

Journal of Universal Language 16-2  
September 2015, 63-93

# On Colexification among Basic Vocabulary

**Vladimir Pericliev**

*Institute of Mathematics and Informatics, Bulgaria*

## Abstract

The paper presents a typological database of colexifications among basic vocabulary, derived from the Automated Similarity Judgment Program (ASJP) database. Some uses of the inventory of colexifications are proposed. Some heuristics are introduced regarding the discrimination of polysemy from homonymy in a typological database, as well as such pertaining to the determination of common membership of two languages in the same language family. In particular, it was found that shared colexifications corroborate the postulation of the Austric stock and the attribution of Sumerian to the Tibeto-Burman language family.

**Keywords:** colexification, basic vocabulary, Austric, Sumerian and Tibeto-Burman

---

Vladimir Pericliev  
Institute of Mathematics and Informatics, bl. 8 1113 Sofia, Bulgaria  
Phone: +3592-9792877; Email: peri@math.bas.bg

Received August 11, 2015; Revised September 14, 2015; Accepted September 22, 2015

## 1. Introduction

We speak of *colexification* (cf. e.g., François 2008) when two distinct word meanings are expressed in one language by a single lexeme (i.e., are *colexified*). In this paper, we shall present a large scale empirical study of colexifications among basic vocabulary meanings in the world languages, based on the Automated Similarity Judgment Program (ASJP) database (Wichmann et al. 2013) covering (parts of) the basic vocabularies for well over half of the world's languages.

Colexification, so defined, is a cover term for both polysemy and homonymy, two well-known linguistic phenomena, and it is no surprise that different aspects of these phenomena have received much attention in recent linguistic work. Some of the basic problems addressed in these efforts include: polysemy and semantic change, cataloguing semantic changes in specific languages or cross-linguistically, cognitive versus culture-specific semantic changes, extensions in specific semantic domains in specific languages or cross-linguistically (body parts, kinship terms, and so on), semantic reconstruction, incorporating polysemy in formal grammar, differentiating polysemy from homonymy (cf. e.g., the collection of papers in Vanhove 2008, Falkum & Vicente (eds.) 2015; also Zalizniak 2008, Urban 2012, Zalizniak et al. 2012, List et al. 2014).

The focus of the present long-term project is the computational construction and subsequent machine investigation of a large-scale typological database of colexifications of basic vocabulary meanings. In this paper we can only briefly sketch some aspects and more interesting results. There are some related efforts in the literature. E.g., Zalizniak (2008) and Zalizniak et al. (2012) describe cross-linguistically recurring semantic shifts (=polysemies) in 319 of the

languages of the world and manually implement a catalogue in the form of a manually searchable computer database, while List et al. (2014) have automatically built an on-line database of colexifications, based on the Intercontinental Dictionary Series (IDS, Key & Comrie 2007) featuring lexical data for 233 world languages. In the first two papers, however, the authors concentrate on polysemy (relevant for semantic change) rather than more generally on colexification and use no automation either for database construction or subsequent search. List et al. (2014) have indeed computationally built their database of colexifications and their networking but they are not concerned with offering more sophisticated techniques for searching the database. In neither of these works is basic vocabulary the specific focus, and in neither of these is the empirical base so large as that of the Automated Similarity Judgment Program (ASJP) database (see below).

The paper is organized as follows. Section 2 briefly reviews our empirical base and programs. Section 3 summarizes some basic results from the automatically constructed database of colexifications and Section 4 shows some ways to use it. In particular, we propose heuristics (or rules of the thumb) for distinguishing between polysemy and homonymy in a typological database and such for assignment of languages with unknown (or suspect) genetic classification to the same linguistic grouping. Section 5 concludes.

## 2. The Data and the Program

We explore for colexifications the Automated Similarity Judgment Program (ASJP) database (Wichmann et al. 2013) including the final 100-item list of basic meanings of Morris Swadesh (1971). In our

study we included all listed languages except creoles, pidgins, mixed and constructed languages, amounting to some odd 6809 languages, included in 244 top-level language families, according to the *Ethnologue* classification. The wordlists are not exhaustive for all languages. Forty basic meanings out of the 100-item lists are very representative and include well over half of all known languages, the other sixty are less exhaustive. Below we give the occurrences of each item in the 100-item list in the database, as this allows one to evaluate the actual empirical base for deriving each colexified meaning pair (see Appendix).

Table 1. The ASJP Wordlist and the Occurrences of Each Item in the Database

1	I	5308	22	louse	5463	43	tooth	6455
2	you	5164	23	tree	5772	44	tongue	6377
3	we	5181	24	seed	262	45	claw	262
4	this	302	25	leaf	5899	46	foot	303
5	that	298	26	root	317	47	knee	5027
6	who	310	27	bark	251	48	hand	5769
7	what	313	28	skin	6162	49	belly	299
8	not	290	29	flesh	303	50	neck	312
9	all	310	30	blood	6417	51	breasts	5840
10	many	286	31	bone	6298	52	heart	281
11	one	6313	32	grease	267	53	liver	5209
12	two	6331	33	egg	280	54	drink	5863
13	big	323	34	horn	4005	55	eat	320
14	long	319	35	tail	315	56	bite	308
15	small	319	36	feather	285	57	see	5995
16	woman	313	37	hair	319	58	hear	5679
17	man	314	38	head	324	59	know	319
18	person	5230	39	ear	6465	60	sleep	317
19	fish	6089	40	eye	6576	61	die	5875
20	bird	315	41	nose	6365	62	kill	298
21	dog	6427	42	mouth	329	63	swim	279

64 fly 293	77 stone 6256	90 white 322
65 walk 293	78 sand 314	91 black 321
66 come 5939	79 earth 301	92 night 6226
67 lie 289	80 cloud 309	93 hot 280
68 sit 320	81 smoke 315	94 cold 320
69 stand 304	82 fire 6461	95 full 3726
70 give 289	83 ash 314	96 new 5445
71 say 307	84 burn 305	97 good 317
72 sun 5754	85 path 5987	98 round 233
73 moon 302	86 mountain 5016	99 dry 301
74 star 6080	87 red 317	100 name 5890
75 water 6506	88 green 296	
76 rain 316	89 yellow 296	

---

The lexical correspondences to the above meanings are given in phonological form in the database.

Before proceeding with the program, we may introduce the following terms:

A *colexified meaning pair* is the pair of meanings that are rendered in one language by the same lexeme. We shall enclose these pairs in angular brackets, e.g., <person=tree>, <I=we>, and so on.

A *colexifier* is the specific word-form, or lexeme, which renders the meaning of a colexified pair. We shall enclose colexifiers in square brackets, as in ASJP these are phoneme strings, e.g., [ti] stands for <person=tree> in Kaningi, a Niger-Congo language.

We shall call a *colexification model* (or simply, a *colexification*) a lexified meaning pair alongside with its colexifier, e.g., <person=tree> → [ti] (in Kaningi).

We implemented a computer program to discover the colexifications in the ASJP database. The program first constructs all (unordered) pairs out of the 100 meaning items in the wordlist. The logically possible such pairs are 4950 ( $n$  choose 2, calculated by the formula  $n*(n-1)/2$ , in our case  $100*99/2=4950$ ). It is then tested, for

each potential colexified meaning pair, whether it is actually realized or not in each of the 6809 languages. These procedures guarantee the exhaustive search for colexifications in the database. For each colexified meaning pair, the supporting languages, their familial affiliation, and colexifiers (i.e., word-forms) are also taken into account, as well as their numbers.

An entry in our colexification database looks like this:

- *Colexified meaning pair*: <person=tree>
- *Colexifier(s)*: [re], [te], [ti]
- *Supporting language(s) with family & colexifier*:
  - KANINGI, Niger-Congo (Bantoid branch), [ti]
  - MBOSHI, Niger-Congo (Bantoid branch), [re]
  - NDUUMO, Niger-Congo (Bantoid branch), [ti]
  - TIENE, Niger-Congo (Bantoid branch), [te]
- *Number of supporting languages*: 4
- *Number of supporting families*: 1
- *Number of supporting colexifiers*: 3

This structuring of our colexification database allows straightforward queries to be made to computationally explore it. Any combination of the database parameters, given in italics above, can be queried to retrieve information of interest to the user of the system. E.g., one can discover all colexified pairs valid in only one language family, in three families or in more than 10 families; such valid in more than two languages and having less than 5 colexifiers; such valid only in Indo-European, and so on and so forth. Additionally, a number of subroutines accomplish various counts we found useful in exploring the database, and, importantly, we can unproblematically add more procedures for search as need arises.

### 3. Summary of the Results

Using our machinery, we constructed a typological database of colexifications in the form shown in the previous section. We show a fragment of the database in the Appendix. For economy of space, we list only the colexified meaning pairs occurring in more than one language family, accompanied by the number of supporting languages and top-level families.

Below we summarize the results, specifically in regard to the distribution of colexified meaning pairs in languages and language families.

1) We found 1098 colexified meaning pairs in all out of 4950 theoretically possible from the 100 items in the basic vocabulary list. In other words, more than one fifth of all thinkable patterns are actually realized in at least one of the inspected languages.

2) Some colexified meaning pairs occur in just one or a small amount of *languages*, while others are somewhat better supported. The number of colexified meaning pairs supported by 1, 2, 3, and so on languages is listed in Table 2.

Table 2. The Distribution of Colexified Meaning Pairs in Languages

No. of Colexifications	No. of Supporting Languages
467	1
155	2
110	3
87	4
56	5
120	6-10
71	11-20
14	21-30

6	31-40
6	41-50
5	51-100
1	>100

As seen from the above table, by far the greatest number of patterns (467 out of 1098) are found in only one language, i.e., are idiosyncratic to specific languages. 155 models are found in two languages, 110 in three languages, and so on. A smaller number of association pairs are supported in more languages, e.g., 71 patterns are found in the interval 11-20 languages. Only one colexified meaning pair is found in more than 100 languages (this is <mountain=stone> found in 140 languages, amounting to 12.7% of inspected languages for this particular pattern, cf. Table 1). Generally, colexification among basic vocabulary turns out to be a linguistic phenomenon that would appropriately be characterized as “rara” (Wohlgemuth & Cysouw 2010)

3) Some colexifications are present in just one *language family*, others in two, three or more families. Below we give the distribution of colexifications according to the number of families each is supported by.

Table 3. The Distribution of Colexified Meaning Pairs in Language Families

No. of Colexifications	No. of Supporting Families
559*	1
220	2
162	3
71	4
37	5
16	6



8	7
4	8
3	9
5	10
13	>10

\*92 in family with more than 1 language

As seen from Table 3, 559 colexified meaning pairs hold in a single language. We will use this fact in the next section as a guideline in assigning genetic membership to a language with unknown or suspect affiliation. 220 meaning pairs occur in 2 families, 162 occur in 3 families, 71 in 4 families and so on. Only 13 meaning association pairs hold in more than 10 language families, the highest family support being for <I=we> occurring in 31 families (see Appendix). Generally, we may say that the number of colexification pairs and the number of families supporting them are inversely proportional, i.e., one decreases as the other increases, a fact which could be expected from the analogical distribution of patterns in individual languages, considered in the previous paragraphs.

#### 4. Interpretation of the Results

A typological database of colexifications among basic vocabularies may be put to a variety of uses in fields like semantic change, semantic reconstruction, and historical linguistics more generally. Here we will limit ourselves to considering the problems of distinguishing between polysemy and homonymy and discovering the genetic affiliation of languages with unknown or suspect classifications.

#### 4.1. Colexification: Polysemy or Homonymy

Colexification covers cases of both polysemy and homonymy. It is important to be able to distinguish between the two phenomena in a case of colexification because, while homonymy is less semantically interesting as the result of mere chance coincidence or merging of forms of distinct words, polysemy is the result of real semantic changes and thus reveals general cognitive processes or culture-specific semantic associations. It is no surprise then that it is polysemy rather than homonymy that is subjected to intensive semantic studies.

Which colexifications are cases of true polysemy and hence semantic shifts, and which are homonymous or mere chance form coincidences? In general, this is a complex problem and should be solved individually in each observed case in a specific language. Thus e.g., Modern English *arms* ‘upper limbs’ and *arms* ‘weapons’ are homonyms because we know that in Middle English they were distinct words: the word for upper limbs with the form *earmes* (from Old English *earm*) and that for weapons *armes* (from Old French *arme*) and some phonological processes converged them to a single modern form, viz. *arms*; whether or not these processes would change the distinct word-forms to an identical form is a fact of purely coincidental nature. In contrast, Modern English *arm*, meaning either ‘upper limb’ or ‘a support (as on a chair) for the elbow and forearm’ is a single polysemous word in which the first meaning developed (via metaphor) into the second. In other words, the different meanings of homonymous words do not need to be related and are generally not so; in cases of polysemy the different meanings are related by semantic processes (shifts), e.g., metaphor as in the above case, metonymy as in *horn* ‘animal horn’ or ‘musical instrument’, and so on.

The question now is this: can our database of colexifications help us discriminate (even if approximately) between polysemy and homonymy without recourse to entering into specific historical knowledge pertaining to the individual language at issue? Below we look at several criteria or general heuristics to guide us in a large-scale comparative investigation.

1) *Strong language/family support*

Language, and especially family, support for a colexified meaning pair is an important indicator of polysemy, the greater this support for this pair, the more likely it reflects polysemy rather than homonymy. The obvious idea underlying this guideline is that a repeated occurrence of a meaning association is more likely to be due to some universal semantic shift (i.e., polysemy) rather than due to mere chance coincidence in each language of the forms expressing the colexified meanings (as homonymy would imply).

Below are examples of some of the most common colexified meaning pairs in the ASJP database, quite probably reflecting polysemy (the association pairs are accompanied by the number of supporting languages, followed by the supporting families in brackets):

- <mountain=stone> 140 (25)
- <I=we> 89 (31)
- <ear=leaf> 67 (14)
- <fire=tree> 64 (13)
- <horn=knee> 48 (3)
- <feather=hair> 44 (19)
- <bark=skin> 42 (21)

- <ear=hear> 41 (18)
- <drink=water> 40 (15)
- <man=person> 33 (18)

There are however other cases, with reasonable language/family support, which are less certainly polysemous and require further study to find an explanation as to the origin of their ambiguity. Below are such examples we found:

- <come=dog> 78 (3)
- <louse=we> 59 (4)
- <name=tooth> 46 (7)
- <liver=two> 43 (3)
- <ear=fish> 32 (3)
- <ear=name> 35 (5)
- <see=we> 34 (3)
- <die=eye> 40 (2)

2) *The presence of several distinct colexifiers in one language family*

The association pairs with less support in terms of languages and families are more challenging with respect to the determination of whether or not there is true polysemy at issue. The support of an association pair in more than one language family is some sign of true polysemy, but even when a colexified meaning pair occurs in just one family there are some indicators of polysemy. Thus, the presence of several distinct colexifiers in one language family is suggestive of polysemy rather than homonymy. The reason is that the

different colexifiers are more likely to be innovations in phonological shape preserving the meaning association (possibly due to some culture-specific traits of the people speaking those languages) rather than be due to the mere chance merging of word-forms of distinct words in each language that have happened in all those languages (as homonymy would imply). In other words, we presume that the meaning association in the pair is preserved across the different languages in the family, only the word-forms expressing them innovate with time.

Below are some examples of this type we found exploring the ASJP database.

E.g., some Niger-Congo languages seem to have polysemy rather than homophony in the following colexified meaning pairs: <fish=tooth> in Kensweinstei (Bantoid branch) we have the form [so], in Vagala (Gur branch) we have the distinct form [ɲin], while in Nupe (Nupoid branch) we have the still different [yĩkã]; <blood=nose> in Fang Meke (Bantoid) is rendered by [tsi], in Guro (Eastern Mande) by the distinct word-form [jæ], and in Lere Gana (Kainji) by the still different [mapau], and analogically with Bolon (Western Mande) [ju]; <eye=horn> in Bum B Cameroon (Bantoid) has the form [sæ], Senuoufo Tagwana (Gur) the distinct form [ɲala].

Similar considerations hold in two Austronesian languages: <foot=hand> in Samoan (Oceanic branch) is denoted as [aʔao] and in Yapese (Yapese branch) as [rifrif], two apparently distinct word-forms. Cf. also <earth=liver> in Aguacateco Aguacatan (Mayan) [tsots] and Uspanteko (Mayan) which has the very different form [ulew].

### 3) *The presence of distinct synonymous colexifiers in one language*

Such presence indicates polysemy rather than homonymy for

reasons similar to the previous heuristic. If a language chooses to coin a new form to express some meaning, this meaning, being firmly associated with another meaning in a colexification pair, carries over to the new word-form. Put differently, the linguistic process at work is an innovation in form preserving the stable meaning association, or polysemy.

E.g., the colexification pattern <all=many> occurs only in Nunggbuyu (Australian) and is expressed by the synonymous pair of apparently distinct forms [warawindi] and [dalun], therefore signaling polysemy in that language.

4) *The possibility to explain a colexification by reference to semantic shift*

The possibility to account for a colexification by appealing to known semantic changes (metaphor, metonymy, narrowing, broadening, and so on), which generally reflect universal cognitive associations, apparently gives a preference to interpreting a colexification as a case of polysemy rather than homonymy.

These heuristics can operate collectively and thus present stronger support for an interpretation of a colexified meaning pair as a case of polysemy.

E.g., we may consider the colexified meaning pair <mountain=stone>, registered in the database in 144 languages from 25 families. This strong support in terms of languages and families indicates polysemy (by heuristic 1). The semantic association between these meanings is a clear case of a semantic shift by metonymy (heuristic 4). Below we give examples further strengthening our interpretation of the colexification as a case of polysemy, based on the other two criteria.

In the Niger-Congo family the language Bua (Adamawa branch)

has the form [ta] for <mountain=stone>, Biali (Gur) has [tali], Ditammari (Gur) [yatāda], Ekpetiama (Ijoid) [ugu], Ziriya (Kainji) [kafau], and so on, all forms apparently non-cognate and derived from different parent forms (heuristic 2).

In the Nilo-Saharan family we have an analogical situation: for <mountain=stone> the language Kaba Deme Sara (Bongo-Bagirmi branch) has the form [ko], Kabba (Bongo-Bagirmi) has the distinct form [jer], Berti (Eastern Saharan branch) the form [wi], Gule (Koman branch) the form [of], Uduk (Koman) [wof], Gbaya (Kresh) [angba], Mangbetu (Mangbeti) [nekopi], Andri (Moru-Maadi) [univa], Miza (Moru-Maadi) [baraŋwā], Anyuak (Nilotic) [kidi], Nuer (Nilotic) [p<sup>h</sup>æm], and so on.

In the Australian family, the language Diyari designates the colexification pattern <mountain=stone> by [mada], the languages Gangulu by [bari], Ganggalida by [kamara], Nyangumarta by [wanku], Yanyuwa by the alternative non-cognate colexifiers [buluruluru] and [danggā] (i.e., the latter also by heuristic 3), and so on.

In general, there are numerous examples of this type in our database of colexifications.

## 4.2. Shared Colexification Models and Genetic Affiliation

The study of colexifications in different languages may have different implications for historical linguistics. E.g., Croft et al. (2009) address the problem of quantifying semantic change for the purposes of choosing among alternative historical reconstructions. To this end, they conduct a typological study of word polysemy among basic vocabulary items in order to find the frequencies of the different meanings in polysemous words. They investigate 22 concepts denoting natural objects in the Swadesh list across a

typological sample of over 50 languages.

While the above approach assumes already available hypotheses regarding genetic affiliation, we explore the possibility, based on our colexification database, to guess membership of two languages in the same language family. We propose a heuristic to this effect, derived empirically from the found colexifications. Thus, we found that there are 6426 *shared colexification models* (or simply, shared colexifications) between a pair of languages, i.e., cases in which two languages have the very same colexifying word-form for the same colexified meaning pair. Only 47, or 0.7% (47/6426) of these shared colexifications were found to hold in two languages belonging to *different* language families, while 99.3%, were found to hold in the same language family. (It should be borne in mind, however, that the matching of colexifiers may be somewhat less precise in some cases, owing to the fact that the ASJP database slightly simplifies the phonetic representation of word-forms.) In consequence, shared colexifications seem like a reliable indicator of genetic affiliation and can be used as a heuristic to this effect: finding shared colexifications signals shared genetic membership with a probability of 99.3%. This high reliability of our criterion for same family membership should not be surprising, as the criterion implies that the two languages share two cognate words plus a colexification for these two meanings.

Below are some shared colexifications between language pairs of different language families out of those we found inspecting our colexification database.

(1) Shared colexification <I=horn> → [ku] in Tai-Kadai & Hmong-Mien

- Proto Kadai (Tai-Kadai) & Phetchabun Hmong (Hmong-Mien)
- Proto Kadai (Tai-Kadai) & Proto West A Hmong (Hmong-Mien)
- Proto Kadai (Tai-Kadai) & Shimenkan Hmong (Hmong-Mien)



- Proto Kadai (Tai-Kadai) & Tak Hmong (Hmong-Mien)

(2) Shared colexification <come=tongue> → [ma] in Tai-Kadai & Austronesian

- Laha (Tai-Kadai) & Selaru (Austronesian, Central Malayo-Polynesian)
- Fangcun Mak (Tai-Kadai) & Selaru (Austronesian)
- Kam (Tai-Kadai) & Selaru (Austronesian)
- Kam Zhanglu (Tai-Kadai) & Selaru (Austronesian)
- Mak (Tai-Kadai) & Selaru (Austronesian)
- Sui Jung Chiang (Tai-Kadai) & Selaru (Austronesian)
- Sui Li Ngam (Tai-Kadai) & Selaru (Austronesian)
- Tiangzhu Shidong Northern Dong (Tai-Kadai) & Selaru (Austronesian)
- Zhanglu Dong (Tai-Kadai) & Selaru (Austronesian)

The above two shared colexifications, suggesting pairwise relatedness between Austronesian, Hmong-Mien and Tai-Kadai do not look like wild guesses but rather seem to corroborate a well-known (but still controversial) hypothesis by Paul K. Benedict (1976), who linked these languages in an Austric stock. The Austric hypothesis was actually first proposed by Wilhelm Schmidt, who initially linked only Austroasiatic and Austronesian, suggesting Hmong-Mien as a further possible member of the stock.

Most of the other found shared colexifications do not seem to reflect any suggested relationships in the literature, but we may mention some for consideration by linguists known as “lumpers”, who spend efforts in trying to group languages into larger genetic groupings, or superfamilies.

E.g., a number of pairwise shared colexifications were found between Niger Congo languages and Austronesian and Sepik languages (these could simply be chance coincidences owing to the very large number of languages in these families (esp. Niger Congo and Austronesian) in the ASJP database).

### (3) Niger-Congo and other families

- <I=we> → [mi] Nyambeenge (Niger-Congo) & Jarawa (Austronesian)
- <I=we> → [ŋa] Baka 2 (Niger-Congo) & Raute (Sino-Tibetan)
- <I=come> → [na] Baatonum (Niger-Congo) & Lore diakarkar (Austronesian)
- <I=come> → [na] Baatonum (Niger-Congo) & Shark Bay (Austronesian)
- <come=fire> → [ja] Bongili (Niger-Congo) & Maprik (Sepik)
- <come=fire> → [ja] Bongili (Niger-Congo) & Ngala (Sepik)
- <come=fire> → [ja] Bongili (Niger-Congo) & Nyaura (Sepik)
- <come=fire> → [ja] Bongili (Niger-Congo) & Wosera (Sepik)
- <come=fire> → [ja] Bongili (Niger-Congo) & Yelogu (Sepik)
- <come=fire> → [ja] C831 Ilebo (Niger-Congo) & Maprik (Sepik)
- <come=fire> → [ja] C831 Ilebo (Niger-Congo) & Ngala (Sepik)
- <come=fire> → [ja] C831 Ilebo (Niger-Congo) & Nyaura (Sepik)
- <come=fire> → [ja] C831 Ilebo (Niger-Congo) & Wosera (Sepik)
- <come=fire> → [ja] C831 Ilebo (Niger-Congo) & Yelogu (Sepik)
- <come=fire> → [ja] Mboshi Bunji (Niger-Congo) & Ngala (Sepik)
- <come=fire> → [ja] Mboshi Bunji (Niger-Congo) & Nyaura (Sepik)
- <come=fire> → [ja] Mboshi Bunji (Niger-Congo) & Wosera (Sepik)
- <come=fire> → [ja] Mboshi Bunji (Niger-Congo) & Yelogu (Sepik)
- <come=fire> → [ja] Bongili (Niger-Congo) & Boikin (Sepik)
- <come=fire> → [ja] C831 Ilebo (Niger-Congo) & Boikin (Sepik)
- <come=fire> → [ja] Mboshi Bunji (Niger-Congo) & Boikin (Sepik)

- <come=fire> → [ja] Mboshi Bunji (Niger-Congo) & Maprik (Sepik)
- <see=you> → [na] Lama (Niger-Congo) & Dafang (Sino-Tibetan)

Other cases of shared colexifications may also be mere coincidences, this time not due the large number of inspected languages, but owing to more common colexified meaning pairs (or ‘universal’ semantic shifts) like <I=we>, <mountain=stone>, <drink=water>. Another possibility for explanation in some cases may be areal closeness, as e.g., the third example involving languages from North America below.

#### (4) Others

- <I=we> → [ni] Hausa (Austro-Asiatic) & Montagnais (Algonquian)
- <I=we> → [ni] Hausa (Austro-Asiatic) & Wintu (Penutian)
- <I=we> → [ni] Montagnais (Algonquian) & Wintu (Penutian)
- <mountain=stone> → [wi] Berti (Nilo-Saharan) & Yavapai (Hokan)
- <mountain=stone> → [kau] Buduma (Austro-Asiatic) & Kanuri (Nilo-Saharan)
- <drink=water> → [ma] Tirma (Nilo-Saharan) & Vilela (Lule Vilela)

### 4.3. Shared Colexified Meaning Pairs and Genetic Affiliation

Can the colexified meaning pairs alone signal genetic affiliation? In Table 3 we saw that there are 559 association pairs that hold in a single language family, 92 of which hold in more than one language in that family. This fact apparently can be suggestive of affiliation: if

a colexified meaning pair is idiosyncratic to a language family, then a language with unknown (or suspect) membership could be assigned to that family if it happens to have the same meaning association. Colexified meaning pairs with greater support in terms of languages will have more weight in making this decision.

Below we list meaning pairs that occur only in one language family, and are supported at least by 5 languages. In order to be more precise regarding membership assignment, we give the full lineage of the groupings rather than only the top-level language membership (indicated in bold in the starting example for a family). Also, in the following list we give the number of supporting languages and the number of all languages in the lineage.

(5) Colexifications occurring in a single family ( $\geq 5$  languages)

- <leaf=path> : [**Niger-Congo**, Atlantic-Congo, Volta-Congo]  
(7 from 1286 languages)
- <ear=person> : [Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Southern, NarrowBantu]  
(13 from 752 languages)
- <ear=tree> : [Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid] (12 from 892 languages)
- <breasts=two> : [Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo] (13 from 1075 languages)
- <die=dog> : [Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Southern] (8 from 876 languages)
- <fire=new> : [Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid] (18 from 892 languages)
- <full=liver> : [Niger-Congo] (7 from 1455 languages)

- <drink=louse> : [**Mayan**] (23 from 121 languages)
- <give=tongue> : [Mayan, Yucatecan-Core Mayan]  
(9 from 114 languages)
- <earth=liver> : [Mayan, Yucatecan-Core Mayan, K'ichean-Mamean]  
(5 from 82 languages)
- <louse=new> : [**Sino-Tibetan**] (9 from 273 languages)
- <nose=one> : [**Khoisan**, Southern Africa, Central]  
(8 from 16 languages)
- <tooth=two> : [**Austronesian**, Malayo-Polynesian, Central-Eastern Malayo-Polynesian, Eastern Malayo-Polynesian, Oceanic, Central-Eastern Oceanic, Remote Oceanic, North & Central Vanuatu, Northeast Vanuatu-Banks Islands, Epi, Lamenu-Baki]  
(6 from 13 languages)
- <liver=new> : [**Hmong-Mien**] (10 from 47 languages)
- <burn=eye> : [**Trans-New Guinea**, Finisterre-Huon, Huon, Western]  
(5 from 10 languages)

It turns out that the Niger-Congo family has 7 idiosyncratic meaning association pairs, Mayan 3, Sino-Tibetan, Khoisan, Austronesian, Hmong-Mien and Trans-New Guinea 1 pair.

In our computational exploration of ASJP, we came across the following intriguing case: the language isolate Sumerian, the language of ancient Sumer, which was spoken in northern Mesopotamia (modern Iraq), turned out to share the colexified meaning pair <blood=die> with Sino-Tibetan, and particularly with

27 languages of Tibeto-Burman origin, and these are the only languages in ASJP that have this meaning association. This is a finding of some interest because the classification of Sumerian is highly controversial and a wide variety of proposals have been made concerning its linguistic affiliations, including Munda, Dravidian, Kartvelian, Uralic, Basque, Nostratic, Sino-Tibetan (for the various proposals, cf. e.g., Zakar 1971, Bombard 1984, Diakonoff 1997, Braun 2004, Parpola 2007). Our finding seems to corroborate the hypothesis of the link of Sumerian to Tibeto-Burman, proposed by Braun (2004).

It should be strongly emphasized here that our guidelines are only suggestive of genetic membership and should not be taken at face value. They only provide hints that require further investigations, possibly with worthwhile outcome.

## 5. Conclusions

In the paper we presented a typological database of colexifications among basic vocabulary, computationally derived from the ASJP database. We found 1098 colexified meaning pairs in all out of 4950 logically admissible from the 100 items in the basic vocabulary list. In other words, more than one fifth of all thinkable patterns are actually realized in at least one of the languages in the database. Some potential applications of the inventory of colexifications were proposed. Some heuristics were introduced pertaining to distinguishing of polysemy from homonymy in a typological database, as well as such related to the determination of common membership of two languages in the same language family. In particular, it was found that shared colexifications corroborate the

postulation of the Austric stock and the attribution of Sumerian to the Tibeto-Burman language family.

Our database of colexifications among basic vocabularies provides a solid starting point for deeper analyses of various problems in semantic change, semantic reconstruction, and historical linguistics more generally.

## References

- Benedict, P. 1976. Austro-Thai and Austroasiatic. In P. Jenner, L. Thompson & S. Starosta (eds.), *Austroasiatic Studies* (Part I) 1-36. Honolulu, HI: University of Hawai'i Press.
- Croft, W. et al. 2009. *Quantifying Semantic Shift for Reconstructing Language Families*. Talk held at the 83rd Annual Meeting of the Linguistic Society of America. Available at URL <<http://www.unm.edu/~wcroft/Papers/Polysemy-LSA-HO.pdf>>.
- Diakonoff, I. 1997. External Connections of the Sumerian Language. *Mother Tongue* 3, 54-63.
- Braun, J. 2004. *Sumerian and Tibeto-Burman, Additional Studies*. Warszawa: Wydawnictwo Agade.
- Bomhard, A. 1984. *Toward Proto-Nostratic: A New Approach to the Comparison of Proto-Indo-European and Proto-Afroasiatic*. Amsterdam: John Benjamins.
- Falkum, I. & A. Vicente. (eds.) 2015. *Lingua (Polysemy: Current Perspectives and Approaches)* 157, 1-172.
- François, A. 2008. Semantic Maps and the Typology of Colexification: Intertwining Polysemous Networks across Languages. In M. Vanhove (ed.), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*

163-215. Amsterdam: John Benjamins.

- Key, M. & B. Comrie. 2007. IDS — The Intercontinental Dictionary Series. Available at URL <<http://lingweb.eva.mpg.de/ids/>>.
- List, J.-M. et al. 2014. CLICS: Database of Cross-Linguistic Colexifications. Marburg: Forschungszentrum Deutscher Sprachatlas Version 1.0. Available at <<http://CLICS.lingpy.org>, accessed on 2015-4-22>.
- Parpola, S. 2007. Sumerian: A Uralic Language. Paper presented at 53e Rencontre Assyriologique Internationale. Moscow.
- Swadesh, M. 1971. *The Origin and Diversification of Language*. Chicago, IL: Aldine.
- Urban, M. 2012. Analyzability and Semantic Associations in Referring Expressions: A Study in Comparative Lexicology. Ph.D Dissertation. Leiden University.
- Vanhove, M. (ed.) 2008. *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*. Amsterdam & Philadelphia: John Benjamins.
- Wichmann, S. et al. 2013. *The ASJP Database* (version 16).
- Wohlgemuth, J. & M. Cysouw. 2010. *Rara & Rarissima: Documenting the Fringes of Linguistic Diversity*. Berlin & New York: Mouton de Gruyter.
- Zakar, A. 1971. Sumerian — Ural-Altaic Affinities. *Chicago Journals, Current Anthropology* 122, 215-225.
- Zalizniak, A. 2008. A Catalogue of Semantic Shifts: Towards a Typology of Semantic Derivation. In M. Vanhove (ed.), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations* 217-232. Amsterdam: John Benjamins.
- Zalizniak, A. et al. 2012. The Catalogue of Semantic Shifts as a Database for Lexical Semantic Typology. *Linguistics* 503, 633-669.



## Appendix

The following Appendix contains the colexified meaning pairs in ASJP database which hold in two or more language families. First we list the colexified pair, then the number of languages it is found in the database, and finally, in brackets, the number of top-level language families these languages belong to. The number of supporting languages for each colexified meaning pair should be evaluated in the context of the number of occurrences of each basic vocabulary item in ASJP database, given in Table 1 in the text.

<I=you> 22 (10)	<I=see> 6 (4)	<one=stone> 3 (3)
<I=we> 89 (31)	<I=hear> 12 (2)	<one=path> 2 (2)
<I=one> 2 (2)	<I=die> 2 (2)	<two=you> 9 (3)
<I=person> 3 (2)	<I=come> 11 (5)	<two=water> 6 (3)
<I=fish> 29 (4)	<I=water> 13 (3)	<big=many> 9 (3)
<I=bird> 2 (2)	<I=stone> 3 (2)	<big=long> 4 (2)
<I=dog> 3 (3)	<I=fire> 3 (2)	<man=person> 33 (18)
<I=tree> 2 (2)	<I=black> 2 (2)	<person=we> 8 (3)
<I=leaf> 4 (3)	<we=you> 19 (12)	<person=two> 5 (5)
<I=horn> 16 (4)	<that=this> 9 (7)	<person=woman> 4 (4)
<I=ear> 2 (2)	<what=who> 15 (10)	<person=skin> 3 (3)
<I=eye> 4 (4)	<one=two> 4 (3)	<person=tooth> 3 (3)
<I=nose> 2 (2)	<one=person> 9 (4)	<person=tongue> 3 (2)
<I=tooth> 6 (2)	<one=tree> 4 (4)	<person=see> 3 (3)
<I=tongue> 4 (3)	<one=skin> 2 (2)	<person=sun> 8 (4)
<I=knee> 3 (2)	<one=tooth> 4 (2)	<person=star> 4 (2)
<I=hand> 3 (2)	<one=see> 2 (2)	<person=stone> 4 (3)
<I=drink> 6 (4)	<one=sun> 4 (4)	<fish=you> 3 (3)
<I=eat> 4 (4)	<one=water> 8 (4)	<fish=we> 5 (3)

<fish=one> 3 (3)	<dog=drink> 11 (4)	<leaf=tooth> 14 (3)
<fish=two> 7 (3)	<dog=see> 3 (3)	<leaf=liver> 6 (3)
<fish=person> 2 (2)	<dog=hear> 2 (2)	<leaf=water> 8 (4)
<fish=louse> 3 (2)	<dog=water> 14 (5)	<leaf=stone> 8 (3)
<fish=leaf> 8 (4)	<dog=stone> 3 (2)	<leaf=night> 4 (2)
<fish=skin> 4 (2)	<dog=fire> 4 (3)	<leaf=new> 7 (3)
<fish=nose> 4 (2)	<dog=path> 9 (4)	<leaf=name> 3 (2)
<fish=tongue> 3 (3)	<dog=mountain> 2 (2)	<bark=skin> 42 (21)
<fish=hand> 2 (2)	<dog=new> 2 (2)	<skin=we> 3 (3)
<fish=see> 4 (3)	<louse=you> 5 (3)	<skin=two> 13 (2)
<fish=hear> 8 (6)	<louse=we> 59 (4)	<skin=tree> 4 (3)
<fish=star> 15 (3)	<louse=tree> 10 (5)	<skin=tooth> 3 (2)
<fish=water> 5 (2)	<louse=skin> 6 (3)	<skin=water> 2 (2)
<fish=stone> 4 (2)	<louse=nose> 2 (2)	<skin=stone> 3 (2)
<fish=mountain> 6 (4)	<louse=tooth> 17 (4)	<blood=you> 5 (3)
<fish=night> 2 (2)	<louse=see> 3 (2)	<blood=person> 2 (2)
<fish=full> 4 (2)	<louse=water> 16 (7)	<blood=fish> 7 (4)
<fish=name> 3 (2)	<louse=stone> 6 (3)	<blood=dog> 4 (2)
<bird=feather> 2 (2)	<louse=mountain> 8 (5)	<blood=louse> 3 (3)
<bird=breasts> 2 (2)	<louse=name> 20 (4)	<blood=tree> 10 (2)
<dog=person> 11 (3)	<tree=you> 5 (4)	<blood=skin> 3 (3)
<dog=louse> 3 (2)	<tree=we> 2 (2)	<blood=flesh> 2 (2)
<dog=tree> 2 (2)	<tree=two> 4 (3)	<blood=bone> 4 (3)
<dog=leaf> 13 (3)	<tree=water> 12 (8)	<blood=ear> 6 (5)
<dog=skin> 7 (3)	<seed=stone> 3 (2)	<blood=tooth> 7 (5)
<dog=horn> 7 (2)	<leaf=you> 7 (3)	<blood=tongue> 7 (3)
<dog=eye> 3 (2)	<leaf=one> 2 (2)	<blood=hand> 2 (2)
<dog=tooth> 4 (2)	<leaf=two> 5 (3)	<blood=liver> 13 (6)
<dog=tongue> 15 (2)	<leaf=louse> 3 (2)	<blood=drink> 3 (3)
<dog=hand> 16 (4)	<leaf=tree> 17 (9)	<blood=see> 2 (2)
<dog=liver> 3 (2)	<leaf=nose> 2 (2)	<blood=hear> 9 (2)

<blood=die> 28 (2)	<bone=water> 6 (3)	<ear=horn> 11 (5)
<blood=come> 5 (3)	<bone=stone> 8 (3)	<ear=eye> 3 (2)
<blood=sun> 3 (3)	<bone=fire> 2 (2)	<ear=nose> 15 (5)
<blood=water> 12 (5)	<bone=mountain> 4 (3)	<ear=tooth> 3 (3)
<blood=rain> 2 (2)	<bone=name> 7 (3)	<ear=tongue> 3 (2)
<blood=fire> 6 (5)	<egg=stone> 5 (3)	<ear=hand> 2 (2)
<blood=red> 8 (6)	<horn=you> 2 (2)	<ear=liver> 2 (2)
<blood=night> 20 (3)	<horn=tree> 2 (2)	<ear=hear> 41 (18)
<blood=new> 7 (3)	<horn=skin> 2 (2)	<ear=star> 3 (2)
<blood=name> 2 (2)	<horn=nose> 3 (3)	<ear=water> 5 (2)
<bone=you> 2 (2)	<horn=tooth> 21 (9)	<ear=stone> 2 (2)
<bone=we> 5 (3)	<horn=knee> 48 (3)	<ear=night> 13 (3)
<bone=one> 3 (2)	<horn=see> 14 (4)	<ear=new> 2 (2)
<bone=two> 3 (2)	<horn=sun> 4 (2)	<ear=name> 35 (5)
<bone=fish> 3 (3)	<horn=star> 4 (3)	<eye=you> 18 (2)
<bone=louse> 2 (2)	<horn=stone> 4 (2)	<eye=we> 11 (5)
<bone=tree> 15 (6)	<horn=path> 4 (2)	<eye=one> 5 (3)
<bone=seed> 8 (5)	<horn=mountain> 4 (3)	<eye=two> 3 (3)
<bone=skin> 5 (3)	<horn=name> 5 (2)	<eye=person> 16 (3)
<bone=horn> 5 (3)	<feather=leaf> 6 (6)	<eye=fish> 5 (2)
<bone=ear> 2 (2)	<feather=hair> 44 (19)	<eye=louse> 2 (2)
<bone=eye> 3 (3)	<hair=man> 2 (2)	<eye=tree> 4 (4)
<bone=tooth> 8 (4)	<hair=leaf> 6 (4)	<eye=seed> 3 (3)
<bone=knee> 2 (2)	<hair=head> 14 (10)	<eye=leaf> 2 (2)
<bone=hand> 3 (2)	<ear=you> 4 (2)	<eye=nose> 4 (3)
<bone=breasts> 6 (3)	<ear=we> 5 (3)	<eye=tooth> 10 (4)
<bone=hear> 3 (3)	<ear=two> 3 (2)	<eye=liver> 2 (2)
<bone=die> 29 (4)	<ear=fish> 32 (3)	<eye=see> 16 (9)
<bone=come> 3 (2)	<ear=louse> 2 (2)	<eye=hear> 3 (3)
<bone=sun> 7 (3)	<ear=leaf> 67 (14)	<eye=sun> 17 (10)
<bone=star> 2 (2)	<ear=skin> 4 (3)	<eye=star> 6 (6)

<eye=water> 15 (4)	<knee=sun> 2 (2)	<breasts=drink> 3 (3)
<eye=stone> 7 (3)	<knee=path> 8 (2)	<breasts=see> 3 (2)
<eye=fire> 17 (7)	<knee=mountain> 5 (4)	<breasts=hear> 4 (4)
<eye=mountain> 11 (3)	<hand=you> 11 (3)	<breasts=water> 6 (2)
<eye=full> 9 (4)	<hand=we> 15 (5)	<breasts=stone> 7 (3)
<eye=name> 23 (6)	<hand=one> 3 (3)	<breasts=path> 2 (2)
<nose=you> 6 (5)	<hand=two> 2 (2)	<breasts=mountain> 2 (2)
<nose=we> 2 (2)	<hand=louse> 5 (2)	<breasts=name> 4 (3)
<nose=person> 3 (2)	<hand=tree> 12 (5)	<heart=seed> 3 (3)
<nose=tree> 2 (2)	<hand=leaf> 21 (11)	<heart=liver> 4 (3)
<nose=skin> 15 (4)	<hand=skin> 4 (4)	<liver=two> 43 (3)
<nose=tooth> 3 (3)	<hand=tooth> 4 (2)	<liver=person> 8 (4)
<nose=see> 7 (2)	<hand=tongue> 2 (2)	<liver=louse> 9 (3)
<nose=sun> 5 (2)	<hand=knee> 4 (4)	<liver=tree> 5 (2)
<nose=star> 5 (3)	<hand=see> 3 (3)	<liver=skin> 2 (2)
<nose=stone> 2 (2)	<hand=stone> 8 (4)	<liver=tooth> 3 (3)
<nose=path> 2 (2)	<hand=mountain> 10 (3)	<liver=tongue> 4 (3)
<mouth=tooth> 4 (3)	<hand=new> 6 (4)	<liver=water> 4 (4)
<tooth=we> 3 (2)	<hand=name> 8 (4)	<liver=stone> 8 (5)
<tooth=tree> 9 (5)	<belly=good> 3 (2)	<liver=name> 4 (2)
<tooth=water> 3 (3)	<breasts=you> 2 (2)	<drink=you> 11 (10)
<tongue=we> 2 (2)	<breasts=one> 4 (3)	<drink=we> 4 (2)
<tongue=two> 2 (2)	<breasts=person> 5 (4)	<drink=one> 5 (3)
<tongue=water> 8 (4)	<breasts=dog> 3 (2)	<drink=person> 4 (2)
<claw=foot> 3 (3)	<breasts=louse> 4 (3)	<drink=leaf> 3 (3)
<foot=path> 2 (2)	<breasts=tree> 3 (2)	<drink=skin> 4 (3)
<knee=louse> 3 (3)	<breasts=skin> 4 (2)	<drink=horn> 2 (2)
<knee=tree> 5 (3)	<breasts=tongue> 2 (2)	<drink=ear> 3 (3)
<knee=skin> 2 (2)	<breasts=hand> 3 (2)	<drink=nose> 4 (4)
<knee=tooth> 3 (3)	<breasts=heart> 3 (3)	<drink=tooth> 14 (6)
<knee=liver> 2 (2)	<breasts=liver> 7 (5)	<drink=tongue> 2 (2)

<drink=knee> 5 (2)	<hear=mountain> 2 (2)	<fly=stand> 2 (2)
<drink=hand> 2 (2)	<hear=night> 3 (2)	<fly=give> 2 (2)
<drink=liver> 4 (3)	<hear=new> 2 (2)	<come=you> 11 (4)
<drink=eat> 10 (5)	<hear=name> 11 (7)	<come=we> 7 (5)
<drink=see> 8 (4)	<know=see> 9 (6)	<come=one> 4 (3)
<drink=hear> 15 (2)	<die=you> 8 (2)	<come=two> 9 (4)
<drink=water> 40 (15)	<die=we> 11 (3)	<come=person> 4 (3)
<drink=fire> 5 (2)	<die=not> 5 (2)	<come=dog> 78 (3)
<drink=full> 3 (2)	<die=one> 5 (4)	<come=tree> 7 (4)
<eat=this> 5 (2)	<die=two> 5 (3)	<come=leaf> 2 (2)
<eat=what> 2 (2)	<die=person> 10 (3)	<come=skin> 6 (5)
<bite=mouth> 3 (2)	<die=fish> 5 (2)	<come=ear> 2 (2)
<bite=eat> 9 (8)	<die=louse> 30 (3)	<come=eye> 7 (2)
<see=you> 7 (4)	<die=tree> 5 (2)	<come=tooth> 8 (4)
<see=we> 34 (3)	<die=leaf> 3 (3)	<come=tongue> 21 (7)
<see=two> 6 (2)	<die=skin> 4 (3)	<come=knee> 5 (2)
<see=tree> 8 (5)	<die=ear> 2 (2)	<come=hand> 4 (3)
<see=tooth> 4 (3)	<die=eye> 40 (2)	<come=drink> 10 (5)
<see=tongue> 2 (2)	<die=tooth> 5 (4)	<come=see> 3 (2)
<see=water> 11 (5)	<die=tongue> 3 (2)	<come=hear> 7 (4)
<see=stone> 4 (2)	<die=hand> 14 (3)	<come=die> 29 (3)
<hear=you> 7 (6)	<die=liver> 17 (3)	<come=walk> 3 (3)
<hear=two> 7 (5)	<die=drink> 5 (4)	<come=sun> 4 (3)
<hear=leaf> 9 (5)	<die=see> 10 (5)	<come=water> 7 (4)
<hear=horn> 2 (2)	<die=hear> 3 (3)	<come=stone> 4 (2)
<hear=nose> 4 (2)	<die=kill> 10 (7)	<come=fire> 20 (5)
<hear=liver> 10 (3)	<die=water> 14 (3)	<come=path> 7 (2)
<hear=see> 12 (6)	<die=fire> 13 (6)	<come=full> 2 (2)
<hear=know> 7 (2)	<die=mountain> 2 (2)	<come=new> 3 (2)
<hear=water> 4 (3)	<die=full> 5 (2)	<lie=mouth> 2 (2)
<hear=stone> 3 (2)	<die=new> 15 (3)	<lie=sleep> 20 (8)

<lie=sit> 3 (3)	<fire=tree> 64 (13)	<mountain=see> 3 (2)
<give=mouth> 2 (2)	<fire=leaf> 6 (3)	<mountain=water> 3 (3)
<give=sleep> 2 (2)	<fire=skin> 7 (3)	<mountain=rain> 3 (2)
<say=tail> 2 (2)	<fire=horn> 2 (2)	<mountain=stone> 140 (25)
<say=sun> 2 (2)	<fire=nose> 4 (3)	<mountain=path> 2 (2)
<sun=two> 11 (3)	<fire=tongue> 5 (3)	<mountain=new> 4 (3)
<sun=tree> 4 (4)	<fire=hand> 3 (3)	<mountain=name> 3 (2)
<sun=tooth> 5 (4)	<fire=see> 2 (2)	<red=yellow> 3 (3)
<sun=tongue> 2 (2)	<fire=hear> 5 (2)	<green=yellow> 13 (6)
<sun=water> 3 (2)	<fire=sun> 5 (5)	<night=you> 2 (2)
<moon=tongue> 2 (2)	<fire=star> 5 (3)	<night=person> 11 (3)
<moon=sun> 18 (11)	<fire=water> 5 (3)	<night=tree> 4 (3)
<moon=white> 2 (2)	<fire=path> 2 (2)	<night=nose> 4 (3)
<star=we> 2 (2)	<fire=mountain> 8 (4)	<night=tooth> 7 (3)
<star=tooth> 2 (2)	<fire=night> 6 (2)	<night=see> 4 (3)
<star=sun> 4 (4)	<fire=hot> 4 (2)	<night=sun> 4 (2)
<star=stone> 9 (4)	<fire=name> 20 (6)	<night=water> 8 (2)
<water=you> 8 (5)	<ash=tongue> 2 (2)	<hot=water> 2 (2)
<water=we> 6 (5)	<burn=fire> 3 (2)	<full=louse> 2 (2)
<rain=water> 22 (14)	<burn=red> 2 (2)	<full=stone> 2 (2)
<stone=you> 2 (2)	<path=person> 3 (2)	<full=mountain> 3 (2)
<stone=we> 2 (2)	<path=skin> 5 (4)	<full=new> 3 (3)
<stone=two> 2 (2)	<path=see> 2 (2)	<new=you> 24 (3)
<stone=tree> 3 (2)	<path=walk> 3 (2)	<new=we> 6 (3)
<stone=tongue> 2 (2)	<path=sun> 17 (6)	<new=two> 4 (4)
<stone=water> 3 (3)	<path=stone> 5 (3)	<new=tree> 23 (3)
<cloud=rain> 4 (4)	<mountain=we> 5 (2)	<new=tongue> 4 (4)
<cloud=smoke> 11 (8)	<mountain=one> 3 (3)	<new=see> 6 (2)
<fire=person> 10 (3)	<mountain=tree> 6 (5)	<new=star> 7 (4)
<fire=fish> 7 (5)	<mountain=skin> 5 (4)	<new=stone> 8 (4)
<fire=louse> 5 (3)	<mountain=nose> 17 (5)	<name=you> 10 (7)

<name=we> 9 (3)	<name=nose> 20 (3)	<name=water> 14 (5)
<name=one> 9 (3)	<name=tooth> 46 (7)	<name=stone> 5 (4)
<name=two> 5 (3)	<name=tongue> 2 (2)	<name=path> 3 (2)
<name=person> 4 (4)	<name=see> 7 (3)	<name=night> 4 (3)
<name=tree> 5 (3)	<name=sun> 16 (6)	<name=new> 4 (4)
<name=skin> 6 (4)	<name=star> 2 (2)	