

Российский государственный гуманитарный университет
Russian State University for the Humanities



Russian State University for the Humanities
Institute of Linguistics of the Russian Academy of Sciences

Journal of Language Relationship

International Scientific Periodical

№19/2



Moscow 2021

Российский государственный гуманитарный университет
Институт языкознания Российской Академии наук

Вопросы языкового родства

Международный научный журнал

№ 19/2



Москва 2021

Advisory Board:

H. EICHNER (Vienna) / Chairman
W. BAXTER (Ann Arbor, Michigan)
V. BLAŽEK (Brno)
L. HYMAN (Berkeley)
F. KORTLANDT (Leiden)
A. LUBOTSKY (Leiden)
J. P. MALLORY (Belfast)
A. YU. MILITAREV (Moscow)
V. F. VYDRIN (Paris)

Editorial Staff:

V. A. DYBO (Editor-in-Chief)
G. S. STAROSTIN (Managing Editor)
T. A. MIKHAILOVA (Editorial Secretary)
A. V. DYBO
M. A. MOLINA
M. N. SAENKO
I. S. YAKUBOVICH

Founded by Kirill BABAEV

Редакционный совет:

Х. АЙХНЕР (Вена) / председатель

В. БЛАЖЕК (Брно)

У. БЭКСТЕР (Анн Арбор)

В. Ф. ВЫДРИН (Париж)

Ф. КОРГЛАНДТ (Лейден)

А. ЛУБОЦКИЙ (Лейден)

Дж. МЭЛЛОРИ (Белфаст)

А. Ю. МИЛИТАРЕВ (Москва)

Л. ХАЙМАН (Беркли)

Редакционная коллегия:

В. А. ДЫБО (главный редактор)

Г. С. СТАРОСТИН (заместитель главного редактора)

Т. А. МИХАЙЛОВА (ответственный секретарь)

А. В. ДЫБО

М. А. МОЛИНА

М. Н. САЕНКО

И. С. ЯКУБОВИЧ

Журнал основан К. В. БАБАЕВЫМ

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2021. — № 19/2. — x + 81 с.

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by V. A. Dybo. — Moscow, 2021. — No. 19/2. — x + 81 p.

ISSN 2219-3820

<http://www.jolr.ru/>
gstarst1@gmail.com

Дополнительные знаки: С. Г. Болотов
Add-on symbols by S. G. Bolotov

Подписано в печать 10.07.2021. Формат 60×90/8.
Усл. печ. л. 11,5. Уч.-изд. л. 7,3. Заказ № 1289.
Тираж 1050 экз.

Издательский центр
Российского государственного гуманитарного университета
125993, Москва, Миусская пл., 6
www.rggu.ru
www.knigirggu.ru

Table of Contents / Содержание

Table of Contents / Содержание	vii
Contributors / Сведения об авторах	viii
Note for Contributors / Будущим авторам	ix

Articles / Статьи

<i>John D. Bengtson, Corinna Leschber.</i> Notes on some Pre-Greek words in relation to Euskaro-Caucasian (North Caucasian + Basque)	71
[Джон Бенгтсон, Коринна Лешбер. О возможном эускаро-кавказском (баскско-северокавказском) происхождении некоторых субстратных лексических элементов в греческом языке]	
<i>George Starostin.</i> Lexicostatistical studies in Khoisan II/1: How to make a Swadesh wordlist for Proto-Tuu (Proto-South Khoisan)	99
[Г. С. Старостин. Лексикостатистические исследования по койсанским языкам II/1: к вопросу о составлении списка Сводеша для пра-ту (праюжнокойсанского) языка]	
<i>Nina Sumbatova, Valentin Vydrin.</i> N-initial nouns in Landuma and their counterparts in Mande	136
[Н. Р. Сумбатова, В. Ф. Выдрин. Существительные с начальным носовым согласным в ландума и их соответствия в языках манде]	

Contributors

John D. Bengtson — Evolution of Human Language Project, Santa Fe Institute; Association for the Study of Language in Prehistory, Minnesota, jdbengt@softhome.net

Corinna Leschber — Ph.D, head of the Institute for Linguistic and Cross-Cultural Studies, Berlin, corinna.leschber@icloud.com

George Starostin — candidate of sciences (Philology); chief research fellow, Institute for Oriental and Classical Studies,

Higher School of Economics (Moscow); external professor, Santa Fe Institute (New Mexico, USA), gstarst1@gmail.com

Nina Sumbatova — doctor of sciences (Philology), senior researcher, Institute of Linguistics, Russian Academy of Sciences (Moscow), nina.sumbatova@gmail.com

Valentin Vydrin — doctor of sciences (Philology), Center for African Linguistics, Languages and Cultures (LLACAN) (Paris), vydrine@gmail.com

Сведения об авторах

Бенгтсон, Джон — Программа «Эволюция языка», Институт Санта-Фе, Нью-Мексико; Ассоциация изучения языка в доисторический период, Миннесота, jdbengt@softhome.net

Выдрин, Валентин Феодосиевич — доктор филол. наук, Центр изучения языков и культур Африки (LLACAN) (Париж), vydrine@gmail.com

Лешбер, Коринна — Ph.D, директор Института лингвистических и кросс-культурных исследований, Берлин, corinna.leschber@icloud.com

Старостин, Георгий Сергеевич — канд. филол. наук, главный научный сотрудник Института классического Востока и античности ВШЭ (Москва), внешний сотрудник Института Санта-Фе (Нью-Мексико, США), gstarst1@gmail.com

Сумбатова, Нина Романовна — доктор филол. наук, ст. науч. сотрудник Института языкознания РАН (Москва), nina.sumbatova@gmail.com

Note for Contributors

Journal of Language Relationship welcomes submissions from everyone specializing in comparative-historical linguistics and related disciplines, in the form of original articles as well as reviews of recent publications. Articles are published preferably in English or Russian, although publication of texts in other major European languages (such as French or German) may be considered by the Editorial Board.

All submissions should be uploaded electronically in MS Word and PDF format, using the online *Manuscript Submission Form* at the official website of the Journal (<http://jolr.ru>). Each article should be accompanied with information about the author(s) (names, affiliations, contact information), an abstract (not exceeding 300 words) and relevant keywords.

For more detailed guidelines on article submission and editorial policies, please see our website (<http://jolr.ru>) or address the Editorial Board directly at gstarst1@gmail.com. Inquiries may also be sent by regular mail to the official address of the Journal:

Journal of Language Relationship
Institute for Oriental and Classical Studies
Russian State University for the Humanities
125993 Moscow, Russia
Miuskaya Square, 6

Будущим авторам

Журнал *Вопросы языкового родства* принимает заявки на публикацию оригинальных научных статей, а также рецензий от всех, кто специализируется в области сравнительно-исторического языкознания и смежных дисциплин. Предпочтительные языки публикации — английский или русский, хотя по согласованию с редакционной коллегией возможна также публикация статей на других крупных европейских языках (французский, немецкий и т. п.).

Подача материалов для публикации (в форматах MS Word и PDF) осуществляется через *Электронную форму подачи рукописей* на официальном сайте журнала (<http://jolr.ru>). К каждой статье обязательно прилагаются сведения об авторах (имена, аффилиации, контактная информация), краткое резюме (не более 300 слов) и список подходящих ключевых слов.

Подробнее о требованиях к оформлению рукописи, редакционной политике журнала и т. п. вы можете узнать на нашем сайте по адресу: <http://www.jolr.ru> или непосредственно от редакционной коллегии по электронной почте (gstarst1@gmail.com). По различным вопросам с редакцией журнала можно также связаться по обычной почте:

125993 Москва
Миусская площадь, д. 6
Российский государственный гуманитарный университет
Институт восточных культур и античности
В редакцию журнала «Вопросы языкового родства»

John D. Bengtson[†], Corinna Leschber[‡]

[†] Santa Fe Institute, Evolution of Human Languages Project; palaeojdb@hotmail.com

[‡] Institute for Linguistic and Cross-Cultural Studies, Berlin; corinna.leschber@icloud.com

Notes on some Pre-Greek words in relation to Euskaro-Caucasian (North Caucasian + Basque)

A “Pre-Greek” substratum underlying the Indo-European Greek language has been suspected for a long time. There is no reason to suppose that there was only one “Pre-Greek” language; the region where Greek was and is spoken may have been multilingual, with languages of diverse origins. In the following study a limited number of etyma are examined that seem to bear witness to a widespread Euskaro-Caucasian language (or language family) associated with the spread of agriculture out of Anatolia. Greek words like ἀκαοί ‘mite’, μαστός ‘breast, teat’, β/μύσταξ ‘upper lip, mustache’, ξύλον ‘wood, timber’, and ψῦχή ‘breath’ are basic and not likely to be cultural loans, and could reflect genuine relics of a Euskaro-Caucasian Pre-Greek language. The examples discussed here are probably part of a much larger subset that a thorough study of Furnée’s and Beekes’ total list of “Pre-Greek” words might yield.

Keywords: Basque language; North Caucasian languages; Euskaro-Caucasian hypothesis; Pre-Greek language; linguistic substrates.

A “Pre-Greek” substratum underlying the Indo-European Greek language has been suspected for a long time. Recently Beekes (2010: xiv) reiterated his rejection of the ‘Pelasgian’ theory (of an earlier Indo-European substratum underlying Greek) and preferred Furnée’s (1972) “elaboration of Kuiper’s 1956 study on Greek substrate words, which opened a new chapter in the research of the field.” Beekes resumed that “Furnée’s book met with fierce criticism and was largely neglected. In my view, this was a major mistake in Greek scholarship.” In his 2010 dictionary Beekes devotes ample attention to Pre-Greek, but the “comparison with Basque or Caucasian languages has not been considered in this dictionary, as this is not my [Beekes’] competence; it is likely that there are such connections, but this must be left to other scholars” (Beekes 2010: xv).

One of the current writers had an early exposure to this topic in Nikolaev’s (1985) “North Caucasian loanwords in Hittite and Ancient Greek” (in Russian). According to a current Moscow colleague, “Ancient Greek dialects possess a number of North Caucasian loanwords, see Николаев, 1985 (some [of] Nikolaev’s connections are highly questionable, but some seem probative)” (Kassian 2010: 404).

It seems that there is no reason to suppose that there was only one “Pre-Greek” language, and that the region where Greek was and is spoken may have been multilingual, with languages of diverse origins. Georgiev (1937, 1941) proposed a Pre-Greek language that was Indo-European of a *satəm* type, with *Lautverschiebung*, and close to Thracian. At present Mihaylova (e.g. 2017) holds firmly with Georgiev’s model. Another hypothesis is that of Eric Pratt Hamp (1983, 1985, 1989a, 1989b), also proposing an IE Pre-Greek language with *Lautverschiebung* and *Lex Grassmann*, but of a *centum* type (e.g., πύργος ‘tower’). Besides the possible IE sources and the Euskaro-Caucasian language proposed here, some of the Pre-Greek words have other, non-IE origins: Semitic or Hurrian are primary candidates.¹ The time span is so

¹ Thanks to notes from V. Blažek (p.c. 09-06-2020).

long that it is probable that there were many influences on the formation of the Greek language, which will never be fully disentangled.

In the following study a limited number of etyma are examined that seem to bear witness to a widespread Euskaro-Caucasian language (or language family) associated with the spread of agriculture out of Anatolia (Ehret 2015: 90; BCR 453–460; Bengtson 2017b).² Some of the examples coincide, more or less, with Nikolaev’s, as indicated. In general, these examples have been selected so that (a) the Greek words are endorsed as ‘Pre-Greek’ (or probably non-Indo-European) by Beekes, (b) there are putative North Caucasian cognates (updated to conform with NCED, published almost a decade later than Nikolaev 1985), and/or (c) there exist putative Basque cognates (most of them as cited in BCR).

As a preface to this study a disclaimer should be issued, that the following list of putative substratal words is preliminary, and it is not expected that all of the examples will eventually prove to be substrate words. All readers are invited to put forth alternative explanations, if these can be found.³

ἄκαοί ‘mite’ / **κόοις** ‘bug, bedbug, *Cimex lectularius*’: “I would rather think that **κόοις** is cognate [with **ἄκαοί**], as a substrate word, with prothetic vowel and *α/ο* interchange” (Beekes 49: 754). | **North Caucasian**: Avar *k’:ará* ‘mosquito’, Andi *k’:ara*, Tindi *k’:ara*, Bagwali *č’:ara* id., Karata *k’:ara* ‘gadfly’, Chamali *c’:ara* id., Akhwakh *k’:ara* ‘ant, bug’, etc.; Bezhta *kälä* ‘mosquito’, Hunzib *kelo* id.; Chechen *gora* ‘gadfly’, Ingush *gor* id. < PEC * (*ḳārā* ~ **ḳārā*) (NCED 719). | **Basque** **kaia*-/**karkar*:- (G) *karrakaldo* ‘beetle’, *karkarraldo*, *karkaraldo*, *kakalarro*, (BN) *karkamalo*, (B, G, AN) *kakalarro*, (B) *kakarraldo*, *karkaraldo*, (G, AN) *kakalerdo*, etc.; obviously many expressive changes have occurred. § Contrary to the note about **ἄκαοί** being substrate and cognate with **κόοις**, Beekes’ lemma on the latter word claims **κόοις** is “identical with Ru. *kor*’ [f.] ‘moth’, and traditionally analyzed as an old verbal noun from *(s)*ker*- ‘shave, split, cut’ seen in ... *κείωω* etc.” From a Sino-Caucasian perspective, cf. also Burushaski **kharú* ‘louse’, Tibeto-Burman **k(h)r[ā]-η* ‘mosquito’ (SCG 119–20).

ἄλωή ‘threshing floor, garden’ (Iliad), ‘halo’ (around sun and moon) ...; also ‘disk’ of the sun or moon, or of a shield; **ἄλοάω**, **ἄλοιάω** ‘to thresh, crush’ (Iliad); etymology unknown (Beekes 78). | **North Caucasian**: Tindi =*elī*- ‘to thresh’; Bezhta =*ol*-; Batsbi *arl*-, Chechen *ār*-, ‘to thresh’, *ēra* ‘threshing-floor; grain lying upon it’, Ingush *ard*- ‘to thresh’; Archi *iλ* (ač:as) ‘to thresh’, *iλ* = *itł* ‘threshing; grain prepared for threshing’; (with many derivatives) Archi *λorom* = *tłorom* ‘threshing board’;⁴ Avar *lol* ‘threshing board’, Andi *loli* ‘threshing; threshing-floor’, Tindi *rali* ‘grain ready for threshing’, Karata *lale* ‘threshing’; Tsezi *reła*-y ‘threshing’, Hinukh *reła*, id., etc. < PEC *=*ṼrLV*

² “I think the ancestors of the Basque people were the first European farmers, bringing agriculture from Asia Minor. The first wave went along the north Mediterranean coast and I would seek its traces in Greece and Italy, plus adjacent islands. The northernmost part of this wave was perhaps the Alpine region, where the tribal languages Rhaetic and Camunic were located, probably related with Etruscan. Till the present time there are traces of Basque-like toponyms and dialect words in Sardinia (V. Blažek, p.c. 09/12–13/2015; also quoted in BCR 453–54, footnote 4).

³ Cf. the disclaimer stated by Witzel in his study of a Central Asian substrate: “Naturally, not all words given below will turn out be substrate words; any initial listing like the present one will be fraught with overcounting in favor of non-IE origins, and also with unintended errors.” (Witzel 2015: 149).

⁴ It may be more accurate (phonetically) to transcribe these Archi words as *ikʰ*, *kʰorom*, since the lateral affricates in that language have velarized onsets, i.e. more like [kʰ], [kʰʷ], [gʰ], and in some East Caucasian subgroups of eastern Dagestan (Lak, Dargwa, Khinalug, and most of Lezgian) the lateral affricates have become pure velars, [k], [kʰ], [g], etc.; e.g. Lezgi *rugun* ‘threshing board’ < **riλ:oma* (NCED 52–55).

‘to thresh’ (NCED 1031). | **Basque** **larain* ‘threshing floor’: (B, G, AN, L, Bzt, Z) *lar-rain* ‘threshing floor’, (AN) *larrein*, (G) *larran*, (B) *larren*, *larrin*, (A) *larrin*, (R) *larren*, (with expressive palatal) *llarren*, *llarne* id. (FHV 165, 195; A&T XIX 315, 316; EDB 262; OEH; BCR Q.18). § East Caucasian has numerous derivatives, only some of which are cited here. Archi *łorom* = *tłorom* ‘threshing board’ (which resembles Basque **larain* ‘threshing floor’) is said to be a derivative by metathesis < **ł:iroma* < Proto-Lezgian **mił:o-ra* (see NCED 1031–33). The PEC structure *=*VrLV* is the result of a common transposition < Proto-Euskaro-Caucasian **rVĒV* ~ **LVrV*. From a Sino-Caucasian perspective cf. Burushaski **daltá-n-* ‘to thresh’ < **rVĒV-n-* (SCG 182).

ἄνθρωπος ‘man’ (Iliad); Mycenaean *a-to-ro-qa* /ant^hrōk^wos/. “As no IE explanation has been found, the word is probably of substrate origin” (Beekes 106).⁵ | **Basque**: **andere* ‘lady; young lady; woman; wife’, (AN, G, BN, Z) ‘doll’, (Z) ‘queen bee; concubine’; var. (Z) *andére* (modern *āndé* ‘dame, demoiselle’), (G, AN, L, BN, Z) *andre*, (AN-Larraun) *anre* ‘lady, young lady’, (A, B) *andra*, (B) *anra*, (B-arc) *andera* ‘lady; woman; doll’, (B) *andrako*, *andreko* ‘little woman; doll’; Aquitanian ANDERE (female name), ANDERE- (element in female names); ANDERE, ANDERENI, ANDEREXSO (male names); ANDOS-, ANDOSS- (element in personal names: ‘lord’?) (A&T III 865–67; OEH ANDERE, ANDRAKO; EDB 93). § Hugo Schuchardt, as reported by A&T, believed the original meaning was ‘young woman’ (‘mujer joven’). Etymologists frequently mention Celtic parallels, such as Middle Irish *aínder*, *aíndir* ‘young woman’, Scottish Gaelic *aínnir* ‘virgin’, Welsh *anner* ‘heifer’, *enderig* ‘bull, ox’ < Old Welsh *enderic* ‘steer’, Breton *ounner*, *onner* ‘heifer’, etc. Michelena, agreeing with Tovar, remarks that the Celtic word (reconstructed as **andera*) is not Indo-European but taken from the Iberian languages, that is, that Basque *andere* does not represent a Celtic loan, but rather the opposite (OEH).⁶ Vennemann (1998) has compared Greek ἄνθρωπος with Basque *andere*, also bringing in other Greek words and names with the components *ανδρα*, *ανδρο-*, and *ανθρ-* (see further below about **σαλαμάνδρα**). Vennemann also cites possible substratal relics in Romance and Germanic: Old French *andre* ‘woman’, French argot *andrimelle* ‘woman or girl’, Occitan *andra*, *landra* ‘woman, prostitute’, Bolognese *landra* ‘slut’, etc. (some forms appear to come from *la andra*, incorporating the article).⁷ As to a possible alternation between Greek *ανδρ-* and *ανθρ-*, Beekes (p. xxiii) mentions that Furnée “found that the stops show variation between voiced, voiceless and aspirated, so that there presumably was no phonemic distinction between voice and aspiration in the [Pre-Greek] language.” Western Basque *andrako*, *andreko* ‘little woman; doll’, with the diminutive suffix *-ko*, is a rather close formal match to Mycenaean *a-to-ro-qa* /ant^hrōk^wos/ (see below about suffixes).

⁵ But cf. Garnier (2008): “... an inherited PIE compound **ǵd^h-r-e-h₃k^w-ó-* («directed downward», whence «earthling, earthman, earthwoman».)”

⁶ “pero seguramente tiene más razón Pokorny al suponer que esta palabra en céltico no es indoeuropea sino tomada de las lenguas ibéricas, es decir, que el vascuence aquí no representa un préstamo céltico, sino más bien lo contrario” (quoting Tovar).

⁷ Vennemann cites loans from Romance to Germanic in which the semantic link becomes ever more attenuated, e.g. Middle High German *landern*, *lendern* ‘to walk about idle’, etc.; see Modern German *schlendern* ‘to stroll, wander, amble, saunter, meander’ – seems to be related (in Pfeifer 1997: 1211–1212 a not very convincing solution). *Sch-* in German often marks a negative connotation. *Landern*, *lendern*, etc., may be related to German *Land* ‘country, countryside, land, ground’, which only has cognates in Germanic, Celtic, Slavic, and has been connected with a hypothetical IE **lendh-* ‘free land, heather, steppe’, Pfeifer 1997: 762–763.

δοκός ‘bearing beam’; **δόκανα** ‘name of two upright beams constructed with a cross-beam’ “Benveniste [1929] thinks that *δοκός* and *δόκανα* are PreGreek” (Beekes 345). | **Basque** **tako*, **tak-et*: (B, G, AN) *taket* ‘stake, post’, (B, AN) *taketa* ‘stake, stick, rod’, (B) *tako* ‘circular piece of wood’, (B-Markina) ‘piece’, (c) ‘wedge, block, chock, stopper’ (OEH; BCR Q.55). | **North Caucasian**: Adyge *tāq:a* ‘stump, block’, Kabardian *dāq:a* id.; Chechen *duq’ū* ‘log, beam’; Dargwa *duk’i* ‘log, beam’; Tabasaran *duq’an* ‘pole, small beam’, etc. < PNC **dwij(w)ṽ* ‘log, stump’ (NCED 408). § Nikolaev 67, no. 10. Cf. also Bulgarian *tok* ‘a long board that is dragged on the ground to smoothen it; a four-meter long board on which a man steps and it is dragged by oxen on the plowed fields to crush the ground, thick board’. BER 8, 99 links it to *tok¹*, which needs to be separated from Slavic *takv* ‘current, flow, stream’ (acc. to Bernard 1982: 276); this would be another Balkan manifestation of the Euskaro-Caucasian substratum (see also **καλιᾶ**, **καλύβη** ~ Bulgarian *koliba*, below). From a Sino-Caucasian perspective, cf. Burushaski **dáko* ‘stick, post’ (SCG 44). The Basque word is probably the source of Spanish *taco*.⁸

ζέφυρος ‘west wind’; personified in the Iliad; Mycenaean *ze-pu2-ro*; Beekes wavers between a derivation from IE **h₃iebh-* ‘futuere’ and “... Pre-Greek, with PG **a* turning up as *ε* after the palatal **d^v*?” (Beekes 499). | **North Caucasian**: Andi *sibiru* ‘autumn’, Akhwakh *c:ōro* ‘autumn’ / *c:ibero* ‘winter’,⁹ Tindi *c:ibar* ‘winter’, Karata *c:ibero* id.; Tsezi *sebi* ‘autumn’, Hunzib *sibər* id.; Lezgi *zul* ‘autumn’, Tabasaran *čul*, Tsakhur *cuwul* / *ciwił* id., Udi *ž’obul* ‘spring (season)’;¹⁰ Khinalug *cuwa-ž* ‘autumn’; Batsbi *st’abo* ‘autumn’, (with metathesis) Chechen *bǵästē* ‘spring’; Abkhaz *á-ʒən* ‘winter’, Abaza *ʒnə* ‘autumn’ (< **ʒə-nə*) < PNC **čōjwīlhV* ‘autumn, winter (rainy season)’ (NCED 327). § Nikolaev (68, no. 13) compared Greek and NC, as the former a loan from the latter. It is hard not to think of other words like Russian север /*séver*/ ‘north’, etc., and in fact Nikolaev mentions PIE **kēi₁uero-* (his transcription) ‘winter, north’ as a loan from PNC to PIE. A similar view was taken by S.A. Starostin (1988, no. 5.10), citing Latin *caurus* ‘north wind’;¹¹ Lithuanian *šiáurė* ‘north’, *šiaurỹs* ‘north wind’; Slavic **sěverb* ‘north’; Old High German *skūr* ‘Ungewitter’ [English *shower*, etc.],¹² but not Greek ζέφυρος. Derksen (2008: 448–449) links Slavic *sěverb* ‘North’ to an IE **kēh₁uer-o-*, and to Latin *caurus* ‘north-western wind’ (< **kēh₁uer-o-*). Discussion in Bezlaj (III, 231); Snoj (2003: 652) sees an unexpected root, linked to PIE *(*s*)*k’éH₂ero-*. Martirosyan (2021) adds PIE **kēh₁uer-* > Arm. *sir* ‘cold wind’ (with an unclear etymology) and links it to the Slavic and Latin word (see above). Derivation from IE **h₃iebh-* ‘futuere’ (cited by Beekes) seems semantically unconvincing. Since all the words cited here are European, they could alternatively be interpreted as independent substratal loans from various Euskaro-Caucasian dialects.

⁸ ‘Thick and short piece of wood or other material, and generally cylindrical or rectangular, for various uses; plug (small, short and elongated piece, usually made of plastic, wood or metal, which is inserted into a hole); cue (for billiards),’ etc.

⁹ The “relation [of Akhwakh *c:ōro* ‘autumn’] to *c:ibero* ‘winter’ is not quite clear: perhaps old interdialectal loans, which gave rise to an etymological doublet?” (NCED).

¹⁰ Transcribed as *ž’obul* in NCED (the *paločka*, or ‘baton’ /*l*/ is not a vowel but is a conventional symbol among Russian Caucasologists that denotes pharyngealization of the preceding consonant or vowel); Klimov & Xalilov (2003: 280) transcribe the Udi word as *žž’obul* ‘вечна’.

¹¹ De Vaan (2008: 100) regards Latin *caurus* and the Balto-Slavic words as IE cognates (< **kēh₁u-er-o-*), but excludes the Germanic words (and Armenian *c’owrt* ‘cold’).

¹² Kroonen (2013: 451), however, derives the Germanic words from PIE **skeh₁-* ‘shade, shadow’ (Greek *σκιά*, etc.) and separates them from the Latin, Baltic and Slavic words.

ἰκτίν ~ ἰκτινός ‘kite’ (bird of prey) (Beekes 585–86) | **Basque** **saie* ‘vulture, eagle’: (B, G, BN, Z) *sai* ‘vulture’, (B-Gernika) *zai* id.; (B-Orozko) /*śái*/ ‘eagle’ (EHHA, map 130); also ‘eagle’ per Voltaire (northern Basque, ca. 1620) (OEH); *saie* is used to denote ‘ostrich’ in Uriarte’s Bible, Lev. 11:16 (1859, Gipuzkoan dialect) (BCR B.13). | **North Caucasian**: Tsezi *ce(y)* ‘eagle, vulture’, Hinukh *coy* ‘eagle’, Bezhta *cuha* id.; Chamali *s’ūy* ‘eagle’, Tindi *c:ū*, Karata *c’:ūyi*, Avar *c’:um* ~ *c’:un* id.; Andi *c’:un* ‘eagle, vulture’, Botlikh *c’:ūzi*, Godoberi *c’:ūzi* id.; Khinalug *c’im-ir* ‘small bird, sparrow’ < PEC **çwäm?V* ‘eagle’ (NCED 370). § Beekes also cites Armenian *c’in* ‘kite’; Old Indic *śyená-* ‘eagle, falcon’, Avestan *saēna-* ‘name of a big bird of prey’ are “rather deviant” (unclear what this means), and regards the Greek word as of IE origin (< **t̥kiH-in-*). Witzel (2015: 167, no. 293) however cites **caina* (> *saēna-*, *śyená-*) among examples of a Central Asian substrate in Old Iranian. Nikolaev (68, no. 14) cited Greek ἰκτίν as a loan from PEC, along with Hittite *ḫaštapi-* ‘oracular bird’ (p. 61, no. 6); Armenian *c’in* ‘kite’ is regarded as an independent loan from North Caucasian. Bouda (1948) compared Basque *sai*, *sae* and Avar *c’:um*. On the phonetics between Basque **saie* and PEC **çwäm?V*, the loss of a nasal before a laryngeal is recurrent in Basque (and convergently in some NC languages), e.g. Basque *(*e=*)*kē* ‘smoke’ = PNC **ḫwīnḥV* ‘smoke’ (Avar *k’:uy*, Bagwali *k’:ūy*, etc.: NCED 738; BCR F.2); Basque **sihi* ‘wedge, skewer, spit’ = PNC **çānHV* ‘arrow, bow’ (Bagwali *c’i* ‘arrow’: NCED 358; BCR Q.42).¹³ The phonetic link between Greek ἰκτίν and PEC **çwäm?V* (for Greek -*v* cf. Andi *c’:un*, etc.) is not as clear; the initial *i-* could correspond to Basque fossilized class prefixes, as in Basque **e=tanha* / **e=rhala* ‘swallow, swift’ (bird) (BCR B.21) or **i=tain* ‘tick’ (BCR B.41); see below under **Morphology: Fossilized class (gender) markers**. The Greek cluster -*κτ-* may be a rare example of Pre-Greek **-kt-* = the PNC/PEC tense affricate **ç*, though more examples would strengthen the case. (Cf., perhaps, Greek ἰκτίς, -ιδος ‘marten’ [Beekes 2010: 586 “no etymology”] if it is related to PNC **çĒrcV* ‘marten, weasel’ > Adyghe *cəza* ‘marten’, etc. [NCED 360]).

καλιᾶ ‘wooden dwelling, hut, barn, granary, bird’s nest’. “Etymological connection with ... καλύπτω [‘to cover’], etc. is extremely doubtful” (Beekes 624); (probable derivatives) καλύβη ‘hut, cabin’; ‘bridal bower’; ‘sleeping tent’; κόλυβος ‘farmstead’ (Hesychius); variant κολυβός: “The variant κολυβός ... shows that the word is Pre-Greek” (Beekes 628). | **Basque** **o=ketu*: (B, G) *okellu* ‘stable, corral’, (B) *ukullu* id., (B) *okolo*, *okolu*, *oko(i)llu* ‘corner’, (G) *okolu* ‘yard’, (BN-Amikuse, Z) *okholü* id., (G) *okullu*, *okollu*, *ikullu* ‘hall’ (FHV 83; EDB 307; OEH; BCR: Q.5). The oldest attestation is {*oquelua*} = /*okelua*/ ‘rincón [corner, nook]’ (with definite article -*a*) in Landucci’s (1958) dictionary. | **North Caucasian**: Dargwic (Akusha, Urakhi) *qali* ‘house, room’, Akusha *qal-č* ‘roof’; Tabasaran, Agul, Rutul *χal* ‘house’, Archi *χal* ‘nest’, *χali* ‘family’ (< *‘household’), Kryz, Budukh *χal* ‘roof’; Avar *hor* ‘mow, hayloft, shed’, etc. < PEC **qālV* ‘house, hut’ (NCED 889).¹⁴ § Nikolaev (69, no. 16) proposed Greek καλιᾶ as a loan from PEC **qālV*.

¹³ Trombetti (1925: 142, no. 289) cites Basque *sahi* ‘avvoltoio’, with internal -*h-*. We have not been able to confirm this form in any other source.

¹⁴ This EC word is not to be confused with another that is quite similar, phonetically and semantically: Lak, Dargwa *qala*, Avar *q:ala*, Lezgi *qele* ‘fort, citadel, fortress, tower’, etc., from Turkic: cf. Azeri *gala* ‘fortress, lock’, Kumyk *qala* id., Old Turkish *qala* ‘fortified part of town’ (Džidalaev 1990: 94). Klimov & Xalilov (2003) clearly show the difference, with two separate lemmata, between: комната [‘room, chamber’] (p. 114): Dargwa *qali*, Tabasaran, Agul, Rutul *χal* ‘room’, also ‘house’ in all languages cited (no note about any borrowing) and крепость [‘fort, castle’] (p. 116): Avar, Andi, Karata (and several other NC languages) *qala*; (with glottals) Lezgi *q’ele*, Tsakhur (and 4 other Lezgian langs. + Khinalug) *q’ala*, etc. Note the oppositions such as Dargwa *qala* ‘fort’ vs. *qali* ‘room, house’; Ta-

The proposed derivation of Basque *okelu* from Latin *locellum* is highly questionable semantically; the specialized meanings of the Romance words derived from *locellum* ‘Sarg, Grab’ are quite distant from the Basque meanings (‘stable, corral, hall, yard’) and are instead associated with burial (e.g., Spanish *lucillo* ‘burial urn’: see Dicc; REW 5095); none of the Basque glosses have anything to do with burial. Basque **o=ketu* matches PEC **qǎlǎ* very well, phonetically and semantically: Basque **k* = PNC **q* and Basque **e* = PNC **ə* are regular.¹⁵ Basque **o=* is the fossilized class prefix (with an allomorph **u=*) seen also in, e.g., Basque **o=hol* ‘board, plank’ (Q.62) ~ Rutul *χil* ‘wooden trough’, etc. < PEC **χulV* (NCED 1078), Basque **u=pe* ‘barrel, cask’ (Q.29) ~ Tsezi *pelu* ‘pipe, reed pipe’, etc. < PEC **HpēŧV̄* ‘pipe; vein’ (NCED 601); and others (BCR 67–68). See also Bulgarian *koliba* ‘hut, cabin, shack’, etc., which is considered a very early loan from Greek, with many cognates in Balkan languages and perhaps beyond (BER 2, 555–556). The word is considered as stemming from an autochthonous Balkan population: see BER 2, 556, Skok 2, 124. (See also *δοκός* ~ Bulg. *tok*, above).

κόμη ‘hair’ of the head, also of the mane of a horse (Iliad), metaphoric: ‘foliage’, also of growth in general ... ‘tail of a comet’ ... “ETYM Not explained with certainty” (Beekes 743–44). | **North Caucasian:** Andi *q’:*āw ‘hairdo’, Avar, Tindi *q’:*ama ‘cock’s comb’, (with suffix) Bagwalal *q’:*am-ča ‘mane’; Dargwa *q’:*ama ‘hairdo; fringe, forelock’; Archi *q’:*am ‘forelock, mane’; Abkhaz a-χ^wá ‘hair’, Abaza *qwa* ‘hair, feather, wool’, etc. < PNC **ǧ(w)ǎmǧǎ* ‘plait, mane; hair’ (NCED 931). | ? **Basque** **kima* ‘mane (of horse); bristles (of swine)’: (G, AN) *kima*, (L, BN) *khima*, (AN, B) *kime*, (BN) *khinba*, (AN, BN, Z) *k(h)uma*, (Z) *gima*, etc. (OEH KIMA; FHV 296; A&T XVIII 1001; EDB 251). § Nikolaev (69–70, no. 23) compared Greek and NC. The Basque forms are rather difficult and questionable: (a) differences of the first vowel (PNC /ǎ/ : Basque /i/; /u/ in some Basque forms may be due to secondary assimilation before /m/); (b) the possibility of borrowing or influence from older Spanish *coma* ‘mane’ (now obsolete in favor of *crin*) < Latin *coma* < Greek; (c) the similar word (G) *zima*, (B, G) *txima* /čima/, (G) *txuma*, *txume*, etc. ‘greña / hair of a person or animal that is long and badly combed, tousled, or tangled’ (OEH TXIMA), which is often discussed in connection with **kima* and may be cross-contaminated with it (FHV 296). But *zima* and *txima*, at least, cannot be derived from Latin *coma*.

μάδρουα ~ **άμάδρουα** ~ **βάδρουα** ~ **ǎδρουα** (< *Ἰάδρουα /wádrua/) ‘plums, sloes’: Beekes (890) explains: “άμάδρουα did not originally mean ‘belonging to a tree’, as tree names in [άμα-] meant ‘blossoming at the same time as’. Rather, initial /h-/ was added by folk etymology to *ǎ-μάδρουα, a form with (non-IE) prothetic vowel. This form (ǎ)μάδρουα must be a Pre-Greek etymon. If βάδρουα is reliable, we also have variation μ/β, to which Ἰ may be added in order to explain ... ǎδρουα” (Beekes 22–23, 191, 890). | **Basque** **ma=dari* / **u=dari*: (A, AN, B, L, BN, Z) *madari* ‘pear’, (G, AN, L, BN) *udare*, (L, BN) *udari* ‘pear’; in some dialects ‘fruit’ (in general); other variants: *udara*, *udere*, *urdare*, *urdere* (OEH); in place names / family names *Madariaga*, *Maltzaga* ‘(place of) wild pear trees’ (FHV 528; A&T XX 651; EDB 354; OEH). § K. Bouda and J. Hubschmid men-

basaran, Agul *q:ala*, Rutul *q’ala* ‘fort’ vs. Tabasaran, Agul, Rutul *χal* ‘room, house’, etc., showing clearly that the etyma are distinct.

¹⁵ The correlations between Basque **l*, **l̥* and PNC **l*, **l̥* (where **l̥* in each family seems to have been a ‘dark’ or velar lateral) are not yet entirely clear (BCR 189–193). Possibly a secondary contrast **l* / **l̥* developed independently in each family.

tioned connections of Basque *madari* with Greek μάδρα, άμάδρα, and Latin *malva* ‘mallow’ (!) (as reported by A&T XX 651). The semantic difference ‘plum’ ~ ‘pear’ is unremarkable; cf. the North Caucasian etymology including Chamali *k’uk’ul* ‘apricot’, Andi *k’urk’ul* ‘plum, damson’, Lak *k:urk:ul* ‘a sort of pear’, etc. (NCED 728).

μάλκη ‘numbness from cold’ in hands and feet, plur. ‘chilblain’; **μαλκίω** ~ **μαλακίω** ‘to become numb with cold, freeze’; “A convincing explanation is still lacking ... The variant spelling *μαλακίω* may indicate that the word is Pre-Greek” (Beekes 898–99). | **Basque** **mal-gor* ‘numb (from cold)’: (BN) *malgor*, (Z) *mālgor* ‘entumecido’ / ‘engourdi par le froid’; (Sal) *malgor-tu* ‘to get moldy, go numb, dry up (a tree) completely’, *malgor* ‘hollow tree’ (A&T XX: 662; OEH); if the Basque word is a compound **mal-gor* in which **mal-* meant ‘cold’ (= PEC **mhēl̥χe* ‘cold’: see below) + **gor* ~ **gogor* ‘hard, cruel; deaf’ (cognate with PEC **gwērV* ‘stone’: NCED 467; BCR R.28).¹⁶ | **North Caucasian:** Tabasaran *merc’-uli* ‘cold’ (adj.), Lezgi *meq’i*, Rutul *miq’di*, Tsakhur *miq’ana* id.; Chechen *mil-la* ‘from cold, with cold’, Batsbi *mil-dar* ‘to get cold’; Avar *mart* ‘hoarfrost’, etc. < PEC **mhēl̥χe* ‘cold’ (NCED 808). § The Basque development **mal-gor* is parallel to Basque (BN, L, Z) *molkho* ‘cluster’ < **mardo* ‘cluster’ + **-ko* [diminutive/expressive suffix], in which **mardo* = PEC **mār[χ]o* ‘handful, armful’ (NCED 798; BCR L.9), i.e., the original resonant+lateral affricate cluster like **-l̥χ-* or **-r̥χ-* resolves as Basque /l/ when stem-final before a suffix or compounded morpheme. As to the loan correspondence of Greek /k/ in μάλκη to PEC **χ* in **mhēl̥χe*, it is parallel to the loan correspondence postulated by Starostin (1988, nos. 1.6, 1.7, 2.2), e.g. PIE **peku-* ‘livestock’ < PEC **bhāχwi* ‘small cattle’ (NCED 293; BCR N.20).

μαστός ~ (Doric) **μασδός** ~ **μασθός** ~ (Ionic, Epic) **μαζός** ‘teat, breast, woman’s breast; (metaphorically) hill, knoll’. “If the form is Pre-Greek, *μαζός* [mazdos] and *μαστός* differ in voice only (and aspiration in Hell. *μασθός*). Since voice and aspiration are not distinctive in PreGreek, all forms may go back to the same Pre-Greek word” (Beekes 912). | **Basque** **mośu*, **mus-ko*: (G) *musu* ‘nose, snout, face, lip, kiss, point, tip’, *musu-zulo* ‘nostril’, (B) *mosu* ‘kiss (on face); lip’; (with suffix) (G) *musu-ko* ‘muzzle’; ‘face, facial, pertaining to the lower half of the face’; (BN) *mos-ko* ‘beak’, (Z) *mūs-ko* id., (Z-Eskiula) *müskúa* ‘(the) nipple’, (Z-arc) *mus-ko* ‘sting’, (B-Oñate) *mus-ki* ‘snot, mucus’, (B, G) *mus-kil* id., (Sal) *titi-mus-ko* ‘nipple’, (AN-Jaurrieta) /*titamúšku*/ id. (AT XXI 936; EHHA; BCR A.17). | **North Caucasian:** Chechen, Ingush *muc’ar* ‘snout, muzzle, trunk’; Avar *móc’u* ‘teat, nipple; tip’; Akhwakh *mic’o* ‘teat, nipple’, Chamali *mis’*, Tindi, Godoberi *mic:i* id.; Inkhokwari *mucu* ‘rib’; Lak *mazu* ‘nipple (of animal)’; Dargwa: Chiragh *muc:e* ‘sting’; Lezgi *murz* ‘blade; edge, verge; narrow side of an object’, Tabasaran *murz* ‘edge, verge’ < PEC **mh̥r̥cū* ‘point, edge, protruding part’ (NCED 811). § Nikolaev (70, no. 29) proposed Greek *μαστός* as a loan from North Caucasian. For the correspondence of Greek *στ* = PNC **ç* /*c’*/ = Basque **ś* see also, below, Greek *σταφ-* = PEC **çib-* = Basque **śap-*. On the semantic side the Greek sense of ‘teat, nipple’ is matched in Avar and Andian idioms, and some local Basque dialects (Eskiula *müsko* [with a common diminutive suffix *-ko*]; in Salazar, Jaurrieta, compounded with *titi* or *tita*, a widespread so-called ‘nursery word’). The original meaning may have been ‘point, tip, edge’ (attested in NC and Basque), with multiple specializations (see below). As to a supposed Romance origin of Basque *musu*, typical is Trask’s (1997: 261,

¹⁶ For the semantic relationship of ‘deaf’ and ‘hard’ (~ PEC ‘stone’), cf. English *hard of hearing*, *stone deaf*; Spanish *duro de oído*, *sordo de piedra*, etc.

284) statement (based on those of earlier scholars) that “late Latin *MŪSU* ‘muzzle’ and its diminutives are widely represented in western Romance ... and it is difficult or impossible to trace the histories of the Basque words [*musu*, etc.] with confidence” (AT XXI 947). In fact “Proto-Romance” **mūsus* ‘snout’ (REW 5784) is only hypothetical; Meyer-Lübke considers the word, with wide attestation in the whole Mediterranean area, as “stemming from northern France” and being an “Urschöpfung,” which is not an etymological explanation; and Trask’s reference to “western Romance” is key: the lack of reflexes in Rumanian may indicate a Vasconic substratum word attested only in the West. But see also Bulgarian *mucūna* ‘snout’, thought to have been loaned, via Modern Greek *μουτσοῦνα* ‘mask, snout from an animal or human mouth, jaw, pig’s snout’ < Venetian *musona* id. (Leschber 2011: 78); further, Bulgarian *mucūna* > Aromanian *muṭună* ‘mask’ (BER 4, 359). Pellegrini (1999) considers the root **musu*, on which the Italian word *muso* ‘snout’ is based, to be an extremely old root, citing various Italian forms, always with the voiced *-s-*, and also makes references to non-European terms by pointing out that words that come from the root **musu-* (and **busu-*) were formed in many languages, not just Indo-European, meaning ‘mouth, lip, kiss, face’, etc. PEC **mhǎrcū* offers a potential cognate that is a phonetic match, and has reflexes with meanings precisely matching those of Basque, specifically:

Basque (G) *musu* ‘snout, nose’, etc., *musu-ko* ‘muzzle’ ~ Chechen, Ingush *muc’-ar* ‘snout, muzzle, trunk’

Basque (Sal) *titi-mus-ko* ‘nipple’, (Z) *müs-ko* id. ~ Avar *móc’:u* ‘teat, nipple’, etc.; Lak *mazu* ‘nipple (of animal)’; Pre-Greek *μαστός* ~ *μασδός* ~ *μασθός* ~ *μαζός* ‘teat, breast’

Basque (G) *musu* ‘point, tip’, etc. ~ Avar *móc’:u* ‘tip’, etc.; Tabasaran *murz* ‘edge, verge’

Basque (Z-arc) *mus-ko* ‘sting’ ~ Dargwa (Chiragh) *muc:e* ‘sting’

To sum up, cognation of Basque **mośu* with PEC **mhǎrcū* seems preferable to a derivation from a hypothetical Latin **mūsus*, which has no Indo-European antecedents.

μέσπιλον ‘medlar, medlar tree, *Mespilus germanica*’; also ‘hawthorn, *Crataegus (orientalis, oxyacantha)*’; “A foreign word of unknown origin. Probably Pre-Greek on account of the suffix *-ιλ-* ... Borrowed as Lat. *mespilum*” (Beekes 935–36). | **Basque** **mahać* ‘grape(s)’: (BN, L) *mahats* ‘grape(s)’, (Z) /*máhãć*/, (G-Bergara, Leintza) *magats*, (B, AN-Larraun) *maats*, (B-Ibarruri, Zeanuri) /*márac*/, (B, G, AN, Bzt, Sal, R) *mats*, (B-Aulestia) *matz* id., etc. (FHV 113; A&T XX 651; EDB 278; OEH; BCR P.17). | **North Caucasian**: Chechen *hamc* ‘medlar’, Ingush *hamis-k* id.; Avar *ŕeč* ‘apple’, Andi *inči*, Akhwakh, Karata *ŕeče* id.; Tsezi *heneš* ‘apple’; Lak *hiwč*; Dargwa *ŕinc* id.; Tabasaran *wič* ‘apple’, Archi *q̄nš* id.; Khinalug *mič* id.; Abkhaz *a-báć* ‘medlar’, Adyge *nāpca* id., etc. < PNC **ǵämčō* ‘apple; medlar’ (NCED 237). § The Basque-NC comparison would require metathesis such as [**maŕaćV*] > Basque **mahać*. Cf. the metathesis in Adyge *nāpca* ‘medlar’ < **banca* < **bVmc:ᵂV* (according to NCED); **banca* is remarkably similar to Michelena’s **banats* ‘grapes’ (FHV 113). If, as Beekes suggests, *-ιλ-* is the suffix of the Pre-Greek word, it leaves *μέσπ-* as the root, also requiring metathesis according to the PNC form; the /p/ is evocative of the /p/ in Adyge *nāpca* ‘medlar’, but these are at best just convergent developments. From a Sino-Caucasian perspective cf. Burushaski **[m]içil* ‘pomegranate’, with a suffix similar to the Pre-Greek suffix *-ιλ-* (SCG 267). Diakonoff & Starostin (1986: 24) suggest a Hurrian cognate, *χιν3-urə* ‘apple’ (cf. Dargwa *ŕinc*), borrowed in Armenian as *χιν3or*. The semantic change of Basque ‘grape’ ~ NC ‘medlar, apple’ ~ Burushaski ‘pomegranate’ should not be surprising: cf. Rumanian *poamă* ‘fruit, apple’, Moldovan *poamă* ‘grape’ ~ French *pomme* ‘apple, potato’, etc.

(Buck 5.71); and other ‘fruit’ etymologies (e.g. Greek μάδουα ‘plum, sloe’ ~ Basque **madari* ‘pear’, above). Nevertheless, this comparison remains difficult, if not implausible: the origin of the -π- in Greek is not well explained. Even if the Basque and NC terms are indeed related, the Greek form is far removed phonetically, and the etymology requires many assumptions.

μῖκρός ~ σμικρός ~ μικκός ~ μικός ‘small, short, little’: “The group of words has a familiar and colloquial aspect, as is shown by the variants μικός and geminated μικκός. The initial interchange in μικρός and (older) σμικρός is unexplained and (also) points to Pre-Greek origin” (Beekes 951–52). | **Basque** **miko*: (BN, L) *miko* ‘a little, a little bit, a pinch’, (AN-Irun, Bzt) *miki* id., (BN-Garazi, Sal) *mikitta* ‘a tiny bit’. This word is traditionally derived from Spanish *miga* ‘crumb’, etc. < Lat. *mīca*; and/or Greek μῖκρός, but these do not quite work phonetically (A&T XXI 926; OEH; REW 5559). | **North Caucasian**: Chamali *mik’u-b* ‘small’, *mač* ‘child’, Karata *mik’i-s*: ‘small’, *mak’e* ‘child’, Godoberi *mik’i-si* ‘small’, *mak’i* ‘child’; Dargwa Chiragh *nik’a-ze* ‘small’; Budukh *mik’e* ‘few; a little, small’, etc. < PEC **mikwV* ‘small, young one’ (NCED 821). § Note also Romanian *mic* ‘small’ (see *nimic* ‘nothing’ < Latin *nēmīca*: REW 5885), normally etymologically linked to a totally hypothetical Latin **miccus* or Latin *mīca* ‘Krümchen’ (REW 5559),¹⁷ which also mentions Basque *mika* and Romanian *mic* ‘klein’; the Romanian and South Italian forms could be based on Greek *mik(k)ós* – no further etymological explanation is given.

μύλλον [n.] ‘lip’ (Beekes 980). | **North Caucasian**: Dargwic **muḥuli* ‘mouth’ (Akusha *muḥli*, Chiragh *mūle*, Kadar, Mekeg, Urakhi, Kharbuk *muḥli*, Gapshima *mūḥli*, Kubachi *mūle*, Tsudakhar *muḥuli* ‘mouth’);¹⁸ (with metathesis) Avar *humér* ‘face’, Akhwakh *hama-ʒal* ‘face’ (< **hVmV-ǰili*) < PEC **mVhwVlī* / **hwVmVlī* (NCED 499).¹⁹ § “Frisk compares a Germanic group with a single consonant: OHG *mūla* [f.], MHG *mūl* [n.] ‘mouth, jaws’ ... It does not seem that μῦθος is connected, nor that the gemination is expressive. Perhaps an onomatopoeia” (Beekes 980). (Onomatopoeia - how?) Kroonen (2013: 374) notes that (possibly apart from μύλλον) the Germanic word is restricted to Germanic and could go back to quasi-PIE **muH-lo-*, if Bavarian *māuen* ‘to chew, rumigate’ is related.²⁰

μύσταξ ~ βύσταξ ‘upper lip, mustache’. “Both the variation μ-/β- and the variant μύττακες point to Pre-Greek origin” (Beekes 249, 986). | **Basque** **bisa-ř* ‘beard’: common Basque *bizar* (definite form *bizarra*), (Z) *bizar*, (AN) *bizer*, *pizer*, *pizar*, (B) *bisar*, *bixar* /bišar/, *bixer* id. (EHHA; BCR A.24). With the frequent fossilized plural ending *-ř (BCR 76–78), and analogous in form to Agul *muž-ur* ‘beard’ (see below). With expressive palatal: (Z) *bixar* /bišar/ ‘goatee’. | **North Caucasian**: Khwarshi *biša-n-de* ‘beard’, Hunzib *bilažba* id. (< **biža-l-ba*), Bezhta *biza-l-ba* ‘mustache’; Tindi *miža-tu* ‘beard’, Akhwakh *miže-tʰu*, Chamali *miza-tʰw*, Bagwali *miža-tʰw* id.; Tabasaran *muž-ri*, Agul *muž-ur*

¹⁷ De Vaan (2008: 378) prefers to separate *mīca* from (σ)μικρός, citing Nyman’s connection with “*micāre* ‘to quiver, dart, flash’, viz. as the ‘glittering’ particle.”

¹⁸ In NCED Dargwa is considered a single language with diverse dialects, but there seems to be a growing consensus that Dargwa is instead a small linguistic group, like Tsezian or Nakh; see **Dargwic** in *Glottolog*: <https://glottolog.org/resource/languoid/id/darg1242>. /u/ denotes a pharyngealized vowel; /h/ is a voiceless emphatic laryngeal [pharyngeal] fricative.

¹⁹ NCED cites **hwVmVlī* as the proto-form, but **mVhwVlī* is implicit in the Dargwic forms, with no indication in NCED which would be original.

²⁰ By “rumigate” *ruminate* must be intended.

‘beard’; Khinalug *mič:-äš* ‘beard’, etc. < PEC **bil̥ʒV* ‘beard’ (NCED 303).²¹ § Nikolaev (71, no. 31) proposed that Greek *μύσταξ* ~ *βύσταξ* was a loan from PEC **bil̥ʒV* (later reconstructed as **bil̥ʒV* in NCED). In some NC languages there are regular changes of the type **bil̥ʒV* > **bin̥ʒ-* > **mi̯ʒ-* > **mu̯ʒ-*, etc. (cf. the Greek variants *μύσταξ* ~ *βύσταξ*). This etymon, via Greek, is the source of European words such as English *mustache*. In Baztanese Basque there is a strange blend, *mustratx* ‘mustache’ (apparently Basque **mu̯ʒtu-r̥* ‘snout’ [BCR A.19] + French *moustache*).

νέκταρ ‘nectar, drink of the gods’; **νεκτάριον** a plant name = *έλένιον* [Dioscorides Medicus], also name of a medicine and several eyesalves [Galenus]; “In contrast with *ἀμβροσία*, which is of related meaning ... *νέκταρ* does not have an ascertained etymology. ... [Furnée 1972]: 320 compares **νικάριον**, an eye-salve. If this is correct, the word may be Pre-Greek. He also points to other Pre-Greek words in *-αρ* (op.cit. 134⁷⁵)” (Beekes 2010: 1004–05). | **Basque** **nega-r̥* / **niga-r̥* ‘tears, weeping’ ~ **nega-l* ‘herpes, scurf’: (B, G, AN, L, Sal) *negar* ‘tears, weeping’, (Sal, B-dial.) *near*, (BN, L, Bzt, Azk) *nigar*, (Bzt) *niger*, (Z) *nīgar*, (R) *nexar* [nešar] id.; (B-Ubidea) *negar* ‘rennet’; (AN-dial., L) *negar* ‘sap, resin (of plants)’²² (A&T XXI 958; OEH NEGAR; BCR A.78). Cf. also (BN, L, Bzt) *negal* ‘skin rash, scurf, herpes’, (AN, BN, L, Z) *negel*, (L) *nagel* id., with a different suffix, **-l*, common in Basque body-part words, and the sense ‘herpes, rash’ is similar to ‘pus’ in the Nakh languages. | **North Caucasian:** Dargwic (Akusha, Chiragh) *ner̥* ‘tear’, (Urakhi) *nir̥*, (Kaitag) *ner̥* ~ *mer̥*, (Tsudakhar) *ner̥*, (Kubachi) *mē̯* id.; Lezgi *nab̯*, Agul *neb̯*, Archi *nabq*, Udi *nek*, Lak *maq*; Avar *máʃu*, Akhwakh *maq’a*; Bezhta *maq’o*, Khwarshi *muq’u* id.; Chechen *not’q’a* ‘pus’, Ingush *nod*, Batsbi *not’q* ‘pus’, *nat’q’-ayr̥* ‘tears’, etc. < PEC [direct stem] **nē̯w̯q̣ũ* ‘tear; pus’ / [oblique stem] **nī̯w̯q̣ũ̯-* (NCED 848). § **Phonetics:** The languages compared here involve a segment *NEK-* or *NIK-* + a suffix *-(A)R*: (Pre-)Greek *νέκταρ*, *νικάριον*, Basque **nega-r̥* / **niga-r̥*, Pre-Proto-Dargwic **ner̥-r̥*. The internal /r/ in Dargwic **ner̥* is thought to come from a former plural suffix, thus **ner̥* < **ner̥-r̥*, parallel in formation to Basque **nega-r̥*. “The medial *-r-* in PD is obviously secondary, probably having penetrated there from an original plural form in **-r*, being later substituted in PD by the **-bi-plural*” (NCED). As to the puzzling *-κτ-* in (Pre-)Greek *νέκταρ* (lacking in the possible variant *νικάριον*) there could be a clue from the Proto-Nakh form **nat̥q̣u* ‘pus’, which NCED explains as “an original plural form (**nat̥q̣u* < **na(w)q̣-tu* < **nē̯w̯q̣-dV*),” if a similar formation could be projected back to Euskaro-Caucasian. The vowel alternation *NEK-* or *NIK-* also occurs in all three language areas studied: (Pre-)Greek *νέκταρ* / *νικάριον*, Basque **nega-r̥* / **niga-r̥*, and PEC **nē̯w̯q̣ũ* / **nī̯w̯q̣ũ̯-*. See below, under **Morphology: Ablaut** for a brief discussion of Euskaro-Caucasian ablaut. **Semantics:** The underlying concept is ‘secretion, exudation (of human and animal bodies, and of plants)’, a typologically common semantic realm:²³ in (Pre-)Greek, ‘nectar; medicine; eye-salve’; in Basque, ‘tear(s); rennet; sap, resin’; in

²¹ Due to multiple possibilities of vowel reconstruction based on the attested vowels, the NCED authors allow for the alternative first vowels **-ō-* or **-ä-* as possibilities (~ PEC **bōl̥ʒV*, **bäl̥ʒV*). External comparison with Basque **bisa-r̥* ‘beard’ supports the PEC form **bil̥ʒV*, with **-i-*.

²² OEH gives references to this meaning in dialectal records by Prince Louis Lucien Bonaparte (1813–1891), and the 19th-century unpublished dictionary by Maurice Harriet, who wrote “*Mahatsaren nigarra*, sève, larmes, pleurs de la vigne.” The 20th-century Basque writer Andima Ibiñagabeitia used the compound *arbola-negar* = ‘resina’.

²³ For semantic typology, cf. Lak *pic* ‘dew, sweat’; Avar *pic*: ‘resin’, Karata *bic’i*; Dargwa *penc* ‘resin’; Chechen *mutta* ‘juice, sap’ (Rus. сок); Ubykh *bzə* ‘water’, etc. < PNC **p̥in̥çw̯ǎ* ‘resin, juice’ (NCED 871); a putative Basque cognate is **pista* ‘fresh rheum; sleep sand (secretion from eyes)’ (BCR A.79). See also Basque **i=serdi* ‘sweat; sap (of trees)’ (BCR A.89), putatively cognate with PEC **cāl̥w̯V* ‘blood; life’ (NCED 376).

East Caucasian, ‘tear(s); pus’. The actual substance of νέκταρ is not discussed by Beekes, but Roscher (1883) deemed both ambrosia and nectar to be forms of honey. The modern English sense of ‘the saccharine secretion of a plant, which attracts the insects or birds that pollinate the flower’ is quite recent, ca. 1545–55 CE (Flexner 2001: 1284).

ξύλον ~ (Attic) σύλον ~ σύλιος ‘wood, timber, firewood, tree, beam, stick; wooden block put around the neck, gallows; bench, table’; also as a measure of length. “It seems to correspond with Lith. *šūlas* ‘post, pole, stave’ < IE **ksulo-*, Ru. *šúlo* [n.] ‘garden-pole’, SCr. *šūlj* [m.] ‘block’ < IE **kseulo-* (?). Germanic words like OHG *sūl* [f.] ‘style, pole’, Go. *sauls* ‘pillar’ have a similar appearance. The relation between the Slav., Balt., and Gm. words has been amply discussed, but hardly explained. Was the word taken from a non-IE substrate language?” (Beekes 1037–38). | **North Caucasian:** Lezgi, Tabasaran *č’ul* ‘ceiling beam’, Agul *č’il* ‘thin log’, Tsakhur *č’il* ‘planking poles’; Avar *č’álu* ‘log, beam’; Lak *č’ula* ‘beam, girder, log’, etc. < PEC **čhwītū* (~ **čhwītū*) (NCED 388).²⁴ § S.A. Starostin (1988, no. 4.11) proposed PIE **kseul-* ‘beam, post, piece of wood’ as a loan from PEC **čhwītū* ‘beam’ (reconstructed then as **čiwotu*, six years before NCED was published). Pfeifer (1997: 1179) regards German *Säule* ‘pillar, pile’ as of “Herkunft ungewiss,” while Kroonen (2013: 491) deems its ancestor, Proto-Germanic **sūli*, “an *i*-stem of unknown origin” with an ablaut variant **sauli-* > Gothic *sauls* ‘pillar’.

ὄγχνη ~ ὄχνη ‘pear tree, *Pirus communis*; pear’. “[Furnée 1972] thinks the word is Pre-Greek, also on account of the by-form ὄχνη” (Beekes 1045). | **Basque** **ok-*(*arhan*) ‘plum, sloe’ (BCR P.16): (B, G-Etxarri-Aranaz, AN-Arakil) *ok-aran* ‘plum’, (AN-Olza) *ok-arin*, (AN-Ilzarbe) *uk-arain* id., (B) *txarri-ok-aran* ‘sloe’ (*txar-* /*čai*/ ‘bad, wild’; cf. Tabasaran *č’uru* ‘bad; wild [of plants]’, etc.: BCR R.5; NCED 555). A compound with **ar=han* ‘plum’ (BCR P.15). **ok-arhan* may originally have designated the cultivar plum (cf. Karata *ače*, Lak *aq* ‘garden’, etc.) as opposed to wild plums and sloes (AT XXI 975; OEH). (B) *txarri-ok-aran* ‘sloe’ reflects the fact that the meaning of **ok-* was forgotten before the element /*čai*/ ‘bad, wild’ was added. | **North Caucasian:** Andi *oči* ‘sweet cherry’, Akhwakh *aqi* ‘grape’, Tindi *ači*, Chamali *ač* id., Karata *ače* ‘garden’; Khwarshi, Inkhokwari *oh* ‘grape’; Dargwa Chiragh *aq* ‘fruit(s)’, Akusha, Urakhi *anq* ‘garden’; Lak *aq* ‘garden’; etc. < PEC **?əqV* ‘grape; fruit; orchard, vineyard’ (NCED 206); “...excessive -n- in [Proto-Dargwic **?anq*] (all other languages reveal absolutely no trace of any medial resonant); it may have penetrated from an oblique base like **?aq-nV-* (or, more probably be a result of contamination with another root: PEC **HēnqwV* ‘meadow, plot’ q.v.)” (NCED). § Nikolaev (71, no. 32) proposed the borrowing of Greek ὄγχνη / ὄχνη from PEC **?ē(N)qV*. If, as NCED suggests, there was a PEC “oblique base like **?aq-nV-*,” it could explain the Greek -v- in ὄγχνη. Compare also Latin *acinus* ‘grape or other berry’, a close phonetic match to the hypothetical PEC **?aq-nV-*, just mentioned. Latin “*acinus* is generally regarded a loanword from an unknown Mediterranean language; since the seeds of grapes are rather bitter, I see no reason to reject a derivation from the

²⁴ There is another very similar NC root: cf. Avar *цIул c’ul* ‘wood, firewood’, Andi *c’ul* ‘stick’, Akhwakh *č’uli*, Karata *c’ule* id., Tindi *c’uli* ‘(shepherd’s) staff’, Chamali *c’uli* ‘whip’, etc.; Hunzib *c’ulu* ‘arrow’, Bezhta *c’ulu-c’a* id., Tsezi *c’eru-c’a* ‘bow’, etc. < PEC **čwīhV* (NCED 374). It is also tempting to think about Basque **sul* ‘wood, timber, lumber’: common Basque *zur*, in parts of Bizkaia and Navarre *zul*, Roncalese *zūr* (with a nasal vowel), etc. (BCR Q.51), compared in BCR with PEC **zwl[ē]h* ‘twig, rod, sheaf’ (Andi *žala* ‘branch, rod’, Avar *žul* ‘broom, besom’, Chamali *zala* ‘rod’, etc.; NCED 1103). However, for phonetic reasons, it seems best to keep these forms separate from Greek ξύλον, etc. PNC/PEC **ž-* and **ž-* correspond to Basque initial **s-* (BCR 151–52), while PNC **č-* and **č-* correspond to Basque initial **č-* (BCR 149–50).

root **ak-* ‘sharp’” (de Vaan 2008: 23). The vowels (*o* or *e*) are a little difficult. The change of PNC **e* > Andian **o* is regular, but this “**o* was preserved only in Andi, and merged with **a* in all other languages” (NCED 74, 108), thus Andi *oxi* but *a-* in the other Andian languages. As to Basque **o-*, the best match for PNC **?e-* = Basque **o-* seems to be PEC **?ěndū* ‘forehead’ (NCED 205), Andi *honno* ‘forehead’ = Basque **ondo₂* ‘side; bottom; proximity, closeness’, Bizkaian *ad-ondo* ‘forehead (of cattle)’ (BCR I.4).²⁵ But there are also alternative solutions of Pre-Greek ὄρχνη. Blažek (2014: 45) mentions a different North Caucasian word that is semantically exact with the Greek word: Avar *gēni* ‘pear’, Andi and Karata *hīhī*, etc. ‘pear’; (Tsezian): Bezhta and Gunzib *hī* ‘pear’; (**hī* ‘pear’ + **?ěš* ‘apple’ >) Tsezi *heneš* ‘apple’, Khwarshi *hīyoš* ‘apple’; (Proto-Nakh: **kam-maka*) > Chechen *kam-maka* ‘peach’, Ingush *kam-maka* ‘apricot’,²⁶ all reconstructed as Proto-East Caucasian **yōn?V* ‘pear’ (NCED 475). PEC **yōn?V* has, in turn, been compared with Basque **-han* in **ar-han* ‘plum’ (BCR P.15), as cited above. Blažek also cites some Semitic words meaning ‘fresh, unripe dates’: Akkadian *uḫinnum*, *uḫinnu(m)*, *uḫe(n)num* ‘fresh / unripened date(s)’ > Jewish Aramaic *ʔāhēnā* ‘nicht voll gereifte Datel’ (> Arabic *ʔahān* ‘bunch of green dates’), Syriac *hēnā* ‘an unripe fruit, especially fig’.

ὄρχις ‘spine, backbone, back; (mountain) ridge’. “However, since *φοᾶχ-/φοᾶχ-* cannot be derived from an IE form (the ablaut in the above reconstructions being impossible), it may instead be Pre-Greek” (Beekes 1277–78). | **Basque** **erēka* or **e=rēka* ‘gully, ravine’: (c) *erreka* ‘gully, ravine, riverbed, arroyo, creek, brook, stream’ (FHV 155; AT XI 571; EDB 177; BCR D.8); toponym *Erreka* (Bizkaia 1093 CE); sporadically written *herreka* or *errheka* (OEH). Romance forms like Gascon *rèc*, *arrèc* ‘brook, stream’ are probably from Vasconic; “Geographische Verbreitung und Bedeutung legen iberischen Ursprung nahe” (REW 7299). | **North Caucasian**: Tindi *rek:a* ‘gorge, ravine’, Karata *rik’:e* id., Godoberi *rek:i-n* ‘valley’; Bezhta *rūq’e-ro* ‘mountain slope’, Hinukh *ruqe-s* ‘plain’; Chechen *duq* ‘mountain ridge’; West Caucasian: Ubykh *q’wa* ‘cavern’, Adyge *q’wə-šha* ‘mountain’, Kabardian *q’wə-šha* ‘cavern’ < PNC **rīq’wā* ‘mountain, rock; cave’ (NCED 953). § Nikolaev (71, no. 34) proposed Greek ὄρχις was a loan from PNC **rVq’q’V* (later revised to **rīq’wā* in NCED). Note the vowels in Pre-Greek ὄρχις vs. PNC **rīq’wā* – metathesis of vowels? The semantic glosses are diverse, from ‘ridge’ (Pre-Greek and Nakh) to ‘slope, plain, valley’ (NC), ‘ravine, gorge, gully’ (Tindi, Karata, Basque), and ‘cavern’ (Ubykh).

σαλαμάνδρα [f.] ‘salamander, kind of newt’; “Given its non-Indo-European structure, σαλαμάνδρα may be Pre-Greek. Cf. also on σαύρα [‘lizard’], which is probably Pre-Greek, as well” (Beekes 1303); σαλαμίνθη [f.] ‘spider’ (Byzantine); “The suffix *-ivθη* is clearly Pre-Greek, but further connections are unknown” (Beekes 1303). σαύρα [f.] ‘lizard’ ... also σαῦρος [m.] ‘lizard’. “Without etymology, like many other words for ‘lizard’. ... As the animal was not a part of the PIE world, the word must be of local, i.e. of Pre-Greek origin” (Beekes 1313). **Basque**: **suge* (or **súhe?*) ‘snake’: (c) *suge* [súye], (G) *suga* [súya], *suba* [súβa], (AN, B-Lekeitio, Ubidea, BN-Aldude, G-Iziar) *sube* [súβe], AN (Zugarramurdi) /súyé/ ‘snake’, /súheá/ ‘the snake’, (Z) *süge* [súye] id. (EHHA,

²⁵ (B) *ad-* seems to be a reduced form of **a=dar* ‘horn’ (BCR A.4). The problem of disentangling Basque **ondo₁* ‘joint’ (A.77) from **ondo₂* ‘side, beside’ (I.4; and from **honda-r* ‘sand’, etc. [D.18], and from reflexes of Latin *fundum*) is discussed in BCR (240–41).

²⁶ The Proto-Nakh form is a compound of **van* + **paka*, the second part of which comes from PNC **pīrqwA*, a word which means ‘apricot’, ‘peach’, ‘plum’, ‘fruit’ (in general) in individual NC languages (NCED 873).

map 114); (in compounds): **suge-lind(il)a*: (L–18th c.) *sugalindila*,²⁷ (B) *sugelinda*, ‘lizard’, (G) *sugalinda*, (B) *sugalindara*, (B, G) *sugelindara*, (B) *sugelandara*, (L) *sugekandela*, (L-Ainhoa) *subekandela*, (L, R) *sugekandera*, etc. (see **lindila* ‘lizard’, BCR B.25); **suhalendil(a)*: northern Basque *suhendil* ‘lagartija / lézard des murailles’ (Pouvreau, 17th c.);²⁸ (L-Murgerre) /*suháindola*/, (BN-Armendaritze) /*suyándoila*/, /*súyeandóil*/, /*suyándola*/ ‘lizard’; **suhangil(a)*: (BN-Gamarte) /*súaŋgila*/, (BN-Ezterenzubi) /*suyáingil*/, (BN-Baigorri) /*suβeáingil*/ ‘lizard’, etc. (EHHA, map 119); also *Sugaar*: a mythical serpent in Basque folklore (FHV 59; EDB 342). | **North Caucasian**: Lezgi *šarat’ul* ‘lizard’, Kryz *šurut* ‘scorpion’; Ingush *šulq’a* ‘lizard’, Chechen *šat’q’am* ‘a kind of lizard (медяница)’ (< **šult-ikV*); Dargwic (Akusha) *šuršut’an* ‘lizard’, (Kharbuk) *š:ilt’a* id.; Avar (Antsukh dialect) *š:ut* ‘lizard’ < PEC **šVIVtV* ‘lizard’ (NCED 987). § Since it is well known that words for small creeping creatures (e.g., reptiles, amphibians, arthropods) are fraught with many kinds of expressive and irregular phonetic changes (Bengtson 2017a: 283) it is quite difficult to unravel the origins of the etyma involved; so this lemma can be regarded as more exploratory than definitive. A quick look at the EHHA maps 115 and 119 shows that words for ‘salamander’ and ‘lizard’ are extremely varied from one region or even community to another. Michelena proposed that some of the numerous variants of Basque ‘lizard’ stem from *suge* ‘snake’ + *andere* / *andera* ‘lady’ (see above under *ἄνθρωπος*) and there likely was influence of other words (*sagu* ‘mouse’, *lindo* ‘clean, without stain’, *kandela* ‘candle’, and *süsker* [a Zuberoan word for ‘lizard’]) that would explain the appearance of some variants.²⁹ In BCR (no. B.25) it is postulated instead that there was Basque **lindila* (an element in some ‘lizard’ words), cognate with PEC **lwitlwiV* ‘lizard’ (NCED 763, attested in only three Daghestanian languages, Chamali *toʔol*, Lezgi *ftfil*, Rutul *xutxul*); the vowels match very well, as do the initial laterals, but in inlaut the Basque cluster *-nd- is matched with the strange PEC cluster *-tlw-, which, as far as we know, does not occur in any other PEC or PNC reconstruction; as expected, there must have been some expressive sound changes on both sides. This **lindila* later contaminated with **andere* ‘lady’ and the Romance word *kandela/-ra* ‘candle’ (apparently from the slim shapes of lizards and candles). Or perhaps **andere* is also original, since it occurs in other Basque animal names.³⁰ Regarding *σαλαμάνδρα*, besides Basque **andere* possibly corresponding to the -άνδρα part, some Basque lizard names have components that resemble -μάνδρα: (Bzt-Aniz, Lekaroz) *subemandil*, (L-Azkaine) *sumandil*, (L-Senpere) *subemandil*, (R) *sugemandila* ‘lizard’ (OEH SUGANDILA; EHHA map 119); and possibly the *σαλα-* component is related to PEC **šVIVtV* ‘lizard’, if *-tV is a suffix.³¹

²⁷ The form *sugalindila* is documented by the 18th century Lapurdian writer Haraneder, who recorded several archaic forms (OEH SUGANDILA).

²⁸ Sylvain Pouvreau (d. 1675) was a priest of French descent who in the course of his studies and jobs learned Latin, Hebrew, Spanish, and Basque, the last of these well enough to write several translations of religious tracts as well as an unpublished Basque-French dictionary (ca. 1650~1660), parts of which are preserved in the Bibliothèque Nationale de Paris. This dictionary is respected and widely quoted by Vasconists (Trask 1997: 48, 50–51).

²⁹ “De **suge* + *andere*/-a. Es probable que haya habido influencia de otras palabras (*sagu*, *lindo*, *kandela*/-ra, *süsker*...) que explicaría el aspecto de algunas variantes” (OEH SUGANDILA).

³⁰ (BN, L) *satandere* ‘weasel, marten’ (carnivora: mustelidae) (< **sag-t-andere*, ostensibly ‘mouse-lady’, but originally this *-andere*, before phonetic distortion, may have been related to Proto-Andian **handa-*, as in Tindi *handa-reḷ:u* ‘weasel’, etc.; cf. Basque **ergu-nedi* ‘weasel’, also possibly containing a distorted morph *-*nedi* related to Tindi *handa-*, Tsezi *madu-* (in *madu-tʔi* ‘weasel’), etc. (BCR B.10; NCED 239).

³¹ Cf. Bezhta *dibi-t’o* ‘drum’, *q’asq’a-t’o* ‘throat’; Khinalug *k’unk’u-t’a* ‘weasel, marten’; Basque *neska-to* ‘little girl’, (Bzt) *eltxe-to* ‘small pot’, *elixa-to* ‘small chapel’, etc. (BCR 55).

Pre-Greek **σαλα-** ~ Lezgi *šara-t'u-l* ‘lizard’; Ingush *šul-q'a* id. < PEC *šVIV-*tV*
 Pre-Greek **-μάνδρα** ~ Basque *-mandil(a)* (in dialect words for ‘lizard’)

As to **σαλαμίνθη** ‘spider’, connections between names of reptiles, amphibians and arthropods is not uncommon: e.g., Basque (R) *arreuli* ‘salamander’, (Z) ‘scorpion’ (BCR B.26); Basque (G) *arrubi* ‘salamander, scorpion’ (BCR B.27), and Lezgi *šarat'ul* ‘lizard’, Kryz *šuruł* ‘scorpion’ (mentioned above). Another twist to these etymologies is the ancient belief that salamanders were associated with fire and immune to fire, so much so that several ancient dignitaries (the emperor of India, Pope Alexander III, Prester John) wore garments made of salamander skins, believing that they protected them from fire (Ashcroft 2000: 112–13). The species *Salamandra salamandra*, known as *Feuersalamander* in German (along with numerous dialectal designations) is widespread in Europe, south of the Baltic and North Seas and west of the Bosphorus. It may be notable that this association may have crept into some of the Basque designations of ‘lizard’, such as (L-Azkaine) *sumandil*, in which *su-* in the folk mind could be associated as much with Basque **šu* ‘fire’ (BCR F.1: cf. PNC oblique stem **čũy-* ‘fire’) as with **šuge* ‘snake’. In a Sino-Caucasian context Basque **šuge* ‘snake’ may be linked with Yeniseian **c[ī]k* ‘snake, fish’ (BCR Z.13). On the other hand, Pre-Greek **σαύρα** [f.], **σαῦρος** [m.] ‘lizard’, if not related to Lezgi *šara-t'u-l*, etc., could have a Semitic origin: cf. Akkadian *šurārû*, *šurāru(m)*, (ἄπαξ) *šurīrû* ‘lizard(s)’ (CDA 341) < Semitic **šauru* ‘lizard’.³²

σιρός ~ **σιρός** ~ **σειρός** ‘pit or vessel for keeping corn, silo’. “Technical word without etymology. The variation between *σιρ-*, *σιρ-*, *σειρ-* is hard to explain from an IE point of view” (Beekes 1335). / **σῦριγξ**, **σῦριγγος** ‘quill, flute, syrinx [shepherd’s pipe]’ (*Ill[iad]*); also of pipe-like objects, e.g. ‘windpipe, blood-vessel, fistula’ (medic., etc.), ‘spear case’ ... ‘hole in the nave of a wheel’ ..., ‘subterranean passage’ ... Arm[enian] *sring* ‘flute, pipe’ was probably taken from the same source. Borrowed as Skt. *suruṅgā* [f.] ‘subterranean passage’ (Beekes 1423–24). / **σωλήν** ‘pipe, channel’; ‘grooved tile’, etc. < hypothetical **σωλος* or **σωλον*; “Etymology unclear; ... [Furnée] suggests that the word is Pre-Greek (giving other such words in *-ην*)” (Beekes 1439).³³ | **Basque** **sulho* ‘hole, cave; (anatomical) tube’: (BN, L) *zulho*, *zilho* ‘hole, burrow’, (B, G, AN) *zulo*, (B-Orozko) *zulu*, (AN-Goizueta) *zolo*, (B-Aulestia) *sulo*, (B, Sal, L-Ainhoa) *zilo*, *zillo*, (Z) *zilo*, *xilo*, (R) *xillo* /šilo/ id., (AN, BN, Z) *zilo-ka* ‘cave’, (AN-Lezaka, Bzt) *ur-zilo* ‘cistern’ (‘water-hole’), etc.; generally, southwestern *zulo* / northeastern *zil(h)o*; (G) *zilo* “Silo, lugar subterráneo donde se guarda el trigo” [Larramendi, 18th c.]; (B-Vergara, Salinas) *silo* “Silo para conservar hierba fresca” (OEH SILO); in anatomical compounds: (G) *ipurt-zulo* ‘anus’, *eztar-zulo* ‘pharynx’, *musu-zulo* ‘nostril’, (Z) *südür-xilo* ‘nostril’, (B) *sama-zulo* ‘gullet’, etc. (FHV 77, 320; EDB 227, 342, 380; BCR I.12). | **North Caucasian**: Avar (Antsukh dialect) *šulu* ‘pipe’, Chamali na-*s:ul* ‘tubular bone’, Andi tom-š:*il*, Karata hani-š:*el* id., Tindi han-š:*al* ‘arm (from hand to elbow)’;³⁴ Tsezi *šilu* ‘horn’, Be-

³² Thanks to suggestions from V. Blažek (p.c. 11/04/2020). He is currently preparing a proposal that Greek *σαύρα* / *σαῦρος* were borrowed not directly from Akkadian but more likely from a Semitic language of an Amomite type, thanks to trade contacts between the Levant and Crete.

³³ The comparison with *σωλήν* is suggested by Giampaolo Tardivo (p.c. 11/22/2020). For *σιρός* Tardivo suggests a Semitic origin: Hebrew *sir* ‘pot, vessel’, Arabic *zir* ‘a large jar’. But “Hebrew *sir* cannot correspond to Arabic *zir* which does correspond to [Egyptian] (Pyramid texts) *zwr* ‘drinking vessel’ ... likely one more [Egyptian] loan in Arabic (I’ve just published a paper on these loans [Militarev 2020]) (A. Yu. Militarev, p.c. 12/03/2020).

³⁴ The four Andian compounds come from **honi-š:ˀVIV* ‘marrow-pipe’ (thus, ‘tubular bone’) or **tomV-š:ˀVIV* ? ‘sinew-tube’ (thus, ‘forearm’).

zhta *šelo*, Hunzib, Inkhokwari *šelu*, Khwarshi *šeru* id.; Lezgi *sulu-r* ‘throat’,³⁵ Kryz *sil* ‘top (of boot)’, (with metathesis) Rutul *lis* ‘gullet’, etc. < PEC **šwōtV* ‘hollow tube’ (NCED 978). § At least from the few examples here, nothing decisive can be said about the development of liquids. Pre-Greek also has a high-front vowel (*i* ~ *ī* ~ *ei*) while some NC languages (Tsezi *šilu* ‘horn’, etc.) and Basque dialects (Z *zilo*, *xilo*) have developed *i*-vowels, alongside back-rounded vowels. Semantically, the meanings denoting tubular body parts are attested in NC (Chamali na-*s:ul* ‘tubular bone’; Tsezi *šilu* ‘horn’, etc.); Basque (G *eztar-zulo* ‘pharynx’, musu-*zulo* ‘nostril’, abo-*zulo* ‘mouth(-hole)’, etc.); and in Pre-Greek (σῶριγξ ‘windpipe, blood-vessel, fistula’). Specializations as ‘subterranean passage / cave’ and ‘pit or vessel for keeping corn, silo’ are attested in Basque and Pre-Greek. σωλήν ‘pipe, channel’ is very close to the semantics of Avar (dial.) *šulu* ‘pipe’. By one route or another, this etymon is the likely ultimate source of English *silo*, and related European words. Skeat (1882: 562) derives it from Spanish *silo* < Latin *sīrum* < Greek σῖρός, and this is still a commonly cited source. Though REW (7955) derives Spanish *silo*, Provençal *sil* and Galician *siro* from Greek σῖρός ‘unterirdische Getreidekammer’, the Real Academia (Dicc) declares the Spanish word as “de origen incógnito.” The web resource *Online Etymology Dictionary* has, in our opinion, a more reasonable theory, that “the Spanish word is from a pre-Roman Iberian language word represented by Basque *zilo*, *zulo* ‘dugout, cave or shelter for keeping grain’.” The entry quotes Barnhart & Steinmetz (1988): “The change from *r* to *l* in Spanish is abnormal and Greek *siros* was a rare foreign term peculiar to regions of Asia Minor and not likely to emerge in Castilian Spain.”³⁶ For the German word *Silo* the origin is unclear, according to Pfeifer (1997: 1292).

σταφυλή ‘bunch of grapes’; ‘grape’ [Iliad]; (metaphorically) ‘swollen uvula, uvula inflammation’. “The similarity with ἀσταφίς ‘dried grapes’ is probably not accidental, but the exact relation of the words is unknown. The group of words is Pre-Greek ... ἀσταφίς ~ ὄσταφίς ~ σταφίς ... ‘dried grapes, raisins’ [is a] typical substrate word, with prothetic vowel and variation *α/ο*–” (Beekes 155, 1391–92). | **Basque** **šapa-i* ‘blackberry-bramble, thicket’: (BN) *sapar* ‘thicket, bramble’, (BN-Amikuse, L-Bardos) *saphar* ‘hedge, fence’, (BN) *saparr-ondo* ‘thicket, bramble’; with expressive palatal /č/: (R) *txapar* ‘kermes oak’ (*Quercus coccifera*), ‘scrub, brush, undergrowth’, (Sal) *txaparro* ‘scrub of evergreen oak or holm oak’ (FHV 54, 296; EDB 258; BCR C.19). | **North Caucasian**: Avar *c’ibí-l* ‘grape’, Avar (Chadacolob) *c’ibí-l* ‘grape’; Rutul *c’ib* ‘juniper’, Tsakhur *c’ib* ‘juniper’, Lezgi *c’p’-az* ‘blackberry’ < PEC **čibV* ‘a kind of berry’ [better: ‘berry, plant with berries’] (NCED 367; a sparsely attested [Avar, Lezgian] isogloss.) § The comparison by Nikolaev (72, no. 37) was actually with Proto-Nakh-Daghestanian *[*c’c’*]V*mbitV* ‘виноград’/ ‘grapes, grapevine’ > Avar *c’ibí-l* ‘grape’ and Proto-Lezgian **t’umbul*. By the time the NCED was published, nine years after the 1985 article, these words had been resorted into two different etymologies, the one cited above and PNC **čūmhV* ‘kernel, stone (of fruit, nut); marrow’ (NCED 1004). The types of berries denoted in the above etymologies are diverse (juniper, blackberry, grape), so the oldest meaning may have been ‘plant with (some kind of) berries’. Such plants tend to be low-lying bushes with a tangle of spiny branches (blackberry, juniper), thus the se-

³⁵ Lezgi *sulur* ‘горло / throat’ is not discussed in NCED, but is cited in Klimov & Xalilov (2003: 71–72). Lezgi *sulur* fits this etymology phonetically (consistent with Proto-Lezgian **sol-* ~ **s:ol-*) and semantically: “the meanings ‘gullet, throat’ and ‘top of boot’ are sometimes interchangeable (cf., e.g. Lezg. *q:uχ* meaning both)” (NCED 979).

³⁶ <https://www.etymonline.com/>

mantic connection with Basque ‘bramble, thicket’. Avar ‘grape’ seems to be a secondary semantic development < ‘berry’. For the correspondence of Greek $\sigma\tau$ = PNC $*\zeta /c'/$ = Basque $*\acute{s}$ see also II: $\mu\alpha\sigma\tau$ - = PEC $*mh\acute{o}r\zeta$ - = Basque $*mo\acute{s}$ -; Basque $*a$ = PNC $*i$ is uncommon, but also occurs, e.g. in Basque $*sathui$ ‘quick, nimble, flexible’, etc. = PNC $*\acute{s}iV / *s\acute{i}V$ ‘light (of weight)’ (BCR R.42). Basque $*\acute{c}apa-r$ is the source of Spanish *chaparro* ‘scrub oak’ (Dicc) > American English *chaparral* and *chaps* (leather leggings).³⁷ Other Romance developments include Aragonese *chaparro* ‘scrub pine’; cf. Latin *sapinus* ‘fir’, of unclear (Celtic?) origin > French, Provençal *sapin*, Old Italian *zappino*, etc. (Hubschmid 1960: 40–41; REW 7592). Basque $*\acute{s}apa-r$ should of course be kept separate from some other superficially similar Basque words for ‘bramble’: $*lapa-r$, $*lahar$, and $*gapa-r$ (see BCR C.15, C.17, C.18), each of which has a distinct NC cognate; but clearly these have all become contaminated in the popular mind.

$\psi\bar{\upsilon}\chi\acute{\eta}$ ‘aspiration, breath, life, vitality, soul (of the deceased), spirit’; $\psi\bar{\upsilon}\chi\omega$ ‘to breathe, blow’ [‘I breathe, blow’]. “I do not find these suggestions [of IE etymologies] convincing. There is hardly any evidence for an IE root $*bhes$ - ‘to blow’ ... Therefore, the word is more probably of Pre-Greek origin” (Beekes 1672). | **Basque** $*bi=si$ (noun) ‘life; lifetime’, (adj.) ‘alive’: Common Basque *bizi* ‘alive, living, lively; life’, (B-Markina) *bixi* /*biši*/ (A&T VII 147; EDB 145; BCR A.87). | **North Caucasian**: Chechen, Ingush, Batsbi *sa* ‘soul’; oblique base $*si$ - (Chechen *si-na*-, Ingush *si-no*, Batsbi pl. *siy-š*.); III-class); Lak *si:h* ‘breath, vapor’ (III-class); Karata *si:uh-an*- ‘to get tired’; West Caucasian: PWC $*p\acute{\alpha}-\acute{s}w$ > Ubykh $p=\acute{s}a-\chi^w\acute{\alpha}$ - ‘to breathe’; Adyge, Kabardian $p=\acute{s}a-n$ ‘to get tired’. Abkhaz *a-ps-rá* ‘to die’, Abaza *ps-ra* id. < PNC $*\acute{s}iHwV$ ‘breath; to breathe’; with III-class prefix $*b=siHwV$ (NCED 961). § Nikolaev (72, no. 40) cites Greek $\psi\bar{\upsilon}\chi\acute{\eta}$ as a loan from North Caucasian, in which ψ /*ps*/ corresponds to /*pś*/, /*ps*/ in the West Caucasian forms, and χ /*k^h*/ to the PNC laryngeal $*H$ (Lak /*h*/). The Basque word is analyzed in BCR as the root $*=si$ (= PNC $*\acute{s}iHwV$) preceded by the fossilized class prefix $*bi=$ (= PNC $*b=/*w=$ III-class [inanimate] singular: note that the Nakh and Lak parallels cited above belong to the III-class). NCED suggests deriving PWC $*p\acute{\alpha}-\acute{s}w$ from an earlier $*p\acute{\alpha}-\acute{s}aHwV$, which is exactly parallel in form with Basque $*bi=si$, and also provides a plausible antecedent to Greek $\psi\bar{\upsilon}\chi\acute{\eta}$ /*psūk^hé*/. “The semantic developments ‘to breathe’ > ‘get tired’ ... > ‘die’ are quite usual.” (NCED 961); cf. Russian *dušá* ‘mind, soul, spirit’ : *dušít* ‘to smother’; Greek $\epsilon\kappa\cdot\psi\bar{\upsilon}\chi\omega$ ‘to breathe one’s last; expire, lose consciousness, die’. It is tempting to suggest Latin *spīrō* ‘I breathe, blow; am alive, am inspired’, *spīritus* ‘breath, breathing; breeze, air; spirit’, *ex-spīrāre* ‘to breathe out, die’, etc., from Proto-Italic $*spīr/s-$ (i.e., $*spīr-$ or $*spīs-$, according to de Vaan 2008: 581), if there was a metathesis of $*psī-$ > $*spī-$. DeVaan only opines “Possibly an onomatopoeic formation imitating the sound of breathing. There are no direct [IE] cognates.”³⁸ Diakonoff & Starostin (1986: 36) thought there were cognates of PNC $*\acute{s}iHwV$ in Hurrian-Urartian: Hurrian *šey-iri* ‘alive’, *šey-ori* ‘fate’ or ‘life’, Urartian *šu/ox-ori* / *šex-eri* ‘alive’.

³⁷ Other developments have penetrated world current events. The Mexican Spanish word *chapo* ‘persona de baja estatura’ (Dicc) < Basque (B) *txapar* ‘persona de pequeña estatura’ (OEH) is widely known as the nickname of drug trafficker Joaquín “El Chapo” Guzmán (now imprisoned in Colorado, USA).

³⁸ In his older dictionary Walde (1910: 731) cites Old Church Slavic *pišta*, *piskati* ‘pfeifen, flöten’, Sanskrit *picchorā* ‘Pfeife, Flöte’, and reflexes of Proto-Germanic $*fisan-$ ‘to blow; to fart’. In the etymological lemma for the latter word in Kroonen (2013: 142) Latin *spīr-* is also cross-referenced, along with Welsh *ffŷn* ‘breath’ (< $*spoi-n-eh_2-$). Kroonen speculates that “the PIE form of the verb may have been $*psēi-s-e-$,” with the first /*s*/ dropped in Germanic due to dissimilation. Again, PIE $*psēi-$ is close to the form of PNC $*(b=)siHwV$. and Basque $*bi=si$.

Phonology

Some sound correspondences have already been remarked upon. In the examples **μαστός** and **σταφυλή** we see the equation Pre-Greek *st-*, *-st-* = PNC **ɕ*/*c'*/ = Basque **ś*. In example III we have Pre-Greek *-st-* = PNC **-ɕ-* = Basque **-s-*. Both represent changes of the type TS > ST (or, less likely, ST > TS) which are reminiscent of shifts within North Caucasian (Nakh languages) and between North Caucasian and Basque. For the former, the authors of NCED remark that

Most difficult to explain are cases of [Proto-Nakh] reflecting PEC hissing (and hissing-hushing ...) affricates and fricatives as a *st (*st̥) cluster (both in initial and non-initial positions) ... probably as a result of distant palatalization) after or before a *j ... or following the resonant *l (NCED 47, 51).

Basque /st/, /st̥/ realizations coincide with Nakh /st/ or /st'/ in only a few cases:³⁹

Basque (B) *beaztun* 'gall, bile' (vs. [L-arc] *behazuñ*, etc. < **beha-sun* A.88) ~ Chechen *stim* 'gall' (PNC **cwǎymě* 'gall, anger': NCED 329)

Basque (R) *aizto* 'knife' (Q.11) ~ Chechen *sto* 'chisel', Ingush *osta*, Batsbi *st'o* id. (PNC **Hǎy3V̄* 'chisel': NCED 542)

Basque **aštun* 'heavy' (R.29) ~ Chechen, Ingush =*arst-* 'to fatten, become fat', Batsbi =*arst'*-id. (PNC *=*HrV̄yśĒ* 'thick, dense, fat': NCED 608)

This suggests that the conditions producing Basque /st/ clusters were at least slightly different from those underlying Nakh /st/ clusters. In putative Basque-NC cognates there seems to be a correlation between Basque /st/ or /st̥/ and PNC/PEC tense sibilants (**s̥*, **ɕ*, **ɕ'*, **ʒ*):

Basque **e=stari* 'throat' (A.32) ~ PNC **ɕwĕri* / **riɕwĕ* 'neck' (Agul *s:ür* 'gullet': NCED 953)

Basque **hestu-n* 'ring, link' (Q.37) ~ PEC **HV̄ɕV* (Khwarshi *ocu* 'ring, hoop; buckle': NCED 612)

Basque **listo-* / **lišto-* 'hornet, wasp' (B.31) ~ PEC **łǎmɕV* (Akhwakh *łac':u* 'ant; bug, bedbug': NCED 766)

Basque **pista* 'rheum, eye secretion' (A.79) ~ PNC **pĩncwǎ* 'resin, juice' (Lak *pic* 'dew, sweat': NCED 871)

Basque **esti* 'honey, sweet' (if < **emsti*: P.21) ~ PEC **mǐzV* 'sweet' / **hwmižū* 'honey' (Archi *ic'*: 'sweet' / *imc* 'honey': NCED 824)

With only two Pre-Greek examples, it is insufficient to demonstrate a firm correlation between its /st/ reflexes and Euskaro-Caucasian. However, it may be a tantalizing hint that more examples could be found with a concerted search.

Morphology

Fossilized class (gender) markers: Beekes and others have remarked on the variations of initials in cases such as **μάδρουα** ~ **άμάδρουα** ~ **βάδρουα** ~ **ǎδρουα** 'plums, sloes'; we saw a similar alternation in Basque **ma=dari* / **u=dari* 'pear'. There is a similar situation in **ἀκαρί** 'mite' / **κόρις** 'bedbug', about which Beekes remarked on the prothetic vowel as one of the signs of a Pre-Greek substrate word: "Pre-Greek had a prothetic vowel, e.g. **άσκάλαφος** ['name of an unknown bird, perhaps an owl'] beside **κάλαφος**. In most cases, the vowel is **ά-**. The numbers

³⁹ Number designations after Basque forms (A.88, etc.) correspond to etymology numbers in BCR, where more complete information on each comparison may be found.

[according to Furnée 1972: 368ff.] are as follows: $\alpha \pm 90$, $o 10$, $\varepsilon 5$, $\iota 3$, $u 0$, $\eta 6$, $\alpha\iota 2$. Note that, generally speaking, α may interchange with o , ε , and $\alpha\iota$. Indeed, we have cases where prothetic o interchanges with α , and the same holds for ε ...” (Beekes xxiii). Yet another example is $\acute{\alpha}\sigma\tau\alpha\phi\acute{\iota}\varsigma \sim \acute{\omicron}\sigma\tau\alpha\phi\acute{\iota}\varsigma \sim \sigma\tau\alpha\phi\acute{\iota}\varsigma$ ‘dried grapes, raisins’, beside $\sigma\tau\alpha\phi\upsilon\lambda\acute{\eta}$ ‘bunch of grapes; grape’. With a possible prefix ι - we have $\acute{\iota}\text{-}\kappa\tau\acute{\iota}\nu \sim \acute{\iota}\text{-}\kappa\tau\acute{\iota}\nu\omicron\varsigma$ ‘kite’.

This brings to mind the list of seven “puzzles” that R.L. Trask thought Basque-Caucasian comparisons should help to solve. First on the list was a statement that “Pre-Basque clearly had an extraordinarily large proportion of lexical items beginning with a vowel, and ... only a very few word-initial consonants. Why is this so?” (Trask 1996: 115–16).

In fact, one of the current authors (e.g. Bengtson 1994) had already offered an explanation regarding Basque nouns, that some of the initial vowels reflected fossilized class prefixes, or “stage III articles,” a solution that Trask repeatedly rejected. This hypothesis is supported by the fact that many, but by no means all, Basque-North Caucasian parallels involve Basque words with initial vowels.

Basque $*a=\acute{c}o$ ‘old woman’ (BCR J.7) ~ Lak $c:u$ - ‘female’, etc. < PNC $*\acute{c}w\acute{o}yV$ ‘woman, female’ (NCED 374)

Basque $*a=kec$ ‘boar’ (BCR N.23) ~ Lak $q:\acute{a}\acute{c}a$ ‘bull-calf’, etc. < PEC $*g\acute{a}\acute{c}V$ (NCED 453)

Basque $*e=k\acute{e} / *k\acute{e}$ ‘smoke’ (BCR F.2) ~ Avar $k':uy$ ‘smoke’, etc. < PNC $*\acute{k}w\acute{i}nhV$ (NCED 738)

Basque $*i=\acute{c}u / *s\acute{u}$ ‘fire’ (BCR F.1) ~ Lak $c'u$ ‘fire’, etc. < PNC $*\acute{c}\acute{y}\acute{i} / *c\acute{y}y$ - ‘fire’ (NCED 354)

Basque $*o=hol$ ‘board, plank’ (BCR Q.62) ~ Rutul $\chi\acute{i}l$ ‘wooden trough’, etc. < PEC $*\chi ulV / *xutV$ (NCED 1078)

Basque $*u=ri[s]a$ ‘female (animal); woman’ (BCR N.15) ~ Akhwakh $reša$ ‘heifer’, etc. < PEC $*r=išw\acute{E}$ ‘heifer; female child’ (NCED 671)

Many more examples are cited in BCR (pp. 58–71). It is further proposed that Basque $*e=$ and $*i=$ are historically the same prefix, likewise with $*o= / *u=$, that is, mid and high vowel allomorphs, as also seen in the hypothesized prefix $*be= / *bi=$ (see below).

Iversen & Kroonen (2017: 517), in their study of a postulated pre-Indo-European substrate that they term as the “Early European Neolithic language” point out that many relic words traced to this substratum “exhibit the same alternation consisting of forms with and without word-initial a -. In all likelihood, this was a productive derivational element—that is, a prefix—in the language from which these words were borrowed.” Some examples cited by Iversen & Kroonen (with putative Basque and NC cognates) include:

Latin *merula* ‘blackbird’ (< $*mesl-$) : Old High German *amsala* id. (< $*a\text{-}msl-$) : cf. (without a prefix) Basque $*mosolo$ ‘(small) owl; buho, mochuelo’: *mozolu*, *mozoilo*, *mosolo*, (expressive) *moxolo*, *motzollo* id.; NC: Archi *mus:al* ‘wild turkey’, Chamali (dial.) *mus:iya* id.⁴⁰

Old English *lāwerce* ‘lark’ (< $*laiwar-$) : Gaulish *alauda* id. (< $*a\text{-}laud-$) : cf. Basque $*e=tanha$ (~ $*e=nhala$) ‘swallow, swift’; Dargwic *laha* ~ *lawha* ~ *lahwa* ~ *lax^wa* ~ *nax^wa* ‘pigeon’⁴¹

⁴⁰ The NC words reconstruct to PEC $*Vms\acute{g}w\acute{e}l\acute{z}\acute{e}$ ‘wild turkey’ (NCED 225); Spanish *mochuelo* ‘a kind of small owl’ looks like a loan from a Basque or Vasconic expressive variant of $*mosolo$. On semantic changes, as can be seen from other bird etymologies, meanings can historically vary quite widely: **A.** Hittite *haraš*, *harašaš* ‘eagle’; Greek $\acute{o}\rho\nu\epsilon\omicron\nu$ ‘bird’ (general), $\acute{o}\rho\nu\iota\varsigma$ ‘bird, cock, hen’; Armenian *oror*, *urur* ‘seagull, harrier’, etc.; **B.** Dargwa *hunuc* ‘eagle’ (poetic); Avar *hinč*: ‘bird’ (in general); Chechen *hōza* ‘sparrow’, etc. (NCED 525); compared with Basque $*hunc/*honc$ ‘owl’ (BCR B.23); **C.** Chechen *šoršal* ‘blackbird, thrush’; Bezhta *šašu* ‘swallow’; Khinalug *čānč* ‘pigeon’ (NCED 987); compared with Basque $*soso / *šošo$ ‘blackbird, thrush’ (BCR B.18).

⁴¹ BCR B.21; NCED 750. The bird species swallow and swift are superficially similar, less so the pigeon (dove). (See the notes to ‘blackbird’, just above.)

Old English *secge* ‘sedge’ (< **sak-*) : Russian *osóka* id. (< **a-sak-*) : cf. Basque **i=śac* ‘broom (plant)’; Chechen *šac* ‘sedge’⁴²

Latin *rāpa* ‘turnip’, Old High German *ruoba* (< **rāp-*), Proto-Slavic **rěpā* ‘turnip’ (BER 6, 387) : Welsh *erfin* < **a-rb^(h)-* id. : cf. Basque *arbi* (< **a=r(V)bi*) ‘turnip’

It can be noted that several of Iversen & Kroonen’s Early European Neolithic words have Basque and/or North Caucasian comparanda, giving weight to the hypothesis that “Early European Neolithic language” was related to Basque and North Caucasian. More weight is supplied by the fact that a similar phenomenon – the presence or absence of these types of vocalic prefixes – can be observed synchronically within Basque. For example, the Basque word for ‘smoke’, cited above as **e=kē* / **kē* ‘smoke’, is attested as standard (EB) *ke*, and variants of this in most of western and northern Basque (*ké, kée, keé, kái, ki, khe*, etc.), but in large parts of High Navarre (AN) and sporadically in Low Navarre (BN) there is a prefix *e-* or *i-* (*éke, eké, iké*, etc.).⁴³ Similarly:

Basque **geHeti* / **i=keta* / **o=keti/a*: (L-arc) *geheli* ‘fresh beef’, (B) *geeli* ‘(fresh) beef, fresh meat’, (B, R, Bzt) *geli* ‘lean meat’ / (BN) *ikhel* ‘fattened ox’, (AN) *ikela* id., (B, Z) *okela* ‘meat’, (L) *okhela, okheli* ‘meat; piece (of meat, cheese), morsel’, etc.; cf. Dargwa *q^{al}*, Lak *ul* ‘cow’, etc. < PEC **qhwētV* / **qwēthV* ‘large female domestic animal (cow, mare)’ (BCR P.12; NCED 917)

Basque **purdi* / **e=purdi*: (Bzt, BN-Aldude) *purdi* ‘buttocks, arse’, (AN) *epurdi*, (L) *iphurdi*, (Z) *iphürdi*, (A, G, Sal, B-Markina, Oñate) *ipurdi*, (B-arc) *ipirdi*, (B) *eperdi*, id.; cf. Archi *p^{art}i* ‘one of the large intestines’, etc. < PEC **pHVr^twV* ‘some inner organ’ (BCR A.45; NCED 871)⁴⁴

Basque **gai* / **e=kai*: (B-arc) *gei* ‘thing’, (B, Z, R) *gei* ‘material, subject’, (G, AN, BN, L) *gai*, (BN-Garazi) *kai*, (BN-arc, L-arc) *ekhai, ekai*, (BN) *ekhei*, (Z) *ekhéi* id.; cf. Avar *q^{ayí}* ‘thing(s), possession(s)’, etc. < PEC **q^wäyē* ‘thing(s), possession(s), household’ (BCR L.13; NCED 930)

These variants seem to reflect a time, long before Basque was a written language, when fossilized class prefixes (stage III articles) were in free variation, and eventually each dialect generalized, in different ways, either the prefixed or unprefixing form, or sometimes both. In other words it can be called the reorganization of allomorphs.

In North Caucasian traces of a similar trend are found sporadically, mainly in the East Caucasian branch. In one of the words for ‘snow’ Lezgian languages (Lezgi *žiw*, Tabasaran *yif*, Agul *ibx*) reflect PEC **yⁱwλV* ‘snow’, which also appears to include an incorporated **y=* (II-class) prefix analogous to **e=* in the Basque word **e=thu-^r* ‘snow’, i.e. **yⁱwλV* < **y(i)=λⁱwV*; on the other hand the synonymous Nakh words (Chechen *lō ~ lūo*, Ingush *lō ~ loa*, Batsbi *law*) stem from the unprefixing PEC form **λⁱwV* ‘snow’ (BCR G.17; NCED 684).⁴⁵ Thus it is proposed that PEC **λⁱwV* / **y(i)=λⁱwV* ‘snow’, with a regional reorganization of allomorphs, is parallel to the Basque cases like **kē* / **e=kē* ‘smoke’ cited above. Consider also:

⁴² BCR C.26; NCED 983. The semantic differences may be based on ‘plant used in making brooms’: some types of sedge are suitable for this. Broom and sedge are both under the order *Poales*.

⁴³ These patterns are shown quite clearly in the Basque dialect atlas (EHHA map 1026).

⁴⁴ For semantic typology cf. Old Indic *gudá-* ‘intestine, bowels, anus’ > Pali *guda* ‘anus’, Sindhi *guī* ‘anus, posterior’, etc. (Turner 1962, lemma 4194).

⁴⁵ Perhaps also in Basque **lu-* / **e=thu-*: (Sal, Bzt, AN-Lezaka) *lauso* ‘avalanche of snow’ / (BN) *elhauso* id., compounds of Basque **lu-* / **e=thu-* ‘snow’ + **auśo* ‘fall (of snow, rain)’ (BCR G.11), the latter related to Basque **e=auśi* ‘to fall’; cf. PEC **ũśV* ‘to descend, fall, be scattered’ (NCED 1011; BCR V.20).

PEC **λwilV* / **y(i)=λwilV* ‘elbow’:⁴⁶ Tsezi *horu* ‘elbow’, Hunzib *hɔru*, Khwarshi *hal*; Agul *q:ar-xil* ‘elbow’⁴⁷ (< Proto-Lezgian **λwil*) / (with prefix) Akhwakh *eλelo* (*et̪elo*) ‘elbow’ (NCED 770); compared with Basque **be=thau*n / **be=thaur-* ‘knee’ (BCR A.74).

Examples of prefixed and unprefixed nouns can also be found involving the fossilized prefix (article) **be=* / **bi=*:

Basque **hac* ‘finger, paw’ / **be=hac* ‘thumb, toe’: (BN, L) *hatz* ‘paw’, *be-hatz* ‘finger, thumb’; (B) *atz* ‘finger, inch’, *be-atz* ‘toe’, etc., with many more meanings depending on dialect (BCR A.68); cf. Avar *kwac̣ʰ* ‘paw’, Batsbi *kʰač* ‘foot, leg’ (a slighting expression), etc. < PEC **kwāčĕ* (NCED 704)

Basque **herde*, **helde-ř* / **bilde-ř* (< **bi=helde-ř*): (BN, L) *herde* ‘drool(ing), slobber, slaver’, (AN, Bzt, Sal) *erde* id.; (with **ř-* suffix and dissim.) **helde-ř* id. > (BN, L) *helder*, *heldor*, (L, BN-Baigorri, R-Uztárróz) *elder*, (Z) *élder* ‘drop of spittle that falls from the lips’; (G-Gabiria, Iziar) *bilder*, (G-Zestoa) *bildar* ‘drool, saliva’ (BCR A.80); cf. Karata *hanlʼa* ‘sweat’, Akhwakh *ātʰʼa* id., etc. < PEC **hāmλă* (NCED 509)

The Basque prefix **m=* / **ma=* / **mo=* is far less frequent than **be=* / **bi=*, and may have been a nasalized variant of the latter. Both Michelena and Trask accepted the reality of the **m=* prefix.⁴⁸ Besides Basque **ma=dari* / **u=dari* ‘pear’, discussed above, consider the following examples.

Basque **mo=kol(o)* / **a=kal* / (reduplicated) **kakol*: (B) *mokol* ‘shell (of egg, nut), husk (of maize)’, *mokolo* ‘husk (of maize)’ / (Bzt) *akal* ‘empty (of a chestnut shell)’ / (B) *kakol* ‘shell’ (BCR C.38); cf. Akhwakh *qʰoli* ‘crust, rind’, Tsezi *qʰul* ‘bark’, Bezhta *qʰeqʰel-ba* ‘birch bark’, etc. < PEC **q̄wātV* ‘bark, crust’ (NCED 931)

Basque **ma=kac*, **ma=keć* / **a=kać* / **o=koc* : (G) *makatz* ‘nick, scratch’, (G) *makets* ‘deformed or defective thing’, / (B, G) *akats* ‘cut, nick, notch, scratch; fault, defect’ / (B) *okotz* ‘chin, snout’ / (with reduplication) (AN) *kokots* ‘chin, nape’, (BN) *kokots*, *kokotz* ‘chin’, (L) *kokots*, *kokotz* ‘chin’, (Z) *kokots* ‘chin’ (BCR A.15, L.1); cf. Lezgi *qʰac* ‘notch, nick’, Khwarshi *qʰacʼa* ‘slice (of bread)’, Rutul, Tsakhur *qʰac* ‘chin’; Lak *qʰac* ‘bite, mouth’. etc. < PEC **q̄āčĕ* / **q̄āčŭ* (NCED 907)⁴⁹

Basque **ma=gal* / **e=gal*: (R, Sal) *magal* ‘wing’ / (AN) *egal* ‘wing, fin’, (BN, L) *hegal*, (Z) *hégal* id., (B) *egal* ‘loin, flank (of cow)’ (BCR A.63);⁵⁰ cf. Lak *qa* ‘wing’, Lezgi, Agul *vil* ‘hand’, Archi *χol* id., Bezhta *χaro* ‘elbow’, etc. < PEC **q̄ilʔi* ‘elbow, arm, wing’ (NCED 895)

Turning now to North Caucasian, there are many cases in which fused or lexicalized class prefixes are attested in some languages, often with a different class prefix, or no prefix, in other languages:

⁴⁶ NCED cites the reconstruction as *(*Hi*)*λwilV* ‘elbow’; **λwilV* / **y(i)=λwilV* is Bengtson’s reinterpretation.

⁴⁷ The Agul word is “a compound with some not quite clear first component (is it a distorted [Proto-Lezgian] **χ:il* ‘hand’? or **q:lun* ‘arm’?)” (NCED).

⁴⁸ “No se puede poner en duda, por el contrario, la realidad de un prefijo nominal *m(a)-* ...” (FHV 271). It was also mentioned by Trask, along with a long list of “expressive” Basque words with initial *m-* (Trask 1997: 257–58; EDB 273–78).

⁴⁹ Semantic changes (‘cut’ or ‘bite’ > ‘notch, nick’ or ‘mouth, chin’) are parallel in Basque and NC. Glosses in Lezgian languages are 1 ‘bit, slice’ (Tabasaran, Agul), 2 ‘notch, nick’ (Lezgi), 3 ‘(biting part) > chin’ (Rutul, Tsakhur).

⁵⁰ It has been suggested that an original Basque **e=gal* ‘wing, fin, loin, flank’ (corresponding to PEC *(*y=*)*q̄ilʔi* ‘elbow, arm, wing’) later contaminated with the originally distinct **hega-* ‘to fly’ (BCR V.43) to produce the blended form *hegal* in northern Basque.

Tsakhur *wu(=)xun* ‘belly’, Rutul *u(=)xun* id.; Avar *ma(=)xá* ‘abomasum’ < **bV=xwVn* (III-class prefix) / Lezgi *ru(=)fun* ‘belly’ (IV-class prefix)⁵¹ / (unprefixed) Agul *fun*, Dargwa *k^(w)ani* ‘belly’, etc. < PEC **λwĩnʔi* (NCED 771)

Godoberi *re(=)vil* ‘leg’, Botlikh *re(=)vil* ‘thigh’ (IV-class prefix) / (unprefixed) Tsakhur *q:el* ‘foot, leg’, Rutul *vil* id., etc. < PEC **gētu* (NCED 455)⁵²

Avar *mi(=)tír* ‘wing’⁵³ / Andi *tíru* ‘feather, wing’, Tsezi *lel* ‘wing’, etc. < PEC **lila* ‘wing’ (NCED 762)

PNC **bēmʔV* (< **b=fwemʔV*) > Hunzib *bət’i* ‘worm’, Bezhta *bet’e-la* id., Lezgi *büt’-rük* ‘larva’, Abkhaz *á-mat* ‘snake’, etc. (NCED 290) / PNC **fiwe(m)ʔi* > Avar *hut’* ‘worm’, Bezhta *hat’o-la*, Lak *yət’i*, etc. (NCED 535)

In the last set the opposition of Bezhta *hat’o-la* ‘worm, helminth’ vs. *bet’e-la* ‘worm’ is typologically parallel to the opposition of Basque (BN, L) *helder* ‘drool, saliva’ vs. (G) *bilder* id. (see above). In each case the second word, with initial *b-*, incorporates the former class prefix. According to NCED there is a color adjective **fiwVmʔV* ‘red’ that is related to the two words for ‘worm’, and there is a familiar pattern here too: Agul *b(=)at’-ar-* ‘beautiful, handsome’, which incorporates the class prefix, vs. Khwarshi *ut’ey* ‘red’, Dargwa *hunt’-ena* id., etc. (NCED 541).⁵⁴

The morphological patterns described for Basque and North Caucasian, of bare noun stems alternating with (fused or lexicalized) CLASS PREFIX + NOUN stem (e.g., Basque **kē* / **e=kē* ‘smoke’; PEC **λiʔwV* / **yĩwλV* < **y(i)=λiʔwV* ‘snow’) are consistent with the Pre-Greek hypothesized by Beekes and others, in which noun stems with no initial vowel alternate with those with prothetic vowels (e.g., Greek κόρις ‘bedbug’ / ἀκαρί ‘mite’), and with the postulated “Early European Neolithic language” which shows a similar pattern, e.g., Latin *merula* ‘black-bird’ < **mesl-* / Old High German *amsala* id. < **a-msl-* (Iversen & Kroonen 2017: 517).

Ablaut: Beekes (49: 754) mentions another Pre-Greek feature in the remark that “I would rather think that κόρις is cognate [with ἀκαρί], as a substrate word, with prothetic vowel and α/o interchange.” Another example of α/o alternation may be found in the apparent derivatives of καλιᾶ, as reported by Hesychius: καλύβη ‘hut, cabin’ and κόλυβος ‘farmstead’. Beekes cites some other examples, e.g. κάβαξ ‘crafty, knavish’ and κόβακτρα ‘pieces of flat-tery, knavery’ (both from Hesychius); λυκάφος ‘name of a poisonous plant’ / λύκοψος id.

Ablaut, according to NCED, was a productive feature of Proto-North Caucasian, including an alternation of **o* and **a*, as in PNC **fiðçV* / **fiǎçV* ‘full, to fill’ (NCED 525), reflected in Proto-Nakh **=uç-* ‘to be filled, satiated’ / **=aç-i(n)* ‘heavy’ (> Chechen =üz-na ‘full’ / =eza ‘heavy’). It has been proposed that there is a relic of this ablaut in the Basque adjective **ośo* ‘whole, complete’ and the verb **aśe* ‘to be filled, satiated’ (BCR R.65, V.66). It was also suggested that this **a*/**o* ablaut could account for some cases in which Basque has **a* versus PNC **o*,

⁵¹ Note that Lezgi retains the fossilized *ru=* (orig. IV-class prefix) even though the language (like Agul and Udi) has lost class or gender as a grammatical category.

⁵² “However, there are two possible reconstructions: a) the one proposed above — in this case we must consider **r-* in PA and PTs as a former class prefix (which raises some doubts); b) we can reconstruct **Gēlu* (with **l*) and a metathesized variant **lēGV* > PA **rivi-*, PC **riçi-* (with a rather frequent *-IV extension). At present it is hard to choose one of these solutions only” (NCED).

⁵³ “The origin of the initial *m(i)-* is not clear (perhaps, analogy with names for body parts like *mehéd* ‘breast’ etc.?)” (NCED).

⁵⁴ The semantic link of ‘red’ ~ ‘beautiful’ is common: cf. Czech *krásný* ‘beautiful’ ~ Russian *krasnyj* ‘red’; Latin *pulcher* ‘beautiful’ ~ Middle Irish *erc* ‘gay-colored, red’, etc. (Buck 16.81); for ‘red’ ~ ‘worm’ cf. Old Church Slavonic *črŭvŭnŭ* (чръвнѣ) ‘red’ < **čirvi* ‘worm’; French *vermeil* < Latin *vermiculus* ‘little worm’ (Buck 15.66).

or vice-versa. Basque **gari* / **gal-* ‘wheat’ = PEC **ǭlʔe* ‘wheat’ (BCR O.1), beside Basque **goše* ‘hunger, hungry’ = PNC **gašē* ‘hunger’ (BCR R.30); if so, this could reflect reorganizations of allomorphs, i.e. that Basque selected one allomorph (with either **a* or **o*) and PNC (or individual NC languages) selected another.

Another possible trace of Euskaro-Caucasian ablaut is seen in the vowel alternation NEK- / NIK- in (Pre-)Greek **νέκταρ** / **νικάρ-**, Basque **nega-ř* / **niga-ř*, and PEC **nēwǭ* / **nīwǭ*Ÿ-. Only in PEC is the variation explained as a morphologically significant ablaut alternation, in which /e/ is associated with direct stems and /i/ with oblique (NCED 81–82). It has been suggested (BCR 105–110) that there are traces of this and other North Caucasian ablaut alternations in Basque; in the case of Basque **nega-ř* / **niga-ř* ‘tears’ the allomorphs have apparently been redistributed as regional variants, generally, (south-)western /e/ vs. (north-) eastern /i/. Beekes (2010: xxx) calls attention to an apparent alternation of /ε/ with /ι/ in Pre-Greek words.

Suffixes: Beekes (xxxvii, xxxix) cites the suffixes **-ιλ-** and **-υλ-** as indicators of Pre-Greek words, and among the words cited above these suffixes figure in two words connected with fruits: **μέσπιλον** ‘medlar, medlar tree’ and **σταφυλή** ‘bunch of grapes, grape’. Greek σταφυλή (cf. ἀσταφίς ~ ὀσταφίς ~ σταφίς ‘dried grapes, raisins’) has a close formal match in Avar *c’ibil* ‘grape’, also with a lateral suffix. Other NC languages have a form with no suffix (e.g. Rutul *c’ib* ‘juniper’) or a form with a different suffix (Lezgi *c’p:-az* ‘blackberry’); Basque **šapa-ř* ‘thicket, bramble’ has been proposed as a cognate, with a common fossilized plural suffix **-ř*. In a Sino-Caucasian perspective Pre-Greek **μέσπιλον** may have a remote cognate in Burushaski **mičil* / **bičil* ‘pomegranate’ (SCG 267), also with a suffix **-il* (cf. Khinalug *mič* ‘apple’ and Abkhaz *a-báč* ‘medlar’ for convergent phonetic developments). Other Euskaro-Caucasian words for plants and trees with fruits or berries and a suffix **-al-/*-il-/*-ul-* include:

Basque (AN) *magauri*, *maguri*, (AN-Erratzu) *mauli* ‘strawberry’, (Bzt) *mauri* id. < ? **mag-uli*;⁵⁵ cf. NC: Akhwakh *muq’ali* ‘blackberry’ < PEC **niwǭV* (EHHA, map 572; BCR P.20; NCED 854)

Basque (B) *zumel* ‘cornel; kermes oak; holm oak’, (B-Gernika) *zumel* ‘Mediterranean buckthorn (Rhamnus alaternus)’; **sumal* in the personal name Pero Gonçalves de Çumalburu (1293 CE); NC: Budukh *žumal* ‘cornel’, Lezgi *č:umal*, Avar *žulám* / *žulán* (< **žum-al*) ‘cornel’, etc. (OEH ZUMEL; BCR C.2; NCED 1107)

NC: Archi *t’ummul* ‘grape’, Rutul *t’imil*, Tsakhur *t’umil* id., Budukh *t’ombul* ‘plum’ (< Proto-Lezgian **tum(:)-ul*), beside suffixless Chechen *t’um* ‘marrow; kernel of fruit, nut’, Abkhaz *a-t’amá* ‘peach’, etc. < PNC **tūmhV* ‘kernel, nut, fruit-stone; marrow’ (NCED 1004; SCG 205).⁵⁶

NC: Batsbi *kumel* ‘raspberry’; Bezhta *gemalo* ‘a kind of berry’; beside suffixless Chamali *g^wab* ‘raspberry’, Karata *gobe* ‘strawberry’, etc. < PEC **gwǭmpV* ‘raspberry’ (NCED 443)

Other possible suffix parallels could be explored, e.g. **-ιγξ**, **-ιγγος**, identified as typical Pre-Greek suffixes by Beekes (xxxvii), in **σῦριγξ**, **σῦριγγος** ‘quill, flute, syrinx, windpipe’, etc., which is evocative of Basque *-inko* ~ *-ingo*, e.g. in eastern Basque *gorrinko*, western Basque *gorringo* ‘egg yolk’ (*gorri* ‘red’); (R) *baratxinko* / *baračinko* ‘cuadrado de un huerto’ < **barace* ‘garden, orchard’ = PEC **bārǭV* ‘enclosure’ (BCR Q.8; NCED 1039). **ἄνθρωπος** ‘man’ = Mycenaean

⁵⁵ Basque regional words for ‘strawberry’ are exceedingly diverse, many with initial *ma-* but probably of diverse origins (BCR P.18, P.19, P.20; EHHA map 572).

⁵⁶ From a Sino-Caucasian perspective, cf. Bur **tumáy* ‘shell of nut, fruit stone’, with the distinctive lateral-glide-retroflex /y/ sound (SCG 205; Bengtson & Blažek 2011: 29–30).

a-to-ro-ḡo /ant^hrōk^wos/ is close in formation to Western Basque *andrako*, *andreko* ‘little woman; doll’, with the frequent diminutive suffix *-ko*; cf. (EB) *otso-ko* ‘wolf cub’ (*o^o ‘wolf’), (G) *musu-ko* ‘muzzle’; in North Caucasian: Avar *yasi-k’o* ‘little girl, doll’ (*yas* ‘girl, daughter’), *wac:a-ko* ‘little brother’ (*wac*: ‘brother’), etc. (BCR 56).

The comparison involving Pre-Greek *νέκταο* also calls attention to a putative suffix *-αο*, which Beekes (again following Furnée) cites as a characteristic Pre-Greek element; note also *-αο(α)*, *-αο(ος)*, *-εο*, *-ηο*, *-ηο-*, *-οο-*, etc. (Beekes 2010: xxxvi–xxxviii). The suffix **-r̥* is very common in Basque nouns, especially those with an underlying plural or collective meaning, e.g., **nega-r̥* / **niga-r̥* ‘tears’, **lega-r̥* ‘small stones, gravel’, **itha-r̥* ‘peas, beans’, **lance-r̥* ‘drizzle’, **moko-r̥* ‘buttocks’ etc.: see BCR 77–78.⁵⁷ In East Caucasian there has been a similar development in which the well-known plural suffix **-r*, attested in all branches of North Caucasian, has been lexicalized, with bleaching of the plural meaning, in a significant number of words, e.g. Avar *bucú-r* ‘fortification, dike’, Tabasaran *marca-r* ‘hearth’ (historical plurals of PNC **bōlcĒ*: NCED 308). In several cases the historical plural form has replaced the original singular, e.g. Agul *ib-ur*, Rutul *ub-ur*, Budukh *ib-ir* ‘ear’, historically ‘ears’ (plural of Proto-Lezgian **?Iam*: < PEC **ɣwānɿV* ‘ear’: NCED 240); Khinalug *cul-oz* ‘tooth’ (< *-or*: PNC **cīfhV̄* ‘tooth’: NCED 326) has replaced the original singular, etc. Besides *νέκταο*, it would be important to discover other Pre-Greek words with these *-ο-* suffixes and Basque and/or North Caucasian cognates. It is tempting to consider, for example, *σπινθήρ* ‘spark’, designated as (Pre-Greek?) by Beekes (2010: 1383), possibly connected with Basque (AN) *pintar*, (BN, L, Z) *p(h)indar*, beside (L) *pinta*, (BN, L, Z) *p(h)inda* ‘chispa, centella / étincelle, flammèche’ (‘spark, flash’) (A&T XV: 910; OEH PINDAR).

Conclusions: As mentioned at the beginning, the putative Pre-Greek examples discussed here were selected with three constraints: (a) Pre-Greek status, or questionable IE etymology, according to Beekes, and the presence of (b) putative North Caucasian cognates, and/or (c) putative Basque cognates. Comparing Pre-Greek specimens only with North Caucasian cognates (as with Nikolaev), or comparing Pre-Greek only with Basque might reveal larger numbers of etymologies. Altogether these would form a corpus of Euskaro-Caucasian etymologies from three branches: the still extant Basque and North Caucasian languages, and the extinct Pre-Greek language recoverable from numerous substratal loanwords. The material analyzed above is summarized below in table form.

Greek	Basque parallel	North Caucasian parallel
ἀκαρί ‘mite’ κόρις ‘bug, bedbug’	*kara-/*karkar- ‘beetle’	PEC *k̄ārā ‘mosquito, gadfly’, etc.
ἀλωή ‘threshing floor, garden’	*larain ‘threshing floor’	PEC *=V̄rĪV ‘to thresh’
ἄνθρωπος ‘man’; Mycenean <i>a-to-ro-ḡo</i> /ant ^h rōk ^w os/	*andere ‘lady; young lady; woman; wife’; (B) <i>andrako</i> , <i>andreko</i> ‘little woman; doll’	–
δοκός, δόκανα ‘beam’	*tako, *tak-et ‘stake, post’, etc.	PNC *dwiq̄(w)V̄ ‘log, stump’ Tabasaran <i>duq’an</i> ‘pole, small beam’
ζέφυρος ‘west wind’	–	PNC *cōjwīhV̄ ‘autumn, winter (rainy season)’

⁵⁷ This suffix has traditionally been described as *-ar*, but the examples cited here show that vowels other than /a/ could precede the **-r̥*, e.g. Basque **moko-r̥* ‘buttocks, backside’ = Hinukh *moko-li* ‘back’ < PNC **bōnq̄ō* ‘back’ (NCED 310; BCR A.44).

Greek	Basque parallel	North Caucasian parallel
ἰκτίν ~ ἰκτινός 'kite'	*śaie 'vulture, eagle'	PEC *čwāmŕǃ 'eagle, vulture'
καλιᾶ 'wooden dwelling, hut'	*o=kefu 'stable, hall, corner', etc.	PEC *qǎlǃ 'house, hut'
κόμη 'hair, mane'	*kima 'mane (of horse); bristles (of swine)'	PNC *ǵ(w)ǎmŕǃ 'plait, mane; hair'
μάδρουα ~ ἀμάδρουα ~ βάδρουα ~ ἄδρουα 'plums, sloes'	*ma=dari ~ *u=dari 'pear'	–
μάλκη 'numbness from cold'	*mal-goī 'numb (from cold)'	PEC *mhēlǎe 'cold'
μαστός ~ μασδός ~ μασθός ~ μαζός 'teat, woman's breast'	*mośu 'nose, snout, face, lip, kiss, point, tip'; *muś-ko 'nipple' (Z) <i>müskúa</i> '(the) nipple'	PEC *mhǎrcū 'point, edge, protruding part'; Akhwakh <i>mic'o</i> 'teat, nipple'
μέσπ-ιλ-ον 'medlar'	*mahać 'grape(s)'	PNC *ǵǎmćo 'apple; medlar'
μικρός ~ σμικρός ~ μικκός ~ μικός 'small, short, little'	*miko 'a little, a little bit'	PEC *miǵwV 'small, young one'
μύλλον 'lip'	–	PEC *mVhwVli / *hwVmVli 'mouth, face'
μύσταξ ~ βύσταξ 'upper lip, mustache'	*bisa-ī 'beard'	PEC *bilǵV 'beard'; Tindi <i>miža-tu id.</i>
νέκταρ 'nectar, drink of the gods' νεκτάριον 'medicine, eye-salve' νικάριον 'eye-salve'	*nega-ī / *niga-ī 'tears, weeping' *nega-l 'herpes, rash, scurf'	PEC *nēwǵū / *niwǵǃ- 'tear(s); pus'
ξύλον ~ σύλον ~ σύλινος 'wood, timber'	–	PEC *čhwǎū 'beam, log, pole'
ὄγχνη ~ ὄχνη 'pear, pear tree'	*ok-(arhan) 'plum, sloe'	PEC *ǵēqV 'grape; fruit; orchard, vineyard'
	*ar-han 'plum'	PEC *ǵonǃV 'pear'
ῥάχis 'spine, backbone, back; (mountain) ridge'	*e=řeka 'gully, ravine'	PNC *riǵwǎ 'mountain, rock; cave' Tindi <i>řek:a</i> 'gorge, ravine' Chechen <i>duq</i> 'mountain ridge'
σαλαμάνδρα 'salamander'	*śuge-mandil 'lizard'	PEC *śVIVǃV 'lizard'; Lezgi <i>šarat'ul</i> 'lizard'
σιρός ~ σιρός ~ σειρός 'pit or vessel for keeping corn, silo' σύριγξ, σύριγγος 'quill, flute, syringe; windpipe, blood vessel' σωλήν 'pipe, channel'; 'grooved tile'	*suffho 'hole, cave; (anatomical) tube' (G) <i>eztar-zulo</i> 'pharynx'; (Z) <i>südü-r-xilo</i> 'nostril', (B) <i>sama-zulo</i> 'gullet'	PEC *śwōǃV 'hollow tube' Chamali <i>na-s:ul</i> 'tubular bone'; Lezgi <i>sulu-r</i> 'throat'
σταφυλή 'bunch of grapes, 'grape'; ἄσταφίς ~ ὄσταφίς ~ σταφίς 'raisins'	*śapa-ī '(blackberry-)bramble, thicket'	PEC *čibV 'a kind of berry'; Avar <i>c':ibil</i> 'grape'
ψυχή 'aspiration, breath, life, vitality, soul, spirit'	*bi=si (noun) 'life; lifetime', (adj.) 'alive'	PNC *(b=)śiHwV 'breath; to breathe'

It is important to emphasize that authentic Pre-Greek words, if they are of a more or less 'basic' nature, are not loans directly from North Caucasian (as framed by Nikolaev), but instead substratal remnants of a Euskaro-Caucasian language related to (Proto-)North Caucasian, but surely not identical with it. These substratal words should be separated from later

cultural loans.⁵⁸ From among the Pre-Greek words discussed above, this caveat seems to apply especially to, e.g., ὄγχνη ~ ὄχνη ‘pear tree; pear’, for which there are several possible sources. Words for fruits and fruit trees are not among the most basic, and there was active trade in such items in the Mediterranean regions. (See also μάδρουα ~ βάδρουα ‘plums, sloes’; μέσπιλον ‘medlar’.)

On the other hand, words like ἀκαρί ‘mite’, μαστός ‘breast, teat’, μύλλον ‘lip’, β/μύσταξ ‘upper lip, mustache’, ξύλον ‘wood, timber’, ῥάχιν ‘spine, back, ridge’, and ψῦχή ‘breath’ are far more basic and much less likely to be counted among cultural loans. They could reflect genuine relics of a Euskaro-Caucasian Pre-Greek language. The two dozen examples discussed here are probably part of a much larger subset that a thorough study of Furnée’s and Beekes’ total list of “Pre-Greek” words might yield.

Abbreviations: languages and dialects

AN	Alto Navarro = High Navarrese (Basque dialect)
arc	Archaic or obsolete form
B	Bizkaian = Biscayan (Basque dialect)
Bzt	Baztanese (Basque dialect)
BN	Bas-navarrais = Low Navarrese (Basque dialect)
EB	Euskara Batua (standard Basque)
EC	East Caucasian (= Northeast Caucasian = Nakh-Daghestanian)
G	Gipuzkoan (Basque dialect)
L	Lapurdián = Labourdin (Basque dialect)
NC	North Caucasian
PEC	Proto-East Caucasian
PNC	Proto-North Caucasian
PWC	Proto-West Caucasian
R	Roncalese (Basque dialect)
Sal	Salazarese (Basque dialect)
WC	West Caucasian (= Northwest Caucasian = Abkhaz-Adyghe[an])
Z	Zuberoan = Souletin (Basque dialect)

References

- A&T = Agud, Manuel, Antonio Tovar. 1988–1995. *Materiales para un diccionario etimológico de la lengua vasca. Anuario del Seminario de Filología Vasca ‘Julio de Urquijo’*.
- Ashcroft, Frances M. 2000. *Life at the extremes*. Berkeley: University of California Press.
- Azkue, Resurrección María de. 1905–1906. *Diccionario vasco-español-francés / Dictionnaire basque-espagnol-français*. Bilbao/Paris: The Author / Paul Geuthner.
- Barnhart, Robert K., Sol Steinmetz (eds.). 1988. *Barnhart Dictionary of Etymology*. New York: H.W. Wilson Co.
- BCR = Bengtson, John D. 2017. *Basque and its Closest Relatives: A New Paradigm*. Cambridge, Mass.: Mother Tongue Press.
- Beekes, Robert. 2014. *Pre-Greek: Phonology, Morphology, Lexicon. Brill Introductions to Indo-European Languages, vol. 2*. Leiden: Brill.

⁵⁸ “The cultural words frequently discussed by Furnée have usually been recorded in relatively late sources (Hesychius, etc.) and so they could represent also relatively late loans from various cultural languages of the East Mediterranean. If there is witness from both North Caucasian and Northwest Semitic & Akkadian, the most probable donor-language was Hurrian, used not only in North Mesopotamia, but also in Syria, Asia Minor and Cyprus” (V. Blažek, p.c. 06/06/2020).

- Beekes, Robert, Lucien van Beek. 2010. *Etymological Dictionary of Greek. Leiden Indo-European Etymological Dictionary Series, vol 10/1*. Leiden: Brill.
- Bengtson, John D. 1994. On the Genetic Classification of Basque. *Mother Tongue Newsletter* 22: 31–36.
- Bengtson, John D. 2017. The Anthropological Context of Euskaro-Caucasian. *Iran and the Caucasus* 21(1): 75–91.
- Bengtson, John D., Václav Blažek. 2011. On the Burushaski–Indo-European Hypothesis by I. Čašule. *Journal of Language Relationship* 6: 25–63.
- Benveniste, Émile. 1929. Le nom d'un animal indien chez Elien. In: St. W. J. Teeuwen (ed.). *Donum natalicium Schrijnen: verzameling van opstellen door oudleerlingen en bevriende yaggenooten, opgedragen aan Mgr. Prof Dr. Jos. Schrijnen bij gelegenheid van zijn zestigsten verjaardag*: 371–376. Nijmegen / Utrecht: N. V. Dekker en Van de Vegt.
- BER = Vladimir I. Georgiev, Ivan Gălăbov, Jordan D. Zaimov. 1971–2017. *Bălgarskii etimologičen rečnik*. Sofija: Bălgarskata akademija na naukite.
- Bernar, Rože. 1982. *Bălgaristični izsledvanija*. Sofia: Nauka i izkustvo.
- Bezljaj, France. 1977–2007. *Etimološki slovar slovenskega jezika*. Ljubljana: Mladinska knjiga.
- Blažek, Václav. 2014. Etymologizing ‘Unetymologizable’ Greek Dendronyms. *Graeco-Latina Brunensia* 19(1): 41–50.
- Bouda, Karl. 1948. Baskisch und Kaukasisch. *Zeitschrift für Phonetik* 2: 182–202, 336–352.
- Buck = Buck, Carl Darling. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*. University of Chicago Press.
- CDA = Black, Jeremy, Andrew George, Nicholas Postage (eds.). 2000. *A Concise Dictionary of Akkadian*. Wiesbaden: Harrassowitz.
- Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.
- Diakonoff, Igor M., Sergei A. Starostin. 1986. *Hurro-Urartian as an Eastern Caucasian Language. Münchener Studien zur Sprachwissenschaft, Beiheft, N. F. 12*. München: R. Kitzinger.
- Dicc = Real Academia Española. 2001. *Diccionario de la lengua española de la Real Academia Española*. 22nd Ed. Available online at: dle.rae.es.
- Džidalaev, Nurislam S. 1990. *T'urkizmy v Dagestanskix jazykax. Opyt istoriko-etimologičeskogo analiza [Turkisms in Dagestani languages. An experiment in historical and etymological analysis.]* Moscow: Nauka.
- EDB = Trask, R. L. 2008. *Etymological Dictionary of Basque. Ed. by Max W. Wheeler*. University of Sussex. Available online at: <http://www.sussex.ac.uk/linguistics/1-4-1-2.html>.
- EHHA = Bidegain, Xarles, Adolfo Arejita et al. N.d. *Euskararen Herri Hizkeren Atlas / Atlas linguistique du Pays basque*. Euskaltzaindia / Real Academia de la Lengua Vasca / Académie de la Langue Basque. Bilbao. Available online at: euskaltzaindia.eus.
- Ehret, Christopher. 2015. Agricultural origins: What linguistic evidence reveals. In: G. Barker, C. Goucher (eds.). *The Cambridge World History. Volume 2: A World with Agriculture, 12,000 BCE–500 CE*: pp. 26–54. Cambridge University Press.
- FHV = Michelena, Luis. 1961. *Fonética histórica vasca*. San Sebastián: Diputación Provincial de Guipúzcoa.
- Flexner, Stuart Berg (ed.). 2001. *The Random House Dictionary of the English Language, Unabridged. 2nd edition*. New York: Random House.
- Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen. Mit einem Appendix über den Vokalismus*. The Hague: Mouton.
- Garnier, Romain. 2008. Nouvelles réflexions étymologiques autour du grec ἄνθρωπος. *Bulletin de la Société de Linguistique de Paris* 102(1): 131–154.
- Georgiev, Vladimir. 1937. *Die Träger der Kretisch-Mykenischen Kultur, ihre Herkunft und ihre Sprache. I Teil. Urigriechen und Urillyrier (Thrako-illyrier)*. *Annuaire de l'Université de Sofia. Faculté Historico-Philologique, Tome xxxiii.4*. Sofia: Imprimerie de la Cour.
- Georgiev, Vladimir. 1941. *Vorgriechische Sprachwissenschaft. Annuaire de l'Université de Sofia. Faculté Historico-Philologique, Tome xxxvi*. Sofia: Imprimerie de l'Université.
- Hamp, Eric. 1983. Prehellenic ἄχρον “chaff, awn”. *Živa Antika* 33(1): 22.
- Hamp, Eric. 1985. Notes on Indo-European dialects. *Indogermanische Forschungen* 90: 70–71.
- Hamp, Eric. 1989a. Prehellenica 5–6. *Živa Antika* 39(1–4): 54.
- Hamp, Eric. 1989b. Prehellenica 7: Words derived from IE. *gher-. *Živa Antika* 39(1–4): 75–76.
- Hubschmid, Johannes. 1960. *Mediterrane Substrate: mit besonderer Berücksichtigung des Baskischen und der westöstlichen Sprachbeziehungen*. Bern: A. Francke.
- Iversen, Rune, Guus Kroonen. 2017. Talking Neolithic: Linguistic and Archaeological Perspectives on How Indo-European Was Implemented in Southern Scandinavia. *American Journal of Archaeology* 121(4): 511–525.

- Klimov, Georgij A., Madzhid Š. Xalilov. 2003. *Slovar' kavkazskix jazykov: Sopostavlenie osnovnoj leksiki* [Dictionary of Caucasian Languages: comparison of the principal lexis]. Moscow: Vostočnaja literatura.
- Kronasser, Heinz (ed.). 1956. *Mνήμης χάριν: Gedenkschrift Paul Kretschmer, 2 Mai 1866 – 9 März 1956*. Wien: Wiener Sprachgesellschaft.
- Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden Indo-European Etymological Dictionary Series 11. Leiden / Boston: Brill.
- Kuiper, Franciscus Bernardus Jacobus. 1956. The etymology of ἄνθρωπος. In: Kronasser (ed.) 1956: 211–226.
- Landucci, Niccolò. 1958. *Dictionarium linguae cantabrigiae (1562)*. Critical ed. by M. Agud & L. Michelena. San Sebastián: Diputación de Guipúzcoa.
- Leschber, Corinna. 2011. Zeitliche Tiefe etymologischer Bezüge. In: *Linguistique balkanique* 50(2–3): 75–82.
- Leschber, Corinna. 2012. Latin tree names and the European Substratum. *Studia linguistica Universitatis Jagellonicae Cracoviensis* 129: 117–125.
- Martirosyan, Hrach. 2021. *Lexical agreement between Armenian and Balto-Slavic in the domain of the physical world*. PowerPoint presentation: Conference Sub-Indo-European Europe – Problems, Methods and Evidence. Leiden University, August 30–31, 2021.
- Mihaylova, Biliana. 2017. The Pre-Greek substratum revisited. In: Bjarne Simmelkjær Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen (eds.). *Etymology and the European Lexicon. Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*: 307–317. Wiesbaden: Reichert Verlag.
- Militarev, Alexander. 2006. Semitic Etymological Database. Available online at: starling.rinet.ru (Tower of Babel project).
- Militarev, Alexander Yu. 2020. Ancient Egyptian – Arabic contacts in lexicon: clue to Arabic Urheimat? *Orientalistica* 3(3): 783–798.
- NCED = Nikolaev, Sergei L., Sergei A. Starostin. 1994. *A North Caucasian Etymological Dictionary*. Moscow: Asterisk Press.
- Nikolaev = Nikolaev, Sergei L. 1985. Severokavkazskie zaimstvovaniya v xettskom i drevnegrečeskom [North Caucasian loanwords in Hittite and Ancient Greek]. In: B. B. Piotrovskij, Vyach. Vs. Ivanov, V. G. Ardzinba (eds.). *Drevn'aya Anatolija*: 60–73. Moscow: Vostočnaja literatura.
- OEH = Michelena, Luis [Koldo Mitxelena]. 1987–. *Orotariko Euskal Hiztegia / Diccionario General Vasco*. Bilbao: Euskaltzaindia. [Edited since Michelena's death by Ibon Sarasola.]
- Pellegrini, Giovan Battista. 1999. A proposito di *musu- e *busu-. *La ricerca folklorica* 39: 131–134.
- Pfeifer, Wolfgang. 1997. *Etymologisches Wörterbuch des Deutschen*. München / Berlin: Zentralinstitut für Sprachwissenschaft.
- REW = Meyer-Lübke, Wilhelm. 1911. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.
- Roscher, Wilhelm Heinrich. 1883. *Nektar und Ambrosia*. Leipzig: Teubner.
- SCG = Starostin, Sergei A. 2005. *Sino-Caucasian: Comparative Glossary*. Available online at: starling.rinet.ru/Texts/glossary.pdf.
- Skeat, Walter W. 1963. *An Etymological Dictionary of the English Language*. (Reprint of revised edition, 1909.) London: Oxford Clarendon Press.
- Skok = Skok, Petar. 1971–1974. *Etimologijski rječnik hrvatskoga ili srpskoga jezika I–IV*. Zagreb: JAZU.
- Snoj, Marko. 2003. *Slovenski etimološki slovar*. Ljubljana: Modrijan.
- Starostin, Sergei A. 1988. Indoeuropejsko-severnokavkazskie izoglossy [Indo-European-North Caucasian Isoglosses]. In: G. M. Bongard-Levin, V. G. Ardzinba (eds.). *Drevnij Vostok: etnokul'turnye sv'azi [The Ancient East: ethnocultural connections]*: 112–163. Moscow: Nauka.
- Starostin, Sergei A. 1996. Comments on the Basque-Dene-Caucasian Comparisons. *Mother Tongue* 2: 101–109.
- Starostin, Sergei A. 2007a. *Trudy po jazykoznaniju [Works on linguistics]*. Ed. by G. S. Starostin. Moscow: Jazyki slav'anskix kul'tur.
- Starostin, Sergei A. 2007b. Indoeuropejsko-severnokavkazskie izoglossy [Indo-European-North Caucasian Isoglosses]. In: Starostin 2007a: 312–358. [Reprint of Starostin 1988].
- Starostin, Sergei A. 2009. Indo-European-North Caucasian Isoglosses. Transl. by Ronald W. Thornton. *Mother Tongue* 14: 77–135. [Translation of Starostin 1988].
- Trask, R. L. 1996. Response to Starostin [1996]. *Mother Tongue* 2: 111–117.
- Trask, R. L. 1997. *The History of Basque*. London / New York: Routledge.

- Trombetti, Alfredo. 1925. *Le origini della lingua basca*. Bologna: Azzoguidi.
- Turner, Ralph L. 1962–1966. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.
- Vaan, Michiel de. 2008. *Etymological Dictionary of Latin and the Other Italic Languages. Leiden Indo-European Etymological Dictionary Series 7*. Leiden / Boston: Brill.
- Vennemann, Theo. 1998. Andromeda and the apples of the Hesperides. In: K. Jones-Bley, A. Della Volpe, M. Robbins Dexter, M. E. Huld (eds.). *Proceedings of the Ninth Annual UCLA Indo-European Conference, Los Angeles, May 23, 24, 1997*: 1–68. Washington, D.C.: The Institute for the Study of Man.
- Walde, Alois. 1910. *Lateinisches etymologisches Wörterbuch*. Heidelberg: Winter.
- Witzel, Michael. 2015. The Central Asian substrate in Old Iranian. *Mother Tongue* 20: 149–178.

Джон Бенгтсон, Коринна Лешбер. О возможном эускаро-кавказском (баскско-северокавказском) происхождении некоторых субстратных лексических элементов в греческом языке.

Широко распространено представление о том, что в греческом языке, относящемся к индоевропейской семье, содержится немало «догреческих» субстратных элементов; при этом нет оснований утверждать, что существовал всего один «догреческий» язык, поскольку область распространения греческого языка вполне могла быть многоязычной. В настоящем исследовании проведен анализ ряда лексических элементов, которые могут свидетельствовать о влиянии некоторого эускаро-кавказского языка (или языковой семьи), носители которого попали в Грецию вместе с распространением земледелия из Анатолии. Такие греческие слова, как *ἄκαρσι* ‘клец’, *μαστός* ‘грудь, сосок’, *β/μύσταξ* ‘верхняя губа, усы’, *ξύλον* ‘древесина, дрова’, и *ψῦχή* ‘дыхание’ относятся скорее к разряду базисной лексики и скорее отражают эускаро-кавказский догреческий субстрат, чем более поздние культурные заимствования. Анализируемые случаи, скорее всего, представляют собой лишь часть более обширного лексического слоя, идентификация которого требует детального анализа списка «догреческих» слов, составленного Фюрне и Беекесом.

Ключевые слова: баскский язык; северокавказские языки; эускаро-кавказская гипотеза; догреческий язык; языковые субстраты.

Lexicostatistical studies in Khoisan II/1: How to make a Swadesh wordlist for Proto-Tuu

The paper is the first in a planned two-part series, whose main goals are to conduct a general lexicostatistical survey of the Tuu, or South Khoisan, family of languages; to reconstruct a reliable approximation of the Swadesh wordlist for Proto-Tuu; and to clarify certain as of yet unresolved issues about the internal classification of Tuu languages. In the first part of the study, I survey the main data sources, identify the main obstacles to historical reconstruction in the Tuu domain, and make observations on some aspects of Tuu diachronic phonology. The main bulk of the paper is actually represented by the Appendix, in which I attempt to reconstruct the equivalents of the first 50 Swadesh list items for the three intermediate nodes of the Tuu family (Proto-!Ui, Proto-Nossob, and Proto-Taa).

Keywords: South Khoisan languages; Tuu languages; click languages; lexicostatistics; basic lexicon; onomasiological reconstruction.

Introduction

Of all the different linguistic lineages commonly united under the umbrella term of “Khoisan”, the Tuu family (originally = Dorothea Bleek’s “Southern Bushman” and Joseph Greenberg’s “Southern Khoisan”, see Güldemann 2005a) shares certain unique properties which simultaneously make it one of the most important and one of the most difficult groupings for any comparative-historical analysis of the Khoisan-speaking area. First, although the overall number of known Tuu languages is smaller than the respective number for Khoe (Glottolog, following Güldemann 2018, currently recognizes 8 different units¹ as compared to 13 for Khoe²), observed grammatical and lexical differences between these languages on the average exceed those observed between the various members of Khoe. Thus, lexicostatistical calculations show that, although the lowest observed percentages of matches within the Tuu family (e.g. 42% between !Xam and !Xóǒ) are comparable to the respective lowest percentages within Khoe (e.g. 41% between Nama and Kxoe), the internal branching of Tuu on the whole is deeper and more complicated than the internal branching of the two major subfamilies of Khoe (Khoekhoe and Kalahari Khoe; see Starostin 2013: 355, 407 for particularities). Among other things, this implies more possibilities for various important diachronic discoveries during the reconstruction of Proto-Tuu, hardly imaginable for Proto-Khoe because of the relatively young age of both of its constituent branches³.

¹ See glottolog.org/resource/languoid/id/tuuu1241.

² See glottolog.org/resource/languoid/id/khoe1241.

³ In fact, the divergence between some of the members of the Tuu family is so impressive that concerns have been voiced in the past about whether one may consider the common ancestry of all its members as an established fact (see e.g. Westphal 1962, 1971). As of today, however, there seems to be a general consensus among all specialists working on Khoisan that the observed phonetical, lexical, and grammatical correlations between all the small sub-branches of Tuu are best interpreted in terms of genetic relationship rather than contact (Güldemann 2005b). In this paper, I proceed from the assumption that this relationship has already been safely established and there is

Second, based on scrupulous phonetic documentation and phonological analysis of those Tuu languages which have survived into the modern age (namely, !Xóǎ and N|uu), this family has emerged as featuring one of the most complex sound systems in the entire Khoisan area. Thus, all known Tuu languages share no fewer than five types of click influxes, including the rare labial type θ (outside Tuu, it is only encountered in the †Hōǎ language of the Ju-†Hōǎ, or Kx'a, family), and at least some Tuu languages have more than 15 phonologically contrasting types of click accompaniments, a number unmatched by any Khoisan language outside of that particular family. Understanding the reasons which underlie this staggering complexity may provide an important insight into the evolutionary mechanisms of click systems in general, yet such an understanding is impossible to gain without a thorough diachronic study of the Tuu family as a whole.

On the negative side of things is the fact that, unlike Khoe, the Tuu family is nearly extinct. The only survivors, as has already been mentioned, are the small dialectal cluster of !Xóǎ (Taa) and N|uu, and even the latter is moribund and has, in fact, up until recently been considered completely extinct (Sands et al. 2007). All of the data that we have on the other languages come from older sources, stretching across about 150 years of ethnographic and linguistic research and widely varying in phonetic and semantic accuracy. Some of these data collections are quite comprehensive, such as Wilhelm Bleek's and Lucy Lloyd's archive on |Xam; other doculets are less lucky, being represented by ultra-short grammatical sketches and minimal wordlists. What is common for most of them, however, is the general unreliability of phonetic notation, grammatical analysis, and semantic glossing — implied by a lack of consistency between recordings of the same language by different researchers (quite often, even by the same researcher over an extended period of time) and by comparison with more recent and more accurate notations by newer and more experienced generations of scholars.

Similar problems are encountered with other Khoisan groupings as well, since data on both Ju (North Khoisan) and Khoe (Central Khoisan) languages often come from the same researchers as data on Tuu (Lucy Lloyd, Wilhelm and Dorothea Bleek, etc.). However, very few languages belonging to either of these stocks are *exclusively* represented by archaic and unreliable data; and even when they are, they usually have very close linguistic relatives with more recent and/or more accurate descriptions against which the questionable data may be cross-checked (e.g. certain Central and Southern sub-dialects of Ju against Ju|'hoan, or the extinct !Ora against its much more prominent neighbor Nama). By contrast, Tuu languages such as |Xam or ||Xegwi, while certainly not linguistic isolates per se, are still quite separate and distinct linguistic units, and cross-checking their data with, for instance, the modern phonetic and lexical descriptions of N|uu would be like trying to ensure the correctness of one's transcription of Czech or Polish by comparing it with Russian (while also having a very vague understanding of the historical phonology and lexicology of Slavic languages in general).

Subsequently, without access to more and better data (which is hardly likely, given the alleged extinction of most of those languages) our ability to properly and definitively reconstruct both the phonological system of Proto-Tuu and its lexical inventory is severely limited, and many problems will likely remain forever unresolved. Nevertheless, approximations are still possible, and any attempt to disentangle the complex web of genetic connections and areal interactions between Tuu languages and their other Khoisan neighbors is liable to shed at least some light on important diachronic processes, some of which may have even chronologically preceded the arrival of Bantu speakers into the area.

no need for special additional validation, allowing us to properly concentrate on issues of reconstruction and internal classification.

In this two-part paper, the next in an ongoing series on Khoisan lexicostatistics, I set up the challenge of conveying a full lexicostatistical survey of those Tuu languages which can actually be used for this purpose, as well as reconstructing Swadesh proto-wordlists for the three major linguistic clusters which constitute this family (!Ui, Nossob, and Taa) and ultimately for Proto-Tuu itself. A first attempt at Tuu lexicostatistics has already been published in Starostin 2013, along with provisional Proto-!Ui and Proto-Taa (but not Proto-Tuu) reconstructions for the 50-item “ultra-stable” subset of the Swadesh wordlist; this publication includes the revised and corrected results of that lexicostatistical survey and expands the reconstructions to include the Swadesh wordlist in its entirety.

The main bulk of both papers will constitute of appendices, containing specific comments on individual Swadesh items (due to volume limitations, the wordlist will be split in two). As for the theoretical parts, the first paper will briefly outline the data sources, the methodology, and the main issues concerning phonetic and lexical reconstruction; the second will deal with the actual internal classification of Tuu and give a brief analysis of the reconstructed proto-wordlists.

The data

Of the eight units currently listed in Glottolog as distinct Tuu languages, relatively complete Swadesh wordlists may be assembled for five, but their respective quality varies significantly depending on the age and thoroughness of the source(s). Additionally, while data from such languages as ||Ku||e, ||Kxau, and others are clearly insufficient to include them in any statistical calculations, they may still be relevant for purposes of etymological study and reconstruction. Below I list first the principal languages (and/or dialects) included in the statistical procedure, and then the list of auxiliary sources which will be consulted in the process of reconstructing wordlists for Proto-!Ui, Proto-Taa, and Proto-Tuu.

A. |Xam

Sources. This formerly widespread language became largely extinct even prior to the extensive field research of Dorothea Bleek in the first half of the 20th century; most of our knowledge on its grammar and lexicon comes from the archival records of Wilhelm Bleek and Lucy Lloyd, many of which were originally published in Bleek & Lloyd 1911 and later included into D. Bleek’s comparative dictionaries (Bleek 1929, 1956).

Dialects. Considering the overall expanse of the territories formerly populated by |Xam speakers and the fact that Bleek and Lloyd worked with a variety of informants (from Achterveld, Katkop, Strandberg, and other locations), dialectal diversity within the language must have been quite notable. However, precise differentiations are impossible without a meticulous study of the entire assembled text corpus. Lexicostatistical analysis of the data shows that there are relatively few Swadesh items transparently represented by two or more synonyms which could be thought of as representing different dialects; as for observed phonetic variation, it is not always clear when it should be ascribed to dialectal diversity or simply errors in transcription. For the purposes of the current study, we treat the entire Bleek-Lloyd corpus as a single “doculect”, while admitting that this is somewhat of a provisional simplification.

Quality. Transcription accuracy is always dubious, especially concerning the system of click accompaniments (see Traill 1995 for insightful comments on how to interpret various elements of Bleek and Lloyd’s transcription system for |Xam). Semantic glossing is frequently questionable as well, but at least in many cases it may be checked against the large assembled text corpus.

B. N||ng – N|uu

Sources. This is a large dialectal cluster which, unlike |Xam, is represented by several very distinct doculects from sources widely varying in space and time. This means that, for lexicostatistical purposes, it is possible and recommendable to build as many as three distinct wordlists: B.1 = ||Ng!ke (the dialect originally described by Dorothea Bleek; data published in Bleek 1929, 1956, and later separately in Bleek 2000), B.2 = †Khomani (the dialect originally described in Doke 1936 and Maingard 1937), B.3 = N|uu (the recently rediscovered variety spoken by several informants, with lexical data published in Crawhall 2004, Sands et al. 2007, Miller et al. 2009, Collins & Namaseb 2011; a complete Swadesh wordlist was kindly provided for the purposes of this study by Bonny Sands). For all of these dialects put together, we reserve the common name of N||ng as suggested in Güldemann 2017: 95.

Dialects. Unlike |Xam, the various attested dialects of this “macro-language” show quite a bit of lexical differentiation, though it is often difficult to tell how much of it is due to inaccurate semantic glossing, how much (especially in the case of N|uu) to very recent borrowings from other languages, and how much to gradual linguistic divergence after the original split of “Proto-N||ng”; for these reasons, as well as the relative incompleteness of the joint Doke/Maingard wordlist for †Khomani, any statistical discrepancies should be viewed with extreme caution.

Quality. Rather predictably, modern N|uu is one of the best transcribed representatives of Tuu; importantly, transcription quality in Doke 1936 and Maingard 1937 also seems superior to D. Bleek’s data (thus, both sources consistently mark the palatal click ʃ, which in most cases remains undistinguished from alveolar ! in Bleek’s records). Semantic glossing is assumed to be accurate for modern N|uu and can sometimes be checked against actual text examples for N||ng and †Khomani.

C. ||Xegwi

Sources. This language, geographically somewhat isolated from the rest of the !Ui continuum, is represented by at least three significantly different doculects, marked respectively as: (a) ||Xegwi-B — the earliest records collected by D. Bleek and published in Bleek 1929, 1956 (in her description the language is usually referred to as Batwa, a local Bantu term); (b) ||Xegwi-Z — as described by D. Ziervogel in a brief grammar sketch (Ziervogel 1955); (c) ||Xegwi-LH — as described by L. W. Lanham and D. P. Hallows in two short papers (Lanham & Hallows 1956a, 1956b).

Dialects. Judging by attested phonetic and lexical differences between the three doculects, a certain degree of dialectal diversity must have been present among ||Xegwi speakers. However, lexicostatistical discrepancies between the three sources are minimal (1–2 entries between ||Xegwi-Z and ||Xegwi-LH; slightly more between each of these and ||Xegwi-B, possibly because of less accurate semantic glossing in Bleek’s earlier records). Given the incompleteness of the sources (for ||Xegwi-Z and ||Xegwi-LH, data have to be extracted from grammar sketches and short text examples rather than actual vocabularies), it makes sense to merge them in one wordlist.

Quality. Transcription quality seems to be surprisingly adequate in the case of ||Xegwi-LH: for instance, Lanham and Hallows are among the first scholars to actually denote the presence of uvular phonemes and click accompaniments in any Khoisan language. Therefore, all data from ||Xegwi-B and ||Xegwi-Z, wherever possible, needs to be cross-checked against ||Xegwi-LH.

D. |'Auni

Sources. This language, which used to represent the westernmost spread area of Tuu, is known exclusively from records by Dorothea Bleek (Bleek 1937; lexical data also printed in Bleek 1929, 1956).

Dialects. Some dialectal variety may be identified from Bleek's records, as the equivalents for various meanings occasionally differ between the earliest ones, collected in 1911 and partially published in 1929, and the later ones, collected in 1936 and published in Bleek 1937 and Bleek 1956. It is, however, often difficult to establish whether these discrepancies (around 4–5 of them are found in items on the Swadesh list) represent true dialectal variation or inaccurate semantic glossing on the part of the researcher. Additionally, it is unclear if there are sufficient grounds to count the idiolect to which Bleek refers to as “Khatia” or “Xatia”, a very small amount of data for which were also collected by her in 1911 and published in Bleek 1956, as anything other than a minor sub-dialect of |'Auni. Finally, the occasional decision to regard |'Auni and |Haasi (see below) as dialects of a single language (e.g. in Glottolog 4.4) is hardly correct due to extremely significant lexical and grammatical differences between the two (e.g. around 20 mismatches on the Swadesh list).

Quality. Transcription quality is generally typical of D. Bleek's recordings for other Khoisan languages; external comparison raises serious doubts about the accuracy of click eflux transcription and slightly less serious about the same for click influxes.

E. |Haasi

Sources. This variety of Lower Nossob is solely known from records made by Robert Story of data from a single informant, Kabala (or Tatabesa), at the same Tweerivieren camp in 1936 where D. Bleek's data on |'Auni were collected; some of the |Haasi data were later published as part of Bleek 1956, but the complete manuscript did not officially see the light of day until Anthony Traill managed to rediscover and edit it for publication (Story 1999). Naturally, there is no dialectal variety to speak of here, but, as mentioned above, neither is there any reason to regard |Haasi as a bona fide “dialect” of |'Auni; both speech forms, as already noted by Traill in his preface to Story 1999, are more closely related to each other than to any other form of Tuu, yet both clearly have to be treated as different languages.

Quality. Although, in his own words, Story was a “complete amateur” and had no formal training in phonetics (Story 1999: 10), the overall quality of his transcription, at least at a rough glance, seems to be no worse than D. Bleek's or almost anybody else's at the time (e.g. he seems to have had a good ear for distinguishing between the palatal and alveolar clicks, with which quite a few other Khoisanologists seem to have struggled back then). The accuracy of his semantic notation can usually be confirmed by specific texts and phrases found in the manuscript. The worst problem is the scarceness of material: thus, as many as 40 Swadesh items cannot be recovered from extant data, which makes it impossible to offer reliable glottochronological datings for the moment of separation between |'Auni and |Haasi. That said, |Haasi data are of vital importance for attempting to at least partially reconstruct the basic lexicon of Proto-Nossob and, in turn, Proto-Tuu itself.

F. Taa (!Xóõ, Kakia, N|u||en)

Sources. Precisely three different varieties of Taa allow for the construction of more or less representative Swadesh wordlists. First and foremost among them is !Xóõ (more precisely, Lone

Tree !Xóõ) as represented in Anthony Traill’s now-classic and extensive dictionary of this particular dialect (Traill 1994, 2018). The other two are much older, dating back to D. Bleek’s brief research on the language of the “Masarwa” (a generic pejorative Bantu term for the San) of Kakia in 1913, and on the language of the N|u||en of Nausanabitz in 1920 (most of the data were subsequently published in Bleek 1929 and Bleek 1956). Both of these speech varieties seem to have become extinct and, to the best of my knowledge, are not directly identified with any of the still living dialectal varieties of !Xóõ (such as described, e.g., in Naumann 2014); concerning the latter, although some research has been carried out on them, no significant amounts of lexical data have been published to allow for a proper lexicostatistical comparison between them and Traill’s Lone Tree !Xóõ.

Dialects. Although all the three varieties of Taa for which it is possible to produce more or less complete Swadesh wordlists show up to about 20% of lexical discrepancies in these wordlists, which would, under normal circumstances, clearly speak of them as three different languages, the widely varying quality of recorded data does not allow to take these discrepancies at face value: Bleek herself admits that data on Kakia and N|u||en were collected in haste, and the probability of semantic and lexical inaccuracies in her records is fairly high. It is, therefore, possible that ultimately these two varieties are not nearly as distant from !Xóõ proper as are !Xóõ’s own 20 or so sub-dialects, tentatively classified in Naumann 2014 on the basis of some phonetic and grammatical isoglosses observed over the course of a general survey. In any case, at this time a detailed lexicostatistically based phylogeny of Taa languages and/or dialects is impossible due to lack of data; a tentative reconstruction of the Swadesh wordlist for Proto-Taa, based on all available evidence, is, however, somewhat within reach.

Quality. Lone Tree !Xóõ expectedly boasts the highest quality of phonetic (and probably semantic) accuracy among all South Khoisan languages, possibly second only to N|uu (for which, however, published data are far more limited) — all due to the extensive research of Anthony Traill. Nevertheless, the huge discrepancy between the quality of Traill’s data and everything else should not lead anyone into the fallacy of conflating Traill’s Lone Tree !Xóõ with Proto-Taa itself, at least not when lexical reconstruction is involved. In terms of phonetics, there is little, if anything, that data from Kakia or N|u||en could contribute in light of Traill’s clearly superior, and extremely detailed, description of Taa phonology (comparison with Bleek’s data shows plenty of unrecognized phonetic features and a lot of mistakes in the transcription of even the basic click influxes). But from a purely lexical point of view, there is no reason to *a priori* consider the Lone Tree !Xóõ equivalent for a particular meaning as more archaic than the corresponding Kakia or N|u||en equivalent whenever the two (or three) are clearly etymologically different from each other.

G. Other !Ui languages

Data from the following languages, unquestionably identifiable as separate linguistic units belonging to the !Ui group, may and should be used for etymological purposes (including reconstruction of Proto-!Ui basic lexicon) but is generally unusable for lexicostatistical goals, making a precise identification of their respective position on the !Ui tree somewhat difficult:

- (a) ||Kxau (small grammatical sketch, a few phrases, and a very short vocabulary in Meinhof 1929; all lexical data reprinted in Bleek 1956);
- (b) ||Ku||e (a small amount of lexical data collected by D. Bleek and published in Bleek 1956);
- (c) “Seroa” and “!Gǎ!ne”, both represented by short, old, and phonetically dubious collections of lexical data by T. Arbousset, C. F. Wuras, and H. Anders (all data reprinted in Bleek 1956).

Forms from some of these doculects will occasionally be quoted below, specifically as additional etymological support for particular reconstructions, but no systematic conclusions about their historical phonology or classification details shall be drawn.

Methodology

For the sake of this paper, I proceed from the following historical assumptions:

- (a) all of the languages listed above are genetically related within a single “Tuu” family;
- (b) all of those languages may be definitively and uncontroversially divided into no fewer and no more than three separate clusters — !Ui (|Xam, N!ng, ||Xegwi); Nossob (|’Auni, |Haasi); and Taa (!Xóǒ and all of its dialects as well as Bleek’s Kafia and N|u|en), each of these representing the result of divergence from its own intermediate protolanguage.

Convincing evidence for both of these assumptions, including (partial) regular phonetic correspondences and numerous sets of lexical and grammatical isoglosses, has been presented in numerous sources, from Bleek 1956 and Westphal 1962 to more modern research (e.g. Hastings 2001, Güldemann 2005b, Starostin 2008), and alternate scenarios, such as trying to explain similarities between !Ui and Taa as a result of areal convergence (a possibility not ruled out by such notable “splitters” in the field of Khoisan studies as E. O. J. Westphal), are unlikely and generally unwarranted.

What remains much less clear is the degree of relationship of these three clusters to each other, or even of some of the individual languages within these clusters to each other. While certain elements of consensus between the various classification schemes offered by researchers do emerge, such as, e.g., the understanding that |Xam and N!ng are closer to each other than to ||Xegwi, a particularly tricky issue rests with the |’Auni-|Haasi cluster, commonly referred to today as the “Lower Nossob”, or simply “Nossob”, languages. Here at least three conflicting schemes have been put forward:

- (a) E. O. J. Westphal (1971: 381) directly groups this cluster with the Taa languages, using the term “Taa” for the entire agglomeration; furthermore, as has already been mentioned, he has forever remained skeptical about the idea of a genetic connection between Taa and !Ui;
- (b) Oswin Köhler (1981: 469) counts the Nossob languages as a part of !Ui, considering them all related to Taa (which he calls “non-!Ui”) on a deeper level; this classification scheme has generally become more popular than Westphal’s until recently;
- (c) Tom Güldemann (2014) has partially reverted to Westphal’s model, arguing for a closer affinity between Nossob and Taa while at the same time not denying that both are ultimately genetically related to !Ui. His arguments are based on a number of lexical and grammatical isoglosses, as well as a strongly supported observation that the similarities between Nossob and !Ui are exaggerated because of extensive areal contact between |’Auni and N!ng (involving elements of bilingualism).

Out of these three, Güldemann is the only author who has actually published detailed linguistic argumentation in favor of his hypothesis, which may be one reason why it is currently accepted as the default phylogenetic scheme for Tuu in Glottolog. Nevertheless, due to the scarceness and sometimes dubious quality of the data, using selective lexical and grammatical arguments in this kind of linguistic investigation (the way it is done in Güldemann 2014) may not be totally free of bias, and it would be reasonable to take a more holistic approach to the matter, if at all possible. This is why an overall lexicostatistical survey, focusing on attested core basic lexicon for all the languages involved, would be a very useful addition to Güldemann’s methods of classification; and in the event of it producing different phyloge-

netic results from Güldemann 2014, analyzing the reasons for such a discrepancy could shed new light on both the historical relations between the various Tuu languages and the methodology of phylogenetic classification as a whole.

The actual results of an initial, preliminary survey based on 100-item Swadesh wordlists for all the languages listed above have already been published in Starostin 2013: 355; they showed that, although cognacy percentages between the Nossob languages and the various !Ui languages sometimes drop to around 46–48%, they are still consistently a little higher than the average percentages between Nossob and Taa, speaking in favor of Köhler's older classification rather than Güldemann's. However, there is a way to both correct and refine those results and make them more visually transparent by shifting from direct comparison of attested languages to comparing *reconstructed* wordlists — for Proto-!Ui, Proto-Nossob, and Proto-Taa, respectively. Condensing lexical evidence from a dozen languages into the shape of evidence from just three reconstructed proto-languages would be useful in helping clear away the “chaff” of identifiably recent innovations and borrowings, and would also make it easier to focus on the analysis of specific lexical isoglosses between the three branches in order to figure out which ones may have more weight for phylogenetic classification.

The general methodology for reconstructing proto-wordlists of the Swadesh type was already described in detail in several of this author's previous publications (Starostin 2013: 153–183, Starostin 2016) and, from a substantial point of view, needs no major modifications when applied to available Tuu material. Most of the specific challenges encountered along the way are of a technical nature — namely, scarceness and phonetic / semantic inaccuracies in the source data. These can sometimes be neutralized through careful scrutiny, but on the whole, of course, it should be well understood that the presented results are only as good as the data that currently support them, and are liable to change with each new significant publication of an additional data source (although, unfortunately, this is not likely for most of the languages involved in this study).

An important tripartite distinction could be introduced between *reconstructions*, *pseudo-reconstructions*, and *zero reconstructions* for each of the Swadesh items within each of the three subgroups. For the wordlist appendix below, the following rules are observed.

(a) A *reconstruction*, marked with an asterisk, is generated when cognates are attested in at least two separate doculects which do not represent close sub-dialects of a single language. In the case of !Ui, this means that the word has to be encountered at least in two out of three main clusters (|Xam; ||Ng!ke – †Khomani – N|uu; ||Xegwi), or, failing that, at least in one of them + one or more supporting languages whose data are not eligible for lexicostatistics (e.g. an isogloss between |Xam and ||Ku!e, or between N|uu and ||Kxau). Technically speaking, since |Xam and N!ng are closer to each other than to ||Xegwi, this does not allow to formally equate a “Proto-|Xam- N!ng” reconstruction with “Proto-!Ui” in the absence of a clear cognate in ||Xegwi; however, considering the scarceness of ||Xegwi data, we do not really have the luxury of downplaying |Xam – N!ng isoglosses, and for the sake of this particular phylogenetic study it seems reasonable to go along with a slightly broader understanding of “Proto-!Ui”.

Accordingly, in the case of Nossob languages “Proto-Nossob” is understood as the common invariant of cognates in |'Auni and |Haasi; in the case of Taa “Proto-Taa” is understood as an isogloss between !Xóõ and either Kakia or N|u!en (or all three).

(b) *Pseudo-reconstructions* can sometimes be substituted for actual reconstructions for both lexicostatistical and etymological purposes. Thus, if out of all the languages belonging to one of the three main subgroups of Taa, the Swadesh item in question is only attested in *one* language, and the form itself is not transparently identifiable as a recent morphological derivation or borrowing, there is a more-than-zero chance that it might actually be a direct reflex of the proto-

item (a very common situation for Nossob languages, where available data on !Auni are much more extensive than data on !Haasi, see ASHES, BARK, BELLY etc. below); naturally, this chance is increased even further if the form has credible external cognates in any of the other two branches.

If there are two or more non-cognate forms for the same equivalent in different languages and it is impossible to make a sound judgement on which one is the lexicostatistical archaism and which ones are the innovations, it is permissible to count them all as “technically synonymous” pseudo-reconstructions (see, e.g., BELLY or BIG in the !Ui list below), in the sense that each of them has a comparable chance of having expressed the required Swadesh meaning in the proto-language (this is more credible than the idea of “absolute” synonymy in the proto-language, with each daughter language retaining only one of the several earlier synonyms). Again, discovery of a potential cognate for one of these “technical synonyms” on the external level of comparison drastically increases its chances and almost (but not quite) raises the item’s status from pseudo-reconstruction to actual reconstruction.

(c) Finally, *zero reconstructions* — implying, among other things, that this particular item has to be excluded from lexicostatistical calculations — appear whenever the required item is either not found at all in any of the languages, or, if found in any of them, is transparently identifiable as a recent innovation or borrowing. In the list below, there are very few genuine zero reconstructions, since most of the Swadesh items are found to have some sort of equivalent in at least some of the discussed languages; the biggest problem is with a very small bunch of concepts whose “near-universality” does not properly apply to Tuu realities (e.g. FISH, notably absent in the area, or LEAF, seemingly a difficult concept for Tuu speakers which is usually expressed by borrowings).

Regarding the highest level of reconstruction (Proto-Tuu), we consider any Swadesh item to be formally reconstructible for Proto-Tuu if it is reconstructible in the exact same Swadesh meaning for both Proto-!Ui *and* Proto-Taa. The lower level reconstructions may be pseudo-reconstructions, i.e. an isogloss between !Xam and !Xóõ (or even an isogloss between !Xam and the far less reliably attested Kakia or N|u||en on the other end) may be taken as strong evidence for a Proto-Tuu reconstruction, unless there are additional obstacles to this interpretation (e.g. both forms may be easily interpreted as recent borrowings from a Khoe source). The Nossob languages, with their phylogenetic status not yet clearly resolved, are currently not very telling: it is extremely important to spot all the exclusive !Ui-Nossob and Taa-Nossob isoglosses, yet directly equating them with Proto-Tuu is impossible before the final conclusions on their position on the genealogical tree of Tuu languages.

With all possible Proto-!Ui, Proto-Nossob, Proto-Taa, and ultimately Proto-Tuu reconstructions on hand, the natural advantage is that it shall be much easier to not only calculate the distances between the specific branches, but also to analyze the possible classification alternatives in terms of individual shared archaisms and innovations, reducing the overall evidence to a small, but objectively attained, number of truly diagnostic etymologies. These results will be presented in the second part of the paper.

Notes on phonetic reconstruction in Tuu

Considering how much emphasis has been placed (and will continue to be placed) on the word “reconstruction” in this paper, some clarifications must be made about how we actually understand this term when applied to Tuu data. At the present state of our knowledge about Tuu languages as a whole, it is extremely difficult, if not downright impossible, to rigorously and rigidly apply the classic Neogrammarian methodology in order to elicit fully regular pho-

netic correspondences between the phonemic systems of these languages — mainly due to the relative scarceness of data from most of them, and to the generally poor transcription quality of those languages which are indeed represented by relatively large corpora (like |Xam or !'Auni). There is plenty of phonetic similarity between them, and there are enough recurring patterns of correspondences to usually (though not always) recognize etymological cognates, but a highly detailed system of correspondences which would fully cover all the subsystems of the complicated Tuu phonologies (click influxes, click effluxes, non-click consonants, vowels and their secondary features, tones, etc.) and reduce them to a parsimonious and typologically credible Proto-Tuu inventory at best requires a much huger research effort than is currently possible, and at worst may turn out to be objectively unreachable.

Nevertheless, even at this stage it is possible to operate on the level of what might be called “lax” reconstructions, along lines already suggested for Tuu languages in Starostin 2008, 2013. What this means is separating the phonological units of Tuu into categories which are found, based on comparative evidence, to be generally both more stable from a historical perspective and more consistently transcribed from a notational perspective — and those which seem to be more fluent over time, as well as less easily defined by inexperienced field workers. “Lax” reconstructions might then latch on to the more reliably established correspondences for the first category, while offering reasonable approximations (for instance, bluntly based on the majority rule) for the second. Such half-way reconstructions are always amendable if more high quality data come along or additional recurring patterns are confirmed statistically, but even without this they can still serve as proper historical evidence, provided that at least a certain “sound skeleton” has been recovered for them based on Neogrammarian-type correspondences.

According to my observations, the generally stable parts of phonological inventories in Tuu can be defined as (a) click influxes; (b) non-click consonants, especially in word-initial position; and, to a slightly lesser degree, (c) main root vowels (not including vocalic codas, correspondences between which are often chaotic, possibly because they represent variable morphological add-ons). The least stable parts, in addition to vocalic codas, are tones (if only because prosody is not marked consistently and reliably in any of the older sources) and click effluxes — which often show tremendous variation not just between different languages, but even between closely related dialects or sub-dialects of the same language. Below I adduce several important notes on each of these subseries, additionally referring the reader to my earlier and more detailed, but also sometimes outdated, observations on the comparative phonology of Tuu as published in Starostin 2008.

A. *Click influxes*. Correspondences between these segments are more often than not regular and trivial (one-to-one), but there are some important exceptions. The principal correlations are listed in Table 1; for some extra details (largely irrelevant when applied exclusively to the 100-item wordlist) see Starostin 2008: 365–370.

	Xam	N!ng	Xegwi	!'Auni	!Xóǒ
*θ	θ	θ	θ	θ	θ
*					
*ǰ	!	ǰ	λ / š	ǰ	ǰ
*!	!	!	∅	!	!
*					
*Ǿ	!		!	ǰ	ǰ

Table 1. Principal correspondences between click influxes across major Tuu languages

Notes.

(1) Labial click (*ǀ): see Starostin 2008: 366 on several examples where labial clicks in Taa may correspond to lateral clicks in !Ui, perhaps indicating secondary labialization. It is still unclear whether this correspondence is truly regular or if all the listed examples are just accidental resemblances; in any case, none of them are relevant to the data subset of the 100-item wordlist.

(2) Dental click (*ǃ): see *ibid.* on such specific correspondences as Taa *ǃq(?) = !Ui *c(?) and Taa *ǃn = !Ui *d. Examples for these are somewhat more reliable than for (1), but, once again, they are only encountered outside the Swadesh wordlist.

(3) Palatal click (*ǂ). This is the least stable of all click influxes in Taa, and it deserves more detailed commentary. First, in such extinct languages as !Xam and (maybe) some of the dialects of N!ng, such as the Bleek-transcribed ||N!ng!ke, it seems to have merged with the alveolar click (*ǂ → !), see below examples such as DOG, EAR, EGG etc.). The reason why I suspect it must have been a real diachronic development rather than a simple transcriptional error is that there are quite a few entries in !Xam which have been transcribed, both by Wilhelm Bleek and Lucy Lloyd, with an initial ǂ (cf. !Xam ǂenn 'to know', ǂa 'to kick', ǂxoa 'elephant' etc.), but many, if not most, of them look like relatively recent borrowings from a Khoe source⁴. This would imply that after the original palatal click had shifted to a different manner of articulation (perhaps merging with the alveolar click or becoming so close to it as to become indistinguishable for the early scholars of Khoisan⁵), it may have been reintroduced into the language/s/ along with lexical loans from their Khoe neighbors.

Second, in ||Xegwi the palatal click undergoes a unique development, shifting toward a non-click lateral affricate articulation. The regular development seems to be *ǂ → ǂ̣ (see DOG, EAR, EGG below), but occasionally post-alveolar fricative reflexes (č, š) are observed as well; this seems to happen when the click has a uvular efflux (cf. N!uu ǂqōē 'short' = ||Xegwi-Z čwe id.; N!uu ǂq^hoe 'wind' = ||Xegwi-LH šwe: id.). Unfortunately, scarceness of available ||Xegwi data prevents us from being able to fully describe the picture here, which must have been typologically somewhat similar to the well-studied behavior of palatal clicks in Eastern Kalahari Khoe languages (Vossen 1997: 285–288).

(4) Alveolar click and lateral click (*ǁ, *ǁ̣). Both of these are typically quite stable, but the alveolar click undergoes seemingly regular deletion in ||Xegwi as well (*ǁui 'person' → ||Xegwi kwi, etc.), again, parallel to similar developments in Kalahari Khoe.

(5) The “sixth click influx” (provisionally marked as *ǂ̣ for lack of a better idea⁶). This reflects the unusual, but seemingly recurrent correspondence “!Xam ! : N!ng || : ||Xegwi ! : (?) Nosob ǂ : Taa ǂ”, established on the data of several basic items on the Swadesh list (BONE, ONE, RED, also FOOT in !Ui) as well as additional basic lexicon (e.g. the root for ‘female breast / milk’, listed in Starostin 2008: 368). The evidence for this extra influx is not overwhelming, but too strong to be brushed away as a mix of accidental lookalikes and incorrect transcriptions; in particular, given the regular deletion of the plain alveolar click in ||Xegwi, it is the only way to account for those cases in which ||Xegwi lexical items still feature the alveolar click (and cannot

⁴ Bonny Sands suggests that the loans may have come specifically from Korana (Sands 2014: 13).

⁵ In this respect, it may be instructive to recall a typological parallel in which the original Ju (North Khoisan) palatal click *ǂ has shifted to a retroflex articulation (!) in Ekoka !Xun (König & Heine 2001: 22–23), already after the original retroflex click *ǂ̣ had merged with lateral *ǁ in that same dialect. Could something of the sort actually have taken place in some of the now-extinct Tuu languages?

⁶ The symbol ǂ̣ is actually borrowed from Clement Doke’s ingenious, but forgotten alphabet for click consonants, where it was reserved for the unvoiced alveolar click (now commonly marked as !).

be explained away as borrowings). Postulation of a phonologically distinct sixth click influx for Proto-!Ui and Proto-Tuu would make these protolanguages typologically unique (no living or attested extinct Khoisan language has more than five), but not theoretically impossible; more work on available material is necessary to understand whether the observed correspondence should be truly traced back to a separate phonological contrast or whether it may be explained by a conditioned split.

B. *Click effluxes*. Very few Tuu languages can be said to have adequate descriptions of their complicated click accompaniment systems. The best ones have arguably been produced by Traill for !Xóǒ (up to 19 different effluxes per influx), Miller et al. for N|uu (up to 10 different effluxes per influx), and by Lanham and Hallows for ||Xegwi (up to 7 different effluxes per influx). Even these descriptions may not be completely accurate and finalized in terms of recognized contrasts, and observed correspondences between different languages are by no means trivial.

Our current “lax” strategy on the matter is simple: for Proto-!Ui and Proto-Taa, unless there is a very strong individual argument about the secondary nature of these effluxes, we provisionally accept the efflux in N|uu and in !Xóǒ (respectively) as representing the proto-state — simply because any discrepancy between these languages and the earlier described ones may be theoretically attributed to incorrect transcriptions in older sources (where the same word may very often be found transcribed in multiple variants with different click effluxes). If this tactical decision somehow contradicts the majority rule, i.e., for instance, the N|uu click efflux is not the same as the efflux in the majority of other !Ui reflexes, such a situation deserves detailed individual analysis⁷.

C. *Non-click consonants*. A staggeringly low percentage of either Proto-!Ui or Proto-Taa Swadesh items are reliably reconstructible with a word-initial non-click consonant (approximately 14–15 items on the Proto-!Ui wordlist and 18–20 items for Proto-Taa), which goes to show how thoroughly integral click phonemes are to these languages (for comparison, the corresponding number for Proto-Ju, even though Ju languages have the second most complex inventory of click phonemes after Tuu, is no fewer than 35 items out of 100). This does not mean that the Proto-Taa system of non-click consonants was necessarily modest — Traill lists more than 40 such consonants for !Xóǒ, of which only very few can be reliably proven as secondary — but it does mean that the issue of an accurate reconstruction of this sub-system for Proto-Taa is not particularly relevant for our current task.

Phonemes encountered in basic lexicon items include *t- (HEAR, LIE), *k- (ALL), *s- (BITE, COME, FAT), the ejective velar affricate *kʰ- (DRINK), and the alveolar affricates *ʃ- (FLY) and *cʰ- (EYE), though for these last two phonemes evidence is more marginal and problematic. Correspondences for the others are largely trivial (arguably the most serious phonetic change is from *t- to palatal *c- in N|ng), though see notes on BITE for a possible affricativization scenario for *s- in certain contexts. Not a single complex consonantal cluster, such as *tkʰ, etc., is reconstructible for this particular subset of the basic lexicon in any of the daughter branches of Tuu.

⁷ It should be kept in mind that click efflux articulation in Tuu, as well as other Khoisan languages, may occasionally be correlated with secondary features of vowel articulation, such as nasalization, pharyngealization, glottalization, and breathiness — both “genuinely” (if vocalic articulation exerts assimilative influence on the efflux, or vice versa) and “virtually” (if, in one of the less than accurate sources, a vocalic feature is transcriptionally mistaken for a back closure release, or vice versa). Unfortunately, secondary vocalic features are quite inconsistently marked in older sources.

D. *Vowels and codas*. Reconstruction of the Proto-!Ui, Proto-Taa, and especially Proto-Tuu systems of vowels and vocalic/consonantal syllabic codas is extremely difficult due to huge amounts of variation, which should be attributed not only to phonetic change (or pseudo-phonetic change, reflecting inaccurate transcription) but also to morphological variation, as the exact same nominal, adjectival, or verbal root may frequently be encountered in different languages (or even within the same language) in combination with different suffixal components – noun class markers, agreement morphemes, or various other clitical elements fused with the root and not recognized as separate morphemes.

The *main vowels* in Tuu languages, as follows from reliable modern data on Nluu and !Xóð, are typically restricted to three unrestricted phonemic units (*a, o, u*), occurring freely and frequently after any consonants; and two highly restricted units (front vowels *e, i*), whose occurrences after click phonemes are exceedingly rare, but who are somewhat more frequently met after non-click phonemes. The original picture may have been more complicated, as there are numerous cases in which the vowel *a* in Taa corresponds to either *e* or *o* in !Ui languages (see examples in Starostin 2008: 372); it is still unclear if such situations reflect additional original phonemes (such as **ε* and **ɔ*) or the results of phonetic contraction of different morphological variants (for a good example, see notes on FIRE below).

The precise inventory of Proto-Tuu *codas* (i.e. second morae of nominal and verbal word forms, which are often morphologically detachable even on the synchronic level, or may be shown to have been fossilized through external comparison) cannot be determined at the moment; on the whole, relatively few bimoraic sequences may be reliably reconstructed by comparing !Ui, Nossob, and Taa data. Given the fact that only !Xóð yields itself relatively well to detailed morphophonological analysis (in Nluu, most of the old derivational morphemes seem to have lost their productivity, and data on all other languages are antiquated and unreliable), reconstruction of nominal and verbal morphological elements for Proto-Tuu may turn out to be an even more challenging task than the reconstruction of its click system. Consequently, in the current paper, the emphasis is always on checking whether a bisegmental (initial click or non-click consonant + main vowel) sequence may be identified as the original root morpheme for Proto-!Ui, Proto-Nossob, Proto-Taa, and, ultimately, Proto-Tuu: by default, discrepancies between codas are provisionally written off as reflecting morphological variation, either already present on the Proto-Tuu level or arising independently in one or more branches after the split of the proto-language.

Notes on transcription

The transcriptional system used in this paper generally follows the transcriptional standard which is currently employed in the Global Lexicostatistical Database and is itself essentially based on IPA, but with a few important modifications.

(1) Clicks: following the system adopted in Vossen 1997, nasalized clicks are transcribed with a superscript tilde sign ($\tilde{\theta}$, \tilde{l} , etc.) while voiced clicks have a subscript tilde (θ , l , etc.).

(2) Affricates: instead of IPA's digraphic combinations, single letters are used to denote alveolar (*c, ʒ*) and palatal (*ç, ʝ*) affricates.

(3) For morphological segmentation, the hyphen sign is used to separate root morphemes from suffixes (*ku-ka*, etc.), while the equation sign is used to separate roots from prefixal components (e.g. !Auni *si=|u* 'bird', etc.).

For a more detailed description of the transcription system, including notes on transliteration of data from old sources, see Starostin 2015.

Appendix. Comparative analysis of Tuu basic lexicon (Items 1–50)

In this Appendix, I list the results of intermediate and Proto-Tuu reconstructions for the first (alphabetically) 50 items on the Swadesh wordlist (more or less closely following the semantic specifications set out in Kassian et al. 2010). Each entry is structured as follows:

(1) Name of the item, together with a formal notation of the presence / absence of lexicostatistical parallels between the three branches: e.g. [!Ui + Taa] [- Nossob] means that the reconstructions for Proto-!Ui and Proto-Taa are cognate, whereas the reconstruction for Proto-Nossob is not (this also includes pseudo-reconstructions). Sometimes, even when all three branches reflect the same root, two out of three may be more tightly connected, for instance, sharing common morphological formations (suffixes, etc.). Such extra proximity is indicated with additional parentheses, e.g. [!Ui + [Nossob + Taa]]: it offers additional evidence for phylogenetic classification. If there are no matches whatsoever between any of the three branches, the word is marked with [-].

(2-4) Reconstructions for Proto-!Ui, Proto-Nossob, and Proto-Taa, accompanied with a list of most of the attested reflexes. If the onomasiological reconstruction is equivocal, two or more roots may be listed instead as (a), (b), etc. The \diamond sign separates listed data from comments on the reconstructions⁸. Note that the Appendix does not necessarily list *all* the attested forms corresponding to the Swadesh items in question, but mainly those that justify the reconstruction. For complete lexicostatistical lists, the reader is advised to refer to the South Khoisan (!Ui and Taa) databases that are separately available online at the Global Lexicostatistical Database (Starostin 2011–2021).

(5) Proto-Tuu reconstruction (where it is at all possible). For reasons described above (in the “Notes on phonetic reconstruction” section), we do not systematically list Tuu protoforms, but rather use the notation “Tuu+” to indicate credible lexicostatistical isoglosses between !Ui and/or Nossob and Taa which almost certainly go back to a common Tuu protoform, and the notation “Tuu–” to indicate the lack of such isoglosses. Note that “Tuu–” also marks situations where one of the branches may have an etymological cognate in the other, but since the meanings are different, this does not qualify as a proper lexicostatistical match (e.g. BIG, etc.).

1. ALL [!Ui + Taa] [-Nossob]

- !Ui: ***ku** (|Xam *ku*; ||Ng!ke *kwa*; Seroa *ku*). \diamond Attestation in ||Ng!ke is somewhat dubious (the word is only found in the earlier source Bleek 1929, not in Bleek 1956), but the |Xam entry is hardly questionable. Isolated equivalents (a) in N|uu: *huni-ki* (= †Kho-M *huni-ǂe*); (b) in ||Xegwi-LH: $\zeta'i \sim \zeta'i$ (only found as part of composite pronominal stems *i- $\zeta'i$* ‘we all’, *u- $\zeta'i$* ~ *u- $\zeta'i$* ‘you all’).
- Nossob: |’Auni *bà* (?). \diamond Cf. the example in Bleek 1956: 13: *tuku bà su !^hǂbati* “men shall all return”. Not clear if this semantic glossing should be trusted, especially given that the word *||ani* is also occasionally glossed as ‘all’, e.g. *ku totos ||ani* “all the people” (on the other hand, the primary meaning of *||ani* is probably ‘many’, cf. below).
- Taa: ***kU-ka**^f (!Xóǎ *kô: kàǎ*, Kakia *ku-ka*^f ~ *ku-ka*). \diamond Clearly a compound, but it is hard to delineate the individual meaning and function of each component. For N|u|en, the

⁸ These comments are sometimes identical with notes on specific items and reconstructions which have already been published as part of the !Ui and Taa databases at the Global Lexicostatistical Database (Starostin 2011–2021). However, the present paper also adds new details and observations that are relevant for reconstruction purposes, while at the same time omitting a large amount of synchronic information on the actual South Khoisan forms which may be found in the database notes.

only attested equivalent is *||árrri*, the same word as ‘many’ q.v.; it is impossible to tell if both meanings truly merged in the same word or if this is just a case of inaccurate semantic glossing.

- Tuu+: A clear isogloss between at least |Xam (+ Seroa) and Taa, allowing to reconstruct **ku* as a basic Tuu morpheme for ‘all’. Specific evolution of this meaning in various daughter languages and its correlation with the related meaning ‘many’ may be obscured by inadequate glossing and insufficient contextual data.

2. ASHES [!Ui + Taa] [-Nossob]

- !Ui: **!qui* (|Xam *!úi ~ !úí*, †Kho-D *!wí*, N|uu *!qui*). ◇ Not attested in ||Xegwi or any of the minor sources. Perhaps phonetically identical with |Xam *!ku:i* ‘to burn, smart, pain’ (Bleek 1956: 449), but without data from other sources it is premature to suggest semantic derivation (may simply be a case of homophony or close phonetic similarity).
- Nossob: |’Auni *!ʰana*. ◇ Phonetically similar to !Xóõ *||q’âna* ‘dirt, rubbish, rust’, but the click influx correspondence would be unprecedented (unless the |’Auni form is inaccurate).
- Taa: (a) !Xóõ *ƒòa*, Kafia *||wa;* (b) N|u||en *!’wi*. ◇ Technically, the form in !Xóõ is more reliable than item (b), and its distribution is confirmed by the parallel in Kafia (with a mistranscribed lateral click, cf. ‘bone’, etc., below). However, the word is also phonetically identical with the widely distributed Proto-Kalahari Khoe root **ƒoa* ‘ashes’ (Vossen 1997: 417), and the lack of etymological parallels in !Ui strongly suggests that we are simply dealing with one of !Xóõ’s many borrowings from the neighboring |Gui. In this light, the form attested by D. Bleek in N|u||en looks more trustworthy as a potential archaism.
- Tuu+: The isogloss between !Ui and N|u||en (West Taa) strongly suggests Proto-Tuu **!qui* ‘ashes’, replaced in !Xóõ by a borrowing and not found in the Nossob subgroup.

3. BARK [-]

- !Ui: Not reconstructible. ◇ The word is attested consistently only within the N|uu cluster, where all forms are identifiable as borrowed from Kalahari Khoe (cf. PKK **||x’ũ* ‘bark’ in Vossen 1997: 421): †Kho-D *||x’ũŋ*, N|uu *||x’ũ:-si*. The only other known form is ||Ng!ke *!o;* somewhat dubious because it is not backed by any textual examples.
- Nossob: |’Auni *||õ:*. ◇ This could actually be the same word as †Kho-D *||x’ũŋ*, etc., i.e. also a borrowing from Kalahari Khoe, although Bleek’s transcription of the click efflux (zero instead of expected *-x’-*) would seem to contradict this.
- Taa: (a) !Xóõ *gú-le* (pl. *gú-n*), Kafia *gu-le;* (b) N|u||en *!um*. ◇ The !Xóõ form has obvious parallels in the |Gui-||Gana cluster of Kalahari Khoe (*gure* ‘bark’ in Tanaka 1978: 10), but in this case, the word seems to be exclusive for that particular cluster rather than reconstructible for PKK, implying possible borrowing from Taa rather than vice versa.
- Tuu-: No proper isoglosses between the three clusters, and the word itself is formally not reconstructible. Its frequent re-borrowing from Kalahari Khoe indicates that the concept itself is not very stable in Tuu languages.

4. BELLY [-]

- !Ui: (a) |Xam *!áu-tu;* (b) ||Ng!ke *||x’ã;* N|uu *||x’ã*. ◇ In most languages, available data do not allow to perfectly distinguish between the meanings ‘belly’ and ‘stomach’, although

at least W. Bleek's notes on |Xam suggest that *!áú-tu* 'belly' may have been opposed to */oa^f* 'stomach'.

- Nossob: |'Auni *!ai*.
- Taa: (a) *!Xóõ !^húma*; (b) N|u|en *!a:ban*.
- Tuu-: Not properly reconstructible. \diamond 'Belly' / 'abdomen' as a concept referring to the external part of the body seems to be fairly unstable in Tuu, with each individual language essentially having its own equivalent (this assuming that the semantic interpretation in older, uncheckable sources actually holds water). 'Internal belly' = 'stomach, bowels' is actually more stable: Proto-Tuu *|oa- can be reliably reconstructed based on the correlation between |Xam */oa^f* 'stomach' and *!Xóõ !^hóã^h* 'innards, bowels, stomach' (cf. also, perhaps, ||Xegwi-Z *!^hu-ga*: 'stomach', ||Xegwi-B *|u-bwa* id., although the second syllable in each of these forms remains unexplained).

5. BIG [!Ui + Taa] [-Nossob]

- !Ui: (a) |Xam *!ui-ya*; (b) N|uu *!xo:*; (c) ||Xegwi-Z *!xeya* ~ *!x'eya*, ||Xegwi-LH *!x'e*. \diamond Not properly reconstructible. The adjectival meaning 'big' in general is unstable and its equivalents seem to be easily reinvented from various verbal stems, e.g. |Xam *!ui-ya* is most likely derived from *!ui* 'to grow'. N|uu *!xo:* is clearly the same as |Xam *!xo:* 'upright, tall', but this does not guarantee that 'big' was the original semantics.
- Nossob: *o-si ~ *u-si (|'Auni *ús* ~ *ú:si* ~ *ú:ši*, |Haasi *â-si*). \diamond For |Haasi, Story also records usage of *!xwa:* 'big' as a free synonym; this may actually be a borrowing from N|uu.
- Taa: *!xa-(i): *!Xóõ !xa-*, Kafia *!xai*, N|u|en *!xai*.
- Tuu-: Since !Ui *o : Taa *a is a recurrent correspondence (possibly indicative of a special Proto-Tuu phoneme such as *ɔ), it is formally admissible to postulate Proto-Tuu *!xɔ- 'big' on the basis of the isogloss between Proto-Taa *!xa- and N|uu *!xo:* (+ |Xam *!xo:* 'tall?'). This is, however, not a perfect onomasiological match in light of the overall instability of the concept and uncertainties about specific semantic glossing in separate languages.

6. BIRD [!Ui + [Nossob + Taa]]

- !Ui: *|q^hui (||Ng!ke */wí* ~ */wi:*, †Kho-M */wi-si*, N|uu */q^hui-si*, ||Xegwi-Z *!^hwi*). \diamond Reconstruction based upon the presumably accurate efflux transcription in N|uu. The original root seems to have been narrowed down to the meaning 'vulture' in |Xam (*/wi:*), whereas two innovative forms are attested for 'bird': (a) *x'arri* ~ *x'ãnni*, usually glossed as 'little bird'; this is clearly related to Proto-Khoekhoe *k'ani 'bird' (ironically, itself ← Proto-Khoe *k'ani 'vulture', see Vossen 1997: 441) and likely reflects a recent borrowing; (b) *!errri-tãn-* 'large bird', transparently derived from *!errri(ya)* 'feather' and thus also clearly innovative.
- Nossob: *si=|u (|'Auni *si=|u*, |Haasi *si=|ǎ:*). \diamond Reconstruction of voiced efflux and *-u is provisional (largely based on external data). Initial *si=* is a nominal prefix commonly observed in other words as well (nothing to do with the copula *si*; possibly the same singulative marker as in N|uu */q^hui-si*, etc., only prefixed rather than suffixed?). Note that Bleek also lists |'Auni */o:* as a free synonym; this may be a prefix-less variant of the same stem, perhaps from a different dialect since in this shape, the word is actually phonetically closer to the variant in |Haasi.
- Taa: *|u(ǃ)- (*!Xóõ !^hú^hu*, pl. *!^hú^hǎ-tê*, Kafia *šǃ=|u*, N|u|en *si=|óu*). \diamond Kafia *šǃ=|u* is probably a typo for *šǃ=|u. It is notable that Kafia and N|u|en both share the prefix *si-* with Nossob forms; *!Xóõ*, however, shows no traces of it.

- Tuu+: Nossob and Taa forms are pretty much identical. The question of how they tie together with !Ui *|q^hui is more problematic, but etymological identity is possible assuming that (a) *-i is a fossilized class marker and (b) the aspirated uvular efflux in N|uu is somehow correlated with strident vowel articulation in Taa (in any case, there are additional examples where uvular efflux articulation in !Ui correlates to a lack thereof in Taa, cf. ‘horn’, etc.). We may tentatively reconstruct Proto-Taa *|u^ʕ- or even *|q^(h)u^ʕ- to account for this alignment.

7. BITE [|!Ui + Nossob] + Taa]

- !Ui: *c'i (|Xam *c'i* ~ *c'i*, ||Ng!ke *ci* ~ *c'i*, †Kho-M, †Kho-D *c'i*, N|uu *c'i*, ||Xegwi-Z *ci*, ||Xegwi-LH *c'i*).
- Nossob: |Haasi *c'i*. ◊ In !'Auni, the form *c'i* is only attested by Bleek in the meaning ‘to ache’, but given that the polysemy ‘bite / ache’ is also attested in †Kho, it is possible that this was the proper equivalent in !'Auni as well.
- Taa: *siʔ- (!Xóõ *síʔi*, N|u||en *ce-ya*). ◊ The N|u||en form is the same as the !Xóõ variable stem *siʔ-JV*. Not attested in Kafia.
- Tuu+: *siʔi may be reliably reconstructed as the original root. !Ui and Nossob forms seem to share the phonetic shift *siʔi → *sʔi → *c'i, in which the intervocalic glottal stop fused with the word-initial sibilant and turned it into a glottalized affricate. In some of the varieties of Taa and Nossob, the same root also serves as the derivational basis for ‘snake’ (see below).

8. BLACK [-]

- !Ui: *!(^h)oe (|Xam *!we:n* ~ *!wèŋ*, ||Ng!ke *!we* ~ *!oe*, N|uu *!^hoe*, ||Xegwi-Z *čwa* ~ *nčwa*, ||Xegwi-LH *žwa*: ~ *žwã*). ◊ The forms in |Xam and the N|uu cluster are clearly related (nasal coda in |Xam is likely of suffixal origin). Relation of these forms to ||Xegwi (*n*)*čwa* or *žwa*: is less certain, but a probable scenario is [1] regular deletion of initial *!- (*!oe → *koe) with [2] subsequent palatalization before a front vowel (→ *čoe) and [3] lowering of the diphthong (→ *čwa). For [1], see TAIL, TWO, WATER below; for [3], cf. ‘one’; no clear examples of [2], but no contradictory cases either. For now, we may count all these forms as cognates.
- Nossob: |Haasi *||e*. ◊ Not attested in !'Auni.
- Taa: *ʔaʔ- (!Xóõ *ʔáʔ-ŋa*, N|u||en *ʔa-na*). ◊ Same root in !Xóõ *ʔā-be* ‘black person’; *-ŋa* is a common adjectival suffix also encountered in other color terms (see RED, WHITE below).
- Tuu-: Not reconstructible. All three branches have their own equivalents.

9. BLOOD [|!Ui + Nossob] [-Taa]

- !Ui: *|xau (|Xam *||xáú-ka* ~ *||xáú-kən* ~ *||xau-ki*, ||Ng!ke *||xau*, N|uu *||xau-ke*). ◊ Solid reconstruction. Probably not related to ||Xegwi-LH *λ'ēũ*, which should reflect something like *ʔ'āũ, without any clear external parallels (phonetic similarity to Proto-Khoe *ʔao ‘heart’ is likely accidental, since click effluxes do not match and semantic connections between ‘blood’ and ‘heart’ are not particularly common in the Khoisan area).
- Nossob: *|xau (!'Auni *||xau(?)u*, |Haasi *||xau*).
- Taa: *|a^ʕ (!Xóõ *ĩá^ʕ* ~ *ĩá^ʕm*, Kafia *ĩá^ʕa*, N|u||en *ĩá^ʕa*). ◊ Provenance of *-m* in *ĩá^ʕm* is unclear; perhaps the result of morphological reanalysis of the plural form *ĩá^ʕ-ma-tê*.
- Tuu-: Not reconstructible. ◊ An obvious isogloss between !Ui and Nossob, on one hand, and Taa, on the other. It is curious that in both of D. Bleek’s dictionaries, she re-

cords a !Ui-like form for ‘blood’ for Kacia: $\|x\tilde{a}\tilde{u}$ (1929), $\|x\tilde{a}\tilde{u}^{\text{f}}$ (1956). However, it is not confirmed by textual examples, not distinguished semantically from $\tilde{!}a^{\text{f}}a$, and is clearly not the principal word for ‘blood’ in Taa as a whole. It may be a borrowing from some variety of !Ui (which is hard to confirm without a systemic analysis of the entire corpus) or, if it is some sort of archaic retention in limited (bound?) contexts, it could be a valuable indication that Common Taa $*\tilde{!}a^{\text{f}}$ is innovative.

10. BONE [!Ui + Taa]

- !Ui: $*\mathcal{C}^{(\circ)}a$ (|Xam *!wá*, ||Xegwi-LH *!a*). ◊ Correspondences between |Xam and ||Xegwi are non-trivial, but regular, reflecting the “sixth click” and extra labialisation in |Xam. A different equivalent for ‘bone’ is seen in the N|uu cluster: ||Ng!ke *||abba*, N|uu *||aba*, evidently cognate with |Xam *||abba* ‘a piece of eland’s bone that forms part of the completed arrow’. It must be noted, however, that Bleek transcribes the plural form for ||Ng!ke *||abba* as *||a||a* (reduplication is typically indicative of plurality), and that similar forms are also found in N|uu compound plurals, e.g. $\dagger q^h a$: *||ai-ke* ‘chest bones’. This may indicate suppletivism and preservation of the original equivalent for ‘bone’ in the collective / plural forms. Since the expected reflex of $*\mathcal{C}^{(\circ)}a$ in N|uu would indeed be *||a*, such a solution is quite likely; it must be noted, however, that *||a* and *||abba*, despite phonetic similarity, can hardly represent the same root due to having different clicks in |Xam.
- Nossob: Not attested in either |’Auni or |Haasi.
- Taa: $*\tilde{!}a$ (!Xóõ $\dagger\tilde{a}$: /poss./, $\dagger\tilde{a}$: /alien./, Kacia *||a*, N|u||en $\dagger\tilde{a}$). ◊ Lateral click in Kacia is probably mistranscribed, as in many other similar cases.
- Tuu+: !Ui $*\mathcal{C}^{(\circ)}a$ and Taa $*\tilde{!}a$ represent a solid etymological and lexicostatistical match; the only discrepancy is nasalization in Taa, which may ultimately go back to a suffixal extension ($*\dagger a\text{-}\eta$).

11. BREAST (CHEST) [!Ui + Taa] [-Nossob]

- !Ui: (?) $*\tilde{!}u\text{in}$ (|Xam *||wain-tu*; ||Ng!ke *||woen* ~ *||woin-tu*; N|uu *||ũĩ-ɕu*, ||Ku||e *||ɕin-tu*). ◊ This word is clearly distinct from the word for ‘female breast / milk’ (|Xam *!hwai*, N|uu *||hũĩ*, etc.), and its semantic properties in individual languages are not always clearly distinguishable from those of close synonyms, e.g. |Xam *||axu* ‘chest’ (possibly the same as *||axu* ~ *||ãxu* ‘side’) or N|uu $\dagger q^h a$: ‘sternum’, ‘breastbone’. The latter directly corresponds to ||Xegwi-Z/LH *ša-gu* ‘chest’ (see the exact same phonetic correspondence in ‘wind’ below), which makes the ||Xegwi form a less probable candidate for Proto-!Ui status (i.e. we reconstruct an original semantic opposition between $*\tilde{!}u\text{in}$ ‘chest’ and $*\dagger q^h a$ ‘breastbone’, with both meanings probably merged in one in ||Xegwi).
- Nossob: |’Auni *†an* ~ *†an*. ◊ Distinct from *||ẽi-si* ‘female breast’.
- Taa: Kacia *||am*. ◊ All three varieties of Taa have different equivalents, of which !Xóõ $\tilde{!}u$: ‘chest’ (distinct from $\dagger q^h \tilde{e}$: ‘female breast’) phonetically coincides with Proto-Khoe $*\tilde{!}u$ ‘chest’ (Vossen 1997: 426) and quite likely represents a borrowing from the |Gwi-||Gana cluster; and N|u||en *||u* ‘chest’ is either a typo for *||u* (see a similar case for BIRD above) or, alternately, could be compared with !Xóõ *||úi* ‘breastbone’. This leaves Kacia *||am* (distinct from *||xa:n-sa* ‘female breasts’) as the only form for which it is difficult to suggest a secondary origin.
- Tuu+: The correspondence between !Ui $*\tilde{!}u\text{in}$ and Kacia (= Proto-Taa?) *||am* is almost exactly the same as in the word for ‘liver’ (see below), likely reflecting Proto-Tuu $*\tilde{!}u\text{-}$

with different suffixal extensions (!Ui *-iŋ, Taa */a/m). !'Auni *ʃan* is incompatible with this stem.

12. BURN (tr.) [?]

- !Ui: *||a (|Xam ||a(:) ~ ||e(:), ||Ng!ke ||a ~ ||e(:)). ◊ Although the forms by themselves are glossed as intransitive in Bleek's sources, textual examples clearly confirm transitive use as well, e. g. ||ōĩ ʃe sa, ha ||a ŋ "the sun comes, it burns me" (Bleek 1956: 545), etc. For modern N|uu, Sands et al. 2006 give !xao as the main equivalent; this root is listed for ||Ng!ke as !xau 'to kindle, make or light a fire', and its cognates in the Nossob languages have the same semantics (!'Auni !xau 'to light (fire)', |Haasi !xau 'to kindle'). Overall, the data are insufficient to reach a certain conclusion, but it is quite possible that this word, originally only taking 'fire' as its object, has widened its scope in N|uu. Examples of transitive usage also attested for ||Ku||e ||a: 'burn'. In ||Xegwi, ||a is only attested in the meaning 'to cook' by Bleek; no other equivalents are known for the meaning 'to burn' in available sources. On the whole, there are sufficient reasons to think that both 'burn (tr.)' and 'burn (intr.)' in Proto-!Ui were expressed by the same root *||a (|Xam and ||Ng!ke ||e represent a secondary morphophonological variant, probably fused with an agreement marker).
- Nossob: Not reconstructible. ◊ For !'Auni, the only attested form is |á 'to burn, light a fire, roast', with one accompanying example: |á n |i 'light the fire', meaning that the semantics could have actually been 'light, kindle'. For |Haasi, Story lists the form ||ɔ: 'to burn (tr.)', but it is not confirmed by textual examples — actually, the only textual example for this form is |i |a ||ɔ: 'the sun is hot', which may, of course, be interpreted as 'the sun burns', but there is no explicit justification for this. Intransitive 'burn' = 'to be cooked' is actually attested as ||a (θwi: k'i |a ||a k'a 'the meat is burning', θwi: k'i ||a 'the meat is cooked').
- Taa: Not reconstructible. ◊ The only solidly attested equivalent for 'to burn (both tr. and intr.)' is !Xóō θ'á:, clearly the same as Kakia θwa ~ θwā 'to make a fire' (e.g. ši a θwa |a 'we will light a fire'). Intransitive ||a 'burn' is also found in Kakia (|a: wa ||a a 'the fire is burning'); in !Xóō, however, the meaning of the cognate ||āha is listed as 'set alight, set on fire, torch (e.g. tobacco, the veld, a hut), singe', indicating transitive use.
- Tuu-: Not reconstructible. ◊ Onomasiological reconstruction in this particular case is seriously hampered by what looks like incomplete and inaccurate semantic glossing in both older and newer sources of data, and by the difficulties in distinguishing between transitive and intransitive usages of verbal stems, as well as subtle semantic distinctions between 'to light, kindle' (= 'to make to begin to burn') and 'to burn down' (= 'to reduce to ashes by burning'). Clearly, the verbal root *||a is in itself well preserved in all three branches of Tuu, but whether it was indeed the basic equivalent of the meaning 'burn' in contexts like 'I burned [down] the house' remains unclear. For now, we should probably exclude this item from any calculations.

13. CLAW (= FINGERNAIL) [!Ui + Nossob + Taa]

- !Ui: *||qo-rV (|Xam ||ur(r)u, ||Ng!ke ||uri-si, †Kho-M ||oro(-si), N|uu ||qoro-si, ||Xegwi-B ||ɔla). ◊ The detachable origins of the second syllable are hinted at by plural forms in |Xam: ||u-||u-(t)tan, where the original expression of plurality is represented by root reduplication. In all other languages, the nominal suffix of the singular form has completely

fused with the root. It must be noted that the form in ||Xegwi-Z is completely different: sg. *!'elo-lorj*, pl. *!'elo-le*, of unknown origin.

- Nossob: *!'Auni* *||ora-sa*. ◇ The situation in |Haasi is not clear. Story lists the form *k'a=fĩi*, correctly identifies it as plural and further connects it with *fĩi* ‘finger’, which seems to be a phonetic variant of the same root. Further etymological connection of this word is clearly with !Xóõ *fĩi* ‘foot, spoor, track, hoof of an ungulate’ and its cognates (see FOOT below). Given that there are no textual examples confirming the semantics of ‘fingernail’, the glossing may very well be erroneous.
- Taa: (a) *||**qu-** (!Xóõ *||qû-le*, pl. *||qû-n-sâ*); (b) *||**aʔm** (!Xóõ *||aʔm*, Kákia pl. *!'am-te*). ◇ For !Xóõ, Traill lists two synonymous equivalents with the meaning ‘fingernail’ without specifying any semantic differentiations. The former is a perfect etymological match for Proto-!Ui *||**qo-rV** (right down to the detachable suffix of the sg. form), but is not supported by older data on Taa. The latter has no parallels in !Ui, but could be equated with Kákia *!'am-te* assuming that the dental click / here is a mistranscription for lateral // (these two symbols seem to be frequently mixed up in Bleek’s materials on Kákia). It must be noted that if *||aʔm* is analyzed as *||aʔ-m* (where *-m* is a fossilized plural marker, as in *||á:* ‘stick’, pl. *||á-m* id. and many other examples), the forms are comparable with Proto-Kalahari Khoe *||*a* ‘fingernail’ (Vossen 1997: 436) and could be interpreted as old borrowings from a Khoe source, leaving only *||**qu-** as a viable etymon.
- Tuu+: The form *||**qu-rV** (where **-rV* is likely to have been a detachable segment, appearing only in sg. forms) may safely be reconstructed for Proto-Tuu based on equidistant evidence from all three branches.

14. CLOUD [-]

- !Ui: Not reconstructible. ◇ Each language or dialect cluster has its own equivalent: (a) |Xam *!waʔ:-gən*; (b) ||Ng!ke *tix-ke* (pl. form) = N|uu *ʒo:-si* (reflecting **to-* or **do-*); (c) ||Xegwi-B *||xe:ŋ* (dubious form).
- Nossob: Not reconstructible. ◇ *!'Auni* *!'hum-sa* cannot be compared with |Haasi *!al=xwai*; the second form is clearly of composite origin, but the two halves are not easily decipherable.
- Taa: Not reconstructible. ◇ !Xóõ and N|u||en employ different periphrastic expressions for the concept: !Xóõ *!qʰà:=qʰũã*, lit. ‘water-hair’ vs. N|u||en *!xwe: ǎarri*, lit. ‘rain-sky’. Kákia *!wé* ‘cloud’ is unclear and without further connections.
- Tuu-: Not reconstructible. ◇ The generic concept of ‘cloud’ is clearly unstable in Tuu, although specific narrow terms denoting various types of clouds are encountered in !Xóõ (e.g. *qò:* ‘fairweather cumulus’) and other languages. This word should probably be excluded from comparison.

15. COLD [-]

- !Ui: Not reconstructible. ◇ Another unstable concept. In |Xam, no fewer than three equivalents are attested: (a) *x'aoʔ* ~ *x'áo* ‘cold’, (b) *sérri* ‘cool, cold’, (c) *||xwe:* ‘to be cold, become cold’. Attested examples are insufficient to draw clear semantic distinctions between these forms. In the N|uu cluster, most sources are in agreement on a single root, cf. ||Ng!ke *!'hu:* = †Kho-M *!'hu* = N|uu *!'hũ:* (but cf. additionally ||Ng!ke *si:-ya* ‘to be cold’, †Kho-D *kāriʔi* ‘cold’). ||Xegwi-Z *||keʔe* ‘cold’ contrasts with ||Xegwi-B *!xoa* ‘cold’. For most of these forms, it is hard to find etymological connections, but neither do they look like recent borrowings from Khoe or other sources.

- Nossob: |’Auni $\|’o\eta a$. \diamond In the early source Bleek 1929: 29, $\|xau$ ‘cold’ is listed instead. Not attested in |Haasi at all.
- Taa: $\|’\tilde{a}\tilde{z}\tilde{u}$ (!Xóõ $\|\hat{a}\tilde{z}\tilde{u}$, Kakia $\|x’we$, N|u|en $\|k’\tilde{a}\tilde{u}$). \diamond The phonetic discrepancy between Traill’s !Xóõ and Bleek’s earlier data is suspicious, but it may be argued that her transcriptions of ejective click effluxes actually reflect the same intervocalic glottal stop as in !Xóõ. In the case of Kakia, she glosses the word as ‘wind, cold’, but it is likely that she confuses here the reflexes of two separate roots, e.g. in $\|š i i a t i \|x’we$: “we are cold” the form $\|x’we$: = !Xóõ $\|\hat{a}\tilde{z}\tilde{u}$ ‘cold’, but in $\|x’we$: !xai “a big wind” the form $\|x’we$: = !Xóõ $\|q^h\tilde{u}e$ ‘wind’ (see WIND).
- Tuu-: Not reconstructible. \diamond It is tempting to connect Taa $\|’\tilde{a}\tilde{z}\tilde{u}$ with at least |Xam $\|xwe$: (and possibly |’Auni $\|xau$ if this is indeed a real form), but the discrepancy in effluxes is disconcerting, with additional examples for such a correspondence being hard to find. In any case, since there are problems with confirming the archaic origins of $\|xwe$: even on the !Ui level, this can hardly be counted as a lexicostatistical match on grounds of poor distribution.

16. COME [!Ui + Nossob + Taa]

- !Ui: $\|’s i \sim \|’s a$ (|Xam $s:e$, ||Ng!ke $si \sim se \sim se-ya \sim sa$, †Kho-M $si \sim si-ya \sim sa$, †Kho-D $s\bar{i}-y\bar{a}$, N|uu $sa: \sim ca:$, ||Kxau $sa: \sim se:$, ||Ku|e $sa \sim si$, ||Xegwi-Z, ||Xegwi-LH sa). \diamond In addition to seemingly random variation of the root vocalism, some sources also register a glottalic articulation of the initial sibilant, e.g. |Xam $s:e \sim ss’e \sim s:a: \sim ss’a:$ (W. Bleek). The reason for these variations is unclear; some of them may represent fusions of the root with agreement markers, but since few other verbal roots with codas in either $-a$ or $-i$ display so much variation, this is clearly not the only reason.
- Nossob: $\|’s^p i \sim \|’s a$ (|’Auni $sa \sim sé \sim sí$, |Haasi $c’i$). \diamond Nossob language data shows more or less the same variation as !Ui.
- Taa: $\|’s i \sim \|’s a$ (!Xóõ $s\hat{i}$, Kakia $si \sim ša$, N|u|en $sa \sim se \sim si \sim ša$). \diamond The precise !Xóõ forms are glossed as follows: $s\hat{i}$: ‘come arrive’, $s\bar{i}$: (var. form $sa-V$) ‘come to, come up to’. Cf. also $s\hat{i}$: ‘go’.
- Tuu+: $\|’s i \sim \|’s a$. \diamond Vocalic variation in this root clearly goes all the way back to Proto-Tuu. If the (presumably accurately defined) situation in !Xóõ is deemed indicative, $\|’s i$ may be thought of as the original unbound form (infinitive, etc.) while $\|’s a$ would be the stem variant used in conjunction with agreement markers. Still, the general issue remains open.

17. DIE [!Ui + Nossob + Taa]

- !Ui: $\|’a$ (|Xam $\|’a$, ||Ng!ke $\|’a$, †Kho-M, †Kho-D $\|’a$, N|uu $\|’a$, ||Kxau $\|’a$, ||Ku|e $\|’a$, ||Xegwi-Z $\|’a$, ||Xegwi-B $\|’a$). \diamond Apart from a strange lack of glottalic articulation in some of the attested varieties of ||Xegwi (cf. also ||Xegwi-LH $\|’a$: ‘dead’), all languages clearly reflect a single form $\|’a$.
- Nossob: |’Auni $\|’\tilde{a}$ ‘dead’. \diamond |Haasi seems to have lost the old root, since Story only records $\|’^h o$ ‘to die’, $\|’^h wa$: ‘dead’ — an innovation without a definitive etymology (cf., perhaps, !Xóõ $\|’^h \tilde{u}$ ‘be old?’).
- Taa: $\|’a$ (!Xóõ $\|’\hat{a}$, Kakia $\|’a \sim \|’a:$, N|u|en $\|’a$).
- Tuu+: $\|’a$. \diamond This is one of the most stable and widely distributed verbal roots in Tuu.

18. DOG [!Ui + Nossob] + Taa]

- !Ui: ***ʃ^hu-** (|Xam *!wínj* ~ *!úínj* ~ *!^hwínj*, ||Ng!ke *!wínj*, †Kho-D *ʃān*, N|uu (W) *ʃ^hun* ~ (E) *ʃ^hun*, ||Kxau *ʃ^huni*, ||Ku||e *!wínj*, !Gā!ne *!inyi*, Seroa *kuenia*, ||Xegwi-Z *λwa* ~ *λweŋ* ~ *ɓwe*, pl. *λume*, ||Xegwi-LH *λ^hwínj*, pl. *λ^hu-min*). ◇ Despite the seemingly chaotic array of reflexes, all forms are related. Original palatal articulation of the click is preserved in N|uu and confirmed by the regular development into a lateral affricate in ||Xegwi. Aspiration of the click is strongly confirmed by the same two languages (for ||Xegwi, only in the LH doculect, which seems to be more phonetically reliable than ||Xegwi-Z). As for the coda, most of the languages reflect the stem ***ʃ^hu-ni** (preserved as such in ||Kxau, losing the final vowel in N|uu, weakened to **ʃ^hu-ĩ* ~ **ʃ^hu-inj* in |Xam and ||Xegwi), but it seems that at least some dialects of ||Xegwi had different stem extensions (*λwa* ← **ʃ^hu-a?*).
- Nossob: ***ʃ^hɔŋ** (|ʔ Auni *ʃɔ̃*, |Haasi *ʃ^hāŋ*). ◇ Reconstruction of the coda is highly approximate.
- Taa: ***ʃ^hq^ha-** (!Xóõ *ʃ^hq^hàì*, pl. *ʃ^hq^hà-ba-tê*, Kakia *ʃ^hxai* ~ *!xài* ~ *!ài*, N|u||en *ʃ^hi* ~ *ʃ^hi* ~ *ʃ^hxi*). ◇ The paradigm in !Xóõ shows that **-i* is a detachable class marker.
- Tuu+: ***ʃ^hɔ-**. ◇ All attested forms are related through recurrent correspondences (for N|uu *-^h* vs. !Xóõ *-q^h-*, see HAIR, TOOTH; rounded vocalism in !Ui vs. unrounded vocalism in Taa is very frequent, see BREAST, etc.). The Nossob forms with their nasal coda are notably closer to !Ui than to Taa; it seems that the full stem ***ʃ^hɔ-ni** separates !Ui and Nossob from Taa ***ʃ^hɔ-i**.

19. DRINK [!Ui + Nossob + Taa]

- !Ui: ***k^ha-** (|Xam *k^hwā* ~ *k^hwī* ~ *k^hwū*, ||Ng!ke *k^ha* ~ *k^hā* ~ *k^hē* ~ *||x^hā*, †Kho-M *k^hā* ~ *k^hē*, N|uu *k^hā*, ||Ku||e *k^hwā* ~ *||x^hwā*, Seroa *k^hā*, !Gā!ne *k^ha*, ||Xegwi-Z pres. *k^hi*, past *k^ha*, ||Xegwi-LH *k^hē*). ◇ Secondary labialisation in |Xam under unclear conditions, as in many other examples.
- Nossob: ***k^ha-** (|ʔ Auni *k^hā* ~ *k^hē*, |Haasi *k^ha*).
- Taa: ***k^ha^h-** (!Xóõ *k^hā^h*, var. *k^ha^h-V*, Kakia *k^hā* ~ *k^hā* ~ *k^he* ~ *||x^hā*, N|u||en *k^ha-a* ~ *k^ha-u*).
- Tuu+: ***k^ha^(h)-**. ◇ An extremely stable basic verbal root, well preserved in every language. At least some of the attested variants, most notably **k^ha-* and nasalized **k^hā*, must go all the way back to Proto-Tuu where they may have been, as in !Xóõ, indicative of free and bound (“variable”) usage. Other variants (*k^hī*, *k^hē*, *k^hā*, etc.) probably represent fusion with various auxiliary particles. It is worth noting that this root represents one of the best known isoglosses between Tuu and Khoe, cf. Proto-Kxoe **k^ha* ‘to drink’ (Vossen 1997: 497), but since in both cases the word is clearly reconstructible to the topmost level, no a priori judgment can be made on the direction of borrowing, or even on whether this is indeed a borrowing or a super-archaic retention from a common linguistic ancestor of both Tuu and Khoe.

20. DRY [!Ui + Taa] [- Nossob] (?)

- !Ui: (?) ***||o** (|Xam *||o* ~ *||ɔ* ~ *||ò*, N|uu *||o*). ◇ This concept is not too well attested for !Ui languages; additionally, it is not easy to distinguish between the required semantics of ‘dry = not wet (e.g. of clothes)’ and ‘dry = dessicated, dried up’. Still, such examples as *||ainj ||ai:e se ||ɔ*: ‘...so that the inside of the house may dry’ (Bleek 1956: 581) confirm reliability of the item in |Xam, and the overall reconstruction is based on the correlation between |Xam, N|uu, and ||Xegwi-B *||o*: ‘thirsty’ (‘dry’ is not attested for ||Xegwi, but the semantic shift or extension from ‘dry’ → ‘thirsty’ is trivial).

- Nossob: !'Auni $\|x'om$. \diamond Not attested in |Haasi.
- Taa: (a) !Xóõ /ò:; (b) !Xóõ $\|úá^f$. \diamond Not attested in either Kákia or N|u|en. For !Xóõ, Traill lists two synonyms without specifying the distinctions. It may, however, be reasonably conjectured that /ò: is an areal root, since it is well attested across Khoe (Vossen 1997: 497), whereas for $\|úá^f$ no immediate source of borrowing can be detected.
- Tuu+: (?) * $\|o^f$ -. \diamond Proto-!Ui * $\|o$ and !Xóõ $\|úá^f$ are formally traceable back to a single source, although pharyngealized articulation in !Xóõ vs. lack thereof in N|uu is somewhat puzzling. The only attested Nossob form, !'Auni $\|x'om$, has no known parallels; if it is phonetically and semantically accurate (no guarantee), it can only be treated as an innovation of unknown origin.

21. EAR [!Ui + Nossob + Taa]

- !Ui: * $\|u$ - (|Xam $\|u-ntu$, ||Ng!ke $\|we:(-ntu) \sim \|u:-ntu$, †Kho-M $\|ui(-si)$, N|uu $\|ui-si$, ||Kxau $\|we:-ntu$, ||Ku||e de , ||Xegwi-Z $\|we$, pl. $\|u-me$, ||Xegwi-LH $\|wĩ$). \diamond All languages show traces of the original root * $\|u$ - (click correspondences are regular; the development * $\| \rightarrow \|$ in ||Xegwi is unique, but not contradicted by any other evidence, and ties in well with the general tendency of loss of palatal click articulation; ||Ku||e d - is also a regular reflex of both the alveolar and the palatal clicks); codas are different across most of major dialect clusters, reflecting such morphological variants as * $\|u-ntu$ and * $\|u-i$.
- Nossob: * $\|u$ - (!'Auni $\|ui$, (?) |Haasi $\eta=k'u=\|a-am$). \diamond The attested form in |Haasi contains the 1st p. possessive prefix $\eta=$ and the plural prefix $=k'u=$. The discrepancy in vocalism between |Haasi and !'Auni is more serious, but if the original suffixed stem was $\|u-a$ (cf. Taa), elision of the labial element in such a complex form could be a possibility (hard to confirm or disprove). In theory, it would be possible to think of $\|a-$ as the original root shape in Proto-Nossob assuming that !'Auni $\|ui$ is a form influenced by or directly borrowed from N|uu, but there is no conclusive evidence for such an assumption.
- Taa: * $\|u$ - (!Xóõ $\|úá^h$, Kákia $\|wa$, N|u|en $\|u-ša$, pl. $\|u-i-te$). \diamond As in !Ui, all forms reflect the base root * $\|u$ - with different suffixal extensions (* $\|u-á$, * $\|u-sa$).
- Tuu+: * $\|u$ -. \diamond Although it is hardly possible to unambiguously reconstruct the original paradigm for this root, given the massive amount of variation across different lineages, all languages (with the possible exception of |Haasi) clearly show that * $\|u$ - was the original root. No specific morphological isoglosses across the three branches.

22. EARTH (= SAND) [-]

- !Ui: *!(q)áũ (|Xam $\|k'áũ$, ||Ng!ke $\|áũ$, N|uu $\|áũ$). \diamond This concept is rather poorly attested in extinct languages and is usually not distinct from 'sand' (typically of the entire Khoisan area). At least the isogloss between |Xam and N|uu is reliable, though reconstruction of the click efflux is ambiguous (Bleek and Lloyd's transcription of the form with $-k'$ suggests something other than a standard glottalized efflux — possibly a uvular release — but this seemingly contradicts the N|uu transcription which shows no signs of uvular articulation).
- Nossob: (?) *!aʔa (!'Auni $\|á:a \sim \|a$ 'ground', $\|a:a$ 'dust'; |Haasi $\|aa$ 'ground'). \diamond Assuming that there is no lexical distinction between 'earth' (as substance) and 'ground' (as surface), which is a rather typical situation for San languages, we can tentatively set up *!aʔa as the Proto-Nossob equivalent; reconstruction of the coda as $-aʔa$ is confirmed by the transcription both in !'Auni (where it is reflected as variation between $-a(:)a$ and $-ʔa$)

and in |Haasi (doubled *-aa*). Additionally, cf. |’Auni *!ãũ* ‘dust’, which could either be an archaic retention with a slight semantic shift, or a borrowing from N|uu.

- Taa: **ɬx’um* (!Xóõ *ɬx’úm*, Kakia *!um ~ ||um ~ ||k’om*, N|u|en *!om-sa* ‘ground’, *ɬum* ‘ground, sand’). ◊ We rely on the accurately transcribed !Xóõ form for the phonological reconstruction; Bleek’s transcriptions of Kakia and N|u|en probably reflect the usual inaccuracies characteristic of items with original palatal clicks.
- Tuu-: Not reconstructible. ◊ Each of the three main branches has its own equivalent. Of these, only Proto-Nossob **!aʔa* has a transparent external etymology in !Xóõ *!ãʔa* ‘below; to lower’, indicating that ‘ground’ (surface) rather than ‘earth’ (substance) was, after all, the original meaning in the Nossob languages.

23. EAT [!Ui + Nossob + Taa]

- !Ui: **ɬã* (!Xam *hã: ~ ha:*, ||Ng!ke *ã ~ ë ~ ëĩ*, ɬKho-M *ã ~ ãĩ*, ɬKho-D *ʔãĩ*, N|uu *ʔã*, ||Kxau *ʔa*, ||Ku|e *ẽ*, ||Xegwi-Z pres. *ʔĩ*, past *ʔã:*, ||Xegwi-LH *ʔĩ: ~ ʔiŋ*). ◊ Vocalic variation here is similar to the situation with COME q.v.; original root vocalism *a* is strongly suggested by external data.
- Nossob: **ɬa ~ *ɬã* (|’Auni *ã ~ hà ~ hàa*, |Haasi *à:*).
- Taa: **ɬã* (!Xóõ *ʔã:*, var. form *ʔa-V*, Kakia *ã ~ a: ~ e: ~ é:*, N|u|en *ã ~ ë*).
- Tuu+: **ɬã*. ◊ All languages preserve the original root. Nasalization must be reconstructed as an intrinsic property of the original root vowel: it is extremely frequent across all three branches, and emerges clearly in such diagnostic forms as, e.g., the !Xóõ nominalization *ʔã:-sà* ‘eating, food’.

24. EGG [-]

- !Ui: (?) **ɬaʔwi* (!Xam *!áúi ~ !àúwi ~ !k’áú:wi*, ||Ng!ke *!ʰãũ*, pl. *!wi-tən*, ɬKho-D *ɬwi* ‘ostrich egg’, N|uu *ɬui*, ||Xegwi-Z *ɬwiŋ*, ||Xegwi-LH *ɬwĩ*). ◊ All of these forms are most likely related, since they all contain regular reflexes of the palatal click **ɬ-* (including the shift to a lateral affricate in ||Xegwi) and of the diphthong *-ui* (*-wi*). The overall shape of the root, however, is less clear. Perhaps the solution is hinted at by the quasi-suppletive paradigm recorded by D. Bleek for ||Ng!ke, which can be historically interpreted as going back to sg. **ɬaʔwi*, pl. **ɬui-ten* with contraction of the singular stem in a long plural form; in this case, *ɬui* in modern N|uu would be a back-formation from the original plural form. The form **ɬaʔwi* (← **ɬaʔbi?*) would also agree perfectly with |Xam data, as well as explain the variation between presence and lack of glottalic articulation in the different doculects of ||Xegwi. Still, the reconstructed shape remains speculative in the absence of similar corroborating examples.
- Nossob: (a) |’Auni *!ũĩ* ‘ostrich egg’; (b) |Haasi *k’ii*. ◊ The |’Auni form is likely related to !Ui **ɬaʔwi*, but it is unclear in which capacity — given the glossed semantics, and the glaring discrepancy with |Haasi, it could actually be a borrowing from N|uu (with misspelled click articulation). As for the |Haasi form, it has no external etymology at all.
- Taa: **ɬu-* (!Xóõ *ɬúú*, dimin. *kâ=ɬú:-bê*, Kakia *||wa:*, N|u|en *ɬwõĩ*). ◊ All listed forms are compatible, given how frequently the palatal click is transcribed as lateral or alveolar in Bleek’s Kakia and N|u|en records (see multiple other examples on this list); the basic root shape without suffixal extensions is **ɬu-* as seen in the !Xóõ diminutive form. It is not quite clear if !Xóõ *ɬú:*, pl. *ɬúã-tê* ‘empty ostrich egg’ is a phonetic variant of the same root (with a voiceless click) or a completely different etymon.

- Tuu-: Not reconstructible. ◊ Despite the obvious phonetic resemblance between !Ui (especially N|uu) and Taa forms, there is no easy scenario that would allow to reconcile them with each other (an original bisyllabic root like **ʃaʔbi* would not be expected to contract to **ʃu-* in Taa, since there are plenty of stems with the shape CV(?)bi in !Xóǝ).

25. EYE [!Ui + Nossob] [- Taa]

- !Ui: **c'a-xu* (|Xam *caxáú*, ||Ng!ke *cáxu* ~ *ca:xem*, †Kho-M *c'axau* ~ *c'axu* ~ *c'axəm*, †Kho-D *c'āxám*, N|uu *c'axam*, ||Kxau *c'axɔʔ*, ||Ku||e *caxu*, ||Xegwi-Z *sagu*, ||Xegwi-LH *c'agu*, pl. *c'a-ŋ*). ◊ Unique example of a bisyllabic stem in which the intervocalic consonant is not a resonant; this implies that the stem is historically a compound formation. The first root is unequivocally reconstructible as **c'a-* (most of the phonetically reliable sources mark glottalic articulation of the affricate); the second alternates between several variants (*-xau* ~ *-xu* ~ *-xam*), of which *-xu* is the most frequent one and is also often encountered as a nominal suffix in various words denoting surfaces (cf. in ||Ng!ke: *ʃa:xu* 'foot', *!a:xu* 'sky', *ʃa:xu* 'side'). It is very tempting to equate it with Proto-!Ui **xu* 'face' (|Xam *xú*, N|uu *xu* etc.), although this still leaves variants like *-xam* without a satisfactory explanation.
- Nossob: (?) **cxo* (|'Auni *cóo* 'eye/s', *c'a:xu(-ke)* 'eyes', |Haasi *cxɔ*, pl. *cxɔɔ*). ◊ The cluster *cx-* is extremely rare in |Haasi, making it all the more probable that the form *cxɔ* is contracted from an earlier bisyllabic form, clearly equatable with !Ui **c'axu*. If so, |'Auni *cóo* may further be regarded as its true cognate (with further simplification: **cxo* → *co*), while the doublet form *c'a:xu*, also attested by D. Bleek, could be interpreted as a re-borrowing from one of the dialects of N|uu (alternately, it could be an archaic preservation of the uncontracted form in some peripheral dialects or a higher register of the language, but this is unverifiable).
- Taa: (?) **ʃũ-* ~ **!ũ-* (!Xóǝ *!ũĩ*, pl. *!ũũ-tê* ~ *!ũũ-nî*, Kakia *ʃx'wĩ*, N|u||en *ʃũ*). ◊ Correspondences are unique: !Xóǝ clearly shows an alveolar click, whereas all other varieties of Taa speak in favor of palatal articulation (N|u||en in particular, but Kakia *ʃ* in Bleek's records very often transcribes an etymological palatal click, and almost never an alveolar one). This could be a serious argument for rejecting cognacy between !Xóǝ and Kakia-N|u||en; however, since there is no evidence in any of these languages for two different roots, and since 'eye' is typically one of the most stable items on the Swadesh list, it seems more prudent to admit the possibility of an irregular development in one of the two clusters (perhaps contamination with some other root).
- Tuu-: Not reconstructible. ◊ For this case, much rides on whether it is possible to demonstrate that !Ui **c'a-xu* (as well as Nossob **cxo*, which looks like a contracted variant of the former) is a compound form of secondary origin. While its composite nature is evident from its structure, the first component is not immediately identifiable, but it is phonetically and semantically possible to equate it with !Xóǝ *sàʔã* 'face, surface'. The optimal, though not the only possible, scenario here would be: (a) Proto-Tuu **saʔ-* with typologically common polysemy 'eye / face'; (b) Proto-Taa: **saʔ-* is retained in the meaning 'face, surface', but replaced by an innovation in the meaning 'eye'; (c) Proto-!Ui: **saʔ-* is replaced in the meaning 'face, surface' by the innovation **xu* (which has no cognates in Taa); the meaning 'eye' is eventually transferred to the new compound **saʔ-xu* → **c'a-xu*. Notably, such a scenario would hardly be compatible with the idea of a common ancestor for Taa and Nossob, but quite compatible with the idea of a common ancestor for Nossob and !Ui.

26. FAT (n.) [!Ui + [Nossob + Taa]]

- !Ui: ***so-** (|Xam *s:wéŋ*, ||Ng!ke *soa* ~ *süŋ*, †Kho-M *sōē*, N|uu *sun* (W) ~ *suŋ* (E), ||Xegwi-LH *swĩ*). ◊ Coda correspondences for this stem are extremely similar to the ones for DOG q.v., speaking in favor of reconstructing ***so-ni** for Proto-!Ui (original root vowel is **o* rather than **u*, which accounts for two different paths of assimilation: → **su-ni*, leading to ||Ng!ke *sü-ŋ*, N|uu *su-n*, or → **so-ne*, leading to |Xam *s:we-ŋ*, †Kho-M *sō-ē*). It is possible that ||Ng!ke *so-a* actually reflects the same root with a different suffixal extension.
- Nossob: ***so-** (|Haasi *cwa*). ◊ Not attested in !'Auni, but cf. *sā:a* 'fat' (adj.; polysemy 'fat /n./' : 'fat /adj./' is quite common for this word in Tuu languages). The |Haasi form regularly reflects **so-a* (with expected affricativization); !'Auni *sā:a* is somewhat strange due to lack of labial articulation, but there are no solid counterexamples for the potential change **-oa-* → **-a-*.
- Taa: ***sā^f** (!Xóō *sā^f*, Kakia *šā*). ◊ Note pharyngealized articulation in !Xóō.
- Tuu+: ***sɔ^f**. ◊ A clear isogloss between all three branches; vocalic correspondences between !Ui / Nossob, on one hand, and Taa, on the other, are recurrent, possibly reflecting **ɔ*. Pharyngeal articulation of the vowel in !Xóō may be archaic (it finds no correlation in N|uu, the only !Ui language where pharyngealization is marked accurately, but pharyngeal articulation seems to be prohibited in this language in structures like **CVn* ~ **CVŋ* anyway). Morphologically, the stem in Nossob seems to be closer to Taa than to !Ui (**so-a* or **so-ã* vs. **so-ni*).

27. FEATHER (= HAIR)

- It is preferable to exclude this word from comparison due to scant and dubious attestation. In both languages which have relatively modern descriptions (N|uu and !Xóō) the equivalent for FEATHER is the same as for HAIR q.v. In many others the word is not explicitly attested (||Xegwi; both Nossob languages; N|u||en), and those few equivalents which are distinct from HAIR are dubious (e.g. Kakia *dohé* 'feather' = !Xóō *dū^he* 'white ostrich plume' and may in reality be a more specialized term; |Xam *||erre* ~ *||árre* 'feather' is concurrent with FEATHER = HAIR and may actually mean 'wing' or a special type of feathers, etc.).

28. FIRE [!Ui + Nossob [+ Taa]]

- !Ui: ***'i** (|Xam */i*, ||Ng!ke */i*, †Kho-M */i*, N|uu */i*, ||Kxau */i*, ||Ku||e */e*, ||Xegwi-Z, ||Xegwi-LH */i*). ◊ Lack of glottalized efflux in ||Xegwi is surprising — it is hardly a transcriptional error, being recorded independently in two doculects — but still probably secondary, given the overwhelming testimony of other languages.
- Nossob: ***'i** (!'Auni */i*, |Haasi */i*). ◊ |Haasi shows the same lack of glottalized efflux as ||Xegwi, but in this case it is not so surprising, since Story very rarely marks ejective articulation in clicks anyway (see PERSON, for example).
- Taa: ***'a-** (!Xóō */ã*, Kakia */ã* ~ */a*, N|u||en */ã*). ◊ Nasalization in the coda is of morphological origin (the word belongs to Class 2 in !Xóō, typically marked by nasal suffixes).
- Tuu+: ***'i**. ◊ Although all three forms are quite likely related, reconstruction of the original root vocalism poses problems due to discrepancy between !Ui-Nossob **-i* (quite unambiguous) and Taa **-a* ~ **-ã*. Purely phonetic reasons are out of the question here, since the correspondence is non-recurrent. From a morphological perspective, a scenario deriving **'i* from **'a-i* is not too likely, since there are numerous examples of

-ai ~ -ae diphthongs in !Ui languages, and it is not clear what might have caused such a tight fusion in Proto-!Ui. The most probable hypothesis, therefore, is that there is an underlying contraction in the Taa form: */i- (root) + -ā (class suffix) → */ā with elision of the original root vowel. Of note is the near-total identity of this item with Proto-Khoe */(a)e ‘fire’ (Vossen 1997: 435), but since in both cases the items are clearly traceable all the way back to the proto-language, no assumptions may be made at this point about the reasons for this similarity (ancient borrowing, common ancestry or even chance resemblance).

29. FISH [-]

- This word is excluded from comparison due to the near-total lack of the corresponding reality in the Tuu-speaking area and, subsequently, in Tuu languages as well. (Curiously, Doke records the form *ǀebē* ‘fish’ for ǀKho-D, but it has no parallels anywhere and its origins are obscure).

30. FLY (v.) [!Ui + Nossob + Taa]

- !Ui: (a) */^h/au (|Xam *ǁau* ~ *ǁ^hau* ~ *ǁ^hóu* ~ *ǁxáú*, ||Ng!ke *ǁóu* ~ *ǁ^hou*); (b) *ze^ʕ (N|uu *ze:ʕ*). ◊ Unfortunately, this word is not attested in many languages, which makes the situation difficult to resolve. On one hand, the isogloss between |Xam and Bleek’s records of ||Ng!ke is fairly strong, despite some phonetic problems (e.g. confusion about the click efflux), and speak in favor of an original !Ui root such as */^h/au. On the other hand, N|uu *ze:ʕ*, attested in a more modern variety of the N|uu cluster, is strongly confirmed as the original word for ‘fly’ by its external parallels in both Nossob and Taa. Unclear if the former is really some sort of secondary synonym (perhaps ‘to fly up, to rise’ as opposed to simply ‘to fly?’), or if, vice versa, the latter was somehow reintroduced into modern N|uu from an outside source (ǀ’Auni?); better to take both as technical synonyms.
- Nossob: ǀ’Auni *zé*. ◊ Not attested in |Haasi.
- Taa: *zōē^ʕ (!Xóō *zāi^{ʕh}*, Kafia *zōī^ʕ* ~ *zwe^ʕ*). ◊ The parallel between !Xóō and Kafia is straightforward enough, right down to pharyngealization, but the labial vs. non-labial articulation of the vowel is surprising.
- Tuu+: *z(o)e^ʕ. ◊ The parallel between N|uu *ze:ʕ*, ǀ’Auni *zé*, and Taa *zōē^ʕ hints at a common Tuu origin for all these forms, despite some issues with vocalism (particularly on the Taa side) and distribution (see notes on !Ui). There are further areal connections to Khoe, cf. the clearly related Naro *cāē^ʕ* ‘to fly’ (Visser 2001: 98); however, this Naro word has no further Khoe etymology, meaning that it might itself be of Taa origin (the Taa word seems to have also made it into ǀHoan, cf. ǀHoan *zòe^ʕ* ‘to fly straight’ in Honken 1988: 65).

31. FOOT [-]

- !Ui: *ča (|Xam *ǀwa*, ||Ng!ke *ǁa(-xu)*, ||Kxau *ǁa-xu-ŋ* ~ *ǁa-xu-si* ‘leg’). ◊ Forms in |Xam, “old N|uu” (Bleek’s ||Ng!ke), and ||Kxau (for which Meinhof lists the meaning ‘Bein’, but there is no separate ‘Fuß’) agree well with each other and point at a protoform with the “sixth click”. Other forms are less clear. For modern N|uu, Sands et al. 2006 list the form *!x’u:-ke* but warn that the recording may be inaccurate; this may, in fact, be the same form as *!u:-ke* ‘shoe’ (Collins, Namaseb 2011: 35), which also puts the semantics

in doubt. ||Xegwi-Z /^hiʔi = ||Xegwi-B /x'e 'foot' has no etymological parallels in !Ui (but see below).

- Nossob: *!Xai (|'Auni /x'ai, |Haasi n=!'ai). ◇ |Haasi n= is a pronominal prefix ('my'). Both forms are clearly related, but the click efflux is ambiguous, probably due to mistranscription in one out of the two cases, or in both.
- Taa: *!fu- (!Xóõ /fũ:, Kákia /fo ~ /fo ~ /fò, N|u|len /fu). ◇ Also frequently encountered in the bisyllabic variant *!fu-ma, most likely a former diminutive: !Xóõ /fũ-ma-tê pl. 'feet', Kákia /fu-ma ~ /fu-mma 'foot', N|u|len /fu-ma 'claws, little feet'.
- Tuu: Not reconstructible. ◇ Each of the three subgroups has its own root to denote the required meaning, with no obvious etymologies in the other ones. It is quite tempting to relate Taa *!fu- with !Ui *!ca, especially considering the labialization in |Xam /wa; however, its secondary nature is strongly hinted at by ||Ng!ke /a- and ||Kxau /a-xu-, and the correspondence !Ui *a : Taa *u finds no reliable confirming examples. The meaning 'foot' does seem to be generally unstable in Tuu; for a possible example of semantic shift, cf. ||Xegwi /^hiʔi 'foot' = !Xóõ /q^hi: 'to walk (pl.)' (very likely the same root, implying a nominalization in ||Xegwi).

32. FULL [-]

- !Ui: *!qauŋ (|Xam /áúin ~ /áúin ~ /áú:enyā, ||Ng!ke /xan, N|uu /qã-ya). ◇ Not attested anywhere other than |Xam and the N|uu cluster. The root is verbal in origin ('to fill / be filled'). Protoform is approximate (*!q- reconstructed based on N|uu as well as occasional |Xam transcriptions with velar ejectives, e.g. /k'ãú 'to fill'; coda *-auŋ could also be *-ain, since labialization frequently turns out to be of secondary origin in |Xam).
- Nossob: |'Auni /x'an-si. ◇ Cf. also /ãú 'to fill', which can only be related if the click efflux in one of the forms is mistranscribed. Not attested in |Haasi.
- Taa: *!um (!Xóõ /ú^hm, Kákia /úm, N|u|len /um). ◇ Another !Xóõ equivalent is the verb /ola, but the latter has no parallels in other varieties of Taa.
- Tuu: Not reconstructible. Each subgroup has its own equivalent for this concept.

33. GIVE [-]

- !Ui: *a (|Xam /á: ~ /à: ~ /ã: ~ /a-a ~ /a-ã, ||Ng!ke /a ~ /a:, N|uu /ã:). ◇ Although the verb is not attested beyond |Xam and the N|uu cluster, it is clearly the most basic equivalent for 'to give' in both of these nodes and is easily reconstructible for Proto-!Ui. Nasalization is infrequent and likely secondary; quality of root vocalism is notably stable. The only other language where the main equivalent for 'to give' is perfectly clear is ||Xegwi, cf. ||Xegwi-Z /sa, ||Xegwi-LH /sa ~ /s- (as in /in za s-e 'I will give'). It correlates with ||Ng!ke /sa: 'to bring, fetch' and with ||Kxau /ŋ-sa 'to give' (where /ŋ- may be 'me'); ultimately, all these forms can be explained away as originally causative formations from *sa ~ *si COME q.v.
- Nossob: Not reconstructible. ◇ The situation in Nossob languages is complicated. For |Haasi, the only recorded equivalent is the monovocalic verb /i; its cognacy with !Ui *a is not excluded, but given the total lack of vocalic alternations in this root in any of the !Ui languages, there is nothing to confirm it. For |'Auni, Bleek records (a) rather marginal /a 'to give', only attested in one or two dubious examples; may be a result of mistaken analysis or a N|uu form; (b) /ã ~ /ɔ ~ /o, most often used in an imperative function and consequently comparable with |Xam /ã 'let, give' (also typically an imperative). Because of this variety and the relative unreliability of Nossob data, it is better to exclude the word from comparison.

- Taa: ***!q^ha-** (!Xóõ *!q^hǎ:*, Kákia *!xa: ~ !xe:*). ◇ Apart from this autonomous verb, the meaning ‘give’ is also expressed in !Xóõ by the auxiliary “verb-postposition” *ǎ:*, e.g. *ǎa ǎ:* ‘to pass to, give to’, where *ǎa* by itself = ‘hold, grab, grasp’. This is probably the same word as N|u|en *ǎi* ‘give’, but its usage in !Xóõ makes it somewhat less eligible for inclusion (and there is no way to verify if it actually displaced ***!q^ha-** in N|u|en or just accidentally happened to be the only recorded variant for GIVE).
- Tuu: Not reconstructible. ◇ The only secure isogloss between more than one branch of Tuu is the auxiliary verb ***ǎ-**, commonly used in the imperative function (‘give!’) in some of the !Ui and Nossob languages, and in a postpositional function in !Xóõ. The principal indicative forms are, however, clearly different between !Ui and Taa, and somewhat obscure in Nossob. Available data do not allow to reconstruct a precise historical scenario.

34. GOOD [-]

- !Ui: Not reconstructible. ◇ Most languages have their own equivalents, including some roots of clearly non-!Ui origin (||Xegwi-Z *luga-ge*, probably a Bantuism; †Kho-D *kǎm-ǎé* ← Khoekhoe **kǎm* ‘right; true’) and some with very weak distribution (|Xam *a:-kǎn* ‘good’; |Xam *twá:i-ǎ* ‘good’; ||Ng!ke *kǎi ~ kǎiʔ*, N|uu *ǎǎ-kǎ* ← **tǎi ~ *tǎiʔ*).
- Nossob: Not reconstructible. ◇ Not attested in |Haasi. For !’Auni, Bleek records the variants *xwe ~ xwoi*, without any etymology.
- Taa: Not reconstructible. ◇ !Xóõ *ǎǎ* has no parallels in Kákia or N|u|en. Not attested in Kákia; N|u|en *ǎǎ* also has no etymology.
- Tuu: Not reconstructible. ◇ The meaning GOOD is clearly very unstable in Tuu; the concept itself seems rather diffuse, and the relations between all these forms and similar forms in other Khoisan groups may reflect a complex network of areal interaction (cf. Proto-Khoe **!ǎǎ* ‘good’; †Hoan *qǎǎ* ‘good’, etc.).

35. GREEN [-]

- !Ui: Not reconstructible. ◇ Highly unstable and poorly documented meaning. In modern N|uu, expressed with *ǎǎo-a*, a borrowing from Khoe (cf. Nama *ǎǎo* ‘to turn green; to grow’).
- Nossob: Not reconstructible. ◇ Not attested in !’Auni. |Haasi *ǎǎu* is probably borrowed from the same Khoe source as the N|uu item.
- Taa: ***!ǎǎiʔ-** (!Xóõ *ǎǎiʔ^h*, Kákia *ǎǎiʔ*). ◇ Cf. !Xóõ *ǎǎiʔ^h-sǎ*, pl. *ǎǎiʔ^hm-sǎ* ‘dung beetle’, most likely containing the same root; morphological structure of the noun suggests that *-i-* is an original class suffix and that the semantics of ‘beetle’ might be primary. One might also speculate about further links with **ǎǎana* (← **ǎǎna* ?) LEAF q.v.
- Tuu: Not reconstructible. ◇ The concept is generally unstable, not very well documented, and most of the languages have their own ways of expressing it.

36. HAIR [!Ui + Nossob + Taa]

- !Ui: ***!ǎǎu** (|Xam *ǎǎ(-kǎn) ~ ǎǎ(-kǎn)*, ||Ng!ke *ǎǎ ~ ǎǎu*, †Kho-M *ǎǎ ~ ǎǎu*, N|uu *ǎǎu:-ke*, ||Xegwi-Z *ǎǎu:-zi*, ||Xegwi-LH *ǎǎu*). ◇ A super-stable word with fairly transparent phonology, though the aspirated articulation of the efflux tends to go unnoticed in older transcriptions.
- Nossob: ***!ǎǎo** (!’Auni *ǎǎo*, |Haasi *ǎǎ*). ◇ Aspirated articulation explicitly marked by Bleek for !’Auni, but noticeably absent in |Haasi.

- Taa: ***q^hu-** ($q^h\ddot{u}\ddot{a} \sim |q^h\ddot{u}\ddot{a}$, Kakia $|wa:-ni$, N|u|en $|uun-te$). \diamond Kakia and N|u|en forms are obviously plurals. It may be tentatively assumed that the complex voiced aspirated uvular click in !Xóõ is primary, although there is variation between voiced and voiceless articulation even within !Xóõ itself. Nasalized coda vowel in !Xóõ is detachable as a class 2 suffix.
- Tuu: ***|^hu-**. \diamond All forms are clearly related. The correspondence N|uu ^{-h} : !Xóõ ^{-q^h} is recurrent (see DOG). Nossob forms are slightly closer to !Ui due to lack of nasalization, but since nasal codas are occasionally attested in !Ui as well (cf. ||Xegwi-LH $|^h\ddot{u}$), this cannot be a classificatory argument.

37. HAND [!Ui + Nossob + Taa]

- !Ui: ***|x'a** (|Xam $|x'a$, ||Ng!ke $|x'a$, †Kho-M $|x'a$, N|uu $|x'a:$, ||Ku|e $|x'a:$, ||Kxau $|x'a$, ||Xegwi-Z pl. $|x'a-\eta$). \diamond Curious replacement (no external parallels) in ||Xegwi: ||Xegwi-Z *kyi* (with assumed suppletivism between singular and plural forms) = ||Xegwi-LH q^hi . Otherwise, a very stable item.
- Nossob: ***|x'a/N/** (|'Auni $|x'a/n/$, |Haasi $n=|xan$). \diamond |Haasi $n=$ is probably the 1st person possessive prefix. Nasality in the coda is either the same as the ||Xegwi plural $-\eta$ or the nasal class suffix in Taa languages.
- Taa: ***|x'a** (!Xóõ $|x'ā:$, Kakia $|x'a$, N|u|en $|x'a$).
- Tuu: ***|x'a**. \diamond One of the most stable and securely reconstructed items on the list.

38. HEAD [!Ui + Taa] [? + Nossob]

- !Ui: ***ǃa** (|Xam $\tilde{a} \sim \tilde{a}:$ $\sim \tilde{ā}:$, ||Ng!ke $\tilde{a} \sim \tilde{a}:$, †Kho-M \tilde{a} , N|uu $\tilde{a}:$, ||Ku|e $\tilde{a}:$, ||Kxau $\tilde{a}:$, ||Xegwi $\tilde{a}:$). \diamond Stable and securely reconstructed. Plural form may have been * $\tilde{a}-\eta$ (as in ||Xegwi, etc.).
- Nossob: (a) ***ǃa** (|'Auni $\tilde{a}:$); (b) ***xu** (|'Auni $x:uu$, |Haasi $\eta=x\ddot{a}$). \diamond Difficult situation. On one hand, Common Nossob **xu* is clearly the same as Proto-!Ui **xu* 'face' (|Xam *xu*, etc.) and reflects a very likely semantic shift 'face' \rightarrow 'head' (especially in light of external evidence from Taa which clearly shows * \tilde{a} as the original equivalent for 'head'). Since Bleek records both the older form $\tilde{a}:$ and the innovative form $x:uu$ in the meaning 'head' for |'Auni, it might be assumed that the Proto-Nossob form was still * \tilde{a} . On the other hand, it is also possible that |'Auni had simply reinstated the original word (at least in some contexts) under the very common influence of N|uu (in fact, this scenario is explicitly advocated by D. Bleek herself, see Bleek 1937: 211). Available evidence does not allow to make a definitive decision, meaning that both items have to be counted as technical synonyms on the Proto-Nossob level.
- Taa: ***ǃa-** (!Xóõ $\tilde{ā}n$, Kakia $\tilde{a} \sim \tilde{a}\eta$, N|u|en $\tilde{a}\eta$). \diamond Status of the nasal coda is unclear, but probably suffixal in light of external data.
- Tuu: ***ǃa**. \diamond Another highly stable Tuu root, albeit largely replaced by 'face' in the Nossob group (see discussion on the |'Auni situation above).

39. HEAR [!Ui + Nossob] [- Taa]

- !Ui: ***tu** (|Xam $t:u \sim t:ú:$, ||Ng!ke $tu \sim tu:i$, †Kho-M $\phi^hu \sim \phi^hu:-wa$, N|uu $\phi u:$, ||Kxau $tú$, ||Xegwi-Z tu). \diamond This stem is attested in many different morphological variants (cf. ||Xegwi-Z past stem $tu-wa$, present stem $tu-bi$, etc.; ||Ng!ke tu : 'hear' vs. $tu-ä$: 'heard' vs. $tu-i$ 'listen' in Bleek 2000: 24), but root vocalism is almost always *u* regardless of the morphological environment.

- Nossob: ʔAuni *tu:* ~ *tu:i*. ◇ Not attested in |Haasi. Curiously, the earliest source on ʔAuni (Bleek 1929: 46) lists the forms *ta:ã ~ ta:a* in the meaning ‘hear’, but in Bleek 1937: 203–206, probably the same word is transcribed as *tiãñ ~ kiãñ* (reflecting palatalized articulation of *t-*) with the meaning ‘feel’. These look like two different etyma rather than morphemic variants of each other.
- Taa: ***tã**^f (!Xóõ *tã*^f, Kákia *tãa ~ ta:*^fη, N|u|len *tãñ*). ◇ Variable form of the !Xóõ stem is *ta*^f-, but nasalization is so pervasive in all doculects that we should probably project it onto the proto-level (or even reconstruct **ta*^fη with an actual velar nasal coda).
- Tuu: (?) ***tu**. ◇ In the !Ui branch (and possibly also in ʔAuni, though the data here are limited and may also reflect !Ui influence), there is a rather clear distinction between the verbs **tu* ‘to hear’ and **ta ~ tã* ‘to feel’, cf. |Xam *tã ~ ta: ~ tã:* ‘to feel, try, seem, beware of’ with multiple text examples in Bleek 1956: 184. In Taa (or at least !Xóõ), both meanings seem to have been merged in the same root **tã*^f. Whether or not **tu* and **tã*^f are etymologically connected cannot be determined at this point, but since there is no strong evidence for grammatical Ablaut of any kind in !Ui or Taa, we should certainly treat them as two different roots, and postulate a probable lexical replacement in Taa. It should be noted that ʔAuni is closer in this respect to !Ui than to Taa.

40. HEART [!Ui + Nossob + Taa]

- !Ui: *|**ʔai** (|Xam *ʔi:*, ||Ng|ke *ʔai ~ ʔe*, †Kho-M *ʔe:-ç^hi*, N|uu *ʔe:*, ||Ku||e *ʔẽ:*, ||Kxau *ʔae ~ ʔai-si*). ◇ A stable item, lost only in ||Xegwi where it is replaced by a Bantuism (||Xegwi-Z *kele*, ||Xegwi-LH *keleñ*). However, click efflux correspondences are unique, with a lot of variation between simple velar and glottalized articulation which cannot be fully ascribed to mistranscriptions; we are either dealing with an original root structure like *|*aʔi* (with metathesis of glottalization) or with the unique reflexes of a rare click type (see below).
- Nossob: *|**ʔe** (ʔAuni *ʔe:* ~ *ʔe:*, |Haasi *n=|a-ʔe*). ◇ Structure of the form in |Haasi, except for the usual 1st p. possessive prefix *n=*, is unclear (reduplication?). Note the same discrepancy in click efflux articulation (glottalic in ʔAuni vs. velar in |Haasi) as in !Ui.
- Taa: *|**q’a-** (!Xóõ *ʔq’ãñ*, pl. *ʔq’ã:*, Kákia *ʔi:*, N|u|len *ʔañ*). ◇ Kákia *ʔi:* is unusual here because of the vocalism, but in light of external cognates in !Ui and Nossob it may actually be seen as more archaic in that respect. Perhaps what we see are the results of morphological variation in Proto-Taa, e.g. *|*q’a-i* (→ Kákia *ʔi*) vs. *|*q’a-n* (!Xóõ *ʔq’ãñ*, N|u|len *ʔañ*).
- Tuu: *|**q’a(-i)**. ◇ All the forms seem related, but reconstruction of click efflux and vocalism runs into problems. There may be a correlation between the glottalized / non-glottalized effluxes seen in !Ui and Nossob, on one hand, and the glottalic uvular efflux *-q’-* in Taa, in which case the latter should be set up for the protoform; clear additional evidence for this correlation is, however, lacking at present. As for the root vowel, variation in Taa suggests *a*, but this is also inconclusive, given the prevalence of front vocalism elsewhere. Only the !Xóõ-N|u|len node suggests *|*q’a-n* as a full stem, so this may be an innovation; on the other hand, grammatical variants *|*q’a-i ~ ʔq’a-n* may also reflect some meaningful opposition in Proto-!Ui, with only the former variant fused and preserved in !Ui, Nossob, and Kákia.

41. HORN [!Ui + Nossob + Taa]

- !Ui: *||**ãi** (|Xam *||^hẽ:* ~ *||^hẽi:*, ||Ng|ke *||ãi*, †Kho-M *||ẽi*, N|uu *||q^hoe-si*, ||Xegwi-LH *||i:*). ◇ In most old sources, the word is hopelessly confused with TOOTH q.v., but N|uu and ||Xegwi

data clearly show that these are two different (albeit phonetically similar) etyma. However, there is additionally a serious incongruence between N|uu $\|q^hoe-$ and the rest of !Ui forms which rather go back to something like $\|āĩ$; the worst problem is the vocalism, since the other languages (as well as external cognates in Nossob and Taa) show no signs of labial vowels. It is possible that modern N|uu $\|q^hoe-$ is not related (although in that case, its provenance is a mystery); in any case, the reconstruction is primarily based on the correlation between |Xam and ||Xegwi as the most distant members of the !Ui branch.

- Nossob: |'Auni $\|ēĩ$. \diamond Not attested in |Haasi.
- Taa: $\|ā-$ (!Xóõ $\|āē$, Kakia $\|ān-ša$, N|u|en $\|ā$). \diamond Kakia $\|ān-ša$ = !Xóõ pl. $\|ān-sâ$.
- Tuu: $\|ā-$. \diamond Nasalization of the vowel tentatively projected onto the proto-level due to its presence almost everywhere. Different codas most likely represent old morphology (e.g. sg. $\|ā-i$ vs. pl. $\|ā-n$, as in Taa, with generalization of the sg. form in !Ui and Nossob?).

42. I [!Ui + Nossob + Taa]

- !Ui: $\|ŋ$ (|Xam η , ||Ng!ke $\eta \sim n$, †Kho-M $\eta \sim na \sim n \sim na$, N|uu η , ||Ku|e η , ||Kxau $\eta \sim n$, ||Xegwi-Z $\eta \sim ?n \sim ?in \sim ?iŋ \sim ?m \sim ?im \sim ?ĩ$). \diamond The primary and most common form of the root is that of a syllabic velar nasal; everything else is the result of contextual assimilations or combinations with various emphatic particles.
- Nossob: $\|ŋ$ (|'Auni $\eta \sim n \sim na \sim m$, |Haasi η). \diamond See notes on !Ui.
- Taa: $\|ŋ$ (!Xóõ n , Kakia $n \sim na \sim \eta \sim \eta a \sim nya$, N|u|en $\eta \sim n \sim na$).
- Tuu: $\|ŋ$. \diamond The complete original paradigm (including stressed forms, clitical forms, emphatic forms, assimilated variants, etc.) is hard to reconstruct, but the monoconsonantal core of the Proto-Tuu 1st p. pronoun was undoubtedly a velar nasal, as still preserved in modern N|uu.

43. KILL [-]

- !Ui: $\|h^a \sim \|h^i$ (|Xam $|á(:) \sim |h^á(:) \sim |i$, ||Ng!ke $|a(:) \sim |i$; $\sim |h^i$, †Kho-M $|x^a$, N|uu $|h^a$). \diamond Reconstructible on the |Xam-N|uu level. Vowel gradation is similar to what is observed in several other cases, but difficult to explain based on extant data (for modern N|uu, only the a -grade form of the root is attested, suggesting analogical leveling in recent times). In ||Xegwi, the equivalent is $\lambda iŋ \sim \lambda eŋ$ 'hit, strike; kill' (Z) = $\lambda^e uŋ$ 'to hit' (LH); etymology is unclear, but the attested polysemy suggests semantic innovation (common semantic shift 'hit' \rightarrow 'kill').
- Nossob: |Haasi $!au$. \diamond Clearly the same word as |'Auni $!au$ 'to beat; to knock down', but not 'to kill', although the word 'kill' is not attested in |'Auni at all, so it cannot be excluded that the meaning 'kill' was expressed by the same form ($\|!au$) in Proto-Nossob.
- Taa: (?) $\|qa-$ (!Xóõ $qâi$, var. form $qa-JV$, Kakia $\|a$, N|u|en $\|wan$). \diamond It is unclear if the lateral click in Kakia and N|u|en is a real reflex of Proto-Taa $\|q-$ or if it represents a failed attempt to transcribe uvular articulation, but similar examples exist (e.g. !Xóõ $qâi$ 'pretty' = N|u|en $\|xai$ id.; !Xóõ $qâla$ 'to dig' = Kakia $|kâlaa$ id., where $|$ is clearly a typo for $\|$), so we tentatively list all these forms as cognate. Note that !Xóõ $qâi$ also conveys the meaning of 'forceful downward movement' (e.g. $qâi$ $\|úm$ 'stamp', etc.), which may be a case of homonymy but may also reflect an original meaning similar to 'hit', as in ||Xegwi or Nossob languages.

- Tuu: Not reconstructible. \diamond |Xam-N|uu $*|^{h}a \sim *|^{h}i$, with its seemingly archaic vowel gradation, looks like a good candidate, but has no recognizable cognates outside of that cluster. On the whole, the etymon ‘kill’ looks unstable and easily replaceable by words with the semantics of ‘hit’.

44. KNEE [-]

- !Ui: $*|^{h}u-$ (|Xam $\tilde{|\acute{o}a\eta} \sim \tilde{|\acute{u}a\eta}$, ||Ng!ke $\tilde{|\acute{o}:} \sim \tilde{|\acute{u}:}$, N|uu $\tilde{|\acute{u}:}-si$, ||Xegwi-Z, ||Xegwi-LH $\tilde{|\acute{o}-ma}$). \diamond Same root in all three languages, but with different suffixation ($*\tilde{|\acute{u}}-/a/\eta$ in !Ui-N|uu, $*\tilde{|\acute{u}}-ma$ in ||Xegwi; we can still see the pure root form in |Xam pl. $\tilde{|\acute{u}}-\tilde{|\acute{u}a}-d:e$, with reduplication).
- Nossob: |’Auni $\|w\grave{e}-\|w\grave{e}$ (Bleek 1937); sg. $\|oe$, pl. $\|we:-\|we$ (Bleek 1956). \diamond Not attested in |Haasi.
- Taa: $*|^{h}x\tilde{u}$ (!Xóõ $\|x\acute{u}:-\tilde{|\acute{a}n}$, Kafia $\| \tilde{|\acute{o}}-\tilde{|\acute{a}n}$, N|u|en $\| \tilde{|\acute{u}}-\tilde{|\acute{i}}$). \diamond All Taa doculets yield a compound in the meaning ‘knee’, where the second part is clearly HEAD q.v. (so it may be assumed that the original meaning of the compound was rather ‘knee-cap’). The root is tentatively reconstructed with a voiced velar fricative efflux based on !Xóõ data (Kafia and N|u|en transcriptions are unreliable).
- Tuu: Not reconstructible. \diamond Superficially, |’Auni $\|oe$ resembles Taa $*|^{h}x\tilde{u}$, and a common origin is not excluded (assuming inadequate click transcription and different suffixation). But it is even more similar to Proto-Kalahari Khoe $*|^{h}oe$ ‘knee’ (Vossen 1997: 457), meaning that borrowed origin is more likely here than inherited. If the Nossob form is excluded, !Ui and Taa forms may reflect the original Proto-Tuu ‘knee’ with comparable probability.

45. KNOW [-]

- !Ui: (?) $*|^{h}xae$ (||Ng!ke $\|ai$, N|uu $\|xae$). \diamond This seems to be the main, if not only, equivalent of the required meaning in the N|uu cluster, cognate with |Xam $\|ai \sim \|ai^f$ ‘to take notice, be(come) aware of smth.’ (Bleek 1956: 550). In |Xam itself, the meaning ‘know’ is usually correlated with the root $\acute{f}en(n) \sim \acute{f}e$ ‘to know, to think’, further related to ||Ng!ke $\acute{f}\tilde{e}$, N|uu $\acute{f}i$: ‘to think’, still further to !Xóõ $\acute{f}\tilde{a}n$ ‘to think’ and ultimately to Khoe $*\acute{f}an$ ‘to think (\rightarrow to know)’, from which this root may have been diffused into different branches of Tuu. ||Xegwi-LH ci , ||Xegwi-Z $ci-ya$ remains without any etymology. In light of all this, the N|uu root with its semantically similar |Xam cognate remains the optimal, if still weak, candidate for Proto-!Ui ‘to know’.
- Nossob: (a) |’Auni $\|xai \sim \|x'e-ki$; (b) |Haasi $\|üma$. \diamond The |’Auni form is clearly connected with N|uu $\|xae$, but it may be a borrowing rather than a genetic cognate. The |Haasi form is just as clearly related to !Xóõ (below), yet this could also be interpreted as an areal isogloss. Unclear.
- Taa: !Xóõ $\|ümā$ (var. form $\|u-BV$). \diamond Kafia $\|a$ ‘to know’, published only in the early source Bleek 1929: 51, is somewhat dubious.
- Tuu: Not reconstructible. \diamond Although there is a clear isogloss between |’Auni and !Ui, on one hand, and between |Haasi and !Xóõ, on the other, both may have areal rather than genetic interpretations, and given the overall unstable nature of this concept in Tuu (cf. the Khoe root $\acute{f}an$ with its wide diffusion over Tuu territory), we should probably exclude this word from comparison for safety reasons.

46. LEAF [-]

- !Ui: Not reconstructible. \diamond ‘Leaf’ is not a basic concept for !Ui speakers, and the word is usually borrowed (N|uu *bla:r-si* ← Afrikaans, ||Xegwi-LH *li=k^hasi-zi* ← Swazi), not attested, semantically questionable (e.g. |Xam *juhm*: ‘leaf, stick’), or completely isolated (||Ng!ke *xerro*: ‘leaves, foliage’ without any external cognates).
- Nossob: Not attested.
- Taa: (?) *|**ana** (!Xóõ *āna*, Kafia *a:na*). \diamond Formally reconstructible for Proto-Taa; however, the root is completely identical with Khoe *|*ana* ‘leaf, grass’ (Vossen 1997: 424), which strongly suggests a borrowed origin (which agrees well with the general instability of this concept in Tuu). Cf. also N|u|len *ābu* ‘leaf’ (Bleek 1929: 52; incorrectly listed as SIV = |’Auni in Bleek 1956), a different root without any external etymology.
- Tuu: Not reconstructible. All correlated items may be of secondary origin.

47. LIE [!Ui + [Nossob + Taa]]

- !Ui: ***ta** (|Xam *t:a* ~ *ta* ~ *t:ē* ~ *t:e:n* ~ *te:ŋ*, ||Ng!ke *tia* ~ *kia*:, †Kho-D *çâ*, N|uu *çax*, ||Kxau *ta* ~ *da* ~ *tn*). \diamond Despite some phonetic variation in the coda (most likely reflecting various morphologic variants), the most common and probably original shape of the root should be reconstructed with final *-a. The situation in ||Xegwi is unclear, with Bleek and LH contradicting themselves and neither of the variants (Bleek *la*: ‘lie’, LH *iŋ=θijne* ‘I lie down’) having external connections.
- Nossob: |’Auni *tàa*. \diamond Cf. also *tūa* ‘to lie curled up’ and transitive *tōā-a* ‘to lay down, to bury’. Not attested in |Haasi.
- Taa: ***tu** (!Xóõ *tū*:, Kafia *tu(:)* ~ *tá* ~ *tā*:, N|u|len *tu*:.). \diamond Note the variation in Kafia, unclear on its own but instructive in light of external comparison.
- Tuu: ***ta** ~ ***tu**. \diamond Although the dominant variant of this root in !Ui is clearly **ta* and in Taa clearly **tu*, scant evidence for the opposite also exists: Bleek 1929: 53 records the variant *tu* for ||Ng!ke, and Bleek 1956 has *tá*: for Kafia. The situation is slightly reminiscent of HEAR (except that the distribution of vocalizations is reversed), but in that case there was additional evidence to argue in favor of two original roots (‘hear’ vs. ‘feel’). Here, it is rather advisable to treat both forms as morphological variants of a single original root. The Nossob form is morphologically closer to Tuu, but could actually represent an “intermediate” variant (*to-a* ← **tu-a*, while !Ui languages go further and contract **tu-a* → **tá*?)

48. LIVER [!Ui + Taa]

- !Ui: *|**aN** (|Xam *āwan* ~ *āwāŋ*, ||Ng!ke *āi:n*, N|uu *āan* ~ *āaŋ* ~ *āaŋi*, ||Kxau *ŋaŋa*). \diamond Not attested in ||Xegwi. Variants fluctuate between *|*aŋ/a/* and *|*a-ni*.
- Nossob: Not attested.
- Taa: *|**am** (!Xóõ *ām*, N|u|len *ām*). \diamond Not attested in Kafia.
- Tuu: *|**aN**. \diamond It can hardly be doubted that all listed forms belong together, but reconstruction of the coda is somewhat problematic, given all the variety between !Ui and Taa. It is possible that the original root was simply *|*a-*, particularly in light of !Xóõ plural forms (possessive *ā*:, alienated *ā*) which seem to drop the labial nasal as a suffix. On the other hand, nasality is such a persistent feature for all reflexes that it is hard to believe it was not, in some way or other, an intrinsic part of the root. The provisional reconstruction *|*aN* reflects that uncertainty.

49. LONG [!Ui + Nossob?]

- !Ui: *|a (||Ng!ke /a:, N|uu /ã:, ||Kxau /a:, ||Xegwi-Z /ã, ||Xegwi-LH /ã:). ◇ The isogloss between N|uu, ||Kxau, and ||Xegwi clearly identifies */a as the optimal candidate for Proto-!Ui ‘long’, although there are some phonetic problems — in N|uu, the coda has a nasalized vowel, and in ||Xegwi we see unpredictable variation between /ã (no glottalisation) and /ã̃ (nasal efflux) depending on the dialect. This may be due to a more complex protoform (e.g. something like */ã̃ã̃ with different types of contractions) or to several original morphological variants (e.g. */a vs. */a-/a/N); the issue requires a better understanding of !Ui adjectival morphology. Curiously, the root is not at all attested in |Xam, where the semantic definition ‘tall / long / high’ is instead attributed to the form !xó:-wa, pl. !xó-!xó-ka — transparently derived from the verb !xo: ‘to grow up, climb up; make upright, make tall’.
- Nossob: |Auni /ã̃-si. ◇ Glossed as ‘big, long, tall’, but ‘big’ is probably incorrect (the proper equivalent for this meaning in |Auni is ús/i/, see BIG). Not attested in |Haasi.
- Taa: (a) !Xóõ /ám; (b) Kakia /úm. ◇ Not reconstructible (the two forms are clearly not related).
- Tuu: Not reconstructible. ◇ The |Auni form is clearly the same as the !Ui form, although it is impossible to determine if it is inherited or historically borrowed from N|uu.

50. LOUSE [!Ui + Taa]

- !Ui: *|u- (|Xam /ũiŋ, ||Ng!ke /ũin-ya, N|uu /ũ-si, ||Xegwi-LH /ũe-zi). ◇ Root vowel reconstruction is provisional (labial variant is the most common, but it could be assimilated to the labial click).
- Nossob: Not attested.
- Taa: !Xóõ /ũ. ◇ Plural: /ũ-tê. Not attested in Kakia or N|u|en. Another synonym is !Xóõ /x’óni ‘louse’; semantic differences between the two forms are unclear.
- Tuu: *|u-. ◇ The lexical match between !Ui and !Xóõ is transparent and allows to reliably project the root onto the Proto-Tuu level, even despite relatively scant attestation of the word in both branches (and a complete lack of attestation in Nossob). It should be noted that the alternate !Xóõ synonym /x’óni bears an uncanny resemblance to the common Kalahari Khoe term for ‘louse’, *k’uni, which in turn is somewhat irregularly connected with Khoekhoe *k’uri id. (Vossen 1997: 462); however, presence of a lateral click in !Xóõ is somewhat baffling, since it does not allow to explain the word as a (quite common) relatively recent borrowing from Kalahari Khoe. Could this be another piece of evidence from a “pre-Tuu / pre-Khoe substrate”?

References

- Bleek, Dorothea F. 1929. *Comparative vocabularies of Bushman languages*. Cambridge University Press.
- Bleek, Dorothea F. 1937. Grammatical Notes and Texts in the |Auni Language. In: J. D. R. Jones (ed.). *Bushmen of the Southern Kalahari*: 195–220. Johannesburg: Witwatersrand University Press.
- Bleek, Dorothea F. 1956. *A Bushman Dictionary*. *American Oriental Series, vol. 41*. New Haven, Connecticut: American Oriental Society.
- Bleek, Dorothea F. 2000. *The ||N!ke or Bushmen of Griqualand West* (ed. by Tom Güldemann). *Khoisan Forum Working Paper No. 15*. Köln: University of Cologne.
- Bleek, Wilhelm H. I., Lucy Lloyd. 1911. *Specimens of Bushman Folklore*. London: George Allen & Co.
- Collins, Chris, Levi Namaseb. 2011. *A Grammatical Sketch of N|uuki with Stories*. Köln: Rüdiger Köppe Verlag.

- Crawhall, Nigel. 2004. *!Ui-Taa language shift in Gordononia and Postmasburg Districts, South Africa*. PhD thesis: Faculty of Humanities, University of Cape Town.
- Doke, Clement M. 1936. An outline of ǀKhomani Bushman phonetics. *Bantu studies* 10: 433–461.
- Güldemann, Tom. 2005a. “Tuu”: A New Name for the Southern Khoisan Family. In: Tom Güldemann (ed.). *Studies in Tuu (Southern Khoisan)*. *University of Leipzig Papers on Africa, Languages and Literatures* 23: 2–9. Leipzig: Institut für Afrikanistik, University of Leipzig.
- Güldemann, Tom. 2005b. Tuu as a language family. In: Tom Güldemann (ed.). *Studies in Tuu (Southern Khoisan)*. *University of Leipzig Papers on Africa, Languages and Literatures* 23: 11–30. Leipzig: Institut für Afrikanistik, University of Leipzig.
- Güldemann, Tom. 2014. The Lower Nossob varieties of Tuu: !Ui, Taa or neither? In: Tom Güldemann, Anne-Maria Fehn (eds.). *Beyond ‘Khoisan’: Historical relations in the Kalahari Basin*: 257–282. Amsterdam: John Benjamins Publishing Company.
- Güldemann, Tom. 2017. Casting a Wider Net over Nǀng: The Older Archival Resources. *Anthropological Linguistics* 59(1): 71–104.
- Güldemann, Tom. 2018. Historical linguistics and genealogical language classification in Africa. In: Tom Güldemann (ed.). *African Languages and Linguistics*: 58–444. Berlin: DeGruyter Mouton.
- Hastings, Rachel. 2001. Evidence for the genetic unity of Southern Khoesan. *Cornell working papers in linguistics* 18: 225–246.
- Kassian, Alexei, George Starostin, Anna Dybo, Vasiliy Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.
- Köhler, Oswin. 1981. Les langues khoisan. In: G. Manessy (ed.). *Les langues dans le monde ancien et moderne. I: Les langues de l’Afrique subsaharienne*: 455–615. Paris: Editions du CNRS.
- König, Christa, Bernd Heine. 2001. *The !Xun of Ekoka. A demographic and linguistic report*. *Khoisan Forum Working Paper No. 17*. Köln: University of Cologne.
- Lanham, Leonard, D. P. Hallows. 1956a. Linguistic relationships and contacts expressed in the vocabulary of Eastern Bushman. *African studies (Johannesburg)* 15(1): 45–48.
- Lanham, Leonard, D. P. Hallows. 1956b. An outline of the structure of eastern Bushman. *African studies (Johannesburg)* 15(3): 97–118.
- Maingard, Louis F. 1937. The ǀKhomani dialect of Bushman: its morphology and other characteristics. In: J. D. Rheinallt Jones, Clement M. Doke (eds.). *Bushmen of the southern Kalahari*: 237–275. Johannesburg: Witwatersrand University Press.
- Meinhof, Carl. 1929. Versuch einer grammatischen Skizze einer Buschmannsprache. *Zeitschrift für Eingeborenen-Sprachen* 19: 161–188.
- Miller, Amanda, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, Chris Collins. 2009. Differences in airstream and posterior place of articulation among Nǀuu clicks. *Journal of the International Phonetic Association* 39(2): 129–161.
- Naumann, Christfried. 2014. Towards a genealogical classification of Taa dialects. In: Tom Güldemann, Anne-Maria Fehn (eds.). *Beyond ‘Khoisan’: Historical relations in the Kalahari Basin*: 283–302. Amsterdam: John Benjamins Publishing Company.
- Sands, Bonny. 2014. *Adoption, Maintenance and Loss of Click Contrasts*. Talk presented at: Sound Change in Interacting Human Systems, 3rd Biennial Workshop on Sound Change, University of California, Berkeley, May 28–31, 2014. Available online at: http://linguistics.berkeley.edu/SCIHS/abstracts/5_FridayAfternoon/Sands.pdf.
- Sands, Bonny, Amanda Miller, Johanna Brugman. 2007. The Lexicon in Language Attrition: The Case of Nǀuu. In: Doris L. Payne, Jaime Peña (eds.). *Selected Proceedings of the 37th Annual Conference on African Linguistics*: 55–65. Somerville: Cascadilla Proceedings Project.
- Starostin, George. 2008. From Modern Khoisan Languages to Proto-Khoisan: the Value of Intermediate Reconstructions. In: Ilya Smirnov (ed.). *Aspekty komparativistiki III [Aspects of comparative linguistics III]*. *Orientalia et Classica, Vol. XIX*: 337–470. Moscow: Russian State University for the Humanities.
- Starostin, George. 2015. *Annotated Swadesh wordlists for the !Wi group (Peripheral Khoisan family)*. Ms. Available online at: starling.rinet.ru/new100/kwi.pdf.
- Starostin, George. 2016. From wordlists to proto-wordlists: reconstruction as “optimal selection”. *Faits de langues* 47/1: 177–200.
- Starostin, George (ed.). 2011–2021. *The Global Lexicostatistical Database*. Moscow: Higher School of Economics / Santa Fe, NM: Santa Fe Institute. Available online at: <http://starling.rinet.ru/new100/>.

- Starostin, Georgij. 2013. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoy klassifikacii. Tom I: Metodologija. Kojzanski je jazyki*. Moscow: Jazyki slav'anskoj kul'tury.
- Story, Robert. 1999. *K'u^ha:si manuscript (MS collections of the Ki^hazi dialect of Bushman, 1937)*. Ed. by Anthony Traill. *Khoisan Forum Working Paper No. 13*. Köln: University of Cologne.
- Traill, Anthony. 1994. *A !Xóǀ dictionary. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 9*. Köln: Rüdiger Köppe Verlag.
- Traill, Anthony. 1995. Interpreting |Xam phonology: the need for typological cleansing. In: Anthony Traill, Rainer Vossen, Megan Biesele (eds.). *The complete linguist: papers in memory of Patrick J. Dickens (Afrikanische Sprachen und Kulturen)*: 509–523. Köln: Rüdiger Köppe Verlag.
- Traill, Anthony. 2018. *A Trilingual !Xóǀ dictionary: !Xóǀ – English – Setswana. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 37*. Köln: Rüdiger Köppe Verlag.
- Vossen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe Verlag.
- Westphal, Ernst O. J. 1962. On classifying Bushman and Hottentot languages. *African Language Studies* 3: 30–48.
- Westphal, Ernst O. J. 1971. The Click Languages of Southern and Eastern Africa. In: Jack Berry, Thomas A. Sebeok (eds.). *Linguistics in Sub-Saharan Africa*: 367–420. The Hague / Paris: Mouton.
- Ziervogel, Dirk. 1955. Notes on the language of the eastern Transvaal Bushmen. In: E. F. Potgieter (ed.). *The disappearing Bushmen of Lake Chrissie. Hidding-Currie publications of the University of South Africa, no. 1*: 34–64. Pretoria: J L van Schaik.

Г. С. Старостин. Лексикостатистические исследования по койсанским языкам II/1: к вопросу построения списка Сводеша для пра-ту языка

В статье, представляющей собой первую из двух частей исследования, представлены результаты общего лексикостатистического обзора языковой семьи ту (= южнокойсанской семьи), в ходе которого частично реконструируется список Сводеша для языка пра-ту и разъясняется ряд сложных моментов, касающихся внутренней классификации языков ту. В настоящей публикации представлен краткий обзор источников, перечислены основные методологические проблемы, связанные с диахроническим изучением языков ту, и приведены комментарии относительно исторической фонологии этих языков. Большую часть статьи занимает Приложение, в котором дается попытка реконструкции первых 50 элементов из списка Сводеша для трех промежуточных узлов семьи ту (пра-!ви, пра-носсоб и пра-та).

Ключевые слова: южнокойсанские языки; языки ту; щелчковые фонемы; лексикостатистика; базисная лексика; ономаσιологическая реконструкция.

N-initial nouns in Landuma and their counterparts in Mande

This paper describes a group of kin terms in Landuma (a Mel language spoken in northwestern Guinea) which have a non-standard phonological structure: they begin with the consonant cluster NC. It is shown that the anomalous properties of these nouns can be explained via their origin: all of them are borrowed from Mande.

In Western Mande languages, nouns for elder kin are also anomalous in that they are often unable to adjoin a definite or referential article. It has been suggested previously that this anomaly could be explained by the presence of an archaic nasal prefix, a grammatical marker of elder kin. At the same time, such a nasal prefix is not attested in any modern Mande language.

Two hypotheses can be advanced on the origin of the initial nasal element in the anomalous Landuma nouns. According to the first, this element goes back to a nasal prefix reconstructed for nouns referring to elder kin in Mande. If so, the Landuma data can be regarded as an argument for the relatively recent disappearance of this prefix in Mande (i.e., subsequent to the start of close contact between speakers of Proto-Landuma and speakers of Proto-Manding and Proto-Susu). Alternatively, the nasal element can be regarded as a reinterpreted Mande 1SG pronoun *ni* which, in its possessive function, appears frequently with kin terms. It cannot be excluded that both sources may have been relevant.

Keywords: Landuma language; Mel languages; Western Mande; kinship terms; language contact.

1. General information on Landuma

Landuma is a language of the Mel family spoken by about 30 000 people (our estimate) in the north-west of the Republic of Guinea, in the vicinity of the city of Boke. The Landuma data analyzed in this paper was obtained in the course of fieldwork in Guinea in 2015–2018. A writing system for Landuma was elaborated by Kirk Rogers (Rogers 2005; 2008).

The phonological system of Landuma includes three front vowels *i, e, ε*; four back vowels *u, o, ɔ, ɒ*; and two central vowels *ə, a*. The consonants are represented in Table 1.¹

	Labial	Dental	Alveolar/palatal	Velar	Labiovelar	Laryngeal
Voiceless plosives	p	t		k ² [k/g]		
Voiced plosives	b	d	j		gb	
Fricatives	f	s				h
Affricate			c			
Nasal	m	n	ɲ	ŋ		
Oral sonorants	w	r, l	y			

Table 1. Landuma consonants

¹ In this paper we use an IPA-based phonological transcription (with modifications typical for African linguistics), rather than the practical transcription introduced by Kirk Rogers.

² The sounds [k] and [g] are allophones of the same phoneme *k*.

Landuma has a relatively simple agglutinative morphology, mainly in the verbal domain. The verbs have a rich derivation system (reflexive, reciprocal, instrumental, etc.). They inflect for TAM categories and agree with the subject in person/number, animacy/ “noun class” (see Section 2). Landuma is a right-branching SVO language.

2. Nominal morphology and agreement

The only regular grammatical affixes on the noun are nominal prefixes. These prefixes, which are present on most nouns, at first sight function as class markers. They show a number opposition between singular and plural (see Table 2). Some prefixes are also associated with other semantic features: for example, the prefixes *tɿ-/mɿ-* (SG/PL) express diminutivity, *pɿ-/nɿ-* (SG/PL) are augmentative, the pair *wə-/a-* is characteristic of human nouns, and deverbal nouns are always marked by the prefix *kə-*. Other prefixes have no obvious semantic value, besides signaling number.

Singular	Plural	Examples
wə-	a-	wə-caməs / a-caməs ‘merchant’
dɿ-	sə-	dɿ-lɔkɔ / sə-lɔkɔ ‘day’
	yɿ-	dɿ-sek / yɿ-sek ‘tooth’
kə-	cə-	kə-babu / cə-babu ‘maize’
	yɿ-	kə-ca / yɿ-ca ‘hand’
ɿ-	yɿ-	ɿ-bat / yɿ-bat ‘river’
tɿ-	mɿ-	tɿ-sar / mɿ-sar ‘small stone’ (diminutive)
pɿ-	nɿ-	pɿ-sar / nɿ-sar ‘big stone’ (augmentative)

Table 2. Nominal prefixes in Landuma

At the same time, Landuma has a significant number of nouns that bear no prefix in the singular³. Prefixes are absent in most recent borrowings, which are mainly from French: *farin* ‘flour’ < French *farine*, *plas* ‘place’ < French *place*, etc. Most proper names, both personal names and place names, also have no prefix: *Barlande* ‘Barlande’ (village name), *Fatu* ‘Fatu’ (woman's name). Finally, and most importantly, 12 to 20% of other nouns are prefixless without there being any obvious reason for this, cf. *bumbi* ‘hare’, *dis* ‘body’, *gbundo* ‘secret’, *jombo* ‘hyena’, *kas* ‘father’, *nɛnc* ‘fire’. These nouns form their plurals by adding one of the plural prefixes: *bumbi* ‘hare’ / PL *yɿ-bumbi*, *dis* ‘body’ / PL *sə-dis*, *gbundo* ‘secret’ / PL *yɿ-gbundo* or *sə-gbundo*, *jombo* ‘hyena’ / PL *yɿ-jombo*, *kas* ‘father’ / PL *a-kas*, *nɛnc* ‘fire’ / PL *yɿ-nɛnc*.

The most intriguing property of Landuma nouns concerns how they control agreement. Agreement is found within the noun phrase (adjectives, demonstratives, the numeral ‘one’ and some other words agree with the nominal head) and within the clause (the verb agrees with the subject). The choice of anaphoric pronouns generally follows the same rules as agreement proper.

Agreement in Landuma is basically agreement in animacy, but there is also a kind of agreement based on the phonology of the controlling noun. The basic (and somewhat simplified) agreement rule is as follows:

³ Every plural noun has a prefix.

- (1) a. semantic agreement: animate nouns trigger special animate prefixes (example 2abc).
 b. phonological (radical alliterative) agreement: the first consonant of the agreement prefix repeats the first consonant of the controller; if the controller begins with a vowel, the agreement prefix begins with η -⁴, example (3abcd).

(2a) *jombo wək-in*
 hyena AN-one
 ‘one hyena’

(2b) *karməkɔ wək-in*
 teacher AN-one
 ‘one teacher’

(2c) *wə-bɛ wək-in*
 NP-king AN-one
 ‘one king’

(3a) *tɪ-lər t-in*
 NP-finger AGR-one
 ‘one finger’

(3b) *lɔkuj l-in*
 week AGR-one
 ‘one week’

(3c) *gbundo gb-in*
 secret AGR-one
 ‘one secret’

(3d) *ɪ-sar ɪ-in*
 NP-stone AGR-one
 ‘one stone’

The rules seen in (1) are strictly observed by all inanimate nouns and by human nouns with the prefix *wə-*. At the same time, there are some groups of nouns that combine semantic and phonological agreement and/or fluctuate between these two options in certain constructions.

Nouns denoting ANIMALS trigger animate prefixes on verbs and anaphoric pronouns (4), but, most usually, alliterative prefixes on the agreement targets within a noun phrase⁵ (5ab).

(4) *Jombo ɪ-yup lɛ wə-sɔntle lɛ dɔ, dɔr dɪ-mɔp kɔ lɛ,*
 hyena IPRS-turn.out ASR 3SG.AN-run.fast ASR there hunger AGR-catch him/her ASR
wə-c-kɔ kə-kɔ dade ɲkɔn fəna, wə-sarɪ.
 3SG.AN-CONS-go NP-go village (s)he.SUBJ also 3SG.AN-carry.baggage
 ‘(It turned out that) the hyena was running very fast, (after a while) it felt hungry, it was running and (finally) came to the village. It, too, was carrying its baggage.’
 [oral text]

⁴ Radical alliterative agreement is a typologically rare phenomenon found in the Kru language family (Sande 2019), the Arapesh languages of New Guinea (Dobrin 2012), and a few others.

⁵ Within the noun phrase, some fluctuations between non-alliterative and alliterative agreement markers are attested.

(5a) *jonbo jə-bi*
 hyena AGR-black
 ‘black hyena’

(5b) *jonbo j-in* or *jonbo wək-in*
 hyena AGR-one hyena AN-one
 ‘one hyena’

The agreement behavior associated with prefixless animate nouns is less predictable: they allow phonological agreement in certain constructions, but ultimately their agreement varies across speakers.

Most human nouns without the prefix *wə-* (*karməkɔ* ‘teacher’, *karandi* ‘student’, *bobo* ‘deaf-mute’, *imamu* ‘imam’; diminutive and augmentative nouns referring to people like *tancay* ‘little girl’) allow alliterative agreement within the NP (6ab) but require semantic (animate) agreement markers on verbs (6c) and are antecedents of animate anaphoric pronouns.

(6a) *karməkɔ kə-tɔt* or (6b) *karməkɔ wə-tɔt*
 teacher AGR-good teacher AN-good
 ‘good teacher’

(6c) *karməkɔ wə-n-der* *lɛ*
 teacher 3sg.AN-FACT-come ASR
 ‘The teacher came.’

A special group of animate nouns will be considered in the next section.

3. A special group of animate nouns (N-nouns)

3.1. Presentation of the N-nouns

In this section, we will turn to the main topic of this paper, namely a small group of animate nouns (henceforth “N-nouns”) which are in several respects non-typical of Landuma. The full set, according to the dictionary (Rogers & Bryant 2012), includes nine nouns referring to close social relations and kinship terms: *nna* ‘mother’, *mbariŋ* ‘friend’, *mbeŋba* ‘ancestor’, *ncɔkɔ* ‘uncle, mother’s brother’, *njatiki* ‘host’, *ntana* ‘father-in-law, mother-in-law’, *ntara* ‘elder sibling’, *ntɛnɛŋ* ‘aunt, father’s sister’, *ntɔkma* ‘namesake’; one noun referring to an animal: *mbərfi* ‘wild boar’; and, finally, two proper nouns (*Nfasori* ‘Infasori’, *Nfali* ‘Enfali’) and also *Nkila*, the title of the prophet Muhammad (*Nkila Mahamadu* ‘the prophet Muhammad’).

All N-nouns begin with a consonant cluster of the type NC (where N is a nasal consonant). These are the only words in Landuma that begin with a consonant cluster.

The initial nasal consonant in all N-nouns agrees with the subsequent consonant in place of articulation. In the Landuma orthography elaborated by Kirk Rogers (2005; 2008), it is represented by the letter *n* in all cases.

Landuma has four nasal consonant phonemes: /m/, /n/, /ɲ/, and /ŋ/. The palatal consonant /ɲ/ is rare; it is only attested word-initially before a vowel, as in *ɲamane* ‘time’ (certainly an Arabic borrowing). The consonants /m/ and /ŋ/ are allowed in various positions; in particular, they are found before consonants (although infrequently, and never word-initially) and do not undergo assimilation: *wəkomsi* ‘midwife’; *camne* ‘carry on the head’. Finally, the consonant /n/ is allowed in various positions, but when it precedes an obstruent it undergoes place assimila-

tion: /nb/ is realized as [mb], /nk/ as [ŋk], etc. For example, the factative marker *n-* is realized as [m] before labial consonants and as [ŋ] before velars:

(7a) *Wə-m-bɔr lɛ m-ɔɔ m-i.*
 3SG.AN-FACT-add ASR NP-rice AGR-DEF
 ‘(S)he added some rice.’

(7b) *Fatu wə-ŋ-kudi mi lɛ.*
 Fatu 3SG.AN-FACT-greet me ASR
 ‘Fatu greeted me.’

This means, in particular, that three different nasal sounds ([m], [n], [ŋ]) can be met before dental and alveolar consonants but [n] is impossible before labial and velar consonants. In the latter case, the phonological oppositions /m/ vs. /n/ and /n/ vs. /ŋ/ are neutralized⁶. Henceforth, we will use a capital N to encode the initial nasal consonant in nouns with an initial consonant cluster (for example, *Nbenba* ‘ancestor’, *Ncɔkɔ* ‘uncle’, etc.).

3.2. Morphology of N-nouns

Postulating a common consonant in a group of semantically close nouns immediately raises the question: does this consonant represent a prefix or a part of the noun stem? Identifying it as a prefix does not seem to be a good decision in our case. For example, the plural forms of the N-nouns feature the plural prefix *a-*, which is quite usual for animate nouns: *Ntana* ‘father-in-law, mother-in-law’ — PL *aNtana*, *Ntara* ‘elder sibling’ — PL *aNtara*. One of our language consultants also allowed forms with the unique prefix *ara-*: *araNtana*, *araNtara*. In any case, the plural forms of the N-nouns are invariably formed by adding a prefix to the singular form, which is quite normal for prefixless nouns. All prefixed nouns form their plurals by replacing the singular prefix, cf. the examples in Table 2.

Another important factor is agreement. As described in Section 2, prefixless animate nouns constitute a zone in which agreement patterns are fluctuating or mixed. This is not the case with the N-nouns (including the animal noun *Nbɔrɔfi* ‘wild boar’), which invariably trigger animate agreement markers:

(8a) *Ncɔkɔ ɲɔn*
 uncle AN.DEF
 ‘the uncle’

(8b) *Ncɔkɔ w-a*
 uncle AN-that
 ‘that uncle’

(8c) *Ncɔkɔ k-ɔn*
 uncle AN-3SG.POSS
 ‘his/her uncle’

(8d) *Ncɔkɔ k-a Mamadu*
 uncle AN-POSS Mamadu
 ‘the uncle of Mamadu’

⁶ Different phonological interpretations of the initial nasal consonants followed by another consonant can therefore be proposed, and we do not plan to discuss the issue in more detail here.

(9a) *Nb̄arfi wək-in*
wild.boar AN-one
'one wild boar'

(9b) *Nb̄arfi wə-pɔŋ*
wild.boar AN-big
'a/the big wild boar'

One of our language consultants allowed phonological agreement and used the consonants [m] and [n] in the agreeing units. This suggests that we ought to interpret the initial clusters of the N-nouns as representing different nasal phonemes. However, this data is not fully reliable and was not endorsed by other speakers.

3.3. Mande cognates of N-nouns in Landuma

It turns out that practically all N-nouns in Landuma have cognates in Mande languages. Let us consider them one by one.⁷

1) Landuma *Nna* 'mother'

Proto-Manding **ná*, **Mandinka** (Creissels, Jatta & Jobarteh 1982; Anonym 1995) *náa* 'mother' (address form), **Xasonka** (Tveit & Dansoko 1993) *ná* 'mother; mother's sister', **Kita Maninka** (Creissels 2009: 55) *ná*, **Maninka** *ná* (without the tonal article) 'mother', **Bamana** *nà* 'mother, mummy' (address form)

Bozo-Soninke: Tiegaxo (Anonyme 1982) *nan* 'mother'; (in combination with the name of a fruit) 'fruit tree', **Sorogama** (Daget, Kanipo & Sanankoua 1953) *ná* 'mother'; (in combination with the name of a fruit) 'fruit tree'

Bobo (le Bris & Prost 1981) *nâ*

Samogo: Duun (Hochstetler 1996) *na*, **Dzuun** (Solomiac, Traoré & Traoré 1998) *ná*

Proto-Eastern (Schreiber 2008) **da* / **nā*, **Boko** (Jones 2004) *dá*

Southern Mande: **Dan-Blo** (Erman & Loh 2008) *dǎ́*, **Dan Gwɛɛtaa**, **Kla-Dan** *dǎ́* 'grandmother, great-grandmother; elder paternal aunt; father's or mother's elder brother's wife; mother's elder co-wife; mother-in-law (for a man)', **Tura** (Idiatov ms.) *dǎ́* 'grandmother; elder paternal aunt; father's or mother's elder brother's wife; elder sister-in-law; mother-in-law, grandmother-in-law', **Guro** (Kuznetsova & Kuznetsova) *dǎ́*, **Yowre** (Hopkins 1982) *dǎ́*, **Mwan** (Perekhval'skaya & Yegbé 2018) *dǎ́-lē* 'mother-in-law (for a woman)', **Wan** (Nikitina) *nà* 'mother', **Ben** (Paperno) *nà* 'mother'

2) Landuma *Nbariŋ* 'friend'

Southwestern Mande: **Liberian Kpelle** (Leidenfrost & McKay 2007) *málèŋ* 'sororal nephew'

Soso (Anonyme n.d.) *bárèn* (?) kin; intimate friend, **Jalonke** (Creissels 2010) *bári-méxè* (?) 'kin', *bári-dì* (?) 'child; compatriot'

Jeri (Tröbs 1998) *béli* 'uncle'

Mokole: Lele (Vydrin 2009a) *bérin*, *beyin*, **Lele** (Mara & Camara 1979) *bayen*, **Koranko** (Kastenholz 1987a) *béř*

Manding: **Mandinka** (Creissels, Jatta & Jobarteh 1982) *báriŋ*, *bárimma*, **Nyokolo Maninka** (Meyer 1983) *bariŋ*, **Xasonka** (Tveit & Dansoko 1993) *báriŋ*, **Kita Maninka** (Creissels 2009) *bárin* (no article), **Kagoro** (Vydrine 2001) *bàri-no* 'nephew, niece (sister's child)', **Maninka** *bárin*, *bári*, *bórin*, *bérin* 'maternal uncle; maternal kin'

⁷ In the following list, when a Mande word is cited without gloss, its meaning is the same as the meaning of the Landuma word.

3) **Landuma** *Nbenba* ‘ancestor’

South-Western Mande: Mende (Innes 1969) *mbémbá* ‘lineage, family’; **Guinean Kpelle** (Leger 1975) *ḥomɔ-kɔɔ*, *ḥemɛ-kɔɔ* ‘ancestor (from the same clan)’

Susu (Diané & Vydrine 2012) *bénbá*, *bénbà* ‘ancestor, forefather’; ‘grandfather; great-grandfather’; **Yalunka** *bénbà-nà*

Mokole: Lele (Vydrin 2009b) *bénba* ‘grandfather’, **Koranko** (Kastenholz 1987a) *bénba* ‘ancestor, forefather’

Manding **bÉnbaa* ‘ancestor’: **Mandinka** *bénbaa*, **Xasonka**, **Guinean Maninka** *bénba*, **Bambara** *bénba*

Bobo (le Bris & Prost 1981) *bēmā* ‘ancestor’

South Mande: Eastern Dan *ḥéḥá*, **Tura** (Idiatov) *ḥéḥá*

4) **Landuma** *Ncɔkɔ* ‘maternal uncle’

Susu *sóxɔ* ‘maternal uncle’, ‘sororal nephew’, **Jalonke** (Creissels 2010) *sóqɔ*, **Yalunka** *sóxɔ*

5) **Landuma** *Njatiki* ‘host’

Mende (Innes 1969) *yàlî* (< Manding)

Susu (Willits n.d.) *yàtigi* ‘protector’, **Jalonke** (Creissels 2010) *jààtigi*, **Yalunka** (Willits n.d.) *yatigi*

Mokole: Lele (Vydrin 2009a) *yatii*, **Lele** (Mara & Camara 1979) *yáyí*

Manding: Kita Maninka (Creissels 2009) *jàtigi*, **Maninka** *jàtii*, **Bamana** *jàtigi*

Soninke-Bozo: Bozo-Tigemaho (Anonyme 1982) *jadi*, **Bozo-Sorogama** (Daget, Kanipo & Sanankoua 1953) *jatigi*

6) **Landuma** *Ntana* ‘father-in-law, mother-in-law’

Manding: Maninka *tàna* ‘totem; nuisance’, **Bambara** *tènɛ* ‘taboo, interdiction’

7) **Landuma** *Ntara* ‘elder sibling’

Soso (Diané & Vydrine 2012; Touré 1989) *tààrá*, *-ø*, **Jalonke** (Creissels 2010) *tààrá* ‘elder sister’

Mokole: Kakabe (Vydrina 2015) *tàata* ‘elder sibling’, **Mogofin** (Polinder, Janse & van Linden 2009) *táata* ‘elder sister’

Manding: Mandinka (Creissels, Jatta & Jobarteh 1982; Anonym 1995) *táataa* ‘elder sibling; husband’, **Xasonka** (Tveit & Dansoko 1993) *tàata* ‘elder sibling’

Soninke (Galtier & Dantioko 1979; Smeltzer & Smeltzer 2001) *taata*

8) **Landuma** *Ntɛnɛɲ* ‘aunt, father’s sister’

Southwestern Mande: Bandi (Grossmann, Rodewald & Covac 1991) *tènà* ‘aunt’; **Guinean Kpelle** (Konoshenko 2019) *télêɲ* ‘paternal aunt; any elder woman of the paternal aunt’s family’

Susu (Diané & Vydrine 2012) *ténèn* ‘paternal aunt’

Vai (Welmers & Kandakai 1974) *ténà* ‘maternal aunt’

Mokole: Lele (Vydrin 2009a) *téne* ‘paternal aunt’, **Koranko** (Kastenholz 1987a) *téne* ‘paternal aunt’

Manding: Kita Maninka (Creissels 2009) *ténen* ‘paternal aunt’ (no article), **Guinean Maninka** *ténen*, **Bambara** *téne* ‘paternal aunt’; **Segu Bambara**, **Beledugu Bambara**, **Kaarta Bambara** ‘aunt’ (either paternal or maternal)

Bozo-Tigemaxo (Anonyme 1982) *tayen* ‘paternal aunt’

South Mande: Tura (Idiatov) *téé* ‘paternal aunt’

9) **Landuma** *Ntɔkma* ‘namesake’

This is a borrowing from Manding, where the form **tɔGɔ-ma* is morphologically transparent: **tɔGɔ* ‘name’ + *-ma*, a suffix of mutual kinship relation. The word also appears in many other Mande languages (where it can be also regarded as a Manding loan):

Southwestern Mande: Liberian Kpelle (Leidenfrost & McKay 2007) *tɔmá*, **Guinean Kpelle** (Konoshenko 2019) *tɔwéi*, *tɔyéi*

Susu (Willits n.d.) *tóxómà*, (Diané & Vydrine 2012; Touré 1994) *tóxómàn*, **Jalonke** (Creissels 2010) *tóqómá*

Southern Mande: Western Dan (Erman & Loh 2008) *tóóá*, **Eastern Dan** *túú*, **Kla-Dan** (Makeeva ms.) *túà*

There is also a similar Soninke form *toxora* where the function of the final element *-ra* is not quite clear.

10) **Landuma** *Nbərfi* ‘wild boar’

Susu (Willits n.d.) *báli* ‘pig’ may have a common origin with the Southwestern Mande forms: **Looma** *boi-g, boĩ-g, buĩ-g*, **Liberian Kpelle** (Leidenfrost & McKay 2007) *bōĩ*, **Guinean Kpelle** (Konoshenko 2019) *bòì* (there are some less reliable forms in other Mande languages which may also be cognates). The final element *-fi* is unclear (however, it may be comparable to Manding *fin* ‘black’, in which case the form would mean ‘black pig’).

11) **Landuma** *Nkiln* ‘title of the prophet Muhammad’

Most probably, a borrowing from Manding, where **kí* means ‘send’ and **-la* is an agentive suffix, giving **kíla* ‘messenger’. This word was borrowed into many modern Mande languages.

Manding: Mandinka (Creissels, Jatta & Jobarteh 1982) *kiilaa* ‘messenger; prophet’, **Guinean Maninka** *kíla, kéla, céla* ‘messenger’, **Bambara** *kíra* ‘prophet’ (with an irregular form of the suffix)

Southwestern Mande: Bandi (Grossmann, Rodewald & Covac 1991) *kèelá* ‘messenger’, **Looma** *kela, keela* ‘messenger’, **Liberian Kpelle** (Leidenfrost & McKay 2007) *kélá* ‘messenger’, **Guinean Kpelle** (Konoshenko 2019) *kélé, kélá*

Susu (Willits n.d.) *xéérá* ‘messenger’, *kiilà* ‘Prophet’ (the latter form is probably borrowed from Manding).

Both proper nouns belonging to this group, *Nfasori* ‘Nfasori’ and *Nfali* ‘Nfali’, may also be borrowed from Mande, and more precisely from Manding, where *Fà Sori* may be a honorific form of the male name *Sòri* (*fà* means ‘father’), and *Fàli* may have been a heathen name meaning ‘donkey’.

3.4. N-nouns in the context of Landuma-Mande language contacts

The history of the Landuma (and their closest relatives the Kogoli⁸) is characterized by very close contacts with Mande people. The earliest written sources mentioning them date back to the beginning of the 16th century, and they appear in these documents as subjects of the king of Manding (Suret Canal 2000: 334). One of the main trade routes connecting the medieval Mali Empire led from Siguiri to Boko (Iffono 2000), and there is even an oral tradition that places the origin of the Kogoli in Siguiri, a Maninka town (Suret Canal 2000: 336–337). Historical documents tell us less about contacts of the Landuma with the Susu and Jalonke. However, it can be taken for granted that these contacts continued for centuries, with the result that today most Landumas are bilingual in Susu. Moreover, the Mogofin people are immediate neighbours of the Landuma, therefore some contact between the two languages is to be expected.

The N-nouns are certainly not the only group of nouns borrowed from Mande. However, the initial N is not found in other borrowed nouns. In particular, it is absent in the borrowed

⁸ The Kogoli are an ethnic group in the area of Kumbia in north-western Guinea and adjacent areas of Guinea-Bissau, speaking a language that is, reportedly, closely related to Landuma. Unfortunately, practically no data on their language is available; the rare publications on the Kogoli that do exist (Suret Canal 2000; Ferry & Sande 2000) provide almost exclusively ethnohistorical data.

kin term *dimbore* ‘cousin’, from Susu *dinbore* (Raimbault 1923; Lacan 1942).⁹ Landuma also has three borrowed kin terms that begin with a nasal consonant, but not with the NC cluster: *mama* ‘grandfather, grandmother’; *nande* ‘father’s second wife’, and *nimokɔ* ‘spouse of one’s elder sibling; younger sibling of one’s wife’. These nouns do not show any phonological differences from other Landuma nouns (whether borrowed or not).

This shows that the peculiarities of N-nouns cannot be explained by the simple fact that these nouns are borrowed: other borrowed nouns (including certain kin terms) do not belong to this group. More than that, in the modern Mande languages which represent the most likely sources of the borrowings into Landuma — that is Susu, Jalonke/Yalunka, Maninka and Mogofin — the cognate words have no initial element N-. In what follows, we consider two alternative sources of the initial consonant cluster in these nouns.

4. N- as a prefix for elder kin terms in Western Mande languages

According to the first hypothesis, the presence of the nasal element at the beginning of kinship terms in Landuma can be explained by the fact that the Mande donor languages had, in earlier periods of their existence, a nasal prefix marking exactly this semantic group of words. A reconstruction of this prefix was advanced in (Vydrin 2006). Since this paper is available only in Russian, let us briefly survey the morphological peculiarities of West Mande languages that provide grounds for this reconstruction.

4.1. Elder kin in Manding languages

As mentioned in (Spears 1972), most terms for elder kin in Guinean Maninka are incompatible with the tonal article (which is normally represented by a floating low tone following the noun): *ná* ‘mother’, *fâ* ‘father’, *téne* ‘paternal aunt’, *bórin* ‘maternal uncle’, *má*, *mámá* ‘grandmother’, *bénba* ‘grandfather, ancestor’, *kè* ‘husband’. Maninka texts written in Nko (where tones are accurately marked) confirm the absence of the articles with these words. Spears also points to the fact that *ná* ‘mother’ and *má* ‘grandmother’ have a preceding floating low tone.

In some Kagoro dialects, at least certain terms (*fâ* ‘father’, *bàabaa* ‘father’, *kòto* or *qòdɔke* ‘elder brother’, *bídan* ‘in-law’) are also used without the article (Vydrine 2001: 104, 121, 128).

4.2. Elder kin in Koranko and Susu

In these languages, nouns appear in most cases with a suffix *-í* or *-é*. In Koranko this is a definite or specific article, and in Susu it has evolved into a nominal morpheme.

However, in both these languages, there is a group of nouns that cannot attach the suffix.

In Koranko (Kastenholz 1987b: 206), this group includes: *bó* ‘comrade’ (age-mate?), *díyenamɔ* ‘friend’, *kàrànmɔ* ‘teacher’, *tóoma* ‘namesake’, *téne* ‘paternal aunt’, *bérí* ‘maternal uncle’, *bénba* ‘grandfather, ancestor’, *fâ* ‘father’, *ná* ‘mother’, *númɔ* ‘younger brother-in-law’.

In Susu (Touré 1989), there are several groups of nouns that appear without the nominal morpheme *-i*: French and Arabic loans; some shifters; a few names for biological species; and, finally, kinship terms and some other words for social relations: *ngá* ‘mother’, *bàábá* ‘father’, *sóxɔ* ‘uncle’, *mààmá* ‘grandmother’, *bánbá* ‘grandfather’, *tánún* ‘grandfather’, *ténèn* ‘paternal

⁹ This term appears in older sources on Susu (Raimbault 1923; Lacan 1942). In more recent ones we find another term for ‘cousin’, *dééxɔ*.

aunt', *nándén* 'mother's co-wife', *tààrá* 'elder sibling', *xúnyàà* 'younger sibling', *dééxó* 'cousin', *nìimóxò* 'younger brother-in-law', *tóxómàn* 'nickname', *kèlé* 'friend; lover', *yààtígí* 'host'. I should be noted that *ngá* 'mother' is the only noun in Susu with an initial prenasalized consonant.

4.3. Elder kin in Southwestern Mande

All the languages of this group are characterized by the phenomenon of initial consonant alternation. As a rule, each content word (noun, verb, adjective) has two forms with different initial consonants.¹⁰ These forms appear in different syntactic contexts. The initial consonant alternation in Southwestern Mande has been widely discussed in the specialist literature; see, among many other publications, (Dwyer 1974; Dwyer 1986; Kastenholz 1997: 100–104, 125–137; Vydrin 2006: 100–114). Historically, the trigger of this consonant alternation is a preceding nasal element. This element can be a syllable coda, as in the word *màsà(η)* 'chief' in (10b), or a syllabic nasal **η-* representing a grammatical morpheme: a 3SG pronoun, as in (11b), or a referential article (12b) going back to the same 3SG pronoun.¹¹ In what follows, the alternant appearing after the nasal element will be referred to as STRONG, and the other as WEAK.

Bandi (Rodewald 1989: 30)

(10a) *ɲàhà* *lókò*
REF \ woman hand
'woman's hand', historically **ɲàhà tókò*.

(10b) *màsà* *tókó*
REF \ chief hand
'chief's hand', historically **màsàη tókó*.

Mende (Innes 1971: 146)

(11a) *Ngí* *tì* *lɔ́-á*.
1SG.BAS 3PL see-PRF
'I have seen them', historically **Ngí tì tɔ́-á*.

(11b) *Ngí* *tɔ́-á*.
1SG.BAS 3SG.INAN \ see-PRF
'I have seen it', historically **Ngí ηtɔ́-á*.

Mende (Innes 1971: 36)

(12a) *Pùù* *vàndè-í* *mìà*.
REF \ European cotton-DEF be
'That is the European cotton', historically **N-pùù fàndè-í mìà*.

(12b) *Fàndè-í* *mìà*.
REF \ cotton-DEF be
'That is the cotton', historically **N-fàndè-í mìà*.

At the same time, in Mende, Loko, and Bandi there is a group of nouns whose initial consonants, contrary to expectations, do not undergo alternation: in any context, they appear with

¹⁰ In most languages of the group, there are some consonants which stand outside the consonant alternation system. They will not be considered here.

¹¹ In all Southwestern Mande languages, there are at least two articles: the "referential article" **η-*, going back to the 3SG pronoun, and the suffix *-i*, a "definite article", which most probably goes back to a demonstrative pronoun/determinative. The latter is very likely to be etymologically identical with the article *-i/-e* in other Western Mande languages as discussed above; the former is specific to the Southwestern Mande languages.

a strong initial consonant. This group includes nouns for elder relatives and some semantically close words, e.g.:

Bandi (Covac 1978: 20)

(13) *ní kéeýè*
1SG.POSS father
'my father' (rather than **ní yéeýè*)

(14) *ngì njée*
3SG.POSS mother
'his mother' (rather than **ngì yéeé*)

These nouns are given here (the lists may be incomplete).

Loko: *kèèýé* 'father', *kepa* 'maternal uncle', *nje* 'mother', *ndéýé* 'elder sibling'.

Mende: *kèké* ~ *kèê* 'father', *kéjá* 'maternal uncle', *njě* 'mother', *ngóò* 'elder sibling', *ndéwè* ~ *ndèè* 'brother', *ndíámó* 'friend', *mbăâ* 'age-mate'.

Bandi: *kèèýé* 'father', *kèýá* 'maternal uncle', *njèè* 'mother', *ndé* 'mother' (address form), *ndià* ~ *ndiyà* 'elder sibling', *mámá* 'grandmother',¹² *kàwálá* 'grandfather', *ténà* 'paternal aunt', *ndiámó* 'friend', *mbàlà* 'age-mate'.

These words usually appear without the definite article *-i* (the available data is insufficient to show whether they are compatible with the article in principle).

As we can see, the forms of the terms for elder relatives look as if they constantly appear with the referential article **ɲ*, even in contexts where the referential article would not normally be expected. In (Vydrin 2006: 139) it was suggested that we have here an archaic noun prefix **Ñ-* (presumably homonymous with the 3SG pronoun and the referential article), which marks the semantic group of elder relatives.

The complementary distribution of this marker with the definite article *-i* is an evident parallel with the incompatibility of the elder kin terms with articles in other Western Mande languages mentioned in 4.1 and 4.2. That is why in (Vydrin 2006) the prefix **Ñ-* for elder kin terms is reconstructed for Proto-Western Mande.

4.4. Elder kin terms borrowing from Mande to Landuma

As shown in 3.3, the N-nouns in Landuma were undoubtedly borrowed from Mande languages. According to the first hypothesis, the source of the initial nasal consonant might be the reconstructed prefix **Ñ-* for elder kin terms.

For sociolinguistic reasons, the main candidates likely to have donated these borrowings are Susu and/or Jalonke (Susu is the dominant language of the littoral zone in Guinea) and Manding. Mogofin and Kakabe (two closely related languages of the Mokole group) also cannot be excluded, although they are less probable candidates, given their low social status: the role of their ancestor in the past was hardly more significant. The Southwestern Mande languages are too distant from Landuma to be considered as probable lexical donors.

The borrowed kinship terms in Landuma confirm this assumption. They can be subdivided into the following groups (see 3.3 for details):

- a Susu loan: *Ncɔkɔ* 'maternal uncle';
- Manding loans: *Nna* 'mother', *Nbariŋ* 'friend', *Njatiki* 'friend', *Ntana* 'father/mother-in-law', *Ntɔkma* 'namesake', *Nkilɔ* 'Prophet';

¹² In Bandi *m* alternates with *w̃*, while in Mende *m* is a non-alternating consonant. For this reason we have no way of knowing whether the Mende word *mámá* 'grandmother' belongs to this group or not.

— Susu or Manding loans: *Nbenba* ‘ancestor’, *Ntara* ‘elder sibling’, *Ntɛnɛŋ* ‘paternal aunt’¹³.

In both the Susu-Jalonke and Manding groups, the kinship terms have no prenasalization, although its presence in the proto-language can be reconstructed. In Southwestern Mande, traces of the nasal element are more tangible (although still elusive), but direct borrowing from these languages to Landuma is hardly probable.

If the “archaic nasal prefix hypothesis” is accepted, two main conclusions can be drawn concerning the history of the Mande languages.

First, the Landuma data confirms the reconstruction of a nasal prefix in Proto-Western Mande advanced in (Vydrin 2006).¹⁴

Second, the disappearance of the prenasalized consonants in kinship terms in Western Mande languages (in particular Manding and Susu-Jalonke) seems to be a much more recent phenomenon than one might suppose, most probably dating back less than 1000 years: it must have followed the period when the kinship terms were borrowed into Landuma.

5. N- as the 1st person pronoun in Mande

Another hypothesis is much more straightforward: the word-initial nasal consonant in the Landuma kinship terms can be interpreted as a reflex of a Mande 1SG pronoun.¹⁵

A semi-vocalic nasal, most probably high-toned (i.e. * \acute{N}), can be reconstructed at least for the Proto-Western Mande level (and very probably for the Proto-Mande level too). Kinship terms, being relational nouns by definition, are rarely used in Mande languages without indication of the anchor (i.e. the individual to whom they stand in a kinship relation). As indicated by Dahl & Koptjevskaja-Tamm (2001), for kin terms, “a common case is for the anchor to be identical to the speaker of the utterance”. In Mande, the kin terms typically appear with a 1SG possessor¹⁶ and, hence, could have been borrowed into Landuma in this form. In this relation, we would like to quote Denis Creissels’ observation (p.c.):

... dans un des textes diola-fogny sur lesquels je travaille actuellement, le terme mandinka *nándiŋ* ‘co-épouse de la mère’ revient au moins une dizaine de fois sans aucune référence à un possesseur particulier, et toujours comme *nmandiŋ* ou *inandiŋ*. S’il y avait un possessif, ce serait forcément un suffixe. Or il s’agit d’un emprunt occasionnel au mandinka, pas de quelque chose qu’on pourrait

¹³ Kinship terms borrowed from Mande and retaining an initial *n-* are occasionally found in other Mel languages, cf. *ntene* ‘aunt’ in Baga Tshi-temu (Lamp 2016); *ndòdmáá* ‘namesake’ in Kisi (Childs 2000). These forms seem to be the only instances of the presence of the roots of our list in Mel outside Landuma (or at least we have not found anything else in the data available for the other languages of the family). These two forms certainly result from independent (and relatively recent) borrowing from Mande; there is no reason to postulate their presence in Proto-Mel (or even a proto-language at any lower taxonomic level).

¹⁴ In (Vydrine 1994; Vydrin 2006), this Proto-Western Mande morpheme was interpreted as an archaic noun class marker. Alternatively, it might be regarded as a kind of grammaticalized “honorific marker”. We are not going to delve here into discussion about its nature; our fundamental concern is simply the fact of the existence of this morpheme in Proto-Western Mande.

¹⁵ We are thankful to Denis Creissels for drawing our attention to the plausibility of this hypothesis.

¹⁶ We have tried to check this assumption in the Bambara Reference Corpus (Vydrin, Maslinsky & Méric 2011–2020) (accessed on April 10, 2020). *Fà* ‘father’ appears in the disambiguated subcorpus 1000 times. It is preceded by the non-emphatic 1SG pronoun *ń* 79 times, and by the emphatic 1SG pronoun *nê* 68 times (147 occurrences in total). For *bá* ‘mother’, we find 1137 occurrences; it is preceded by *ń* 87 times, and by *nê* 41 times (128 in total). These numbers may seem not very convincing, but the relatively weak cooccurrence of these kin terms with 1SG pronouns can be explained by the predominance of narratives in the Bambara Corpus. In dialogs the figures would certainly be much higher.

faire remonter à un contact ancien avec une hypothétique langue mandé, puisque les dictionnaires diola n'enregistrent pas ce terme. Le *n-* initial ne peut donc s'expliquer que comme le figement du possessif de première personne du mandinka.

6. Discussion

We have to admit that, at the present state of our knowledge, we do not see decisive arguments in favour of either of the two hypotheses. Certainly, the cognitively sound “1SG pronoun hypothesis” looks highly attractive and convincing. There are, however, some minor arguments which can be interpreted in favor of the “archaic prefix hypothesis” too.

First, all kin terms in the N-group refer to elder relations, which conforms with the proposed meaning of the archaic prefix. The absence of the nasal element in the Landuma word *dimbore* ‘cousin’ borrowed from Susu can be regarded as such an argument.

Moreover, the “1SG pronoun hypothesis” does not explain the initial nasal in the word *Nkila* ‘prophet’.¹⁷ In this particular case, the “archaic prefix hypothesis” fits better, if we assume that this prefix had some kind of honorific semantics.

Meanwhile, neither hypothesis provides any reasonable explanation for the prenasalization in the Landuma word *Nbərɸi* ‘wild boar’, unless we embark on speculations concerning the role of wild boars in the spiritual life of Landuma and/or ancient Manding speakers.

It is quite probable that both sources of prenasalization were pertinent. For some of the prenasalized Landuma forms which are not true kin terms (*Njatiki* ‘host’, *Ntɔkma* ‘namesake’), a pronominal origin for the nasal element seems more plausible. The same is true for the noun *Nna* ‘mother’, which is mainly used in the appellative function (the standard referential term for ‘mother’ is *karɔ*). At the same time, for some other nouns (e.g., *Ntara* ‘elder sibling’, *Ntɛnɛɲ* ‘aunt, father’s sister’ and the other kin terms), the prefixal hypothesis appears quite reasonable.

Abbreviations

1, 2, 3 — first, second, third person	INAN — inanimate
AGR — agreement prefix	IPRS — impersonal
AN — animate	NP — nominal prefix
ASR — assertive	POSS — possessive
BAS — basic pronominal series	PRF — perfect
CONS — consecutive	REF — referential article
DEF — definite	SG — singular
FACT — factative	SUBJ — subject

References

- Anonym. 1995. *Mandinka–English Dictionary*. Banjul: W.E.C. International.
 Anonyme. 1982. *Guide de Transcription et Lexique Bozo — Teyaxo sawananbaana yee a xarabuye*. Bamako: DNAFLA.
 Anonyme. n.d. *Lexique susu-français*. Rep. de Guinée: Service Alphabétisation et Education des Adultes.

¹⁷ *Kíra*, *kéla*, *céla* ‘prophet’ in the Manding languages is not a relational noun: it is normally separated from the possessor by a possessive marker. In the Bambara Reference Corpus there are 1915 occurrences of *kíra* ‘prophet’, and among these there is not a single one immediately preceded by the 1SG pronoun *ń* or 1PL pronoun *án*. The same result obtains for the Maninka Reference Corpus, which gives 8761 occurrences of *kéla* ‘prophet’, not a single one of which is immediately preceded by the 1SG (*ń*) or 1PL (*ń*, *án*) pronoun.

- Bris, Pierre le, André Prost. 1981. *Dictionnaire bobo-français, précédé d'une introduction grammaticale et suivi d'un Lexique français-bobo*. Paris: SELAF.
- Childs, George Tucker. 2000. *A dictionary of the Kisi language, with an English-Kisi index*. Köln: Rüdiger Köppe Verlag.
- Covac, Don. 1978. *A preliminary phonology of Bandi*. Ms. Monrovia.
- Creissels, Denis. 2009. *Le malinké de Kita (Mande languages and linguistics 9)*. Köln: Rüdiger Köppe Verlag.
- Creissels, Denis. 2010. Liste lexicale du dialonké de Faléya. *Mandenkan* 46: 49–71.
- Creissels, Denis, Sidia Jatta, Kalifa Jobarteh. 1982. Lexique mandinka-français. *Mandenkan* 3: 1–207.
- Daget, Jacques, J. Kanipo, M. Sanankoua. 1953. *La langue bozo (Études soudaniennes 1)*. Bamako: IFAN.
- Dahl, Östen, Masja Koptjevskaja-Tamm. 2001. Kinship in grammar. In: Irène Baron, Michael Herslund, Finn Sørensen (eds.). *Dimensions of possession*: 201–225. Philadelphia, PA: John Benjamins Publishing Company.
- Diané, Mamadi, Valentin Vydrine. 2012. Le système des termes de parenté soso: une étude ethnolinguistique. *Revue Horizons Faculté des Lettres et Sciences du langage* 10: 61–75.
- Dobrin, Lise. 2012. *Concreteness in grammar. The noun class systems of the Arapesh languages (Stanford Studies in Morphology and the Lexicon)*. Stanford: CSLI Publications.
- Dwyer, David J. 1974. The historical development of Southwestern Mande consonants. *Studies in African Linguistics* 5(1). 59–94.
- Dwyer, David J. 1986. Evolutionary morphology of definite articles in Southwestern Mande. In: Gerrit Dimendaal (ed.). *Current approaches to African linguistics Vol. 3*: 149–170. Dordrecht: Foris Publications.
- Erman, Anna, Japhet Kahoué Loh. 2008. *Dictionnaire Dan-Français (dan de l'Ouest) avec un index français-dan*. St Pétersbourg: Nestor-Istoria.
- Ferry, Marie-Paule, Lansana Sande. 2000. Le passé des langues: Tyapi autrefois, Kogoli aujourd'hui. In: Gérald Gaillard (ed.). *Migrations anciennes et peuplement actuel des Côtes guinéennes*: 343–352. Paris: Harmattan.
- Galtier, Gérard, Makan Dantioko. 1979. *Lexique soninké-français*. Bamako: DNAFLA.
- Grossmann, Rebecca, Michael K. Rodewald, Don Covac. 1991. *Bandi Dictionary*. Ms. Monrovia.
- Hochstetler, Lee. 1996. Enquête linguistique sur le duungoma: une langue samogo parlée au Burkina Faso et au Mali. *Mandenkan* 31: 1–57.
- Hopkins, Bradley. 1982. *Etude tonologique du yaouré (Cahiers ivoiriens de recherches linguistiques 11)*. Abidjan: SIL.
- Idiatov, Dmitry. n.d. *Dictionnaire toura-français*. Ms.
- Iffono, Aly Gilbert. 2000. Migrations mandingues dans la zone littorale guinéenne du XV^{ème} au XIX^{ème} siècle. In: Gérald Gaillard (ed.). *Migrations anciennes et peuplement actuel des Côtes guinéennes*: 77–80. Paris: Harmattan.
- Innes, Gordon. 1969. *A Mende-English dictionary*. Cambridge University Press.
- Innes, Gordon. 1971. *A practical introduction to Mende*. London: University of London, School of Oriental and African Studies.
- Jones, Ross. 2004. *Boko dictionary with reversed English-Boko finderlist*. München: Lincom Europa.
- Kastenholz, Raimund. 1987a. Materialien zum Koranko. *Afrikanistische Arbeitspapiere* Sondernummer: 1–181.
- Kastenholz, Raimund. 1987b. *Das Koranko. Ein Beitrag zur Erforschung der Nord-Mande-Sprachen*. Köln: Universität zu Köln Ph.D. dissertation.
- Kastenholz, Raimund. 1997. *Sprachgeschichte im West-Mande. Methoden und Rekonstruktionen*. Köln: Rüdiger Köppe Verlag.
- Konoshenko, Maria. 2019. *Dictionnaire kpele de la Guinée (guerzé) – français avec un index français – kpele*. *Mandenkan* 62: 3–164.
- Kuznetsova, Natalia, Olga Kuznetsova. n.d. *Dictionnaire gouro-français*. Ms.
- Lacan, Ph. L. 1942. *Grammaire et dictionnaire français-soussou et soussou-français*. Bordeaux / Conakry / Kindia: Procure des PP. du Saint-Esprit – Mission Catholique.
- Lamp, Frederick John. 2016. *Baga Tshi-tem dictionary (Guinea)*. Ms.
- Leger, Jean. 1975. *Dictionnaire guerzé*. Nzérékoré.
- Leidenfrost, Theodore E., John S. McKay. 2007. *Kpelle-English dictionary with a grammar sketch and English-Kpelle finderlist*. Moscow (USA): Palaverhut Press.
- Makeeva, Nadezhda. n.d. *Dictionnaire kla-dan – français*. Ms. (Toolbox file).
- Mara, Facely, Camara Facely. 1979. *Étude descriptive du Lélé*. Kankan: Institut polytechnique “Julius Nyerere” Mémoire de master.
- Meyer, Gérard. 1983. *Lexique élémentaire Malinké-Français, suivi d'éléments de grammaire. Dialecte maninkaa du Niokolo (sous-préfecture de Bandafassi)*. Kédougou.
- Nikitina, Tatiana. n.d. *Dictionnaire wan-français-anglais-russe*. Ms.

- Paperno, Denis. n.d. *Dictionnaire beng-français*. Ms.
- Perekhval'skaya, Elena, Moïse Yegbé. 2018. Dictionnaire mwan-français. *Mandenkan* 60: 3–122.
- Polinder, Wim, Jacqueline Janse, Kees Jan van Linden. 2009. *Lexique Mògòfin-Français*. Garama: Mission Évangélique Réformée Néerlandaise.
- Raimbault, R.P. 1923. *Dictionnaire français-soso et soso-français. 2nd edition*. Rome: Sodalité de St. Pierre Claver.
- Rodewald, Michael K. 1989. *A grammar of Bandi and Mende tone*. Arlington: University of Texas M.A. Thesis.
- Rogers, Kirk. 2005. *Introduction to Landuma Language and Culture Lessons*. Ms.
- Rogers, Kirk. 2008. *Phonemic Analysis of the Landuma Language*. Ms.
- Rogers, Kirk, Daniel Bryant. 2012. *Dictionnaire landouma — français*. Boké: Mission Évangélique de Boké.
- Sande, Hannah L. 2019. Phonologically determined nominal concord as post-syntactic: Evidence from Guébie. *Journal of Linguistics* 55(4): 831–878.
- Schreiber, Henning. 2008. *Eine historische Phonologie der Niger-Volta-Sprachen (Mande Languages and Linguistics / Langues et Linguistique Mandé 7)*. Köln: Rüdiger Köppe Verlag.
- Smeltzer, Brad, Susan Smeltzer. 2001. *Lexique Soninké-Français-Anglais. Soninke-French-English Dictionary. Index Français-Soninké, English-Soninke*. Bamako.
- Solomiac, Paul, Ali Traoré, Fabé Traoré. 1998. *Dzùùngoo dzuon tsùnyfyeè jàn ko à sebeun'n kur'la. Lexique orthographique dzùùngoo-français*. Ouagadougou: SIL.
- Spears, Richard. 1972. The form of Mandingo kinship terms. *Anthropological Linguistics* 14(7): 281–285.
- Suret Canal, Jean. 2000. Note sur les Kogoli. In: Gérald Gaillard (ed.). *Migrations anciennes et peuplement actuel des Côtes guinéennes*: 333–341. Paris: Harmattan.
- Touré, Aboubacar. 1989. *Éléments de phonologie et morphologie de la langue sòsò (Études mandé 3)*. Grenoble: Université Stendhal.
- Touré, Aboubacar. 1994. *Éléments de description de la langue soso*. Grenoble: Université Stendhal-Grenoble III Ph.D. dissertation.
- Tröbs, Holger. 1998. *Funktionale Sprachbeschreibung des Jeli (West-Mande) (Mande Languages and Linguistics / Langues et Linguistique Mandé 3)*. Köln: Rüdiger Köppe Verlag.
- Tveit, Harald, Gagny Dansoko. 1993. *Petit dictionnaire khassonké-français*. Oussoubidiagna: Mission Protestante Norvégienne.
- Vydrin, Valentin. 2006. К реконструкции фонологического типа и именной морфологии пра-манде [Toward reconstruction of the phonological type and nominal morphology of Proto-Mande]. In: Nikolaj Kazanskiy (ed.). *Mandeica Petropolitana (Acta Linguistica Petropolitana. Trudy Instituta lingvosticheskikh issledovaniy RAN [Acta Linguistica Petropolitana. Transaction of the Institute for linguistic studies] II (2))*: 3–246. St. Petersburg: Nauka. Online at: https://alp.iling.spb.ru/static/alp_II_2.pdf.
- Vydrin, Valentin. 2009a. Esquisse de la langue lélé (groupe mokolé). *Mandenkan* 45: 29–104.
- Vydrin, Valentin. 2009b. Lele: predstavlenije malogo jazyka semji mande [Lele: presentation of a minor Mande language]. In: Valentin Vydrin (ed.). *African Collection – 2009*: 235–268. St. Petersburg: Muzej antropologii i etnografii RAN (Kunstkamera).
- Vydrin, Valentin, Kirill Maslinsky, Jean-Jacques Méric. 2011–2020. *Corpus Bambara de Référence*. Online at: <http://cormand.huma-num.fr/index.html>.
- Vydrina, Alexandra. 2015. *Dictionnaire de la langue kakabé suivi d'un index français-kakabé. Mandenkan: Bulletin d'études linguistiques mandé. N 53*. Ms., online at: <https://llacan.cnrs.fr/PDF/Mandenkan53/53vydrina.pdf>.
- Vydrine, Valentin. 1994. Traces of Nominal Classification in the Mande Languages: the Soninke Evidence. *St. Petersburg Journal of African Studies* 3: 63–93.
- Vydrine, Valentin. 2001. *Esquisse contrastive du kagoro (Manding) (Mande languages and linguistics 4)*. Köln: Rüdiger Köppe.
- Welmers, W. E., C. K. Kandakai. 1974. *A Vai-English dictionary (preliminary draft)*. Ms. Monrovia.
- Willits, Brad. n.d. *Dictionnaire susu-français-anglais*. Ms. (Toolbox file).

Н. Р. Сумбатова, В. Ф. Выдрин. Существительные с начальным носовым согласным в ландума и их соответствия в языках манде

В статье анализируются термины родства в языке ландума (семья мел), распространённом на северо-западе Республики Гвинея. Эти существительные характеризуются не-

стандартной фонологической структурой: они начинаются с консонантного кластера структуры NC. Делается вывод, что эта аномальная особенность может быть объяснена происхождением этих существительных: все они оказываются заимствованиями из языков манде.

В западных языках манде названия старших родственников тоже нередко отличаются аномальным поведением, что проявляется в невозможности присоединения определённого или референтного артикля. Ранее уже высказывалось предположение, что такая аномалия может объясняться присутствием архаичного назального префикса, который был грамматическим маркером этой семантической группы слов. Однако такой назальный префикс не был обнаружен ни в одном современном языке манде.

Для объяснения происхождения начального носового элемента в аномальных существительных ландума можно выдвинуть две гипотезы. В соответствии с первой из них, этот элемент восходит к архаичному назальному для старших родственников, реконструируемому для языков манде. В таком случае данные ландума может считаться свидетельством того, что этот префикс в языках манде исчез сравнительно недавно (уже после начала интенсивных контактов между носителями пра-ландума, пра-манден и прасусу-джалонке). Альтернативная гипотеза возводит начальный назальный элемент ландума к местоимению 1 лица единственного числа *í*, которое существует в соседних языках манде и часто употребляется с терминами родства. Можно также допустить, что могли быть задействованы оба эти источника.

Ключевые слова: язык ландума; западные манде; термины родства; языковые контакты.

