

# Towards establishing a new basic vocabulary list (Swadesh list)

(Version 2)

Carsten Peust, 2013

## Abstract

Basic vocabulary lists are an important tool in comparative and historical linguistics. They provide the base for estimating the time depth of language families by the technique of glottochronology<sup>1</sup> or by other statistical methods.<sup>2</sup> The lists are composed of vocabulary that is intended to be as stable as possible diachronically. Several such lists have been proposed, the most famous ones being those by Morris Swadesh.<sup>3</sup> Swadesh did not elaborate on how and why he made exactly this selection of words.<sup>4</sup> Swadesh's original lists have not convinced everyone, so that various modifications of his lists were proposed by others.<sup>5</sup> Also for all of the modified lists, the choice of items has either not been justified at all or can be criticized for methodological reasons.

I am proposing here a methodology as well as (limited) empirical data for ranking meanings according to their diachronic stability, in order to construct a revised basic vocabulary list. As a result, I arrive at a list of 54 items at the end of this paper. More empirical data can and should be added in the future in order to further improve on the list.

## Lexical stability

A basic vocabulary list is defined by a set of meanings in a meta-language (such as English). Based on this template list, corresponding lists can be established for any language by translating each meaning into the target language. In the translation, according to Swadesh, the most frequent and most basic or general term of the target language must be chosen.<sup>6</sup> The glottochronological method then involves counting the cognate terms in the basic vocabulary lists of two languages or of two diachronic stages of a single language. Under the assumption that the average replacement rate per time is largely language independent for a given list of meanings, the cognate count allows for an estimation of the time distance between both languages. In order to calibrate the glottochronological model, i.e. to determine the replacement rate per time for a given vocabulary list, languages should be chosen whose history is well known and which allow for good cognacy judgments, whereas cognacy will be harder to judge, and perhaps rely only on sound similarity, as the method is applied to languages with no attested history or to long distance relationships. When the observed cognacy rate falls below a critical level, it may therefore become indistinguishable from random similarities between unrelated languages, so that the glottochronological method can no longer be applied. As the list is composed of more stable items, the limit for the applicability of glottochronology can be pushed further into the past.

Diachronic stability of a term in a language during a certain time interval means that the most frequent and basic term for the given meaning is not replaced by any possible competitor term during that interval.<sup>7</sup> On the other hand, a replacement of a term takes place when a competitor term raises its frequency and generalizes its meaning to the degree that it in turn becomes the primary term for the given meaning. The competitor term may either be a native term with an originally different meaning or a loan word from another language. It can be assumed that two factors in particular contribute to the stability of a term:

(1) Frequency. A term that is itself frequent is difficult to challenge, in terms of frequency, by a competitor term. In addition, frequent terms are firmly rooted in the memory of speakers, and known to all speakers of the language community, which favours their stability.<sup>8</sup>

---

1 A method developed by Morris Swadesh which tries to measure the degree of language relationship based on the hypothesis that the lexical replacement rate of a given vocabulary list is approximately constant for all languages and ages.

2 A basic vocabulary list (in that case, Swadesh's 200-item list) is used in Kessler's (2001) Monte Carlo approach to decide upon language relationship.

3 Swadesh (1955).

4 "Swadesh apparently selected items for his lists by a combination of intuition and experience (...). Swadesh calculated a percent persistence factor for each item, based on eight old-world languages, but these factors were not used in deciding what items to keep and what to drop (...)" (Oswalt 1971: 422).

5 E.g. by Bender (1983: 266-281), Dolgopolsky (1986: 34f.), Elbert (1953: 150f.), Halayqa (2007), Holman (2008 *et al.*), Starostin (2000: 257 note 25), Tadmor (2009: 68-75), Woodward (1993: 17) and Yakhontov (cited in Starostin 1991: 59f.).

6 The choice of the best term for a meaning in a given language can, of course, sometimes be disputable, which forms one of the major points of criticism on Swadesh's use of vocabulary lists. While uncertainty about the most adequate translation adds some statistical noise on the results, it does not, in my view, invalidate the glottochronological method in any fundamental way.

7 The gradual phonetic evolution, which all words of a language continually undergo, does not count as a replacement. Also expansions of a term by affixes are not normally counted as a replacement.

8 This relationship is widely acknowledged, cf. e.g. Dyen (1960: 37): "... it is reasonable to suppose that the more common a word is, the less likely it is to be replaced"; van Hout & Muysken (1994: 53): "The more frequent a word in the Quechua data base, the less the chance that it is Spanish. This suggests indirectly that frequency in the recipient language may operate as an inhibiting factor [for borrowing, C.P.]"; Tadmor (2009: 74): "It seems logical that frequently used words would also be highly resistant to borrowing, because more time and effort would be needed for the borrowing to become established." A study which confirms the correlation for Indo-European languages by statistical methods is Pagel *et al.* (2007).

(2) Semantic distinctness. A term whose meaning is unsharp and highly conventional is apt to change more easily than a term whose meaning is clear-cut and expresses a concept that exists (more or less) a-priori. This is the reason why more nouns than verbs can be found among the most stable lexical items. The world of nouns tends to reflect notional concepts which have a more or less a-priori existence, whereas the world of verbal ideas often involves concepts whose definitions are more vague and arbitrary.<sup>9</sup> As a result, the most stable lexical items should be such that are both frequent and stand for concepts with clear-cut meanings.

(3) In addition to these language-independent factors, there can be factors specific to a term in a given language which influence its prospects of remaining stable. If a word happens to be in some respect special, e.g. because it has an irregular inflexion, or if – either through shortness or through accidental similarity – it is in danger of homonymic clash with other terms, the pressure will be high for it to be replaced in many daughter languages even if the meaning itself is a stable one.

## Selection of language couples

It is evident that the stability of a meaning can only be determined empirically.<sup>10</sup> To this purpose, I use a data table which indicates for several candidate meanings how many cognates they share in a number of language couples. I pose three requirements on the selection of the language couples: (1) All the couples are independent from each other, (2) both languages of the couple are actually attested languages, (3) the languages of the couple have a well-known history so that (relatively) safe cognacy judgments are possible.

A fourth potential requirement could be that the chosen language couples should be genetically and geographically diverse. I believe that this requirement, which in practice often contradicts requirement (3), is of lesser importance under the assumption that the glottochronological hypothesis of a language-independent replacement rate, as assumed by Swadesh, is correct.

There have been studies where, as I do here, meanings were ranked according to the cognate preservation count in a number of language couples.<sup>11</sup> In all studies I am aware of, however, the couples were chosen so that the three requirements mentioned above were not all met, particularly not the first one. The former studies typically used data from several interrelated couples out of a single genetic stock. I believe that this can seriously flaw the results. The independence requirement is important for at least two reasons. First, a word can be instable in a language for a language specific reason (as explained under 3 in the preceding section), so that it is at risk to be replaced in many daughter languages even though the semantic concept as such is a stable one. Second, some of the daughter languages may form an unrecognized subgroup within the language family. If a word happened to get lost in the proto-language of that subgroup, it would appear to be missing in all daughter languages although only a single loss occurred.

## Selection of lexical entries

The lines of the table contain the candidate meanings. These are 180 meanings for which I considered it possible that they might end up in a reasonable basic vocabulary list. The candidate list includes almost all members of Swadesh's 100-item list with the exception of "claw", which I replaced by "(finger)nail"<sup>12</sup>, and "to walk", which I replaced by "to go"<sup>13</sup>, as well as several items picked from competing basic vocabulary lists. I have also put to test some words which Swadesh rejected as being "cultural vocabulary", such as "brother" or "house".

## Cognate judgements

Entries are considered cognates if they are etymologically identical at least for their greater part. I accept different affixes or compounding with another element, provided that there is still a substantial part in common. The symbols "]" and "[" indicate prefixed or suffixed additional material.

Although I have attempted to select language couples whose mutual historical relationship is relatively well-known, the judgment on the cognacy of words is not always straightforward, and I have certainly not been able to

---

9 To give just one example, the borderline between meanings such as "man" and "woman", or between "dog" and "cat", has a higher a-priori reality than the borderline between "to go" and related meanings such as "to run", "to come", "to move", etc.

10 It might become possible in the future to predict the stability of a meaning from, e.g., its frequency and its semantic distinctness, but there is so far no known way of measuring the latter. Frequency would, again, have to be measured empirically.

11 Dolgopolsky (1986); Dyen (1964: 242f.); Dyen & James & Cole (1975: 185f.); Holman *et al.* (2008); Kruskal & Dyen & Black (1973: 38f.); Oswalt (1971); Swadesh (1955: 133-137); probably also Lohr (1998), which was not accessible to me. Tadmor (2009: 68-75) provides a ranked 100-item basic vocabulary list which was created on a large statistical basis but considers diachronic stability only as one among several criteria.

12 Both are synonyms in many languages, but in case of divergence I decided to prefer the human term, as is generally so for the other body part terms of the Swadesh list.

13 As other users of the Swadesh lists have already done, because "to walk" has no obvious elementary translation in many languages.

avoid errors completely. Apart from uncertainty about the linguistic history of the word, the judgment can be a matter of definition even where we are informed perfectly. I have adopted the following principles:

- (1) When one of both languages has borrowed a term directly from the other, the terms are considered non-cognate.<sup>14</sup>
- (2) When both languages borrowed their terms independently from a third language, they are considered non-cognate.<sup>15</sup>
- (3) When language A borrowed a term from C where again it is cognate to the term of B, the terms of A and B are considered non-cognate.<sup>16</sup>
- (4) When both A and B borrowed a term from a third source C so early that the borrowing may well have taken place in the common ancestor of A and B, the terms are considered cognate<sup>17</sup>.

## Ranking the items and extracting a basic vocabulary list

Based on the cognate counts of the list, the meanings can be ranked according to their diachronic stability. My measure of the stability of a meaning is simply the number of language couples within my sample that preserve it as cognates. This measure makes sense although the couples differ in their degrees of relationship: Some of them are related much more closely (e.g. English – German) than others (e.g. Finnish – Hungarian), as can be seen in the cognate summations at the bottom of the table. Nevertheless, one can assume that a meaning with a higher count is always likely to be more stable than a meaning with a lower count, irrespectively of which individual couples contribute to the counts.

An intuitive argument for this could be the following: In many cases, a meaning will show up as a cognate in a close couple but not so in a more distant couple. If we encounter, for another meaning, the opposite case, namely the preservation as a cognate only in the distant couple but not in the close couple, one could argue either that this latter meaning is more stable (since it was preserved even in the distant couple) but also that it is less stable (since it was lost even from the close couple) than the first meaning. A more formal proof could look as follows:

*Proof.* Under current glottochronological assumptions, for any concept  $w$  there will be a fixed probability  $p(w)$  for it to survive over a given time interval, say a millennium. Given a language couple  $l$  separated by  $m$  millennia and a word list  $w_1, w_2, w_3, \dots$ , the expected number of surviving cognates  $C(l)$  will be  $p(w_1)^m + p(w_2)^m + p(w_3)^m + \dots$ . Given another couple  $l'$  separated by  $m'$  millennia, we expect  $C(l') = p(w_1)^{m'} + p(w_2)^{m'} + p(w_3)^{m'} + \dots$  cognates. It is obvious that observed cognate counts  $C(l') > C(l)$  imply that  $m' < m$  (and vice versa), irrespectively of which individual cognates contribute to the counts.

Once the meanings have been ranked, an  $n$ -item list can be extracted by selecting the top  $n$  items from the list. There is a tradeoff between the desire to maximize the average stability on the one hand and to have a long list (in order to reduce statistical noise in the application of the list) on the other. There is no known way of how to ideally balance these competing desires, and the purpose for which the list is going to be used may be relevant here as well.

In any case, it must be emphasized that the items of any list will not all have the same degree of stability,<sup>18</sup> so that any stability rate that can be estimated for a given list is only an average value over all list items.

## The data table

*First column:* Description of the word meaning

*Second column:* Indicates for a number of important basic vocabulary lists whether the given word was included in them: “1” = Swadesh 100-item list; “2” = Swadesh 215-item list (both in Swadesh 1955); “B” = first 100-item list by Bender (1983: 266ff.); “b” = 10-item list devised by Herman Bell, which is provided for each language article in the *Encyclopaedia Aethiopica*; “β” = 20-item list by Brinton 1891 (cited from Hymes 1973: 129); “D” = 15-item list by Dolgopolsky (1986: 34f.); “H” = 40-item list by Holman *et al.* (2008); “S” = 55-item list by S. Starostin (2000: 257 note 25); “Σ” = 50-item list by G. Starostin (2010); “T” = 100-item list by Tadmor (2009: 68-75); “Y” = 35-item list by Yakhontov (cited from Starostin 1991: 59f.).

<sup>14</sup> Among the language couples chosen here, this situation arises particularly often for Hindi which has borrowed a lot of words, including basic vocabulary, from Persian.

<sup>15</sup> E.g. English *round* and German *rund*, both from Old French.

<sup>16</sup> E.g. English *flower* < French *fleur* = German *Blume*, or Amharic *ṭägur* (older *ṣägwr*) “hair” < Cushitic and here probably related to Hebrew *se'ar*. A borderline case, which I likewise count as non-cognate, is Engl. *fruit* < French *fruit* < Latin *fructus* and German *Frucht* < Latin *fructus*.

<sup>17</sup> E.g. Irish *chlúmh* = Welsh *plu* “feather”, both from Latin *pluma*, or Finnish *sata* = Hungarian *száz* “hundred”, both from an early Indo-European language (cf. Sanskrit *śatam*).

<sup>18</sup> As was clear already to Swadesh (1952: 457): “A stability score for individual items could be calculated, and this score taken into account in constructing [an] improved test list.”

Cells are marked by “–” when either the entries would not be cognates, or when one of the languages lacks an obvious unmarked term for that meaning.

This is work in progress. I intend to add more language couples to the list in order to expand the empirical basis in the future.

gloss	presence in previous lists	Amharic = Modern Hebrew <sup>19</sup>	Bahasa Indonesia = Malagasy <sup>20</sup>	Bulgarian = Latvian <sup>21</sup>	Egyptian (Old Kingdom) = Coptic (Bohairic dialect) <sup>22</sup>	English = German	Finnish = Hungarian	French = Romanian	Hindi = Persian	Irish = Welsh <sup>23</sup>	Kabyle = Tuareg (Ahaggar dialect)	Oromo = Somali <sup>24</sup>	Swahili = Zulu <sup>25</sup>	Turkish = Yakut <sup>26</sup>
all	12B	<i>hullu=kol</i>	–	<i>vséki=viss</i>	<i>nb=nib[en</i>	<i>all=alle</i>	–	<i>tout=tot</i>	–	<i>uile=holl</i>	–	–	<i>-ote=-onke</i>	–
ant	T	–	–	–	–	<i>ant=Ameise</i>	–	<i>fourmi=furni cǎ</i>	–	–	–	–	–	–
ash(es)	12BΣT	–	–	<i>pépel=peľni</i>	–	<i>ashes=Asche</i>	–	<i>ceindre=cen uşǎ</i>	–	<i>luaith=lludw</i>	<i>işed=eşed</i>	–	–	<i>kül=kül</i>
to ask	–	–	<i>ber]tanya=m anon]tány</i>	–	<i>şni=şini</i>	–	–	–	<i>pūčh=porsīd an</i>	–	–	–	–	–
back (of body)	2T	–	–	–	<i>s3=soi</i>	–	–	–	<i>pīḥ=pošt</i>	–	<i>aşrur=årori</i>	–	–	–
bad	2	–	–	–	–	–	–	–	–	–	<i>yir=erk</i>	<i>hamaa=xun (xum-)</i>	–	–
bark (of tree)	12BS	<i>qarfit=klipa</i>	<i>kulit=hodi[k ázo<sup>27</sup></i>	–	–	–	–	–	–	<i>rúsc=rhisgl</i>	–	–	–	–
to bear / to give birth	–	<i>wállädä=yal ad</i>	–	–	<i>msi=misi</i>	<i>bear=ge]bär en</i>	–	–	<i>janm=zāyīda n</i>	–	<i>arew=aru</i>	<i>ḍala=dhal</i>	<i>-zaa=-zala</i>	–
belly	12BS	–	–	–	–	–	–	–	–	<i>bolg=bol</i>	–	–	–	–
big	12BST	–	–	–	–	–	–	–	–	<i>mór=mawr</i>	<i>meqqweř=mǎ qqaran</i>	–	–	–
bird	12BΣT	–	<i>burung=vórona</i>	–	–	–	–	–	–	<i>éan=edn</i>	–	<i>simbirroo=sh imbir</i>	–	–
to bite	12BT	<i>nākkāsā=naş ax</i>	<i>menggigit=m anáikitra</i>	–	–	<i>bite=beißen</i>	–	–	–	–	–	<i>činiina=qanii n</i>	<i>-uma=-luma</i>	<i>ısr=ıtır</i>

19 Cf. Leslau (1969) who compared the same pair of languages. My transcription of Hebrew refers to the modern Israeli pronunciation.

20 I consider the “Austronesian Basic Vocabulary Database” (<http://language.psy.auckland.ac.nz/austronesian/>).

21 I consider the Latvian Swadesh list with etymological annotations by Holst (2001: 213-222).

22 Data from personal knowledge. I cite both languages in their conventional transliterations which, as should be noted, must not be taken as a phonological rendering. In fact, the conventional transliteration of Egyptian suggests a greater phonetic similarity to Coptic than was actually the case (note in particular that <3> = /r/, <’> = /ð/, <š> = /x/). In some cases where the meaning is not yet attested in sources from the Old Kingdom, I have supplied words used in the Middle Kingdom (“dream”, “flea”, “leaf”, “mouse”, “root”, “tear”).

23 I consider Lucht (2007).

24 Somali is given in its standard orthography (note in particular <c> = /ʃ/, <dh> = /d/, <x> = /ħ/), Oromo in a common orientalist transcription.

25 I consider “Bantu Lexical Reconstructions 3” (<http://www.africanmuseum.be/collections/browsecollections/humansciences/blr>).

26 I use a transcription of Yakut close to the orthography of modern Turkish.

27 Basically the same cognate pair as for “skin”.

bitter	T	<i>mārara=mar</i>	–	–	–	<i>bitter=bitter</i>	–	<i>amer=amar</i>	–	<i>searbh=chwe rw</i>	–	<i>haḍḍaa=qad haadh</i>	–	<i>aci=ahu</i>
black	12BSΣT	–	–	–	<i>kmm=khame</i>	–	–	<i>noir=negru</i>	–	<i>dubh=du</i>	<i>aṣṣṭaf=sāṭṭāf ān</i>	–	–	<i>kara=xara</i>
blood	12BbHΣ TY	<i>dām=dam</i>	<i>darah=ra</i> <sup>28</sup>	–	<i>znf=snof</i>	<i>blood=Blut</i>	<i>veri=vér</i>	<i>sang=sânge</i>	–	–	–	<i>ḍiiga=dhiig</i>	–	<i>kan=xaan</i>
bone	12BHΣT Y	<i>aṭant=étsem</i>	<i>tulang=táola na</i>	–	<i>qs=kas</i>	–	–	<i>os=os</i>	–	–	<i>iḡess=eḡās</i>	<i>lafee=laf</i>	–	–
breast <sup>29</sup>	1BHT	–	–	–	<i>mnd=mnot</i>	<i>breast=Brust</i>	–	<i>sein=sân</i>	–	–	–	–	–	–
brother	2	–	–	<i>brat=brālis</i>	<i>sn=son</i>	<i>brother=Bru der</i>	–	<i>frère=frate</i>	<i>bhāī=berāda r</i>	<i>deJartháir=b rawd</i>	<i>egma=āḡḡa</i>	–	–	–
to burn (intr.)	12BST	–	–	–	–	<i>burn=brenne n</i>	–	–	–	–	<i>ṛeḡ=ārāḡ</i>	<i>guba[ḍḍa=gu bo</i>	–	–
to carry	T	–	–	<i>nósja=nest</i>	–	–	–	–	–	–	<i>awi=awi</i>	–	–	–
child	2T	<i>lāḡ=yéled</i>	–	–	–	<i>child=Kind</i> <sup>30</sup>	–	–	–	–	–	–	<i>mwana=inga ne</i>	–
cloud	12B	–	<i>awan=ráhon a</i>	–	<i>jgp=čhēpi</i>	–	<i>pilvi=felhő</i>	–	–	–	<i>asigna=aḡḡon na</i>	–	–	<i>bulut=bilit</i>
cold	12B	–	–	–	<i>qbb=khbob</i>	<i>cold=kalt</i>	–	–	–	<i>fuar=oer</i>	<i>asemmaḍ=is maḍ</i>	<i>qabbanaa=q abow</i>	–	–
to come	12BHT	–	–	–	<i>jwi=i</i>	<i>come=komm en</i>	–	<i>venir=veni</i>	–	–	<i>as=as</i>	–	<i>-ja=-za</i>	<i>gel=kel</i>
to cut	2B	–	–	–	<i>š´=šōt</i>	–	–	–	–	–	–	–	–	–
day(=not night)	2	–	<i>hari=ádro</i> <sup>31</sup>	<i>den=diena</i>	<i>hrw=ehoou</i>	<i>day=Tag</i>	–	<i>j[our=zi</i>	–	–	<i>azal=ahāl</i>	–	–	<i>gün=kün</i>
to die / dead	12BDHΣΣ	<i>motā=met</i>	<i>mati=máty</i>	<i>u]míram=mi rt</i>	<i>mwt=mou</i>	–	<i>kuolla=hal</i>	<i>mourir=muri</i>	<i>marnā=mord an</i>	–	<i>emmet=ämm āt</i>	–	<i>-fa=-fa</i>	<i>öl=öl</i>
to dig	2	<i>qwäffärä=xaf ar</i>	<i>menggali=mi hády</i>	–	–	–	–	–	–	–	<i>eḡz=āḡāh</i>	<i>qota=qod</i>	<i>-ch]imba=- mba</i>	<i>kaz=xas</i>
to do	T	–	–	–	<i>jri=iri</i>	<i>do=tun</i>	<i>tehdā=tesz</i>	<i>faire=face</i>	<i>karnā=karda n</i>	<i>déanaim</i> <sup>32</sup> = <i>g wneud</i>	<i>eg=āḡ</i>	–	–	–
dog	12BHΣT Y	–	–	–	–	–	–	<i>chien=câine</i>	–	–	<i>aydi=eydi</i>	–	<i>mbwa=inja</i>	–
dream	–	<i>əlm=xalom</i>	<i>mi]mpi=nófy</i>	<i>sən=sapnis</i>	<i>rsw.t=rasoui</i>	<i>dream=Trau m</i>	–	–	<i>xāb=sapnā</i>	<i>bruadar=bre uddwyd</i>	<i>targit=tāharḡ it</i>	–	–	<i>düş=tüül</i>

28 Correspondence Indones. *d* = Malag. *r* as in “leaf”, “two”, “winter”.

29 In case of conflict I prefer words for “female breast(s)”.

30 This etymology is not generally accepted, but I consider it to be correct in view of identical semantics, gender and plural formation, with only an unexplained *n~l*-variation.

31 Despite some uncertainties I consider it probable that these words are cognate. The initial is as in “liver” or “rain”, Malagasy *-ndr-* can be the reflex of a former *\*-r-* following an *-n-* (which is missing from the Indonesian form), cf. “to spit” for a similar situation. Tagalog *āraw* “day” is probably related as well.

32 Old Irish *do-gní-*.

to drink	12BHΣT	–	–	–	<i>zwr=sō</i>	<i>drink=trinke n</i>	<i>juoda=iszik (iv-)<sup>33</sup></i>	<i>boire=bea</i>	–	–	<i>sew=əsəw</i>	–	–	<i>iç=is</i>
dry	12BSΣ	–	–	<i>suh=saus</i>	<i>šw=šōoui</i>	<i>dry=trocken</i>	–	–	<i>sūkhā=xošk</i>	–	<i>aqufan=yəqq urān</i>	<i>gogaa=enge g</i>	–	<i>kuru=kuraan</i>
ear	12BβHΣ TY	–	–	<i>uhó=auss</i>	<i>msḍr=mašč</i>	<i>ear=Ohr</i>	–	<i>oreille=urec he</i>	–	<i>cluas=clust</i>	<i>amezzuγ=ta māzzuk</i>	–	–	<i>kulak=kulga ax</i>
earth / soil	12T	–	<i>tanah=tány</i>	<i>zemjá=zeme</i>	–	<i>earth=Erde</i>	–	–	–	–	<i>akal=ākal</i>	–	–	–
to eat	12BSΣT	–	<i>makan=mihí nana</i>	<i>jam(jad-)=ēs t(ēd-)</i>	<i>wnm=ouōm</i>	<i>eat=essen</i>	<i>syödä=eszik<sup>34</sup></i>	<i>manger=mân ca</i>	–	–	<i>eçç=ākš</i>	–	<i>-la=-dla</i>	<i>ye=sie</i>
egg	12BSΣTY	–	<i>telur=atódy<sup>35</sup></i>	–	<i>swḥt=sōouhi</i>	<i>egg=ei</i>	–	<i>œuf=ou</i>	–	<i>ubh=wy</i>	–	<i>hanqaaquu= ugax<sup>36</sup></i>	–	<i>yumurta=sim ut</i>
eight	2	<i>səmmənt=šm one</i>	–	<i>ósem=astoŋi</i>	<i>ḥmnr=šmēn</i>	<i>eight=acht</i>	–	<i>huit=opt</i>	<i>āḥ=hašt</i>	<i>ocht=wyth</i>	–	<i>saddeet=sidd eed</i>	–	<i>sekiz=ayıs</i>
eye	12BβDHS ΣTY	<i>ayn=áyin</i>	<i>mata=máso</i>	<i>okó=acs</i>	–	<i>eye=Auge</i>	<i>silmä=szem</i>	<i>œil=ochi</i>	–	–	<i>tiṭ=teṭ</i>	<i>ija=il</i>	<sup>37</sup>	–
to fall / to drop	2T	–	–	–	–	<i>fall=fallen</i>	–	–	–	–	–	–	–	<i>düš=tüs</i>
far	2T	<i>ruq=raxok</i>	–	–	–	–	–	–	<i>dūr<sup>38</sup>=dūr</i>	–	–	<i>fagoo=fog</i>	–	–
fat / grease	12B	–	–	–	–	<i>fat=Fett</i>	–	<i>graisse=gräs ime</i>	–	–	–	–	<i>mafuta=amaf utha</i>	<i>yağ=sia</i>
father	2	<i>abbat=av</i>	–	–	<i>jtj=iōt</i>	<i>father=Vater</i>	–	–	<i>pitā=pedar</i>	–	<i>baba=abba</i>	<i>abbaa=aabb e</i>	<i>baba=ubaba</i>	–
to fear/be afraid	2	–	<i>takut=ma]tá hotra</i>	<i>bojá=baidītie s</i>	–	<i>fear=fürchte n<sup>39</sup></i>	<i>pelätä=fěl</i>	–	–	–	–	–	–	–
feather	12B	–	<i>bulu=volo[m bórona<sup>40</sup></i>	–	–	<i>feather=Fede r</i>	–	–	–	<i>clúmh=plu</i>	–	<i>baalle=baal</i>	–	–
finger	–	<i>ṭat<sup>41</sup>=etsba</i>	–	<i>prəst=pirksts</i>	<i>ḍb´=tēb</i>	<i>finger=Finge r</i>	–	<i>doigt=deget</i>	<i>uŋglī=angošt</i>	–	<i>aḍad=aḍaḍ</i>	–	–	–
fire	12BbβHS ΣTY	<i>əsat=eš</i>	<i>api=áfo</i>	<i>ógən=uguns</i>	–	<i>fire=Feuer</i>	–	<i>feu=foc</i>	–	<i>tine=tân</i>	<i>timess=temse</i>	–	–	–
fish	12BHTY	–	–	–	–	<i>fish=Fisch</i>	<i>kala=hal</i>	<i>poisson=pešt e</i>	<i>maçhlī=māhī</i>	–	<i>aslem=āsulm āy</i>	<i>qurṭummii=k alluun<sup>42</sup></i>	–	<i>balık=balık</i>

33 Root \**juγ*-.

34 Root \**sev*-.

35 Correspondence Indones. *l* = Malag. *d* as in “five”, “skin”.

36 Probably cognate although the sound correspondences are not entirely clear. The form *anqoqəho* “egg” of Gəʿəz seems to be a borrowing from a related older Cushitic language.

37 But the plural forms are cognate: Swahili *macho* = Zulu *amehlo*.

38 This word could formally be a borrowing from Persian, but it is common in most Indo-Aryan languages and thus probably inherited.

39 Assuming that there is a connection between the Germanic roots \**fār*- and \**furh-t*-, which is not uncontroversial.

40 Lit. “hair of bird”; *volu* in isolation changed its meaning to “hair”.

41 From Gəʿəz *äšba t*.

42 Somali has *kalluum*- in derivatives. The geminate *-ll-* points to an original consonant cluster which was probably *-lṭ-* as still in Sidamo *qilṭim e* “fish”.

five	2β	<i>amməst=xam eš</i>	<i>lima=dímy</i>	<i>pet=pieci</i>	<i>djw=tiou</i>	<i>five=fünf</i>	<i>viisi (viite-)=öt</i>	<i>cinq=cinci</i>	<i>pānč=panj</i>	<i>cúig=pump</i>	–	<i>šan=shan</i>	<i>-tano=-hlanu</i>	<i>beş=bies</i>
flea	–	–	–	<i>bəlxá=blusa</i>	<i>py=phēi</i>	<i>flea=Floh</i>	–	<i>puce=purice</i>	–	–	–	<i>tafkii=takfi</i>	–	–
flower	2	–	<i>bunga=voni[nkázo]</i>	–	–	–	–	<i>fleur=floare</i>	–	<i>bláth=blodyn</i>	–	–	–	–
fly (animal)	BT	<i>zəmb=zvuv</i>	<i>lalat=lálitra</i>	<i>muxá=muša</i>	<i>'ff=af</i>	<i>fly=Fliege</i>	–	<i>mouche=muscă</i>	<i>makkhī=magas</i>	<i>cuileog=cleren</i>	<i>izi=ehi</i>	<i>titiisa=diqsi</i> <sup>43</sup>	–	–
to fly	12	–	–	<i>letjá=lidot</i>	–	<i>fly=fliegen</i>	–	–	–	–	–	–	–	–
foot	12BβSΣT	–	–	–	–	<i>foot=Fuß</i>	–	<i>pied=picior</i>	<i>pair=pā</i>	–	<i>aḏaṭ=aḏār</i>	–	–	<i>ayak=atax</i>
four	2β	<i>aratt=arba</i>	<i>empat=éfatra</i>	<i>čétiri=četri</i>	<i>fdw=ftoou</i>	<i>four=vier</i>	<i>neljä=négy</i>	<i>quatre=patru</i>	<i>čār=čahār</i>	<i>ceathair=pedwar</i>	–	<i>afur=afar</i>	<i>-nne=-ne</i>	<i>dört=tüört</i>
fruit	–	<i>fəre=pri</i>	<i>buah=voa[nkázo]</i>	–	–	–	–	<i>fruit=fruct</i>	–	–	–	–	–	–
full	1HSY	<i>mulu=male</i>	<i>penuh=féno</i>	<i>pálen=pilns</i>	<i>mḥ=meh</i>	<i>full=voll</i>	–	<i>plein=plin</i>	<i>pūrā=por</i>	<i>lán=llawn</i>	<i>açaṭan=ăḏkār</i>	–	–	<i>dolu=toloru</i>
to give	12BTY	–	–	<i>dávam=dot</i>	<i>rḏi=ti</i>	<i>give=geben</i> <sup>44</sup>	<i>antaa=ad</i>	<i>donner=da</i>	<i>dēnā=dādan</i>	–	<i>efk=ăkf</i>	–	<i>-pa=-pha</i>	<i>ver=bier</i>
to go	BT	–	–	–	<i>šm=še</i>	<i>go=gehen</i>	<i>mennā=megy</i>	–	–	–	–	–	–	–
good	12BT	–	–	–	–	<i>good=gut</i>	–	<i>bon=bun</i>	–	–	<i>elhu=alāy</i>	–	–	–
grass	2B	–	–	–	<i>smw=sim</i>	<i>grass=Gras</i>	–	<i>herbe=iarbă</i>	–	<i>féar=gwair</i>	–	–	–	<i>ot=ot</i>
green	12	–	<i>hijau=má]its</i>	<i>zelén=za]š</i>	<i>w3ḏ=ouotouet</i>	<i>green=grün</i>	–	<i>vert=verde</i>	–	<i>glas=glas</i>	–	–	–	–
hair (of head)	12BβSΣT	–	–	–	–	<i>hair=Haar</i>	–	–	–	–	<i>anzad=emzād</i>	–	<i>nywele=iz]inwele</i>	–
hand	12BβHΣTY	<i>əḡḡ=yad</i>	<i>tangan=tána</i>	<i>rəká=roka</i>	–	<i>hand=Hand</i>	<i>kāsi=kéz</i>	<i>main=mānā</i>	<i>hāth=dast</i> <sup>45</sup>	<i>lámh=llaw</i>	<i>afus=ăfus</i>	–	–	<i>el=ilii</i>
head	12BΣ	<i>ras=roš</i>	–	<i>glavá=galva</i>	–	–	<i>pää=fej</i>	–	<i>sir=sar</i>	<i>ceann=pen</i>	–	<i>mataa=mada</i>	–	<i>baş=bas</i>
to hear	12BHΣT	<i>samma=šama</i>	–	–	<i>sḏm=sōtem</i>	<i>hear=hören</i>	<i>kuulla=hall</i>	–	–	<i>cluim=clywed</i>	<i>sel=əsəl</i>	–	–	<i>išit=ihit</i>
heart	12BDΣ	<i>labb=lev</i>	–	<i>sərcé=sirds</i>	<i>ḥ3tj=hēt</i>	<i>heart=Herz</i>	<i>sydän=szív</i>	–	–	–	<i>ul=ul</i>	<i>onnee=wadne</i>	–	<i>yürek=sürex</i>
heavy	T	<i>kābbad=kaved</i>	<i>berat=ma]vé</i>	–	–	–	–	–	–	<i>trom=trym</i>	<i>aḏayan=iḏay</i>	<i>ulfaa]taa=culus</i>	–	<i>aḡir=iar</i>
hedgehog	–	–	–	<i>taral]éz=ezis</i>	–	–	<i>siili=sün</i>	<i>hérisson=arici</i>	–	<i>gráinneog=draenog</i>	<i>inisi=tekānesit</i>	–	–	–
honey	–	–	–	<i>med=medus</i>	<i>bj.t=ebiō</i>	<i>honey=Honi</i>	–	<i>miel=miere</i>	–	<i>mil=mêl</i>	–	–	–	–

43 Somali *-q-* is here a development from *\*-ḥ-* (cf. Rendille *ḏaḥassi* “fly”), which was regularly lost in Oromo. The initial *ti-* of Oromo must be the result of a reduplication.

44 The true English cognate is an earlier English form *yive* which was reshaped under Scandinavian influence. This is a borderline case which I count as related.

45 Sanskrit *hasta-*.

horn	1BHΣTY	<i>qānd=kéren</i>	<i>tanduk=tánd roka</i>	<i>rog=rags</i>	<i>‘b=tap<sup>46</sup></i>	<i>horn=Horn</i>	<i>sarvi=szarv</i>	<i>corne=corn</i>	–	–	<i>iccew=isək</i>	<i>gaafa=gees</i>	–	<i>boynuz=muos</i>
house	βT	<i>bet=báyit</i>	–	–	–	<i>house=Haus</i>	–	–	–	<i>teach=tȳ</i>	–	–	–	–
hundred	–	<i>māto=me’a</i>	<i>sejratus=zát o</i>	<i>sto=simts</i>	<i>šn.t=še</i>	<i>hundred=hundert</i>	<i>sata=száz</i>	–	<i>sau=sad</i>	<i>céad=can</i>	–	–	–	<i>yüz=süüs</i>
hunger / (to be) hungry	–	<i>rabä=ra’ev</i>	–	–	<i>ḥqr=hko</i>	<i>hunger=Hunger</i>	–	<i>faim=foame</i>	–	–	<i>laḏ=laḏ</i>	–	<i>njaa=indlala</i>	<i>aç=aččik</i>
I	12BDHΣTY	<i>āne=ani</i>	<i>aku=áho</i>	<i>az=es</i>	<i>jnk=anok</i>	<i>I=ich</i>	–	–	<i>maiṛ=man</i>	<i>mé=mi</i>	<i>nekk=nāk</i>	<i>ani=ani[ga]</i>	<i>mimi=mi[na]</i>	<i>ben=min</i>
ice	2	–	–	<i>led=ledus</i>	–	<i>ice=Eis</i>	<i>jää=jég</i>	<i>glace=gheatǎ</i>	–	<i>oighear=iâ</i>	<i>agris=eyāres</i>	–	–	<i>buz=muus</i>
to kill	12BΣ	–	<i>membunuh=mamóno</i>	–	–	–	–	–	–	–	<i>ney=āny</i>	–	<i>-ua=-bulala</i>	<i>öldür=ölör<sup>47</sup></i>
knee	1BHST	–	–	<i>koljáno=celis</i>	–	<i>knee=Knie</i>	–	<i>genou=genu nchi</i>	–	<i>glúin=glin</i>	–	<i>jilba=jilib</i>	–	–
to know	12BTY	–	–	<i>znam=zināt</i>	–	–	–	–	<i>jānnā=dānes tan</i>	–	<i>issin=əssən</i>	–	<i>-j[ua]=-azi</i>	<i>bil=bil</i>
to laugh	2T	<i>saqä=tsaxak</i>	–	<i>sméja=smieties</i>	<i>zbt=sōbi</i>	<i>laugh=lachen</i>	–	<i>rire=râde</i>	–	–	<i>edṣ=äṣ</i>	<i>kolfa=qosol<sup>48</sup></i>	<i>-cheka=-hleka</i>	<i>gül=kül</i>
leaf	12BHΣT	–	<i>daun=ravína</i>	–	<i>g3b.t=čōbi</i>	–	–	–	–	<i>duilleog=dalen</i>	<i>iferr=afraw</i>	–	–	–
to lie (down)	12	–	–	–	–	<i>lie=liegen</i>	–	<i>coucher=culca<sup>49</sup></i>	–	–	–	<i>čičisa<sup>50</sup>=jiif[so]</i>	<i>-lala=-lala</i>	<i>yat=sit</i>
to live/be alive	2	–	–	<i>živéja=dzīvot</i>	<i>‘nh=ōnx</i>	<i>live=leben</i>	<i>elää=él</i>	–	<i>jīnā=zende</i>	–	<i>edder=əddār</i>	–	–	–
liver	12BHST	–	<i>hati=áty</i>	–	–	<i>liver=Leber</i>	<i>maksa=máj</i>	<i>foie=ficat</i>	–	<i>ae=afu</i>	<i>tasa=awsa</i>	–	–	–
long	12BST	–	–	<i>dālæg=ilgs</i>	–	<i>long=lang</i>	–	<i>long=lung</i>	–	–	–	<i>ḏeeraa=dheer</i>	<i>-re[fu]=-de</i>	<i>uzun=uhun</i>
louse	12BDHΣTY	–	–	–	–	<i>louse=Laus</i>	<i>tāi=tetű</i>	<i>pou=păduche</i>	–	–	<i>tilkit=tillik</i>	<i>injiraan=inji r</i>	–	<i>bit=bit</i>
man (male)	12β	–	<i>laki=lehiláhy</i>	–	–	<i>man=Mann</i>	–	–	–	<i>fear=gw̄r</i>	–	–	–	<i>erkek=erkihi</i>
many	12BS	–	<i>banyak=béts aka</i>	–	–	–	–	–	–	–	–	–	<i>-ingi=-ningi</i>	–
meat / flesh	12BSΣT	–	–	–	<i>jf=af</i>	–	–	<i>chair=carne</i>	–	–	–	–	<i>nyama=inya ma</i>	<i>et=et</i>

46 Although the Egyptian consonant ‘ was normally lost by Coptic, there are some instances of preservation as a dental (also in “to cut”).

47 Causative of “to die”.

48 Regular sound shift *s > f* as well as a metathesis of adjacent consonants in Oromo (which still has *koḥla* as a variant).

49 Both have causative meaning: “to lay down”; the concept “to lie” is expressed by passive forms of this verb.

50 Oromo has *čičif-* before consonantic suffixes (e.g. *čičifta*), *s > f* before *C* being a regular alternation pattern in the language. However, as Somali and other cognate languages show, the original root should be *\*čičif-* and the forms in *-s-* were created by false analogy with verbs of the alternating type.



milk	–	–	–	–	<i>jr̥t=erōti</i>	<i>milk=Milch</i>	–	<i>lait=lapte</i>	–	–	–	<i>aannan=caano</i>	–	<i>süt=üüt</i>
moon	1BbβSΣY	–	<i>bulan=vólana</i>	–	<i>j'h=joh</i>	<i>moon=Mond</i>	<i>kuu=hold</i>	<i>lune=lunã</i>	–	–	<i>aggur=ǎyor</i>	<i>ji'a=dayax</i>	–	<i>ay=iy</i>
mother	2	–	–	<i>májka=māte</i>	<i>mw.t=mau</i>	<i>mother=Mutter</i>	–	–	<i>mātā=mādar</i>	–	<i>yemma=ma</i>	–	<i>mama=umame</i>	–
mountain	12BH	–	–	–	<i>ḡw=tōou</i>	–	–	<i>mont[agne]=munte</i>	–	–	<i>adrar=adrar</i>	–	–	–
mouse	–	–	–	–	<i>pnw=phin</i>	<i>mouse=Maus</i>	<i>hiiri=egér</i>	<i>souris=šoarece</i>	–	<i>luch=llygode n</i>	–	–	–	–
mouth	12BβSΣT	–	–	–	<i>r=ro</i>	<i>mouth=Mund</i>	<i>suu=száj</i>	–	–	–	<i>imi=emi</i>	<i>af[aa]=af</i>	<i>mdomo=umlomo</i>	–
(finger)nail / claw <sup>51</sup>	1BDΣ	<i>ṭəfər=tsipor</i>	<i>kuku=hóho</i>	<i>nókət=nags</i>	–	<i>nail=Nagel</i>	–	<i>ongle=unghie</i>	–	<i>ionga=ewin</i>	<i>iccer=eskār</i>	<i>qeensa=ciddi</i> <sup>52</sup>	–	<i>trnak=tujrax</i>
name	12BDHΣTY	<i>səm=šem</i>	–	–	<i>rn=ran</i>	<i>name=Name</i>	<i>nimi=név</i>	<i>nom=nume</i>	<i>nām=nām</i>	<i>ainm=enw</i>	<i>isem=isəm</i> <sup>53</sup>	<i>maqaa=magac</i>	–	–
narrow	2	–	–	–	<i>g3w=čēou</i>	–	–	<i>étroit=strâmt</i>	–	–	–	–	–	–
navel	T	–	<i>pusat=fóitra</i>	–	<i>hp3=xelpi</i>	<i>navel=Nabel</i>	–	<i>nombril=buric</i> <sup>54</sup>	<i>nābhi=nāf</i>	–	–	<i>handuuraaxundhur</i>	–	–
near (adj.)	2S	<i>qərb=karov</i>	–	–	–	<i>nea[r</i> <sup>55</sup> <i>=nahe</i>	–	<i>proche=apropae</i>	–	–	–	<i>ḡi'oo=dhow</i>	–	–
neck	12BST	–	–	–	–	–	–	–	–	–	–	–	–	<i>boyun=mooy</i>
nest	–	–	–	<i>gnezdó=ligzdas</i>	–	<i>nest=Nest</i>	<i>pesä=fészek</i>	–	–	<i>nead=nyth</i>	–	–	–	<i>yuva=uya</i>
new	12BHSETY	<i>addis=xadaš</i>	<i>baru=váo</i>	–	–	<i>new=neu</i>	<i>uusi=új</i>	<i>nouveau=nouu</i>	<i>nayā=now</i>	<i>nua=newydd</i>	–	–	<i>-pya=-sha</i>	<i>yeni=saja</i>
night	12BHSETY	<i>lelit=láyla</i>	–	<i>nošt=nakts</i>	<i>grh=čōrh</i>	<i>night=Nacht</i>	<i>yö=éj[szaka</i>	<i>nuit=noapte</i>	–	–	<i>iḡ=ehăḡ</i>	–	<i>usiku=ubusuku</i>	–
nine	2	–	–	<i>dévet=deviṇi</i>	<i>psḡw=psit</i>	<i>nine=neun</i>	–	<i>neuf=nouã</i>	<i>nau=noh</i>	<i>naoi=naw</i>	–	<i>sagal=sagaal</i>	–	<i>dokuz=toyus</i>
nose	12BβHSETY	–	<i>hidung=órona</i>	–	<i>šr.t=šai</i>	<i>nose=Nase</i>	–	<i>nez=nas</i>	–	<i>srón=trwyn</i>	<i>tizert=tinhar</i>	–	–	<i>burun=murun</i>
not	12DΣT	–	–	<i>ne=ne</i>	–	<i>not=nicht</i> <sup>56</sup>	–	–	<i>nahīṁ=na</i>	<i>ní=ni</i>	<i>ur=wār</i>	–	–	–
old <sup>57</sup>	2T	–	–	–	<i>jz=ap]as</i>	<i>old=alt</i>	–	<i>vieux=vechi</i>	–	<i>sean=hen</i>	–	–	–	–

51 As “claw” in Swadesh’s list.

52 Dialectal Somali also *cinji*. I assume both words to be cognate despite an irregular correspondence in the initial.

53 This could either be a loan from Arabic *ism* “two” or a native Berber term cognate to Arabic. I assume the second alternative here because this noun is not treated grammatically like Arabic loans (at least not like more recent ones).

54 From Latin *umbilicus*, with strong reshaping of the word form in French.

55 Originally a comparative, the base form *nigh* now being obsolete.

56 Both are independently created compounds from the same original elements *\*ne* + *\*wiht*. I count this as etymological identity.

57 In case of conflict “old (of things)”.

one	12BbβHΣ TY	<i>and</i> <sup>58</sup> = <i>exad</i>	–	<i>ed]in=viens</i>	<i>w'w=ouai</i>	<i>one=ein</i>	<i>yksi=egy</i>	<i>un=un</i>	<i>ēk=yek</i>	<i>aon=un</i>	<i>yiwen=iyǎn</i>	–	–	<i>bir=biir</i>
to open	–	–	<i>membuka=m</i> <i>amóha</i>	<i>otvárjam=atv</i> <i>ěrt</i>	<i>wn=ouōn</i>	<i>open=öffnen</i>	–	–	–	–	–	–	–	<i>aç=as</i>
other	2B	–	–	–	<i>kj=ke</i>	<i>other=ander</i> <i>er</i>	<i>muu=más</i>	<i>autre=alt</i>	<i>dūsṛā=diga</i> <i>r</i> <sup>59</sup>	<i>eile=ar]all</i>	<i>iḍen=hăḍǎn</i>	–	–	–
person / human being	12BH	–	–	–	<i>rmṭ=rōme</i>	<i>man=Men[sc</i> <i>h</i>	–	<i>homme=om</i>	–	<i>duine=dyn</i>	–	–	<i>mtu=umuntu</i>	–
rain	12BΣT	–	<i>hujan=órana</i>	–	<i>ḥwy.t=moun]</i> <i>hōou</i>	<i>rain=Regen</i>	–	<i>pluie=ploaie</i>	–	–	–	–	<i>mvua=imvul</i> <i>a</i>	<i>yağmur=sam</i> <i>ur</i>
red	12BT	–	<i>mérah=ména</i>	–	<i>dšr=throšreš</i>	<i>red=rot</i>	–	<i>rouge=roșu</i>	–	–	<i>azeggway=h</i> <i>ăggăyǎn</i>	–	–	<i>kızıl=kıhıl</i>
right (side)	2	–	<i>kanan=havá</i> <i>nana</i>	–	<i>wnm.j=ouina</i> <i>m</i>	<i>right=recht</i>	–	<i>droite=dreap</i> <i>tă</i>	–	<i>deas=de</i>	–	<i>mirga=midig</i>	<i>kulia=ukudl</i> <i>a</i> <sup>60</sup>	–
river	2	–	–	–	<i>jtrw=iaro</i>	–	–	–	–	<i>abhainn=afo</i> <i>n</i>	–	–	–	–
road=path= way	12BH	–	<i>jalan=lálana</i>	–	–	<i>way=Weg</i>	–	–	–	–	–	–	<i>njia=indlela</i>	<i>yol=suol</i>
root	12BT	<i>sār=šóreš</i>	<i>akar=fáka</i>	–	<i>mnj.t=nouni</i>	–	–	<i>racine=rădă</i> <i>cină</i>	–	<i>fréamh=gwr</i> <i>aidd</i>	–	<i>hundee=xidi</i> <i>d</i>	–	–
round	1S	–	–	–	–	–	–	<i>rond=rotund</i>	–	<i>cruinn=crwn</i>	–	–	–	–
salt	2TY	–	–	<i>sol=sāls</i>	<i>ḥm3.t=hmou</i>	<i>salt=Salz</i>	–	<i>sel=sare</i>	–	<i>salann=hale</i> <i>n</i>	–	–	–	<i>tuz=tuus</i>
sand	12BT	–	–	–	<i>š'j=šō</i>	<i>sand=Sand</i>	–	–	–	–	–	–	–	<i>kum=kumax</i>
to say	1BT	–	–	–	<i>ḍḍ=čō</i>	<i>say=sagen</i>	–	–	–	–	<i>ini=ǎnn</i>	–	–	<i>de=die</i>
sea	2	–	–	–	–	–	–	<i>mer=mare</i>	–	–	–	–	–	–
to see	12BHT	–	–	–	–	<i>see=sehen</i>	–	<i>voir=vedea</i>	–	–	–	<i>arga=arag</i>	<i>-ona=-bona</i>	<i>gör=kör</i>
seed <sup>61</sup>	12B	<i>zār=zéra</i>	–	<i>sé]me=sē]kla</i>	–	<i>seed=Saat</i>	–	<i>semence=să</i> <i>mânṭă</i>	–	–	–	–	<i>mbegu=imbe</i> <i>wu</i>	–
seven	2	<i>sābatt=šéva</i>	–	<i>sédem=septi</i> <i>ṇi</i>	<i>sḥw=šašf</i>	<i>seven=sieben</i>	–	<i>sept=șapte</i>	<i>sāt=haft</i>	<i>seacht=saith</i>	–	<i>torba=toddo</i> <i>ba</i>	–	<i>yedi=sette</i>
to sew	2	–	<i>menjahit=ma</i> <i>njâitra</i>	<i>šja=šūt</i>	–	–	–	<i>coudre=coas</i> <i>e</i>	–	–	–	–	–	<i>dik=tik</i>
shadow	T	<i>ṭḷa=tsel</i>	–	–	–	<i>shadow=Sch</i> <i>atten</i>	–	<i>ombre=umbr</i> <i>ă</i>	<i>čhāyā=sāye</i>	<i>scáth=cy]sgo</i> <i>d</i>	<i>tili=tele</i>	–	–	<i>gölge=külük</i>

58 The *-n-* is an irregular compensation of a lost *-ḥ-* (perhaps via *\*add*).

59 Both are derivatives from the word for “two”.

60 This root is derived from the root for “to eat” because the right hand is used for eating.

61 I attempt to choose words which mean both “semen” and “vegetable seed / grain”.

short (of things)	2S	<i>aččər=katsa</i> <i>r</i> <sup>62</sup>	–	–	–	<i>short=kurz</i> <sup>63</sup>	–	<i>court=s]curt</i>	–	–	<i>awezlan=iğh</i> <i>al</i>	<i>gabaabaa=g</i> <i>aaban</i>	<i>-fupi=-fuphi</i>	–
to sing	2	–	–	–	<i>ḥsi=hōs</i>	<i>sing=singen</i>	–	<i>chanter=cânt</i> <i>a</i>	–	<i>canaim=can</i> <i>u</i>	–	–	–	–
to sit	12B	–	–	<i>sedjá=sēdēt</i>	<i>ḥmsi=hemsi</i>	<i>sit=sitzen</i>	–	<i>as]seoir=şed</i> <i>ea</i>	–	<i>suighim=eist</i> <i>edd</i> <sup>64</sup>	<i>qqim=γaym</i>	–	<i>-kaa=-hlala</i>	<i>otur=olor</i> <sup>65</sup>
six	2	<i>səddəst=šeš</i>	<i>enam=énina</i>	<i>šest=šeši</i>	<i>sjsw=soou</i>	<i>six=sechs</i>	<i>kuusi=hat</i>	<i>six=şase</i>	<i>čhah=šeš</i>	<i>sé=chwech</i>	–	<i>ja'a=lix</i> <sup>66</sup>	–	<i>altı=alta</i>
skin	12BT	–	<i>kulit=hóditra</i>	–	–	–	–	<i>peau=piele</i>	–	<i>craiceann=cr</i> <i>oen</i>	<i>agwlim=elä</i> <i>m</i>	–	–	<i>deri=tirii</i>
sky	2	<i>sāmay=šamá</i> <i>yim</i>	<i>langit=lánitra</i> <i>a</i>	<i>nebé=debess</i>	<i>p.t=phe</i>	–	–	<i>ciel=cer</i>	–	–	<i>igenni=ağən</i> <i>na</i>	–	–	–
to sleep	12B	–	<i>tidur=ma]tór</i> <i>y</i>	–	–	<i>sleep=schlaf</i> <i>en</i>	–	<i>dormir=dor</i> <i>mi</i>	<i>sōnā=xābīda</i> <i>n</i>	–	–	–	<i>-lala=-lala</i> <sup>67</sup>	<i>uyu=utuy</i>
small	12BT	–	–	–	–	–	–	–	–	<i>beag=bach</i>	–	–	–	–
smoke	12BΣT	–	–	<i>dim=dūmi</i>	–	–	–	<i>fumée=fum</i>	<i>dhuām=dūd</i>	–	<i>abbu=āhu</i>	–	<i>moshi=umusi</i>	–
snake	2BS	–	–	–	<i>ḥʒw=hof</i>	–	–	<i>serpent=şarp</i> <i>e</i>	–	<i>nathair=neid</i> <i>r</i>	–	<i>bofa=mas</i>	<i>nyoka=inyok</i> <i>a</i>	–
snow	2	–	–	<i>snjag=sniegs</i>	–	<i>snow=Schne</i> <i>e</i>	–	–	–	–	–	–	–	<i>kar=xaar</i>
son	–	–	<i>anak=z]ának</i> <i>a</i> <sup>68</sup>	–	–	<i>son=Sohn</i>	<i>poika=fiú</i>	<i>fiis=fiu</i>	–	<i>mac=mab</i>	–	<i>ilma=inan</i>	–	<i>oğul=uol</i>
to spit	2	–	<i>meludah=ma</i> <i>ndróra</i>	<i>pljívam=s]pl</i> <i>aut</i>	<i>tf=hi]thaf</i>	<i>spit=spucke</i> <i>n</i> <sup>69</sup>	–	–	–	–	<i>susef=sutəf</i>	<i>tuf=tufa</i>	<i>mate=amath</i> <i>e</i> <sup>70</sup>	–
to stand	12BT	<i>qomā=kam</i>	–	<i>stojá=stāvēt</i>	<i>‘ḥ’=ohi</i>	<i>stand=stehen</i>	–	–	–	–	<i>ebded=əbdəd</i>	–	–	<i>dur=tur</i>
star	12BHSΣT	<i>kokāb=koxav</i>	<i>bintang=kínt</i> <i>ana</i> <sup>71</sup>	<i>zvezdá=zvaig</i> <i>zne</i>	<i>sb3=siou</i>	<i>star=Stern</i>	–	<i>étoile=stea</i>	<i>tārā=setāre</i>	–	<i>itri=atri</i>	<i>urjii=xiddig</i> <sup>72</sup>	–	<i>yıldız=sulus</i>
stone	12BHSΣT Y	–	<i>batu=váto</i>	<i>kámək=akme</i> <i>ns</i>	<i>jnr=ōni</i>	<i>stone=Stein</i>	<i>kivi=kõ</i>	<i>pierre=piatr</i> <i>ă</i>	–	–	–	<i>ḍagaa=dhag</i> <i>ax</i>	<i>jiwe=itshe</i>	<i>taş=taas</i>
to suck	2T	<i>mātātā=mats</i> <i>ats</i>	–	–	<i>snq=sōnk</i>	<i>suck=saugen</i>	–	<i>sucer=suge</i>	–	<i>súigh=sugno</i>	–	–	–	<i>em=em</i>

62 Probably related despite an irregularity in the initial (which in Hebrew is an original *q*-). Gəʿəz has *ḥäššir*.

63 The real German cognate is an older form *scurz* which seems to have been reshaped under the influence of Latin *curtus*. Cf. a similar variation between French and Romanian.

64 Both words seem to contain the root \**sed-*, cf. Lucht (2007: 345f.).

65 Cf. Uighur *oltur* “to sit”.

66 Regular loss of *ḥ* in Oromo as well as a development *l > j* as in “eye”.

67 Same word as “to lie down”.

68 Both terms mean “child”, which is the normal equivalent for English “son” in these languages.

69 Derivatives from an underlying root \**spi-*.

70 These are nouns for “spittle”.

71 I assume both words to be cognate despite an irregular correspondence in the initial.

72 Correspondence Oromo *-r-* = Somali *-dd-* as in “seven”.

summer	–	–	–	–	šmw=šōm	summer=Sommer	–	–	–	samh[radh=haf	–	–	–	yaz=sayın
sun	12BbβHΣY	–	matahari=masoáandro <sup>73</sup>	slá[nce=saul e	r'w=rē	sun=Sonne	–	soleil=soare	sūraj=xor[šī d	–	–	–	–	güneş=kün <sup>74</sup>
sweet	T	–	manis <sup>75</sup> =mány	sládak=salds	–	sweet=süß	–	doux=dulce	–	milis=melys	azidan=yäze dǎn	mi'aawaa=macaan	–	–
to swim	12BS	–	–	plúvam=pe[l dēt	nbi=nēbi	swim=schwimmen	uida=úszik	–	–	snámh=nofio	–	–	–	–
tail	12BSΣTY	–	–	–	sd=sat	–	–	queue=coadǎ	–	–	–	–	–	mkia=umsila kuyruk=kuturuk
tear(drop)	D	–	–	–	rmy.t=ermē	–	kyynel=könn y	larme=lacrimǎ	āñsū=aš[k n	deoir=deigryn	imeṭṭi=āmeṭṭi	immimaan=ilmo	–	–
ten	2	assar=éser	se]puluh=fól o	déset=desmit	mḏw=mēt	ten=zehn	–	dix=zece	das=dah	deich=deg	–	–	kumi=ishumi	on=uon
that (far demonstrative)	12	–	–	–	–	–	–	–	–	sin=hwnnw	-ihin=-hen	–	-le=l-	o=ol
thin (of things)	2BS	–	tipis=ma]nífy	–	–	thin=dünn	–	–	–	–	–	–	–	ince=sinnyiges
this (near demonstrative)	12BSTY	–	–	–	p[n=phai	this=dieser	–	ce=acest	yah=īn	–	–	kana=kan	–	bu=bu
three	2Bbβ	sost=šaloš	–	tri=trīs	hmtw=šomt	three=drei	kolme=három	trois=trei	tīn=se	trí=tri	–	sadii=saddex	-tatu=-thathu	üç=üs
to tie/bind	2T	–	–	–	–	–	–	lier=lega	bāndhnā=ba stan	–	eqqen=äqqən	hiḏa=xidh	–	bağla=baay
tongue	12BbDHSΣTY	–	lidah=léla	–	ns=las	tongue=Zunge	–	langue=limbǎ	jībh=zabān	teanga=tafo d <sup>76</sup>	iles=iləs	arraba=carrab	ulimi=ulimi	dil=tl
tooth	12BbβDHΣTY	–	–	zəb=zobs	–	tooth=Zahn	–	dent=dinte	dānt=dandān	–	–	ilkaan=ilig	jino=izinyo	diş=tiis
tree	12BHΣ	–	–	–	–	–	puu=fa	arbre=arbor e	–	crann=pren	–	–	mti=umuthi	–
two	12BbβDHSΣY	–	dua=róa	dve=divi	sn.wj=snau	two=zwei	kaksi=kettő	deux=doi	dō=do	dó=dau	sin=assin	lama=laba	-wili=-bili	iki=ikki
warm	12B	–	panas=ma]fána	–	–	warm=warm	–	chaud=cald	–	–	–	–	–	–
to wash	2	–	–	–	j'i=iōi	wash=waschen	–	–	–	–	ssired=sirəd	miičča=maydh	–	yka=suuy

73 Both literally “eye (of the) day”, a compound that probably already existed in the common ancestor of both languages.

74 Same root as “day”.

75 From < \*mamis, cf. Acehnese mamēh “sweet”.

76 Welsh -f- from \*-gw-.

water	12BbβDH ΣTY	–	–	<i>vodá=ūdens</i>	<i>mw=mōou</i>	<i>water=Wasser</i>	<i>vesi=víz</i>	<i>eau=apǎ<sup>77</sup></i>	–	–	<i>aman=aman</i>	<i>bis[aa=biy o<sup>78</sup></i>	<i>maji=amanzi</i>	<i>su=uu</i>
we	12BHSEΣ	<i>əñña=anáxn u</i>	–	–	– <sup>79</sup>	<i>we=wir</i>	<i>me=mi</i>	<i>nous=noi</i>	<i>ham=mā</i>	<i>sinn=ni</i>	<i>nekwni=nākk āneḏ</i>	<i>nu=a/inna[ga</i>	–	<i>biz=bihigi</i>
wet	2B	<i>ərtəb=ratov</i>	–	–	–	–	–	–	–	<i>fliuch=gwlyb</i>	<i>ebzeg=əbdəg</i>	–	–	–
what?	12BSΣTY	<i>mən=ma</i>	–	–	–	<i>what=was</i>	<i>mikā=mi</i>	<i>quoi=ce</i>	<i>kyā=če</i>	–	–	<i>maa[l=max</i>	<i>nini=-ni</i>	–
white	12BS	–	<i>putih=fōtsy</i>	<i>bjal=balts</i>	–	<i>white=weiß</i>	–	–	–	–	<i>amellal=mäll ān</i>	<i>adii=cad</i>	–	–
who?	12BDSSET Y	<i>man=mi</i>	–	<i>koj=kas</i>	<i>m=ni]m</i>	<i>who=wer</i>	<i>kuka=ki</i>	<i>qui=cine</i>	<i>kaun=kī</i>	<i>cé=pwy</i>	–	–	–	<i>kim=kim</i>
wind	2TY	–	–	<i>vjá[tər=vējš</i>	<i>tʒw=thēou</i>	<i>wind=Wind</i>	–	<i>vent=vânt</i>	–	–	<i>aḏu=aḏu</i>	–	–	–
wing	2T	<i>kənf=kanaf</i>	–	–	<i>ḏnḥ=tenh</i>	–	–	<i>aile=ari[pǎ</i>	–	–	<i>iferr=afraw<sup>80</sup></i>	–	–	<i>kanat=kīnat</i>
winter	–	–	<i>musim dingin=rirīni na<sup>81</sup></i>	<i>zīma=ziema</i>	<i>pr.t=phrō</i>	<i>winter=Winter</i>	<i>talvi=tél</i>	<i>hiver=iarnă</i>	–	<i>geimhreadh= gaeaf</i>	–	–	<i>masika=ubus ika<sup>82</sup></i>	<i>kış=kīhin</i>
woman	12BβS	–	–	–	<i>ḥm.t=s]himi</i>	–	–	<i>femme=femei e</i>	–	–	<i>tameḏtut=tam āḏ</i>	–	–	–
work	2B	–	–	–	–	–	–	–	<i>kām=kār<sup>83</sup></i>	–	–	–	–	–
worm	2S	<i>təl=tola</i>	–	–	<i>fnḏ=fent</i>	<i>worm=Wurm</i>	–	<i>ver=vierme</i>	–	–	<i>tawekka=taw əkke</i>	–	–	–
year	2SY	–	<i>tahun=táona</i>	<i>godína=gads</i>	<i>rnp.t=rompi</i>	<i>year=Jahr</i>	–	<i>année=an</i>	–	<i>bliain=blwyd dyn</i>	–	–	<i>mwaka=unya ka</i>	<i>yıl=sil</i>
yellow	12S	–	–	<i>žalt=dzeltens</i>	–	<i>yellow=gelb</i>	–	<i>jaune=galbe n</i>	–	–	<i>awḏay=āḏyā n</i>	–	–	–
yesterday	T	<i>təlant<sup>84</sup>=etmo l</i>	<i>ke]marin=oma mály</i>	<i>včéra=vakar</i>	<i>sf=saf</i>	<i>yester[day=g estern</i>	–	<i>hier=ieri</i>	–	–	–	<i>kalee=shalay</i>	–	–
you (sg.)	12BDHSEΣ TY	<i>antā=ata</i>	–	<i>ti=tu</i>	<i>ntk=nthok</i>	–	<i>sinā=te</i>	<i>toi=tu</i>	<i>tū=to</i>	<i>tú=ti</i>	<i>keçç=kāy</i>	<i>ati=adi[ga</i>	<i>wewe=we[na</i>	<i>sen=en</i>
you (pl.)	2B	–	–	–	<i>ntn=nthōten</i>	<i>you=ihr</i>	<i>te=ti</i>	<i>vous=voi</i>	–	<i>sibh=chwi</i>	<i>kunwi=kāwā neḏ</i>	<i>isin=idin[ka</i>	<i>ninyi=ni[na</i>	<i>siz=ehi[gi</i>
Σ	180	59	66	75	103	131	48	118	53	82	91	67	55	95

77 Latin *aqua*.

78 A more conservative form has been preserved in Rendille *bičče* “water”.

79 Cf. Quack (2002).

80 Same word as “leaf”.

81 Indones. *dingin* is the word for “cold”, *musim dingin* = “cold season”. Malagasy *ririnina* seems to be a similar composition of a hypothetical related term *\*rinina* “\*cold” plus an unidentified initial element.

82 The principal rainy season.

83 Same root as “to do”.

84 < *\*təməlt*.

## Conclusion

Based on the, admittedly, limited set of language couples evaluated here, the word candidates can be grouped as follows with decreasing degree of stability:

Survives in 13 couples: –

Survives in 12 couples: *five, four, two*

Survives in 11 couples: *I, six, three, you (sg.)*

Survives in 10 couples: *to die, fly (animal), full, hand, horn, one, star, ten, tongue*

Survives in 9 couples: *dream, to eat, eight, to give, hundred, to laugh, nail, name, new, seven, stone, water, we, who?, winter, you (pl.)*

Survives in 8 couples: *blood, eye, heart, moon, night, nine, to sit, year*

Survives in 7 couples: *all, to bear (give birth), brother, day, to do, dry, ear, egg, father, finger, fire, fish, head, to hear, hunger, ice, nose, other, right (side), shadow, son, to spit, sun, sweet, tear, tooth, what?, yesterday*

Survives in 6 couples: *ashes, to bite, bitter, bone, to come, to dig, to drink, green, heavy, to live, liver, long, louse, mother, mouth, navel, rain, red, root, salt, short, sky, to sleep, to stand, to suck, this*

Survives in 5 couples: *black, cloud, cold, flea, foot, grass, hedgehog, honey, knee, to know, to lie (down), milk, mouse, nest, not, to open, person, to see, seed, skin, smoke, snake, to swim, to tie, to wash, white, wind, wing, worm*

Survives in 4 couples: *earth, fat, to fear, feather, to kill, leaf, man, meat, near, old, road, to say, to sew, to sing, summer, tail, that, tree, yellow*

Survives in 3 couples: *to ask, back, bark, bird, breast, to burn, child, dog, far, flower, fruit, to go, good, hair, house, mountain, sand, snow, thin, warm, wet, woman*

Survives in 2 couples: *ant, bad, big, to carry, to fall, to fly, many, narrow, river, round*

Survives in 1 couple: *belly, to cut, neck, sea, small, work*

It turns out that certain items which figure prominently in existing basic vocabulary lists are rather bad, such as “belly”, “dog”, “neck”, “small”, whereas some good words are rarely included in such lists, such as “finger”, “fly (animal)”, “hunger”, “winter”.

Based on these data, a good (= diachronically stable) basic vocabulary list, which I herewith wish to propose, could be the following 54 item-list:

*all, to bear (give birth), blood, brother, to die, to do, dream, dry, ear, to eat, egg, eye, finger, fire, fish, five, fly (animal), four, full, to give, hand, head, to hear, heart, horn, hunger, I, ice, to laugh, moon, nail, name, new, night, one, right (side), shadow, to sit, son, star, stone, sun, sweet, ten, three, tongue, tooth, two, water, we, who?, year, yesterday, you (sg.).*

These are all items of my table that occur as cognates at least 7 times, with the exception of:

(1) some items that in many languages depend from other list items (numbers from 5 to 9 may be composed of lower numbers; “hundred” may be related to “ten”; “other” may be related to “two”; “you (pl.)” may be derived from “you (sg.)”; “what?” often from the same root as “who?”; “day” often from the same root as “sun”; “tear” often expressed as “water of eye” or the like);

(2) three items which tend to be onomatopoeic and can therefore be misleading when used as evidence in historical linguistics (“father”; “nose”, which often contains a nasal consonant; “to spit”);

(3) one item which, despite showing a good stability rate where it occurs, does not exist as a concept in a large part of the world (“winter”).

## Bibliography

- Bender, M. Lionel 1983: "Proto-Koman Phonology and Lexicon", *Afrika und Übersee* 66: 259-297
- Dolgopolsky, Aaron B. 1986: "A probabilistic hypothesis concerning the oldest relationships among the language families in Northern Eurasia", in Shevoroshkin, Vitalij V. & Markey, Thomas L. (eds.), *Typology, Relationship and Time*, Ann Arbor, 27-50
- Dyen, Isidore 1960: Comment on D.H. Hymes "Lexicostatistics so far", *Current Anthropology* 1: 34-39
- 1964: "On the Validity of Comparative Lexicostatistics", in Lunt, Horace G. (ed.), *Proceedings of the Ninth International Congress of Linguistics*, The Hague, 238-252
- (ed.) 1973: *Lexicostatistics in genetic linguistics*, The Hague
- Dyen, Isidore & James, A.T. & Cole, J.W.L. 1975: "Language Divergence and Estimated Word Retention Rate", in Dyen, Isidore, *Linguistic Subgrouping and Lexicostatistics*, The Hague, 181-207
- Elbert, Samuel H. 1953: "Internal Relationship of Polynesian Languages and Dialects", *Southwestern Journal of Anthropology* 9: 147-173
- Halayqa, Issam K.H. 2007: "Swadesh List (Basic Vocabulary List) for Ugaritic, Phoenician, Biblical Hebrew, Syriac and Classical Arabic", *Ugarit-Forschungen* 39: 319-380
- Holman, Eric W. *et al.* 2008: "Explorations in Automated Language Classification", *Folia Linguistica* 42: 331-354
- Holst, Jan Henrik 2001: *Lettische Grammatik*, Hamburg
- van Hout, Roeland & Muysken, Pieter 1994: "Modeling lexical borrowability", *Language Variation and Change* 6: 39-62
- Hymes, Dell 1973: "Lexicostatistics and glottochronology in the nineteenth century (with notes toward a general history)", in Dyen (1973: 122-176)
- Kessler, Brett 2001: *Significance of Word Lists*, Stanford
- Kruskal, Joseph B. & Dyen, Isidore & Black, Paul 1973: "Some results from the vocabulary method of reconstructing language trees", in Dyen (1973: 30-55)
- Leslau, Wolf 1969: *Hebrew Cognates in Amharic*, Wiesbaden
- Lohr, Marisa 1998: *Methods for the genetic classification of languages*, Diss. Cambridge [non vidi]
- Lucht, Martina 2007: *Der Grundwortschatz des Altirischen*, Dissertation Bonn
- Oswalt, Robert L. 1971: "Towards the Construction of a Standard Lexicostatistic List", *Anthropological Linguistics* 13: 421-434
- Pagel, Mark *et al.* 2007: "Frequency of word-use predicts rates of lexical evolution throughout Indo-European history", *Nature* 449: 717-720
- Quack, Joachim F. 2002: "Die erste Person Plural des selbständigen Personalpronomens im Mittellägyptischen", *Lingua Aegyptia* 10: 335-337
- Starostin, George [Georgij] 2010: "Preliminary lexicostatistics as a basis for language classification: A new approach", *Journal of Language Relationship* 3: 79-116
- Starostin, Sergei 1991: *Altajskaja problema i proisxoždenie japonskogo jazyka*, Moskva
- 2000: "Comparative-historical linguistics and lexicostatistics", in Renfrew, Colin *et al.* (eds.), *Time Depth in Historical Linguistics*, Cambridge, vol. I: 223-265
- Swadesh, Morris 1952: "Lexico-Statistic Dating of Prehistoric Ethnic Contacts with Special Reference to North American Indians and Eskimos", *Proceedings of the American Philosophical Society* 96: 452-463
- 1955: "Towards Greater Accuracy in Lexicostatistic Dating", *International Journal of American Linguistics* 21: 121-137
- Tadmor, Uri 2009: "Loanwords in the world's languages: Findings and results", in Haspelmath, Martin & Tadmor, Uri (eds.): *Loanwords in the world's languages. A comparative handbook*, Berlin: 55-75
- Woodward, James 1993: "The Relationship of Sign Language Varieties in India, Pakistan, & Nepal", *Sign Language Studies* 78: 15-22