Lexical Matches between Sumerian and Hurro-Urartian: Possible Historical Scenarios

Alexei Kassian

Institute of Linguistics of the Russian Academy of Sciences, Moscow

The paper deals with lexical matches between two ancient Near Eastern languages: Sumerian and Hurrian (Hurro-Urartian); namely, several basic terms (like 'hand,' 'rain,' etc.), that demonstrate phonetical similarities in both languages, are discussed. Four possible scenarios are evaluated from the typological, etymological and statistical points of view: (1) chance coincidences; (2) lexical borrowings from Sumerian into Hurro-Urartian or vice versa; (3) genetic relationship between Sumerian and Hurro-Urartian; (4) prehistoric language shift: adoption by a Hurro-Urartian (or closely related) group of the Sumerian language or vice versa. Out of these four, two scenarios—lexical borrowings and genetic relationship—are typologically unlikely. The statistical probability of chance coincidences is low, although formally this explanation cannot be excluded. The fourth scenario—language shift—fits linguistic evidence and does not contradict archaeological data.

Keywords: Hurrian, Sumerian, Ancient Near East, language contacts, language shift, loanwords, lexicostatistics

§1. *Introduction*

§1.1. *The languages*

§1.1.1. Sumerian is a language spoken in southern Mesopotamia (modern Iraq). Its earliest cuneiform attestations date from the late 4th or early 3rd millennium BC, and it functioned as a living language until the late 3rd or early 2nd millennium BC. Later, until the late 1st millennium BC, Sumerian was widely used by Babylonians as a language of scholarship and cult. The genealogical affiliation of the Sumerian language is unclear. Sumerian readings and meanings adduced below are quoted from the *Electronic Pennsylvania Sumerian Dictionary* (ePSD), the *Cuneiform Digital Library Initiative* (CDLI) and the *Electronic Text Corpus of Sumerian Literature* (ETCSL), as well as from Jagersma 2010.

§1.1.2. The Hurro-Urartian (in the following: HU) linguistic family consists of two closely related languages: Hurrian (with several dialects) and Urartian. Historical Hurrian was spoken in the southeast of present-day Turkey, in northern Syria and northern Iraq at least from the 2nd half of the 3rd millennium to the end of 2nd millennium BC.[1] Urartian is attested in the 1st millennium BC as a language of the Urartian empire (present-day Armenia and neighboring areas).[2] For the preliterate period, it is natural to associate the HU people with the Kura-Araxes (Early Trans-Caucasian) archaeological culture (Kassian 2010: 423-428 with further references). The HU languages are poorly documented as compared with Sumerian. The genealogical affiliation of the HU languages is likewise uncertain, although I suspect that it is possible to treat HU as a separate branch of the hypothetical Sino-Caucasian (Dene-Caucasian) macro-family, that is, that the HU group is a distant relative of the North Caucasian, Yeniseian and Sino-Tibetan protolanguages; see Kassian 2011 for discussion.

§1.2. Preliminary Methodological Remarks

§1.2.1. I will not discuss in detail what kind of facts can *prove* the genetic relationship between the two lects. The modern view is that two languages can be considered genetically related if there exist (1) an appreciable number of etymological matches between their basic vocabularies,[3] and (2) an appreciable number of etymolog-

---

[1] Cuneiform and Ugaritic alphabetic sources from *ca.* the 23rd century to the late 2nd millennium BC (Salvini 1998; Wegner 2007: 21-32).

[2] Cuneiform (and apparently hieroglyphic) sources of the 9th-7th centuries BC; see two recent editions of the Urartian corpus: KUKN and CdTU.

[3] It is not always stated explicitly, but intuitively understood by professional comparativists that basic vocabulary not

ical matches between their main grammatical exponents (number, case, person); see Campbell & Poser 2008: 4, Burlak & Starostin 2005: 7-24. Following Burlak & Starostin 2005, *pace* Campbell & Poser 2008, I believe that condition (1) is essential, while condition (2) can serve as additional proof. Empirically, any pair of languages conventionally assumed to be genetically related at a reasonable time depth possesses a significant number of etymological matches with identical meanings between the basic vocabularies of these languages, most importantly, between words of their *core* vocabularies, summarized as the Swadesh wordlist.[4] That is, lexicostatistics is a reliable tool for language relationship tests and, moreover, the presence of etymological matches with coinciding semantics between Swadesh wordlists of two languages (or protolanguages) is a *necessary condition* of recognizing a genetic relationship between them.

§1.2.2. As stated in G. Starostin 2010a, *classical* and *preliminary* lexicostatistics are two very different procedures. The former should be used in a situation when a group of genetically related languages is sorted out, and regular phonetic correspondences between the languages are established. In such a case, *classical* lexicostatistics helps to determine the internal genealogical classification of the linguistic group in question. On the other hand, *preliminary* lexicostatistical verification/falsification is used when genealogical affiliation of the examined language is not yet established. This means that, lacking knowledge of *regular* phonetic correspondences, we are compelled to resort to the phonetic *similarity* between the semantically corresponding lexical items of the compared languages.

§1.2.3. Phonetic similarity can be formalized as the method of consonant classes, which was proposed by A. Dolgopolsky (1964; English version: 1986) and successfully tested by various authors, e.g., Baxter 1995; Baxter & Manaster Ramer 2000; Kessler 2007; G. Starostin 2008; Turchin, Peiros & Gell-Mann 2010. This method implies that the phonetic alphabet used in our studies can be divided into several non-intersecting subsets (classes) so that phonetic mutations between the sounds of one class during the natural language development are typologically more normal than mutations between sounds of different classes. Typology of sound changes

is not sufficiently advanced yet (but cf. Brown, Holman & Wichmann 2013 for progress in this area), therefore such a division can only be based on the intuition and experience of individual linguists. Below, I operate with classes currently accepted in the Global Lexicostatistical Database project (GLD)[5]:

P-class (labials): p b ɓ f v ɸ β ʋ
T-class (dentals): t d ɗ θ ð ʈ ɖ
S-class (front affricates & fricatives): c ʒ č ǯ ɕ ʑ s z š ž
Y-class (palatal glides): y
W-class (labial glides): w ʍ
M-class (labial nasals): m ɱ
N-class (non-labial nasals): n ɳ ɲ ŋ ɴ
Q-class (lateral affricates): λ ƛ
R-class (liquida): r ɹ ɾ ʈ ɭ ʀ l ɫ ļ ʎ ɮ ɬ
K-class (velars & uvulars): k g ɠ ɰ q ɢ x ɣ χ ʁ
zero-class or H-class: ħ ʕ ʜ ʔ h ɦ ʡ and any vowels.

Using this simplified transcription system (*P T S Y W M N Q R K H*) we can code any real wordforms or morphemes included into comparison. Note that elements of the zero-class and such features as coarticulation, prosody and phonation are deleted from the structure. Vocalic or laryngeal onsets and vocalic or laryngeal finals, however, are coded as *H*. Thus both hypothetical forms *tasa* and *dʰüʒo* are coded as *TSH*; *alaq* and *ʔärx = HRK*; *na* and *ŋoʔ = NH*; *pkʰot* and *baqʼaθ = PKT*; *wahat* and *ʍad = WT*. Non-initial *Y* and *W* (weak glides) are treated as *H*, thus *ka*, *kay*, *kawa = KH*, whereas *kat* and *kayat = KT*.

§1.2.4. As follows from the above, two forms from compared languages possessing identical simplified transcriptions have a better chance of appearing to be etymological cognates than forms whose simplified transcriptions differ.[6]

5   <http://starling.rinet.ru/new100/sound.pdf> [last visited 25.12.2013]. My system of transcription, in which all the Sumerian, Hurrian and related data are encoded, is normally adapted to the unified transcription system of the *Global Lexicostatistical Database* project, that is generally based on the IPA alphabet, with just a few specific discrepancies (see <http://starling.rinet.ru/new100/UTS.htm>).

6   If we confine ourselves to two first consonants of each word form under study, such a consonant classes test comes closest to modeling real comparative-historical research, at least as far as the criteria for what constitutes an etymological lexical match between two languages are concerned. First, historical linguists implicitly understand that cross-linguistically, the most common root shape is *CVC(V)* (where *C* may be a zero), both consonants of which should correspond to a *CVC(V)* root in the compared language. Second, although exceptions are

    cultural words must be etymologically investigated in the first place, if two languages are suspected to be relatives.

4   To be precise: neither am I personally nor are any of my colleagues from the Moscow school aware of a *single* reliable exception to this phenomenological rule.

§2. *The Problem of the Genealogical Affiliation of Sumerian*

A great number of hypotheses about genetic relationship between Sumerian and various languages of Eurasia have already been proposed and will be proposed in the future. Among those, two deserve special attention in my opinion: I. Diakonoff's Sumerian-Munda comparison and J. Bengtson's Sumerian–Sino–Caucasian comparison.

§2.1. Diakonoff's Sumerian-Munda Hypothesis (Diakonoff 1997)[7]

§2.1.1. The Munda linguistic family consists of *ca.* 20 languages currently spoken in eastern and central India and Bangladesh (apparently Munda and Mon-Khmer are to be treated as two separate branches of the Austro-Asiatic (macro)family; see Sidwell 2010 with references). Diakonoff proposed a theory that the Sumerian and Munda languages could have been fairly close relatives and offered a convincing historical scenario for a prehistoric migration of the Sumerians from India.

§2.1.2. Implicitly using the same consonant classes method as described above, Diakonoff offers 34 Sumerian-Munda *CVC*-root etymologies and several grammatical parallels. *A priori*, the main problem of Diakonoff's theory is that the author normally restricts himself to two Munda languages, Santali and Mundari, that form a separate group within the North Munda branch (Anderson 2008).

§2.1.3. Below, I apply the lexicostatistical test to Diakonoff's data, that is, I single out Sumerian roots with Swadesh meanings and compare them to the corresponding Swadesh terms that could be reconstructed for proto-Munda. A general proto-Munda reconstruction is not completed yet, so I am guided by the Munda data collected in Pinnow 1959 and some other publications. My general criterion for the reconstruction of proto-Munda Swadesh meanings is the distribution of individual roots

within the Munda family. Phonetic shapes of the reconstructed proto-Munda forms below are approximate.

§2.1.4. Formally, the best Sumerian-Munda match among Diakonoff's etymologies is:

1) Sum. *ku* or *kua* ⟨KU_6⟩ 'fish.'[8] In seems that the main candidate for the status of the proto-Munda term for 'fish' is **qa* (Pinnow 1959: 77, 199).

The next etymology could also be very convincing, although formally it does not answer the principle of consonant classes:

2) Sum. *ŋe-* ⟨$ĜE_{26}$⟩ 'I.' Cf. the proto-Munda personal pronoun **iŋ* ~ **iɲ* 'I' (Pinnow 1959: 186, 208).

The next two etymologies are more problematic.

3) Sum. *gaʒ* ⟨GAZ⟩, with polysemy 'to kill, strike dead, slaughter / to beat / to grind, grate / to thresh (grain) / to break.' The main candidate for the status of the proto-Munda term for 'to kill' is the labile verb **goǯ-* 'to die / to kill' (Pinnow 1959: 203, 258). The Sumerian-Munda comparison is phonetically, but not semantically likely, because Sumerian polysemy 'to kill / to beat' should point to the original proto-Sumerian meaning 'to beat.'[9]

4) Sum. *mu* ⟨MU⟩ 'name' (Diakonoff groundlessly reads it as *ŋu* ⟨$ĜU_{10}$⟩). Cf. proto-Munda **ȳimu* (~ **yimu* ~ **ɲimu*) 'name' (Pinnow 1959: 141, 187, 189, 253; Sidwell 2010: 125).[10] The comparison is possible if one assumes the reduction of the first syllable in Sumerian.

The rest of Diakonoff's Sumerian words with Swadesh meanings demonstrate no semantic or phonetic matches with Munda:

5) Sum. *gal* ⟨GAL⟩ 'big,' compared by Diakonoff to Munda forms with the meaning '10.' One of the possible candidates for the status of the proto-Munda term for 'big' is **maraŋ*, which is well attested in North Munda (Pinnow 1959: 73).

---

common and almost inevitable, the bulk of assumed phonetic shifts should be typologically trivial, i.e., the shifts should happen within the limits of phonetically justified consonant classes (assumption of a great number of unusual phonetic shift leads to regrettable results; cf., e.g., the critical overview of an Indo-European-Basque hypothesis in Kassian 2013).

[7] In the following, I conventionally transcribe the two series of Sumerian stops as voiced ~ voiceless (i.e., *d* ~ *t*, although the real opposition was t^h ~ *t* or *t:* ~ *t* or the like), I do not discriminate between the Hurrian phonemes *u* & *o* (both are transcribed as *u*), and so on, because all these peculiarities are irrelevant for my arguments and do not affect my conclusions.

[8] On the reading, see Englund 1990: 227-230.

[9] Semantic development 'to beat' > 'to kill' is typologically normal, whereas vice versa 'to kill' > 'to beat' is odd. It is also possible that the more archaic Sumerian expressions for 'to kill' are the labile verbs 'to die / to kill': *uš* ⟨$UŠ_2$⟩ (sg. subj./obj.) and *ug* ⟨UG_7⟩ (pl. subj./obj.).

[10] This widespread Munda word indeed resembles Indo-Aryan **naːman-* 'name,' but the hypothesis of the borrowing from Indo-Aryan languages into Munda faces phonetic difficulties (namely the palatalization of the initial consonant in Munda). Note that Munda **ȳimu* possesses good Mon-Khmer cognates.

6) Sum. *giggi* or *gig* ⟨GE$_6$⟩ 'black,' incorrectly read by Diakonoff as *ŋi(g)* and compared to some North Munda forms with the meaning 'night.' One of the possible candidates for the status of the proto-Munda term for 'black' is **Kende ~ *hende*, which is attested in North Munda (Pinnow 1959: 103, 201, 294).

7) Sum. *ŋiri* ⟨ĜIRI$_3$⟩ 'foot / leg,' compared by Diakonoff to Munda 'to run.' The best candidate for the status of the proto-Munda term for 'foot' is **ʒVŋ* (Pinnow 1959: 169, 218, 223; Sidwell 2010: 126; Anderson 2004: 163).[11]

8) Sum. *ur* ⟨UR⟩ 'dog,' incorrectly read by Diakonoff as *sur* ⟨SUR$_x$⟩[12] and compared to some Munda forms that originate from proto-Munda **sV* 'dog' (normally attested with suffixes or as an element in compounds; see Pinnow 1959: 112, 210, 242, 242, 350; Anderson 2004: 163).

Thus, the preliminary lexicostatistical test yields rather poor results: Diakonoff's data fail to provide a substantial number of matches between Sumerian and Munda basic vocabularies. Intuitively, it seems that the two best Sumerian-Munda matches ('fish' and 'I') can be coincidental from the statistical point of view. Does it mean that Diakonoff's Sumerian-Munda hypothesis failed? The answer is no. First, the full Swadesh 100- or 110-item wordlists for Sumerian and proto-Munda should be compiled and compared. Statistical tests (one of which is described below) are also necessary. Second, phonetic correspondences between Sumerian and Munda could actually be less trivial than the consonant classes described above. Third, Sumerian could theoretically represent a separate branch of the Austro-Asiatic (macro)family, and a Sumerian-Mon-Khmer comparison might yield better results.

§2.2. Bengtson's Sumerian–Sino–Caucasian Hypothesis (Bengtson 1997)

§2.2.1. In its current state, the theory of the Sino-Caucasian macro-family has been partially substantiated by the late S. Starostin. According to the modern view of the Moscow school, the Sino-Caucasian (or Dene-Caucasian) macro-family consists of three main branches: North Caucasian-Basque, Yeniseian-Burushaski and Sino-Tibetan-Na-Dene. For a brief sketch of the history of Sino-Caucasian studies, see now G. Starostin 2010b and esp. Bengtson & G. Starostin forthcoming. For the comparative phonetics of the Sino-Caucasian macro-family, see Starostin n.d. (this work was not finished and therefore remains unpublished). The highly preliminary

Sino-Caucasian etymological dictionary by S. Starostin is available as Sccet.dbf (see the list of abbreviations below for references to all online database files). Some other papers by the same author, dedicated to the Sino-Caucasian problem, can be found in S. Starostin 2007 (in both Russian and English). A comparative grammar overview of the Sino-Caucasian macro-family can now be found in Bengtson & G. Starostin forthcoming. A formal (lexicostatistical) verification of the Sino-Caucasian theory is currently in preparation for publication as part of the Moscow-based *Global Lexicostatistical Database* (GLD) and *Tower of Babel* projects, and the broader *Evolution of Human Language* project, centered around the Santa Fe Institute. For comparative data of individual Sino-Caucasian branches, see the following publications: North Caucasian – NCED; Caucet.dbf. Yeniseian – S. Starostin 1982/2007 and Yenet.dbf (the latter is based on S. Starostin 1995; Werner 2002 with additions and corrections). Sino-Tibetan – Stibet.dbf, based on Peiros & Starostin 1996, but seriously emended. Basque – Basqet.dbf and corresponding sections in Bengtson 2008. Burushaski – Buruet.dbf and such recent publications as, e.g., Bengtson 2008a; Bengtson & Blažek 2011. Proto-Na-Dene reconstruction is not completed (or not published) yet; cf. some rather preliminary publications on the supposed Sino-Caucasian affiliation of the Na-Dene family: Nikolaev 1991; Bengtson 2008b.[13] It is also possible that two ancient Near Eastern languages belong to this macro-family as additional branches: Hattic (Kassian 2010) and Hurro-Urartian (Kassian 2011).

§2.2.2. Bengtson's (1997) hypothesis is that Sumerian could be a separate member of the Sino-Caucasian macro-family.[14] Besides some typological similarities, Bengtson proposes various Sino-Caucasian cognates for 41 Sumerian words of basic vocabulary (mostly of the Swadesh list). Below, I quote Sumerian words etymologized by Bengtson fulfilling the following conditions: (a) they belong to the Swadesh 100-item wordlist, i.e., indeed represent default expressions for the corresponding basic meanings in Sumerian; (b) their transcription corresponds to modern views; (c) they are connected by Bengtson to the roots that can be reconstructed as Swadesh items at least for one of the protolanguages of the linguistic families

---

11   Another word for 'foot,' attested in some Munda languages, is **kaʈa* (Pinnow 1959: 72, 197, 285).

12   See George 2003, 1: 150 with fn. 56 for a criticism of this reading.

13   For a criticism of the so-called "Dene-Yeniseian" hypothesis, see G. Starostin 2010b; 2012, with E. Vajda's (2012) reply.

14   Of course, this idea cannot be considered fully innovative, because various attempts to uncover a relationship between Sumerian and individual linguistic groups currently included in the Sino-Caucasian macro-family (e.g., Sino-Tibetan or Basque) have been made since the early 20th century.

included in Sino-Caucasian macro-family (i.e., proto-North Caucasian, proto-Yeniseian, and so on). Of four such Sumerian words extracted from Bengtson's list, at least two are etymologized quite convincingly, since they represent Common Sino-Caucasian roots:[15]

1) Sum. *ŋa-* ⟨ĜA₂-⟩ 'I.' Comparison to Sino-Cauc. *$*ŋV$* 'I' suggests itself readily. *$*ŋV$* is one of the two Common Sino-Caucasian stems of the pronoun of the 1st p. sg., see G. Starostin 2010b: 112-113.

2) Sum. *uʒu* ⟨UZU⟩ 'meat,'[16] that is compared to Yeniseian *$*ʔise$* 'meat.' In turn, the Yeniseian form could be compared to Sino-Tibetan *$*sʸa$* (*$*śa$*) 'meat'—one of the two equivalent candidates for the proto-Sino-Tibetan term for 'meat.'[17] In sum, the Yeniseian-Sino-Tibetan match should yield the proto-Sino-Caucasian root for 'meat,' which is phonetically compatible to Sum. *uʒu*.

Two other Sumerian etymologies offered by Bengtson are less convincing:

3) Sum. *naŋ* ⟨NAĜ⟩ 'to drink,' compared to Na-Dene *$*naN$* 'to drink,' which is indeed a Common Athapaskan-Eyak-Tlingit verb (cf. Athapaskan *$*naːŋ_2$* ~ *$*naːŋ^w$* ~ *$*naːm$* ~ *$*naːw̃$* 'to drink,' Krauss & Leer 1981: 21, 39, 70, 133, 139, 151), but note that the final nasal in the Athapaskan root can be a fossilized (perfective?) suffix, because the Eyak (*la* 'to drink') and Tlingit (*naː* 'to drink') cognates demonstrate no traces of nasality and/or labiality. Sino-Caucasian etymology of Na-Dene *$*na(N)$* is unclear, but formally this is one of the several equivalent candidates for the Sino-Caucasian verb 'to drink' in absence of appropriate etymological matches between various root for 'to drink' in other Sino-Caucasian daughter families. Nevertheless, the Sumerian – Na-Dene comparison is formally acceptable.

4) Sum. *iʒi* ⟨IZI⟩ 'fire,' compared to North-Caucasian *$*cʼăyɨ$* 'fire' and Basque *$*sʸu$* (*$*śu$*) 'fire.' The North-Caucasian-Basque root is indeed one of the several equivalent candidates for the Sino-Caucasian term for 'fire,' but the Sumerian – Sino-Caucasian comparison is formally problematic, because the initial syllable in Sum. *iʒi* is inexplicable.

One must conclude that available lexicostatistical evidence for the Sumerian – Sino-Caucasian hypothesis is not stronger than arguments for the above-discussed Sumerian-Munda relationship. It goes without saying, however, that further research may provide more data in support of Bengtson's theory.

§3. *Sumerian and Hurro-Urartian*
§3.1. The Wordlist
§3.1.1. Surprisingly, the best formal results are achieved when comparing the Sumerian 110-item wordlist to the Hurro-Urartian data.[18] Due to the scantiness of known HU vocabulary, only *ca.* 65 slots of the HU 110-item wordlist are filled; one of them does not have a Sumerian counterpart (the original Sumerian personal pronoun of the 1st p. pl. 'we' seems unknown). My Sumerian list presented below is tentative; it is possible that further detailed research will enable us to define some positions more exactly (cf., e.g., the problematic item 'blood'), but it is not likely that such changes would seriously affect the overall statistics. The 65 slots filled for both Sumerian and Hurrian (the poorly attested Urartian, naturally, plays a minor role here) are as follows:

---

15 Below, all reconstructed forms from Sino-Caucasian languages are generally cited after the *Tower of Babel* project databases (Sccet.dbf, Caucet.dbf, Stibet.dbf, Yenet.dbf, Basqet.dbf, Buruet.dbf—see the list of references), unless mentioned otherwise. For the system of transcription see http://starling.rinet.ru/new100/UTS.htm.

16 Apparently *uʒu* is the basic Sumerian term for 'meat as food,' while the word *su* ⟨SU⟩ primarily means 'flesh' and 'body.'

17 To be separated from North Caucasian *$*yǝːmcoː$* 'bull, ox' and Sino-Tibetan *$*cʰu$* 'cow, bull.'

18 The 110-item wordlist accepted in the *Global Lexicostatistical Database* project (GLD) consists of the standard Swadesh 100-wordlist plus 10 additional words from S. Yakhontov's wordlist (taken from the second part of the Swadesh initial 200-wordlist); see Burlak & Starostin 2005: 12-13 for details. The Hurro-Urartian 110-item wordlist is discussed in detail in Kassian 2011. For the Sumerian language, besides various lexicographic and grammatical publications, the preliminary unpublished version of the Sumerian 110-item wordlist by prof. Vl. Emelianov has been used.

#	Word	Sumerian	Hurrian
1	all (*omnis*)	NOUN REDUPLICATION	sua-lːa ⟨šua-lla⟩
2	ashes	dedal ~ didal ⟨DE$_3$-DAL⟩	sal-mi ⟨šal-mi⟩
5	big	gal ⟨GAL⟩	tal-mi ~ tal-a-mi
6	bird	mušen ⟨MUŠEN⟩	eradi
8	black	giggi ⟨GE$_6$⟩	time-ri ~ tima-ri
9	blood	mud ⟨MUD⟩, umun ⟨U$_3$-MUN⟩	cur-gi ⟨zur-gi⟩
11	breast	gaba ⟨GABA⟩	neɣer-ni ⟨neḫer-ni⟩
12	to burn tr.	bil ⟨BIL$_2$ ~ BIL$_3$ ~ BIL⟩	am-
16	to come	ŋen ⟨ĜEN⟩ (perf.), du ⟨DU⟩ (imperf.)	un-
18	dog	ur ⟨UR⟩[19]	ervi ~ erbi
19	to drink	naŋ ⟨NAĜ⟩	al-
21	ear	ŋeštug- ⟨^{ĝeš}TUG$_2$ = ^{ĝeš}TU$_2$ ~ ^{ĝeš}TUG⟩	nui ~ nuɣi ⟨nui ~ nuḫi⟩
22	earth	saxar ⟨SAḪAR⟩	ese ⟨eše⟩
23	to eat	gu ⟨GU$_7$⟩	ul-
25	eye	igi ⟨IGI⟩	si ~ siɣi ⟨ši ~ šiḫi⟩
26	fat n.	i ⟨I$_3$⟩	ase ⟨aše⟩
28	fire	iʒi ⟨IZI⟩	tari
31	foot	ŋiri ⟨ĜIRI$_3$⟩	uri ~ ur-ni
33	to give	šum ⟨ŠUM$_2$⟩	ar-
34	good	dug- ⟨DUG$_3$ = DU$_{10}$⟩	faɣri ~ faɣr-usi ⟨waḫri ~ waḫr-uši⟩
37	hand	šu ⟨ŠU⟩	su-ni ⟨šu-ni⟩
38	head	saŋ ⟨SAĜ⟩	paɣi ⟨paḫi⟩
39	to hear	ŋeš tuku ⟨ĜEŠ TUKU⟩ 'to acquire the ear(?)'	xas- ⟨ḫaš-⟩
40	heart	šag- ⟨ŠAG$_4$ = ŠA$_3$⟩	tisa ⟨tiša⟩
42	I	ŋe ⟨ĜE$_{26}$⟩	is- ⟨iš-⟩ (dir. stem), su- ⟨šu-⟩ (obl. stem)

---

[19] The fact that 'dog' is also frequently designated as a compound *ur-gi* ⟨UR-GI$_7$⟩, lit. 'domestic' *ur*,' does not prove that *ur* originally meant generic 'animal' or 'beast.' First, simple *ur* ⟨UR⟩ is well attested with the meaning 'dog,' whereas, to the best of my knowledge, there are no Sumerian contexts, where plain *ur* ⟨UR⟩ is to be translated as 'animal' or 'beast.' Second, the semantic derivation 'dog' as 'domestic beast' seems typologically odd.

#	Word	Sumerian	Hurrian
45	to know	ʒu ⟨ZU⟩	pal-
48	liver	ur ⟨UR$_5$⟩, ba ⟨BA$_3$ = EŠ⟩[20]	ur-mi
49	long	gid ⟨GID$_2$⟩	keri ~ ker-aːi ⟨keri ~ ker-ašši⟩
50	louse	ex ⟨EḪ⟩	apxe ⟨apḫe⟩
51	man	lu ⟨LU$_2$⟩	taɣe ~ tae ⟨taḫe ~ tae⟩
52	many	šar ⟨ŠAR ~ ŠAR$_2$⟩	te-u-na
53	meat	uʒu ⟨UZU⟩	uʒi ⟨uzi⟩
54	moon	itid- ⟨ITID = ITI ~ I$_3$-TI⟩	kusuɣ ⟨kušuḫ⟩
55	mountain	kur ⟨KUR⟩	pab-ni ~ pab-a-ni
56	mouth	kag- ⟨KAG$_2$ = KA⟩	fasi ⟨faši⟩
57	name	mu ⟨MU⟩	tiye
58	neck	gu ⟨GU$_2$⟩	kudu-ni
59	new	gibil ⟨GIBIL ~ GIBIL$_4$⟩	suɣe ⟨šuḫe⟩
61	nose	kiri ⟨KIRI$_3$⟩	punɣi ~ puxːi ⟨punḫi ~ puḫḫi⟩
62	not	nu- ⟨NU⟩	=u-, =kːV-
63	one	diš	su-kːi ~ su-kːu ⟨šu-kki ~ šu-kku⟩
64	person	lu ⟨LU$_2$⟩	tarsuva-ni ⟨taršuwa-ni⟩
65	rain	šeŋ ⟨ŠEĜ$_3$⟩ 'to rain; rain (n.)'	isena ⟨išena⟩
67	road	kaskal ⟨KASKAL⟩	xari ⟨ḫari⟩
71	to say	dug- ⟨DUG$_4$ = DU$_{11}$⟩ (perf.), e ⟨E⟩ (imperf.)	xil- ~ xill- ⟨ḫil- ~ ḫill-⟩
72	to see	igi du ⟨IGI DU$_8$⟩ 'to spread the eye'	fur-

---

[20] Apparently both terms are attested with the anatomic meaning 'liver' (for *ur* ⟨UR$_5$⟩ cf. *Lugalbanda in the Mountain Cave*, 381: "He put the knife to the flesh of the brown goats, and he roasted the black livers (UR$_5$) there"). Because, however, the normal synchronic meaning of ⟨UR$_5$⟩ is metaphoric 'organ/center of feeling' (glossed as Akkadian *kabattu* 'mood, temper, center of feeling' in lexical lists), it is natural to posit ⟨UR$_5$⟩ as the original Sumerian term for 'liver (anatomic),' synchronously retained as a metaphoric expression, having been superseded by ⟨BA$_3$⟩ as an anatomic term (the original meaning of ⟨BA$_3$⟩ is unclear). Two facts speak in favor of such a solution. First, the semantic shift 'liver' > 'organ/center of feeling' is typologically normal, but probably not vice versa. Second, the assumed semantic evolution of ⟨UR$_5$⟩ is paralleled by Akkadian *kabattu*, which originates from the best candidate for the status of the proto-Semitic term for 'liver' (SED 1: 126), having been superseded by Akkadian *amuˑtu* ⟨amūtu⟩ in the direct anatomic meaning (SED 1: 168).

#	Word	Sumerian	Hurrian
74	to sit	tuš ⟨TUŠ⟩ (perf.), dur ⟨DUR₂⟩ (imperf.)	nax:- ⟨naḫḫ-⟩
75	skin	kuš ⟨KUŠ⟩	asxe ⟨ašḫe⟩
78	smoke	ibi ⟨I-BI₂⟩	xivri ⟨ḫiuri⟩
82	sun	ud- ⟨UD = U₄⟩	simigi ⟨šimigi⟩
85	that	=še	a-ni
86	this	=e[21]	an-ni
87	thou	ʒe ⟨ZE₂⟩	fe-
88	tongue	eme ⟨EME⟩	irde
89	tooth	ʒu ⟨ZU₂⟩	seri ~ sir-ni ⟨šeri ~ šir-ni⟩
90	tree	ŋeš ⟨ĜEŠ⟩	tali
91	two	min	sini ⟨šini⟩
92	to go	ŋen ⟨ĜEN⟩ (perf.), du ⟨DU⟩ (imperf.)	us:- ⟨ušš-⟩
94	water	ay ⟨A⟩	sive ~ siye ⟨šiwe ~ šiye⟩
96	what	ana ⟨A-NA⟩	av-
98	who	aba ⟨A-BA⟩	ab-i ~ av-i
99	woman	munus ⟨MUNUS⟩	asti ~ asta ⟨ašti ~ ašta⟩
106	snake	muš ⟨MUŠ⟩	apsi ⟨apši⟩
107	thin	sal ⟨SAL⟩	niga-le
110	year	mu ⟨MU⟩	savali ⟨šawali⟩

§3.1.2. Out of these 65 pairs, we see five or six cases where the Sumerian *CC*-structure[22] is phonetically compatible with its Hurrian counterpart (these are shadowed in the above table):

1) Sum. *ur* ⟨UR⟩ ~ Hur. *ervi* 'dog' = *HR*.

No appropriate Sino-Caucasian etymology for the HU term (Kassian 2011: 393).

2) Sum. *šu* ⟨ŠU⟩ ~ HU *su- 'hand' (Hur. *su-ni* ⟨šu-ni⟩, Urart. *su-* ⟨šu-⟩) = *SH*.
No appropriate Sino-Caucasian etymology for the HU term (Kassian 2011: 399).

3) Sum. *ur* ⟨UR₅⟩ ~ Hur. *ur-mi* 'liver' = *HR*.
No appropriate Sino-Caucasian etymology for the HU term (Kassian 2011: 402).

4) Sum. *uʒu* ⟨UZU⟩ ~ Hur. *uʒi* ⟨uzi⟩ 'meat' = *HS*.
Can be compared to Yenis. **ʔise* 'meat' and Sino-Tib. **sʸa* 'meat' (the main candidate for the basic Sino-Caucasian term for 'meat'), see §2.2 above and Kassian 2011: 405.

[21] The second Sumerian attributive demonstrative pronoun 'this' is =*be*. Synchronously the opposition between =*e* and =*be* is dialectal (Jagersma 2010: 222). Demonstrative =*be* is apparently secondary, however, originating from the non-human possessive pronoun =*be* 'its, of it,' whereas =*e* seems to be the original attributive demonstrative pronoun 'this' (Jagersma 2010: 224).

[22] That is, the first two consonants in the simplified transcription are taken into account.

5) Sum. *šeŋ* ⟨ŠEĜ₃⟩ ~ Hur. *isena* ⟨*išena*⟩ 'rain' = *SN*. Note that, formally speaking, the Hurrian *CC*-structure is to be analyzed as *HS* (*is*[*ena*]), but in our situation it seems safe to eliminate the initial *i-* from the Hurrian form ([*i*]*sena*). In any case, below I double all calculations for Sum. *šeŋ* ~ Hur. *isena* as both positive (*SN = SN*) and negative (*SN ≠ HS*) pairs.

   As noted in Kassian 2011: 410 f., the Hurrian word can be compared to Sino-Caucasian **Hʾəːrč̌ʷVŋ* 'to be cloudy, to rain (vel sim.)' >
   North Cauc. **Həːrč̌ːʷVn* 'to become cloudy (of weather),'
   Basque **ḫorci / *ḫosʸti* 'sky; storm; thunder; Thursday; rainbow; cloud,'
   Sino-Tib. **ʒʸaːŋ* 'shower, rain.'[23]

6) Sum. *aba* ⟨A-BA⟩ ~ Hur. *ab-i ~ av-i* 'who?' = *HP*.
   No appropriate Sino-Caucasian etymology for the HU term (Kassian 2011: 425).

Strictly speaking, there exists a seventh match:

7) Sum. *ŋen* ⟨ĜEN⟩, which is phonetically compatible with the Urartian verb *nun* 'to come' = *NN*. The difficulty is that the Hurrian verb for 'to come' is *un* and the etymological and morphological relationship between Urart. *nun* and Hur. *un* is unclear (a unique reduplication pattern **un-un > nun*?). Note that Hur. *un* 'to come' may be compared to Sino-Caucasian **=VʔʷʾVŋ*, which is a possible candidate for the status of the Common Sino-Cauc. verb for 'to go' (Kassian 2011: 392-393). Because of this and because my formal statistical comparison is actually Sumerian-Hurrian, I prefer to exclude the Urartian verb from consideration. Note that treating *ŋen ~ nun* as a positive pair will not contradict my general conclusions; to the contrary, it would seriously improve the statistical results.

§4. Explanation of the Sumerian-Hurrian matches

In this section, I discuss four possible explanations of the aforementioned Sumerian-Hurrian lexicostatistical matches: null hypothesis (§4.1), lexical borrowings (§4.2), genetic relationship (§4.3), language shift (§4.4).

§4.1. Null Hypothesis

§4.1.1. It is obvious that the phonetic similarity of six (or five) Sumerian-Hurrian matches in question can actually be coincidental. The question is, what is the probability of such a scenario? Two valid algorithms for calculation of the probability of phonetic matches between formalized wordlists are known.[24] One of them was described by

Ringe (1992); see especially Baxter & Manaster Ramer's (1996) review for a summary and important amendments (further, see Ringe 1998 and Baxter 1998). The second one—the so-called permutation test—was outlined and tested by W. Baxter & A. Manaster Ramer (2000) and some other authors.[25] Below, the Sumerian-Hurrian lexicostatistical matches will be tested with the help of Baxter & Manaster Ramer's (2000) algorithm, that is currently implemented as a plug-in for the StarLing software. The principle of the permutation test is simple and elegant. If we have two bi-unique and uniformly transcribed wordlists with *X* lexical phonetic matches, we can start to shuffle one of the lists, checking the number of matches for each new configuration. If the number of random configurations is great enough, it is possible to establish how many matches are statistically normal and, additionally, to calculate the probability of *X* and more than *X* matches between our original lists.

§4.1.2. For my statistical test, the Sumerian and Hurrian 65-item wordlists have been transcribed according to the simplified notation of consonant classes, as described in §1.2. Two forms constitute a positive pair if the first two consonants (*CC*) of the Sumerian form are identical to those of the Hurrian form. 1,000,000 random (strictly speaking, pseudorandom) trials have been performed in each case described below. If we consider Sum. *šeŋ* ~ Hur. *isena* 'rain' a positive pair (= *SN*), there are 6 *CC*-matches between the original lists (see above). The results of the test are given in figure 1.

§4.1.3. The most statistically common values are 1 match, 2 matches and 3 matches—their probability *P* is 0.234262, 0.277375 and 0.210287, i.e., 23.4262%, 27.7375% and 21.0287%, respectively. The total number of trials with 6 or more matches is 16,058 + 4,282 + 1,034 + 189 + 32 + 8 + 1 = 21,604. This means that the probability *P* of getting at least six matches (as we have in the case of the original Sumerian-Hurrian list) is 0.021604, i.e., slightly higher than 2%.

---

[23] Note that the basic Sino-Caucasian root for 'rain' is **=ŭɢʷʾV* > North Cauc. **=ŭɢʷV* 'to rain; rain,' Yenis. **xu-r* 'rain,' Sino-Tib. **qʰʷăH* 'rain.'

[24] I will not discuss here the statistical algorithms suggested by J. Nichols (see the summary in Nichols 2010, with application to the Dene-Yeniseian hypothesis), because Nichols' approach seems not to be formalized, and pos-

sesses certain logical loops. As a result, her final calculations of probability do not seem reliable (at least, they seriously contradict my linguistic and mathematical intuition).

[25] The general idea goes back to Oswalt 1970; further, see McMahon & McMahon 2005: 66-68 for an overview. See also Justeson & Stephens 1980; Baxter 1995; Kessler & Lehtonen 2006; Kessler 2007; Dunn & Terrill 2012 for an application of the permutation test to lexical lists of specific languages. A very similar bootstrap procedure was described and successfully applied to various languages of Eurasia by Turchin, Peiros & Gell-Mann 2010.
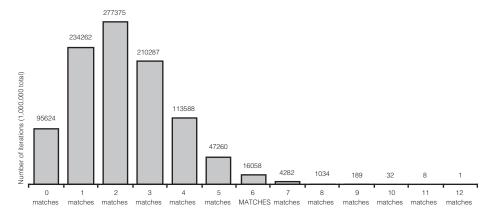
Fig. 1. Sumerian-Hurrian permutation comparison: GLD consonant classes (see §1.2), šeŋ ~ isena *is a positive pair*
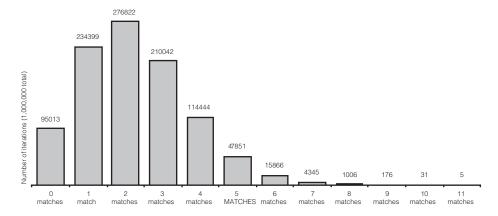

Fig. 2. Sumerian-Hurrian permutation comparison: GLD consonant classes (see §1.2), šeŋ ~ isena *is a negative pair*

§4.1.4. The most frequently accepted level of statistical significance is 5% (it means that the null hypothesis should be rejected if the *P*-value is less than 0.05); another popular significance level, used for more precise calculations, is 1% ($P = 0.01$). The probability of the Sumerian-Hurrian matches (0.021604 = 2.1604%) is lower than the 5% level, although higher than the 1% level. The picture certainly changes if we treat Sum. *šeŋ* ~ Hur. *isena* 'rain' as a negative pair ($SN \neq HS$), that is, if we only proceed with 5 Sumerian-Hurrian matches (fig. 2).

§4.1.5. The total number of trials with 5 or more matches is 47,851 + 15,866 + 4,345 + 1,006 + 176 + 31 + 5 = 69,280. This means that the probability *P* of getting at least five matches is 0.06928 = 6.928%. It is indeed higher than the 5% level, that is, the five Sumerian-Hurrian matches can formally be treated as coincidental. It must be noted, however, that the six (or five) Sumerian-Hurrian matches in question demonstrate very precise phonetic correspondences—not only consonantal, but even vocalic; cf. Sum. *ur* ~ Hur. *ur-mi* 'liver,' Sum. *uʒu* ~ Hur. *uʒi* 'meat,' Sum. *aba* ~ Hur. *ab-i* 'who?.' The correspondence Sum. *š* ~ Hur. *s* (Sum. *šu* ~ HU **su-* 'hand';

Sum. *šeŋ* ~ Hur. *isena* 'rain') is easily explained by the fact that Hurrian, as well as proto-HU, apparently possessed the only sibilant row s^{26} (as opposed to the Sumerian language, that discriminated between *s* ~ *š* phonologically). The same concerns the correspondence Sum. *ŋ* ~ Hur. *n*—there was no *n* ~ *ŋ* opposition in Hurrian and proto-HU, as opposed to Sumerian. The main vocalic discrepancies are Sum. *ur* ~ Hur. *ervi* 'dog' (but even so, the Hurrian form demonstrates the labial element) and the different onsets in Sum. *šeŋ* ~ Hur. *isena* 'rain.'

§4.1.6. This suggests that the simplified transcription described in §1.2 might be too rough for our purposes. The S-class can be divided into the S-class proper (front fricatives: s z š ž …) and the ʒ-class (front affricates: c ʒ č ǯ …); in turn, the R-class can be divided into the R-class proper (r ɾ …) and the L-class (l ḷ ł …). After that, the consonant classes run as follows (new classes are marked with an asterisk *):

P-class (labials): p b ɓ f v ɸ β v
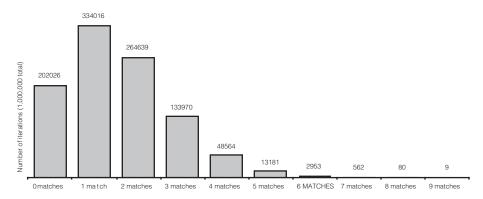T-class (dentals): t d ɗ θ ð ṭ ḍ

---

26   See Yakubovich 2009.

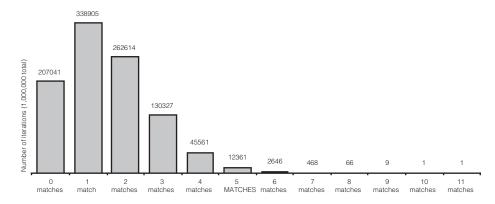*Fig. 3. Sumerian-Hurrian permutation comparison: more precise consonant classes,* šeŋ ~ isena *is a positive pair*

Fig. 4. Sumerian-Hurrian permutation comparison: more precise consonant classes, šeŋ ~ isena *is a negative pair*

S-class (front fricatives): s z š ž
*ʒ-class (front affricates): c ʒ č ǯ ɕ ʑ
Y-class (palatal glides): y
W-class (labial glides): w ʍ
M-class (labial nasals): m ɱ
N-class (non-labial nasals): n ɳ ɲ ŋ ɴ
Q-class (lateral affricates): λ ɬ
R-class: r ɹ ɾ ɽ ʀ
*L-class: l ɫ ɭ ʎ ɮ ɬ
K-class (velars & uvulars): k g ɠ ɰ q ɢ x ɣ χ ʁ
zero-class or H-class: ħ ʕ ʜ ʢ ʔ h ɦ ʔ and any vowels.

§4.1.7. If we use the above transcription, the permutation test will yield the results given in figure 3 (Sum. *šeŋ* ~ Hur. *isena* 'rain' is considered a positive pair = *SN*; in total, there are 6 *CC*-matches between the original lists). The total number of trials with 6 or more matches is 2,953 + 562 + 80 + 9 = 3,604. It means that the probability *P* of getting at least six matches is 0.003604 = 0.3604% (lower than the 1% level).

§4.1.8. If Sum. *šeŋ* ~ Hur. *isena* 'rain' is considered a negative pair (*SN ≠ HS*), i.e., in total there are 5 *CC*-matches

between the original lists, the results are as given in figure 4. The total number of trials with 5 or more matches is 12361 + 2646 + 468 + 66 + 9 + 1 + 1 = 15552. It means that the probability *P* of getting at least five matches is 0.015552 = 1.5552% (lower than the 5% level, although higher than the 1% level).

§4.1.9. The next logical step should be to include vowels in the simplified transcription (e.g., as the following classes: {o, u}, {i, e}, {a, ə} and so on) and compare not the *CC* chains, but the *CVC* ones. Due to technical difficulties, I have not performed this test, but it is obvious that Sumerian-Hurrian *CVC*-comparison will additionally decrease the probability of coincidences.

§4.1.10. Summing up, the statistical probability that the observed Sumerian-Hurrian matches are chance similarities varies from 0.069280 = 6.9280% (a rough approach) to 0.003604 = 0.3604% or lesser (a more sophisticated approach). This means that the null hypothesis is not very plausible.

§4.2. Lexical Borrowings

§4.2.1. Theoretically, the aforementioned Sumerian-Hurrian matches can be considered relatively late Sumerian loanwords in proto-HU or, vice versa, Hurrian loanwords in Sumerian.[27] Such an assumption, however, seriously contradicts the typology of language contacts.

§4.2.2. The general rule says that, among lexical items, cultural vocabulary is always borrowed first, whereas basic vocabulary is generally more resistant to borrowing (Thomason & Kaufman 1988: 74-76; Thomason 2001: 70-71). More precisely, this maxim is complied with in all cases where the sociolinguistic history of relevant peoples and languages is known to us. Traditionally, the Swadesh 100-item wordlist[28] is regarded as a core of basic vocabulary, that is, the Swadesh words are expected to be not only the most stable during natural language development, but also the most resistant to borrowing. It is intuitively likely, however, that it would be necessary to substitute certain, more stable and resistant words for a couple of Swadesh items (e.g., such Swadesh terms as 'seed' or 'person, human being' seem very dubious to me). Nevertheless, it is hardly possible to reform the Swadesh wordlist at the current stage of research.[29]

§4.2.3. If a language has foreign items in its Swadesh wordlist, this language is bound to have borrowings from the same source in other parts of basic vocabulary, and especially a great number of loanwords of the same origin in its cultural vocabulary (cf., e.g., modern English lexified by French and Scandinavian, or various Lezgian languages lexified by Azerbaijani). This is not the case of Sumerian–Hurro–Urartian contacts, because there are virtually no candidates for lexical or grammatical borrowings between these languages besides the six (of five) discussed Swadesh words. In addition to these, I can only quote one Hurrian cultural term possibly borrowed into Sumerian: Hur. *tab-i-ri* 'caster, (copper)smith' > Sum. *tibira*, *tabira* 'sculptor', *scil.* 'metal furniture-maker / craftsman working in metal and wood'[30] and a couple of dubious similarities such as Sum. *ur* ⟨UR₂⟩ 'root, base; limbs; loin, lap' ~ Hur. *uri* (suffixed *ur-ni*) 'foot; leg'[31] and the Sum. verb ⟨NUD = NU₂ = NA₂⟩ 'to lie, lie down (intr., subj. = person)' with the zero-derived substantive ⟨ᵍᵉˢNUD = ᵍᵉˢNU₂ = ᵍᵉˢNA₂⟩ 'bed' ~ Hur. *natxi* ⟨nathi⟩ 'bed.'[32] There are also a number of Hurrian cultural terms of Sumerian origin (see, e.g., Diakonoff 1971: 77 ff.; Wilhelm 2008: 103), but all of them seem to be borrowed via Akkadian (Kassian 2011: 435 with further references).[33] Thus, the absence of a substantial number of cultural borrowings between Sumerian and Hurro-Urartian makes the hypothesis of loanwords very unlikely.

§4.3. Genetic Relationship

§4.3.1. If we observe a number of phonetically similar words between basic vocabularies of two languages, it is reasonable to hypothesize that these languages are genetically related. Thus one could suppose that Sumerian and Hurro-Urartian are linguistic relatives, which means they are descendants of a Sumerian–Hurro-Urartian protolanguage and the discussed lexical matches represent a common heritage. In a sense, any pair of human languages are indeed genetically related (if we accept the monoglottogenesis conception); the question is, what is the date of split of the protolanguage assumed for this pair?

---

[27]   Certainly, these words could have been borrowed not directly from Sumerian, but from an undocumented Sumerian relative that was in contact with proto-HU (or, vice versa, not from Hurrian proper, but from a language related to Hurrian that was in contact with Sumerian).

[28]   For the semantic definitions of the extended Swadesh 110-item wordlist accepted in the *Global Lexicostatistical Database* project (GLD), see Kassian, et al. 2010.

[29]   A recent attempt to revise and modify the Swadesh wordlist (especially in connection with resistance to borrowing) has been undertaken by M. Haspelmath and U. Tadmor within the framework of the *World Loanword Database* project; see Haspelmath & Tadmor 2009, and Tadmor, Haspelmath & Taylor 2010. Instead of the traditional Swadesh wordlist, the so-called Leipzig-Jakarta 100-item list of basic vocabulary was proposed by the authors, differing from the classical Swadesh 100-item list in 38(!) items. Despite the sound theoretical approach, however, the actual results of the WOLD project unfortunately appear to be neither factually nor statistically reliable; see Kassian & M. Zhivlov's forthcoming review of WOLD for details.

[30]   = Akkad. *k'urk'urru* ⟨qurqurru⟩ 'metal-worker, esp. coppersmith'; see Wilhelm 1988: 50-52 and, e.g., Wilcke 2010: 10; cf. contra Waetzoldt 1997; P. Attinger *apud* Hazenbos 2005: 135 fn. 6; Richter 2012: xxviii, 439.

[31]   Cf. also Hur. *ugri* 'leg of table' and Urart. *kuri* 'foot (anatomic).' The relationship between *uri*, *ugri* and *kuri* is unclear (Kassian 2011: 397).

[32]   It is not entirely clear how to read this Sumerian root: *nud-* (thus, e.g., ePSD) or rather *nu-* (thus Jagersma 2010: passim). The final *-xi* in the Hurrian word can indeed be the common nominal suffix *-x(ː)i* (for which see Wegner 2007: 54) modifying the hypothetical root **nat-*, but even though the Sumerian root is to be read *nud-*, the vocalic correspondence Sum. *-u-* ~ Hur. *-a-* is inexplicable in the case of borrowing.

[33]   A. Fournet (2011: 56-57) offers a list of Sumerian-Hurrian lexical matches (Sumerian loanwords in Hurrian, according to Fournet) consisting partly of some of the Swadesh items discussed in Kassian 2011: 434-435 and in the present paper, and partly of several new etymologies that look very dubious semantically and/or phonetically.

§4.3.2. The current version of the StarLing software (May 2012) generates 12,000 BC as the approximate glottochronological date of the Sumerian-Hurrian split, proceeding from the 65 available Sumerian-Hurrian Swadesh pairs (for convenience, I date the Sumerian list to 2000 BC and the Hurrian one to 1500 BC). This is extremely distant dating—ten millennia separate attested Sumerian from its hypothetical ancestor.[34] Of course, such a large gap between empirical data and a reconstructed protolanguage makes further discussion rather vague, but, nevertheless, some conclusions can be proposed.

§4.3.3. First, as one can see, five of the six Sumerian-Hurrian Swadesh matches fall within the most stable half of the Swadesh 100-item wordlist:[35] 'dog,' 'hand,' 'liver,' 'rain,' 'who?.' Only the sixth item—'meat'—falls within the second half, although its stability index is, at 61, still high. The probability of such a distribution (5 : 1) is relatively low: 0.1478 = 14.78% (here and below, the binomial distribution is used). If we treat Sum. *šeŋ* ~ Hur. *isena* 'rain' as a negative pair, the probability of the 4 : 1 distribution is 0.2239 = 22.39%.[36] The fact that the majority of our

---

[34] For example, the same glottochronological calculations yield the late 5th millennium BC as the approximate date of Indo-Hittite split into two branches: Anatolian and Narrow IE; that is, *ca.* 2500 years separate the Indo-Hittite protolanguage and attested Anatolian languages (the distance between the Indo-Hittite protolanguage and the reconstructed Narrow IE protolanguage is even shorter). The next-level taxon is the Indo-Uralic protolanguage glottochronologically dated back to the early 9th millennium BC, i.e., the gap between the Indo-Hittite and Indo-Uralic protolanguages is less than 5 millennia.

[35] The Swadesh list is not homogeneous, but its entries possess different degrees of stability. This factor was called the relative index of stability by S. Starostin, who calculated it for each element of the Swadesh 100-item (strictly speaking, 110-item) list proceeding from typological data of various language families of the Old World (see S. Starostin 2007a; G. Starostin 2010a; Kassian 2011: 430-431 for details, with references to other approaches advocated by Pagel, Atkinson & Meade 2007 and Holman, et al. 2008).

[36] We know 37 Sumerian-Hurrian pairs from the stable 50-item subset: 'two,' 'I,' 'eye,' 'thou,' 'who,' 'fire,' 'tongue,' 'name,' 'hand,' 'what,' 'heart,' 'drink,' 'dog,' 'louse,' 'moon,' 'blood,' 'one,' 'tooth,' 'new,' 'liver,' 'eat,' 'this,' 'water,' 'nose,' 'not,' 'mouth,' 'ear,' 'that,' 'bird,' 'sun,' 'smoke,' 'tree,' 'ashes,' 'give,' 'rain,' 'neck,' 'breast.' Also, 28 Sumerian-Hurrian pairs are known from the "weak" 60-item subset: 'come,' 'foot,' 'sit,' *'thin,'* 'hear,' 'skin,' 'long,' 'meat,' 'road,' 'know,' 'say,' 'black,' 'head,' 'burn tr.,' 'earth,' *'year,'* 'fat n.,' 'man,' 'person,' 'all,' *'snake,'* 'see,' 'walk (go),' 'woman,' 'big,' 'good,' 'many,' 'mountain' (the 10 Yakhontov's words, that serve as a supplement to the classical 100-item wordlist, are italicized; actually we have but three of Yakhontov's words in

potential Sumerian-Hurrian cognates occur among the most stable Swadesh items can be due to chance (both probability values are greater than 0.05) or can be an argument in favor of the hypothesis of Sumerian-Hurrian genetic relationship: the weak items have been eliminated during separate development of proto-Sumerian and proto-Hurro-Urartian, whereas the most stable ones have survived. But it must be emphasized that such a distribution can be alternatively treated as an equally strong argument in support of a very different scenario discussed in the next section—language shift (see §4.4 below).

§4.3.4. Second, there are two objections to the hypothesis of a Sumerian-Hurrian protolanguage:

1) Despite the assumed substantial time gap (ten millennia) between the attested languages and their hypothetical Sumerian-Hurrian ancestor, one could expect a number of cognates (in our case, phonetic consonant matches) between Sumerian and Hurrian basic vocabularies outside the Swadesh 100-item wordlist. I am not aware, however, of appropriate candidates for such inherited retentions in the known Sumerian and Hurrian lexicon, except for a couple of dubious cases like Sum. *ur* 'root, base; limbs; loin, lap' ~ Hur. *uri* 'foot; leg' and Sum. ⟨NUD = NU₂ = NA₂⟩ 'to lie (down),' ⟨^{ĝeš}NUD = ^{ĝeš}NU₂ = ^{ĝeš}NA₂⟩ 'bed' ~ Hur. *nat-xi* 'bed,' discussed in §4.2.

2) It is reasonable to suppose that both proto-Sumerian and proto-Hurro-Urartian languages underwent heavy sound mutations during the millennia of their separate development, and that true Sumerian-Hurrian etymological cognates are currently invisible to the "unaided eye." Such a supposition, however, sharply contrasts with the fact noted in §4.1 above: six (or five) discussed Sumerian-Hurrian Swadesh matches are almost identical phonetically (with *š* & *ŋ* present in Sumerian and absent from Hurrian), and even vocalic segments normally coincide. Linguistic typology is aware of language families with ultra-stable consonant systems: the best

---

the known Hurrian list). Thus, we calculate the probability of the 5 : 1 (or 4 : 1) distribution among two subsets at 37 : 28.

I would like to take this opportunity to correct my miscalculation in Kassian 2011: 430-431. There are 12 Hurro-Urartian Swadesh items for which I suggested Sino-Caucasian etymologies. Out of them, 10 items ('new,' 'I,' 'thou,' 'blood,' 'louse,' 'we,' 'one,' 'this,' 'tooth,' 'ear') fall within the stable 50-item subset, whereas 2 items ('meat,' 'black') fall within the "weak" 60-item subset. Because we know 38 Hurro-Urartian items from the stable subset ('we' is added to the aforementioned words) and 29 Hurro-Urartian items from the weak subset (Urartian 'small' is added to the aforementioned words), the probability of the 10 : 2 distribution is 0.032 = 3.2%. This is much greater than 0.0003 (which I incorrectly cited), but is nevertheless lower than the significance level 0.05.

instance known to me is Semitic. Glottochronologically, the split of the Semitic protolanguage occurred in the early 4th millennium BC,[37] i.e., the time gap between a modern Semitic language and its ancestor constitutes *ca.* 6 millennia. Despite this, a simple browse through the first volume of SED shows that it is fairly easy to find a substantial number of phonetically similar roots that are in fact etymological cognates, e.g., between Modern South Arabian and Modern Ethiopian languages.[38] This is certainly not the Sumerian-Hurrian case. If one advocates for a Sumerian-Hurrian genetic relationship, it is necessary to make a methodologically impossible supposition that several inherited Sumerian-Hurrian basic terms were preserved phonetically intact, whereas the rest of basic vocabulary has mutated and lost visible phonetic similarity between the two languages.

§4.3.5. Summing up, the hypothesis of a common Sumerian-Hurrian protolanguage appears to be very unlikely, first, due to virtual absence of a substantial number of appropriate etymologies between basic vocabularies of the languages in question (not necessarily with direct semantic matches), and, second, due to the suspicious phonetic similarity of the discussed Sumerian-Hurrian Swadesh pairs.[39]

§4.4. Aborted Language Shift
§4.4.1. The fourth scenario to be discussed is an aborted language shift. As noted above, cultural vocabulary is always borrowed first among lexical items, whereas the Swadesh wordlist (the core of basic vocabulary) is gen-

---

[37]   StarLing dating, that coincides with Kitchen, et al. 2009.

[38]   Or—if we proceed from another bifurcation of the Semitic tree—between Modern Arabic and any other modern Semitic language, according to the wordlists quoted in Kitchen, et al. 2009 (see, however, Militarev 2010: 44 fn. 2 for some criticism of Kitchen, et al.'s data analysis).

[39]   In order to get the glottochronological date of the Sumerian-Hurrian split in the StarLing software, the percentage of Sumerian-Hurrian positive pairs (six or five items) within the available 65-item list has indeed been extrapolated to the standard 100-item matrix. But the assumption that the real percentage between the full Sumerian-Hurrian 100- or 110-item wordlist could be different does not change the picture, however. If we suppose that the residual 35 (or 45) Hurrian terms (being uncovered) will yield a great number of forms phonetically compatible with the corresponding Sumerian Swadesh terms (which would mean that the split of the Sumerian-Hurrian protolanguage would acquire a later date), the first counterevidence would become stronger. If rather the residual 35 (or 45) Hurrian terms demonstrate no similarity with their Sumerian counterparts (the split of the Sumerian-Hurrian protolanguage becomes even more distant), the second counterevidence would become stronger.

erally most resistant to borrowing. It is reasonable to suppose that this rule concerns not only trivial language contacts, but is also applicable to certain situations of language shift when the culturally dominated group gives up its language and shifts to the language of the dominant group. If language shift is not an abrupt process (in 1-2 generations), but a gradual replacement of the inherited linguistic material by the borrowed one, it would be reasonable to expect that, at the penultimate stage, the vocabulary of the shifting nondominant group retains only some Swadesh (or similar) items as a remnant of the original language. Theoretically, if the contact between the dominant and subordinate groups is lost (for some historical reasons), the language of the subordinate group should stabilize in a very unusual state: grammatically and lexically, it represents the language of the dominant group, whereas some retained basic terms synchronically look like loanwords.

§4.4.2. Such an aborted or simply unfinished language shift is poorly documented among the world's languages due to natural enough reasons: first, a language shift is normally completed, second, the early history of many tribes or ethnic groups around the world is unknown to us. Nevertheless some probable instances of aborted/unfinished language shift, when basic vocabulary is fragmentarily retained, can be uncovered. Two of them are treated below.

1)   As described by D. C. Laycock (1973: 252) and M. D. Ross (1991: 124), the Malol language (< Oceanic < Austronesian) is very close to the Sissano language spoken in the same or neighboring coastal villages (usually both lects are considered to be dialects). Oral history, however, indicates that the Malol people were originally one of the One clans (non-Austronesian languages of the Torricelli family) that fled from the One territory to the coast during a communal dispute in the first half of the 19th century. Currently, vocabularies of Sissano and Malol generally coincide, with the exception of a few lexical items, for which old One terms are retained in Malol. Two such words are documented by Laycock and Ross: 'dog' (a Swadesh item) and 'coconut' (belongs to the basic vocabulary in this region).

2)   Another instance can be the language of the Polynesian island Niuafo'ou. According to Collocott 1922, Dye 1980, Belikov 1989: 49, synchronically, Niuafo'ou can be considered a dialect of the Tongan language (< Tonga < Polynesian < Austronesian), that is the dominant lect in the region, but some peculiarities of the pronominal system (such as non-Tongan personal pronouns 'we [excl.],' 'you [du.],' 'you [pl.],' and the interrogatives 'when, where') and of basic vocabulary point out that, historically, Niuafo'ou is a Nuclear Polynesian language

(another branch of the Polynesian group), almost completely been supplanted by Tongan. Collocott provides the following Niuafoʻou lexical items, that are cognate to the corresponding Tongan words, but demonstrate Nuclear Polynesian phonetic development: 'to come,' 'road,' 'what?' (together with the aforementioned pronoun 'we,' these are Swadesh items), 'sea' and also such function words as 'up,' 'down.' As noted by Collocott (1922: 189), "[t]he dialectal peculiarities of Niua Foʻou are fast disappearing before the political and cultural authority of Tonga." In his turn, Dye (1980: 350) reports that at least some of the aforementioned Niuafoʻou words have already shifted towards Tongan phonology within the last decades.

§4.4.3. Probably such "intertwining" languages as Ainu/Ejnu (an Iranian language dominated by Uyghur) or Mbugu/Maʻa (a Cushitic language dominated by Bantu) are following suit, although they still retain the major portion of inherited basic vocabulary (Persian and Cushitic, respectively).

§4.4.4. As one can see, the symptoms of aborted or unfinished language shift are very similar to the Sumerian-Hurrian situation, where we have two languages with very different grammars and very different lexica, but with several similar phonetically Swadesh items shared by both lects. In other words, the correlation between the historical Sumerian and Hurrian languages is formally the same as, e.g., between One (Torricelli family) and modern Malol (Austronesian family), treated above.

§4.4.5. Another case of the retention of a certain specific part of an inherited lexicon is retention of the so-called native cultural vocabulary. Such a scenario is typically to be expected in the situation of a language shift unaccompanied by a cultural shift. Two instances are treated below.

1) As described by Dimmendaal (1989: 21-22, 27) and Heine (1980: 175-178), El Molo, or Elmolo, is a small tribe of fishermen in Kenya heavily dominated by the neighboring Nilotic-speaking pastoralists. In the first half of the 20th century, the El Molos still spoke their own language, that belongs to the Cushitic family, but subsequently they have shifted to the Samburu language (< Nilotic < Nilo-Saharan). Currently, El Molo represents a dialect of Samburu. This newborn dialect, however, retains the original El Molo vocabulary concerned with lake bio-nomenclature and fishing.

2) Another probable example is provided by two pygmy tribes—Yaka (Aka) and Baka—that live in the rainforests of Central Africa. Yaka and Baka are neighbors, although there is minimal interaction between the two peoples. The languages in question belong to very different linguistic groups: Yaka is Bantu C10, Baka is Ubangian. Despite this, Yaka and Baka are close not only physiologically, but also culturally and economically: both tribes are hunter-gatherers, as opposed to the neighboring non-pygmy farmer tribes. As described by S. Bahuchet (1992; 1993; 2012: 28-31), Yaka and Baka share more than 20% of their vocabulary, concerning especially food-gathering and other specific rain-forest activity (some shared terms are also related to society, music and religion). An important fact is that these words are apparently unetymologizable within Bantu or Ubangian languages. The rest of the lexicon of Yaka and Baka (including the majority of basic terms), however, differs according to its genetic affiliation (Bantu C10 and Ubangian). There are also some grammatical elements and features of neither Bantu nor Ubangian origin shared by Yaka and Baka, e.g., specific demonstrative pronouns (Duke 2001: 74-78). In such a situation, the most tempting solution is to treat these specific cultural terms as the remains of the pygmy protolanguage (the so-called proto-Baakaa) that were retained due to socio-economic factors after the Yaka and Baka tribes had shifted to the languages of the neighboring farmers (thus Bahuchet). An alternative solution, which seems less likely, is to assume that Yaka and Baka originally spoke Bantu and Ubangian languages, respectively, whereas the discussed common words represent parallel borrowings from a language of extinct rain-forest dwellers into Yaka and Baka. The third, more complex, solution is discussed by Blench (1999; 2006: 173-175).

§4.4.6. Despite typological interest of the El Molo and Yaka-Baka instances, such a scenario is certainly not the case of Sumerian and Hurrian due to the virtual absence of cultural lexical matches between the two languages in question.

§5. *Conclusions*
§5.1. The Sumerian and Hurrian languages demonstrate several Swadesh items that are phonetically very similar, but no lexical matches of the same level of phonetic similarity in other parts of vocabulary and no striking grammatical parallels. Four possible explanation of such a situation are discussed above. Two of them—lexical borrowing (§4.2) and genetic relationship (§4.3)—are unlikely and should be rejected due to typological objections.

§5.2. The null hypothesis that the observed Sumerian-Hurrian matches are chance coincidences (§4.1) is problematic. According to the described permutation test, the probability of such coincidences ranges from $0.069280 = 6.9280\%$ (a rough approach) to $0.003604 = 0.3604\%$ or less (a more sophisticated approach). In my opinion, the

most correct value is 0.015552-0.003604, i.e., 1.5552%-0.3604% (with the more precise consonant classes used; see §4.1, figs. 3-4), but, in any case, the majority of the obtained probabilistic values are less than the most popular significance level 0.05.

§5.3. Does it mean that the null hypothesis *must* be rejected? Certainly not, because nature is actually full of various phenomena the probability of whose emergence is low. The current version of the *Global Lexicostatistical Database* project (GLD) provides us with a substantial number of high-quality 110-item wordlists of various languages from around the world.[40] Most pairs of unrelated lects successfully pass the permutation test, i.e., the amount and probability of phonetic matches between two lists appear to be statistically expected. On the other hand, one can observe a couple of pairs of definitely unrelated languages with a high number of phonetic matches and a low probability of such a configuration. I am currently aware of two such instances.

1) The first pair is Abidji (< Kwa < Niger-Congo, Africa)[41] and Maidu (< Penutian, USA)[42]. The 110-item wordlists of the two aforementioned languages possess 7 *CC*-matches, if we proceed from the GLD consonant classes described in §1.2 (the first form cited is Abidji, the second one is Maidu):

   - *tì ~ dʼo-* 'to bite' = *TH*
   - *hí ~ ʔi-yʼe-* 'to come' = *HH*
   - *ínè ~ ʔonʼo* 'head' = *HN*
   - *pì ... été ~ ɓi-dʼoy-* 'to sit' = *PH*
   - *bɔ̀-dí ~ pi-yʼeto-* 'to swim' = *PH*
   - *ĩné ~ ʔˌen-ʼi* 'tongue' = *HN*
   - *ʔà ~ ʔi-kʼʼoy-* 'to go' = *HH*

   The probability that these Abidji-Maidu *CC*-matches are due to chance is 0.036136, i.e., 3.6136% (1,000,000 random trials have been performed). The picture does not materially change if the more precise consonant classes (see §4.1) are used: we have the same 7 matches whose probability is 0.032043 = 3.2043%.

2) The second case is more interesting: Modern English (< Germanic < Indo-European) and Ari (< South Omot-

ic < Omotic, Africa)[43] yield 8 *CC*-coincidences in the 110-item wordlist:

   - [daɪ] ~ *deʔ-* 'to die' = *TH*
   - [händ] ~ *ʔaːni* 'hand' = *HN*
   - [aɪ] ~ *ʔi* 'I' = *HH*
   - [neɪm] ~ *naːmʼi* 'name' = *NM*
   - [gəu] ~ *kay-* 'to go' = *KH*
   - [wiː, wi] ~ *woʰ, woːʰ* 'we' = *WH*
   - [huː] ~ *aʰy* 'who?' = *HH*
   - [šɔːt] ~ *cʼeːdʼi* 'short' = *ST*

The probability that these English-Ari *CC*-matches are due to chance is extremely low: 0.00044 = 0.044% (1,000,000 random trials have been performed). Again, the picture does not seriously change if the more precise consonant classes (see §4.1) are used: we only have 7 matches ([šɔːt] ~ *cʼeːdʼi* is now a negative pair), but the total probability is 0.000945 = 0.0945%.

§5.4. Nevertheless, despite such unique instances as Abidji-Maidu or English-Ari, the low probability of the Sumerian-Hurrian matches impel us to search for more appropriate explanations.

§5.5. The fourth solution is the hypothesis of aborted language shift (discussed in §4.4), that implies one of two equivalent scenarios.

1) In the preliterate or early literate epoch (say, the second half of the 4th millennium BC), a tribe that spoke a language of the Hurro-Urartian family (not necessarily the Hurro-Urartians proper) migrated from the southern Caucasus to southern Mesopotamia, where it entered into interaction with the Sumerian community. The Sumerians appeared to be the dominant group and the Hurro-Urartian newcomers began gradually to give up their language. At the penultimate stage of that language shift, the process was for unknown reasons interrupted, whereas the Sumerians proper were eliminated. If so, the historical Sumerians were actually a Hurro-Urartian-like people that shifted to the Sumerian language, having retained several Swadesh terms of Hurro-Urartian origin.[44]

---

[40] http://starling.rinet.ru/cgi-bin/main.cgi?root=new100&morpho=0 [last visited 02.06.2012]

[41] G. Starostin 2011a, http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\kwa\agn&limit=-1 [last visited 02.06.2012].

[42] Zhivlov 2012, http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\pen\mai&limit=-1 [last visited 02.06.2012].

[43] G.Starostin 2011b, http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\omo\som&limit=-1 [last visited 02.06.2012]..

[44] The full analogy is a hypothetical scenario in which the Malol people (§4.4 above) would assimilate or murder all the neighboring Sissanos. In such a case, we would deal with Malol as the only known dialect of Sissano and the "Papuan" "loanwords" in the Malol Swadesh list would represent a typological mystery.

2) The second scenario mirrors the first one. A Sumerian-like tribe migrated to the southern Caucasus and then learned the proto-Hurro-Urartian language. If so, the historical Hurrians and Urartians are actually a Sumerian (or related) people that shifted to the Hurro-Urartian language, having retained several Swadesh terms of Sumerian origin.

§5.6. I am aware of no historical or archaeological counterevidence for the theory of aborted language shift between Sumerian and Hurro-Urartian peoples in the preliterate or early literate epoch, as described above. It should be noted that if Hurro-Urartian can indeed be considered a separate branch of the Sino-Caucasian macro-family (see Kassian 2011 for a lexicostatistical discussion) and if such terms as 'meat' and 'rain,' shared by Sumerian and Hurro-Urartian, are indeed etymologically Sino-Caucasian (see §3), the first scenario (the Hurro-Urartian language superseded by Sumerian) is preferable. Since the Kura-Araxes (Early Trans-Caucasian) archaeological culture seems the best counterpart of the proto-Hurro-Urartian language (and, vice versa, the proto-Hurro-Urartian language seems the best counterpart of the Kura-Araxes culture; see Kassian 2010: 423-428 with further references), the hypothetical migration of a Hurro-Urartian-like group to southern Mesopotamia should be connected to the rapid spread of the Kura-Araxes culture along the eastern slopes of the Zagros at least as far as west central Iran in the last centuries of the 4th millennium BC (for which see Kohl 2009: 245-246, 252-255).[45] On the other hand, the sound correspondences like Sum. ŋ—HU n and Sum. š—HU s are more easily explainable under the assumption of the second scenario (Sumerian superseded by Hurro-Urartian).

---

[45]   Alexander Nemirovsky has suggested to me (personal communication) that another theoretical possibility is to attribute the pre-Sumerian substratum (the so-called proto-Euphratic or Banana language, although see the criticism by Rubio 1999; 2005) to the Hurro-Urartian linguistic family.

BIBLIOGRAPHY

Abbreviations

Basqet.dbf = *Basque etymological database* by John Bengtson. Available online at the *Tower of Babel* project: http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl [last visited 25.12.2013].

Buruet.dbf = *Burushaski etymological database* by S. Starostin (based on H. Berger's data). Available online at the *Tower of Babel* project: http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl [last visited 25.12.2013].

Caucet.dbf = *North Caucasian etymological database* by S. Starostin and S. Nikolayev (published as *NCED*). Available online at the *Tower of Babel* project: http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl [last visited 25.12.2013].

CDLI = *Cuneiform Digital Library Initiative*. Available at: http://cdli.ucla.edu/ [last visited 25.12.2013].

CdTU = Salvini 2008.

ePSD = *Electronic Pennsylvania Sumerian Dictionary Project*. Available at: http://psd.museum.upenn.edu/epsd/index.html [last visited 25.12.2013].

ETCSL = *The Electronic Text Corpus of Sumerian Literature*. Available at: http://etcsl.orinst.ox.ac.uk [last visited 25.12.2013].

GLD = G. Starostin, ed., *The Global Lexicostatistical Database*. Available online at: http://starling.rinet.ru/new100/main.htm [last visited 25.12.2013].

KUKN = Harouthiounyan 2001.

NCED = Nikolayev & Starostin 1994.

Sccet.dbf = *Sino-Caucasian etymological database* by S. Starostin. Available online at the *Tower of Babel* project: http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl [last visited 25.12.2013].

SED = Militarev & Kogan 2000-.

Stibet.dbf = *Sino-Tibetan etymological database* by S. Starostin (= Peiros & Starostin 1996, but with serious improvement). Available online at the *Tower of Babel* project: http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl [last visited 25.12.2013].

WOLD = M. Haspelmath & U. Tadmor, eds., *The World Loanword Database*. Available online at: http://wold.livingsources.org/ [last visited 25.12.2013].

Yenet.dbf = *Yenisseian etymological database* by S. Starostin (= S. Starostin 1995; Werner 2002, with additions and corrections). Available online at the *Tower of Babel* project: http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl [last visited 25.12.2013].

Anderson, Gregory D. S.

2004	"Advances in proto-Munda reconstruction." *Mon-Khmer Studies* 34, 159-184.
2008	"Introduction to the *Munda Languages*." In D. Anderson, ed., *The Munda Languages*. London / New York: Routledge, pp.1-10.

Bahuchet, Serge

1992	*Dans la forêt d'Afrique Centrale: les pygmées Aka et Baka*. Paris: Peeters-Selaf.
1993	"History of the inhabitants of the central African rain forest: perspectives from comparative linguistics." In C. Hladik, et al., eds., *Tropical forests, people, and food: Biocultural interactions and applications to development*. Paris: Unesco/Parthenon, pp. 37-54.
2012	"Changing language, remaining pygmy." *Human Biology,* 84/1, 11-43.

Baxter, William H.

1995	"'A stronger affinity ... than could have been produced by accident': A probabilistic comparison of Old Chinese and Tibeto-Burman." In W. Wang, ed., *The Ancestry of the Chinese Language*. Berkeley: University of California Press, pp. 1-39.
1998	Response to Oswalt and Ringe. In J. Salmons & B. Joseph, eds., *Nostratic: sifting the evidence*. Amsterdam: Benjamins, pp. 217-236.

Baxter, William H. & Manaster Ramer, Alexis

    1996   Review of: D. Ringe. *On Calculating the Factor of Chance in Language Comparison*. In *Diachronica* 13, 371-384.

    2000   "Beyond lumping and splitting: Probabilistic issues in historical linguistics." In C. Renfrew, et al., eds., *Time Depth in Historical Linguistics*. Cambridge: McDonald Institute for Archaeological Research, pp. 167-188.

Belikov, V. I.

    1989   "Drevnejshaya istoriya i real'nost' lingvogeneticheskikh dendrogramm." In *Lingvisticheskaya rekonstrukciya i drevneyshaya istoriya vostoka: materialy k diskussiyam na Mezhdunarodnoy konferencii (Moskva, 29 maya—2 iyunya 1989 g.)*, vol. 1. Moscow: Nauka, pp. 44-54.

Bengtson, John D.

    1997   "The riddle of Sumerian: a Dene-Caucasian language?" *Mother Tongue* 3, pp. 63-74.

    2008   *Linguistic Fossils: Studies in Historical Linguistics and Paleolinguistics*. Calgary: Theophania Publishing.

    2008a   "The Problem of "Isolates" II: Burushaski." *Bengtson 2008*, 55-70.

    2008b   "Materials for a Comparative Grammar of the Dene-Caucasian (Sino-Caucasian) Languages." In *Aspects of Comparative Linguistics*, vol. 3. Moscow: RSUH Publishers, pp. 45-118.

Bengtson, John D. & Blažek, Václav

    2011   "On the Burushaski-Indo-European hypothesis by I. Čašule." *Journal of Language Relationship* 6, 25-63.

Bengtson, John D. & Starostin, George

    forthcoming   "The Sino-Caucasian (Dene-Caucasian) hypothesis: State of the art and perspectives."

Blench, Roger M.

    1999   "Are the African pygmies an ethnographic fiction?" In K. Biesbrouck, S. Elders & G. Rossel, eds., *Central African hunter-gatherers in a multi-disciplinary perspective: challenging elusiveness*. Leiden: Centre for Non-Western Studies, pp. 41-60.

    2006   *Archaeology, Language and the African Past*. Lanham, Maryland: AltaMira Press.

Brown, Cecil H., Holman, Eric W. & Wichmann, Søren

    2013   "Sound correspondences in the world's languages." *Language* 89/1, 4-29.

Burlak, Svetlana A. & Starostin, Sergei A.

    2005   *Sravnitel'no-istoricheskoe yazykoznanie* [*Comparative Linguistics*]. 2nd ed. Moscow: Academia.

Campbell, Lyle & Poser, William J.

    2008   *Language Classification: History and Method*. Cambridge, UK: Cambridge University Press.

Collocott, E. E. V.

    1922   "The speech of Niua Fo'ou." *The Journal of the Polynesian Society* 31/4 (124), pp. 185-189.

Diakonoff, I. M.

    1971   *Hurrisch und Urartäisch. Münchener Studien zur Sprachwissenschaft*, Bh. 6 N.F.. Munich.

    1997   "External connections of the Sumerian language." *Mother Tongue* 3, 54-62.

Dimmendaal, G. J.

    1989   "On language death in eastern Africa." In N. C. Dorian, ed., *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge, UK: Cambridge University Press, pp. 13-32.

Dolgopolsky, A. B.

    1964   "Gipoteza drevnejshego rodstva yazykov Severnoj Evrazii s veroyatnostnoj tochki zreniya." *Voprosy yazykoznaniya* 2, 53-63.

    1986   "A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia." In V. Shevoroshkin & T. Markey, eds., *Typology, Relationship, and Time: A Col-*

lection of Papers on Language Change and Relationship by Soviet Linguists. Ann Arbor: Karoma, pp. 27-50.

Duke, Daniel J.

2001   *Aka as a contact language: sociolinguistic and grammatical evidence*. University of Texas, Arlington, MA Thesis. Available at: www.sil.org/Africa/Cameroun/bydomain/linguistics/theses/Complete%20Thesis-DDuke.pdf

Dunn, Michael & Terrill, Angela

2012   "Assessing the lexical evidence for a Central Solomons Papuan family using the Oswalt Monte Carlo Test." *Diachronica* 29/1, 1-27.

Dye, Tom S.

1980   "The linguistic position of Niuafo'ou." *The Journal of the Polynesian Society* 89/3, 349-357.

Englund, Robert K.

1990   *Organisation und Verwaltung der Ur III-Fischerei. BBVO*, 10. Berlin: Dietrich Reimer.

Fournet, Arnaud

2011   "About some features of loanwords in Hurrian." *Aramazd: Armenian Journal of Near Eastern Studies* 6/1, 43-59.

George, Andrew R.

2003   *The Babylonian Gilgamesh Epic. Introduction, Critical Edition and Cuneiform Texts*. 2 vols. Oxford: Oxford University Press.

Harouthiounyan, Nicolay V.

2001   *Korpus urartskikh klinoobraznykh nadpisey* [*Corpus of Urartian cuneiform inscriptions*]. Yerevan: Gitutyun.

Haspelmath, Martin

2008   "Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability." In Th. Stolz, et al., eds., *Aspects of Language Contact. New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*. Berlin: Mouton de Gruyter, pp. 43-62.

Haspelmath, Martin & Tadmor, Uri (eds.)

2009   *Loanwords in the World's Languages. A Comparative Handbook*. Berlin: Mouton de Gruyter.

Hazenbos, Joost

2005   "Hurritisch und Urartäisch." In M. Streck, ed., *Sprachen des Alten Orients*. Darmstadt: Wissenschaftliche Buchgesellschaft, pp. 135-158.

Heine, Bernd

1980   *The Non-Bantu Languages of Kenya*. Berlin: Dietrich Reimer.

Holman, Eric W., et al.

2008   "Explorations in automated language classification." *Folia Linguistica* 42, 331-354.

Jagersma, Abraham H.

2010   *A Descriptive Grammar of Sumerian*. PhD thesis, Leiden University.

Justeson, John S., and Stephens, Laurence D.

1980   "Chance cognation: a probabilistic model and decision procedure for historical inference." In E. Traugott, R. Labrum & S. Shepherd, eds., *Papers from the Fourth International Conference on Historical Linguistics, Stanford, March 26-30 1979*. Herndon, Virginia: J. Benjamins, pp. 37-45.

Kassian, Alexei

2010   "Hattic as a Sino-Caucasian language." *Ugarit-Forschungen* 41, 309-447.

2011   "Hurro-Urartian from the lexicostatistical viewpoint." *Ugarit-Forschungen* 42, 383-451.

2013   "On Forni's Basque-Indo-European Hypothesis." *JIES* 41/1-2, 181-201.

Kassian, Alexei, et al.

    2010   "The Swadesh wordlist. An attempt at semantic specification." *Journal of Language Relationship* 4, pp. 46-89.

Kessler, Brett

    2007   "Word similarity metrics and multilateral comparison." *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*. Prague: Association for Computational Linguistics, pp. 6-14.

Kessler, Brett & Lehtonen, Annukka

    2006   "Multilateral comparison and significance testing of the Indo-Uralic question." In P. Forster & C. Renfrew, eds., *Phylogenetic Methods and the Prehistory of Languages*. Cambridge, UK: McDonald Institute for Archaeological Research, 33-42.

Kitchen, Andrew, et al.

    2009   "Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East." *Proceedings of the Royal Society: Biological Sciences* 276, 2703-2710.

Kohl, Philip L.

    2009   "Origins, homelands and migrations. Situating the Kura-Araxes Early Transcaucasian 'culture' within the history of Bronze Age Eurasia." *Tel Aviv* 36, 241-265.

Krauss, Michael E. & Leer, Jeff

    1981   *Athabaskan, Eyak, and Tlingit Sonorants. Alaska Native Language Center Research Papers* 5. Fairbanks: ANLC.

Laycock, Don C.

    1973   "Sissano, Warapu, and Melanesian pidginization." *Oceanic Linguistics* 12, 245-277.

McMahon, April & McMahon, Robert

    2005   *Language Classification by Numbers*. Oxford: Oxford University Press.

Militarev, Alexander

    2010   "A complete etymology-based hundred wordlist of Semitic updated: Items 1-34." *Journal of Language Relationship* 3, 43-78.

Militarev, Alexander & Kogan, Leonid

    2000-   *Semitic Etymological Dictionary* (*AOAT* 278). Vol. 1: *Anatomy of Man and Animals*. Münster: Ugarit-Verlag, 2000. Vol. 2: *Animal Names*. Münster: Ugarit-Verlag, 2005.

Nichols, Johanna

    2010   "Proving Dene-Yeniseian genealogical relatedness." *The Dene-Yeniseian Connection. Anthropological Papers of the University of Alaska* 5/1-2, 299-309.

Nikolaev, Sergei L.

    1991   "Sino-Caucasian languages in America. Preliminary report." *Dene-Sino-Caucasian Languages: Materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8-12 November 1988*. Bochum:Brockmeyer, pp. 42-66.

Nikolayev, Sergei L. & Starostin Sergei. A.

    1994   *A North Caucasian Etymological Dictionary*. Moscow [reprinted: 3 vols. Ann Arbor: Caravan Books, 2007]. Available online as Caucet.dbf and http://starling.rinet.ru/Texts/caucpref.pdf.

Oswalt, Robert L.

    1970   "The detection of remote linguistic relationships." *Computer Studies in the Humanities and Verbal Behavior* 3, 117-129.

Pagel, Mark, Atkinson, Quentin D. & Meade, Andrew

    2007   "Frequency of word-use predicts rates of lexical evolution throughout Indo-European history." *Nature* 449, 717-720.

Peiros, Ilia I. & Starostin, Sergei. A.

    1996   *A Comparative Vocabulary of Five Sino-Tibetan Languages*. 6 vols. Melbourne: Melbourne University Press.

Pinnow, Heinz-Jürgen

    1959   *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Harrassowitz.

Richter, Thomas

    2012   *Bibliographisches Glossar des Hurritischen*. Wiesbaden: Harrassowitz.

Ringe, Donald A.

    1992   *On Calculating the Factor of Chance in Language Comparison. TAPS* 82/1. Philadelphia: American Philosophical Society.

    1998   "A probabilistic evaluation of Indo-Uralic." In J. Salmons & B. Joseph, eds., *Nostratic: sifting the evidence*. Amsterdam: Benjamins, pp. 153-197.

Ross, Malcolm D.

    1991   "Refining Guy's Sociolinguistic Types of Language Change." *Diachronica* 8/1, 119-129.

Rubio, Gonzalo

    1999   "On the alleged pre-Sumerian substratum." *Journal of Cuneiform Studies* 51, 1-16.

    2005   "On the linguistic landscape of early Mesopotamia." In W. van Soldt, et al., eds., *Ethnicity in Ancient Mesopotamia. Papers Read at the 48th Rencontre Assyriologique Internationale, Leiden, 1-4 July 2002. PIHANS* 102. Leiden: Nederlands Instituut voor het Nabije Oosten, pp. 316-332.

Salvini, Mirjo

    1998   "The earliest evidences of the Hurrians before the formation of the reign of Mittanni." In G. Buccellati & M. Kelly-Buccellati, eds., *Urkesh and the Hurrians Studies in Honor of Lloyd Cotsen. Urkesh/Mozan Studies* 3. Malibu, pp. 99-115.

    2008   *Corpus dei testi urartei*. Vol. 1-3. Rome.

Sidwell, Paul

    2010   "The Austroasiatic central riverine hypothesis." *Journal of Language Relationship* 4, 117-134.

Starostin, George S.

    2008   *Making a Comparative Linguist out of your Computer: Problems and Achievements*. Presentation at the Santa Fe Institute, August 12, 2008. Available at: http://starling.rinet.ru/Texts/computer.pdf

    2010a   "Preliminary lexicostatistics as a basis for language classification: A new approach." *Journal of Language Relationship* 3, 79-116.

    2010b   "Dene-Yeniseian and Dene-Caucasian: Pronouns and other thoughts." *Working Papers in Athabaskan Languages 2009: Alaska Native Language Center Working Papers* 8. Fairbanks: ANLC, pp. 107-117.

    2011a   *Annotated Swadesh wordlists for the Agneby group (Kwa family)*. Database compiled and annotated by G. Starostin (last version: October 2011). Available at GLD: http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\kwa\agn&limit=-1

    2011b   *Annotated Swadesh wordlists for the South Omotic group (Omotic family)*. Database compiled and annotated by G. Starostin (2011). Available at GLD: http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\omo\som&limit=-1

    2012   "Dene-Yeniseian: a critical assessment." *Journal of Language Relationship* 8, 117-138.

Starostin, Sergei A.

    1982/2007   "Praeniseyskaya rekonstrukciya i vneshnie svyazi eniseyskikh yazykov." *Starostin 2007*, pp. 147-246 (first publ.: *Ketskiy sbornik*. Leningrad [1982] 144-237).

    1995   "Sravniteľnyj slovar' eniseyskikh yazykov." *Ketskiy sbornik (Studia Ketica)* 4. Moscow, pp. 176-315.

    2007   *Trudy po yazykoznaniyu* [*Works in Linguistics*]. Moscow: LRC Publishing House.

2007a	"Opredelenie ustojchivosti bazisnoj leksiki [Defining the stability of basic lexicon]." In *Starostin 2007*, pp. 827-839.
n.d.	*Sino-Caucasian*. Unfinished MS. Available online at the *Tower of Babel* project: http://starling.rinet.ru/Texts/scc.pdf

Tadmor, Uri, Haspelmath, Martin & Taylor, Bradley

2010	"Borrowability and the notion of basic vocabulary." *Diachronica* 27/2, 226-246.

Thomason, Sarah G.

2001	*Language Contact*. Edinburgh: Edinburgh University Press.

Thomason, Sarah G. & Kaufman, Terrence

1988	*Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Turchin, Peter, Peiros, Ilia & Gell-Mann, Murray

2010	"Analyzing genetic connections between languages by matching consonant classes." *Journal of Language Relationship* 3, 117-126.

Vajda, Edward J.

2012	"The Dene-Yeniseian connection: a reply to G. Starostin." *Journal of Language Relationship* 8, 138-150.

Waetzoldt, Hartmut

1997	"Die Berufsbezeichnung *tibira*." *NABU* 1997/96.

Werner, Heinrich

2002	*Vergleichendes Wörterbuch der Jenissej-Sprachen*. 3 vols. Wiesbaden: Harrassowitz.

Wegner, Ilse

2007	*Hurritisch. Eine Einführung*. 2nd rev. ed. Wiesbaden: Harrassowitz.

Wilcke, Claus

2010	"Sumerian: What we know and what we want to know." L. Kogan, et al., eds., *Proceedings of the 53e Rencontre Assyriologique Internationale* 1/1. *Babel und Bibel* 4. Winona Lake: Eisenbrauns, pp. 5-76.

Wilhelm, Gernot

1988	"Gedanken zur Frühgeschichte der Hurriter und zum hurritisch-urartäischen Sprachvergleich." In V. Haas, ed., *Hurriter und Hurritisch. Xenia* 21. Konstanz: Universitätsverlag, 43-67.
2008	"Hurrian." In R. Woodard, ed., *The Ancient Languages of Asia Minor*. Cambridge: Cambridge University Press, pp. 81-104.

Yakubovich, Ilya

2009	"Phonetic Interpretation of Hurrian Sibilants in the Light of Indo-European Evidence." Talk given at the conference *The Sound of Indo-European: Phonetics, Phonemics, and Morphophonemics*, Copenhagen, April 2009.

Zhivlov, Mikhail

2012	*Annotated Swadesh wordlists for the Maiduan group (Penuti family)*. Database compiled and annotated by M. Zhivlov (March 2012). Available at GLD: http://starling.rinet.ru/cgi-bin/response.cgi?root=new100&morpho=0&basename=new100\pen\mai&limit=-1 [last visited 25.12.2013].