ – ‘torrent’ (8.33%), ‘dust’ – ‘sand’ (8.09%), ‘lightning’ – ‘gleam/lighten’ (7.97%)

Thus, common metaphorical transfers, few as they are when compared with common associations by contiguity, mostly either have an additional component of biological analogy (i),¹⁷ or are transfers that take place within the same broader semantic domain as in groups (ii), (iii), (iv), or likely have mythological or cultural underpinnings (vi). As for the associations in (vii), one could either say that they generalize the reference of a concrete referent to more abstract referents (which would be an account in terms of *grundbedeutung* in line with the well-known account of metaphor as making accessible more abstract domains by way of more concrete ones), or one could say that general terms may be narrowed down to more concrete referents (an account in terms of *gesamtbedeutung*).

These are, of course, therefore not any less metaphorical than those metaphors that cross domains, but are likely to be, from a cognitive point of view, preferred because they may be easier to process and there is no danger of communicative failure given the anchoring within the same domain (which also accounts for the fact that many of these, unlike those in group (v), occur without a contiguity anchor). The human body is likely a coherent semantic domain in psychological terms: Neely (1977) finds priming effects in a lexical decision task for parts of the body with *body* itself as prime using a short SOA (stimulus-onset-asymmetry, that is, the time elapsing between presentation of the prime and the target) even when subjects had been told before to expect mostly targets unre-

¹⁷ Note that within this group, there are many overlaps. For instance, the association between ‘eyelid’ and ‘leather, hide’ is due to complex terms for ‘eyelid’ one of the constituents of which means ‘skin,’ but also ‘leather, hide’ (compare Appendix E, 113). Thus, in effect, the number of associations in this group may be conceived of as being smaller than reported above.

lated to the prime,¹⁸ and priming effects are also observable for the domains of artifacts as well as living things (Moss et al. 1995, McRae and Boisvert 1998).

Camac and Glucksberg (1984), furthermore, report that there are no priming effects for pairs of words metaphorically related to each other when compared with random pairings, while there are such effects for associatively related words when compared with random pairings. Their conclusion from this is that metaphor does not operate by exploiting already existing associations, but rather is a means to create new ones.

In the light of this the evidence for common associations by metaphor can be reviewed: how would one, apart from colexification with 'star' by configurational contiguity, call the 'Milky Way' if not by some metaphorical transfer? Most languages do not feature a general term for longish accumulations of smaller entities. What would a contiguity-like conceptualization look like for meanings like 'eyeball,' apart from simple colexification with 'eye'? Most languages do not have abstract general terms for 'small round object' that does not at the same time also denote a particular small object (though Oneida and Rotokas, for instance, do). And if 'round object' is available, it is likely to have come into being through gradual extension of terms for referents which in fact are round objects (such as 'seed,' prototypically) to further referents of similar shape and size, until the term is so bleached that it does not make sense to provide an extensional definition anymore (as is evidenced by the lexical sources for classifiers with this meaning).

Brown (1999: 50-51) provides an account for generalizations as to the common semantic associations revealed in terms for items of acculturation in languages of Native America summarized in Brown (1999: 45, table 4.1.; there are both items that would be called metonymy- and metaphor-driven in the present framework on this list) by proposing a so-called "rich cognition model" sharing some aspects with Chomsky (1975), who proposed that there are innate information processing mechanisms forming a part of the language faculty:

[W]hen different human groups are faced with the problem of giving a name to the same newly encountered object or concept, information-processing mechanisms shared by all humans are utilized to accomplish the task. This involves, at least in part, analysis of both the item to be named and of sources from which a label for the item might be retrieved. Since information-processing devices are panhuman, similar, if not the same, analyses will tend to be made, resulting in similar names for the item in question especially if sources for labels are similar.

Unfortunately, this account is only informative insofar as one presupposes knowledge about what these (not unreasonably) alleged panhuman "information-processing mechanisms" are like qualitatively. At least for the common metaphor-based denominations just mentioned, then, no grand yet unspecified theories on processing mechanisms à la Brown

¹⁸ However, in this particular experimental setting, with longer SOA, the obtained result is the other way around, presumably because subjects require time to override the "hard-wired" associations triggered between the prime 'body' and its category members (see also Neely 1991: 285 for review).

are needed. Such theories are surely interesting, but if an explanation on a lower, less spectacular, level that takes into account established findings from psychology is available, it is to be preferred by Ockam's razor (and in fact, quite similar remarks to that in the above paragraph are found in Brown and Witkowski 1981: 606-607).

In spite of the perhaps limited contribution (typological) linguistics can make to issues concerning the workings of the human mind (see Sandra 1998), a tentative generalization would be that WHEN CONTIGUITY-BASED ASSOCIATIONS ARE READILY AVAILABLE, THEY ARE CROSS-LINGUISTICALLY PREFERRED (though "cross-linguistically preferred" does not mean that individual languages cannot not have a metaphor-driven term instead: they can). Further, when there are few possible associations by contiguity, either because there are few contiguously related entities in the real world, or because lexical counterparts for these are lacking, and IF METAPHOR HAS TO BE RESORTED TO, A WITHIN-DOMAIN TRANSFER IS PREFERRED, and only if this is unavailable as well, "bold" domain-transgressing metaphors are resorted to. This would, for instance, explain why the association between 'vein' and 'tendon,' operating within the same semantic domain, is much more frequent than that of either with 'root,' for which a crossing of domain boundaries is necessary. This, as alluded to above, is fully in line with the prevailing view of metaphor as making available less palpable semantic domains by way of more basic and familiar ones. As noted in § 3.6.2.5., however, just what a semantic domain actually is is not as straightforward as it may seem. The above discussion should not least for this reason be read as an assembly of informal thoughts on the results rather than as a full-blown theory deriving from them.

6.6. CHAPTER SUMMARY

This chapter dealt with the semantic side of lexical motivation. It provided summarizing accounts of associations within certain semantic fields, as well as discussion of common extensions of the body-part terms 'eye,' 'mouth,' 'faeces,' and also kinship terminology, to other not closely related referents. The chapter also sought for sources of non-random variation, namely in associations that can be explained by the make-up of the environment in certain regions of the world, and by examining briefly whether there is a genealogical signal within language families in semantic associations, which could not be demonstrated for at least one of the language families examined. Moreover, the data were evaluated (using a preliminary methodology) with regard to areal clusterings. This suggested that areality in lexico-semantic associations, in particular pattern of colexification, is more pervasive cross-linguistically than currently acknowledged in the relevant literature, and that they provide a huge repository for areal linguistics that is presently only scarcely exploited.

Finally, brief evaluation of cross-linguistically common associations shows that common associations are more frequently contiguity- than metaphor-driven, which can be construed as a piece of evidence for the primacy of contiguity in language.

In general, what a comparison with Buck (1949) shows is that many of the cross-linguistically attested semantic associations are mirrored in individual languages of one particular family (compare cross-references in Appendix E). Still, it becomes clear that to

really be able to make solid inferences about likely associations from a cross-linguistic point of view, it is not enough to just investigate one particular family and that doing so may lead to a distorted picture of the possible cross-linguistic variation, as for instance the absence of lexical associations of 'milk' and 'breast' and 'tear' and 'water,' which are both extremely frequent world-wide, in Indo-European show (compare relevant data in Buck 1949).

