

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

N° 20/1



Moscow 2022

Российский государственный гуманитарный университет
Институт языкознания Российской Академии наук

Вопросы языкового родства

Международный научный журнал

№ 20/1



Москва 2022

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., 2022. — № 20/1. — x + 80 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, 2022. — No. 20/1. — x + 80 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 30.03.2022. Формат 60×90/8.
Усл. печ. л. 11,3. Уч.-изд. л. 7,6. Заказ № 1592.
Тираж 1050 экз.

Издательский центр
Российского государственного гуманитарного университета
125047, Москва, Миусская пл., 6
www.rggu.ru

Table of Contents / Содержание

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

Articles / Статьи

<i>Blanca María Prósper</i> . The sibilant sounds of Hispano-Celtic: phonetics, phonology and orthography	1
[Б. М. Проспер. Сибиллянты в испано-кельтских языках: вопросы фонетики, фонологии и орфографии]	
<i>George Starostin</i> . Lexicostatistical studies in Khoisan II/2: Towards a more precise phylogeny for the Tuu family	25
[Г. С. Старостин. Лексикостатистические исследования по койсанским языкам II/2: к уточнению филогенетической классификации языков семьи ту]	
<i>Mikhail Zhivlov</i> . Studies in Yukaghir etymology I	71
[М. А. Живлов. Исследования в области юкагирской этимологии I]	

Contributors

Blanca María Prósper — faculty member, University of Salamanca, indoling@usal.es

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

Mikhail Zhivlov — candidate of sciences (Philology), lecturer, Institute for Oriental and Classical Studies, Russian State University for the Humanities; senior researcher, Institute for Oriental and Classical Studies, HSE (Moscow), zhivlov@gmail.com

Сведения об авторах

Живлов, Михаил Александрович — канд. филол. наук, доцент Сектора компаративистики Института восточных культур и античности РГГУ, старший научный сотрудник Института классического Востока и античности ВШЭ (Москва), zhivlov@gmail.com

Проспер, Бланка Мария — сотрудник Университета Саламанки, indoling@usal.es

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

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
125047 Moscow, Russia
Miusskaya 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 прикладываются 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:

125047 Москва
Миусская площадь, д. 6
Российский государственный гуманитарный университет
Институт восточных культур и античности
В редакцию журнала «Вопросы языкового родства»

The sibilant sounds of Hispano-Celtic: phonetics, phonology and orthography

On the strength of both recent discoveries concerning the use of the Latin alphabet in different places of Central Hispania to write official documents in Celtiberian and the new readings of indigenous names on inscriptions belonging to Latin, not Celtiberian, epigraphy, some novel reflections on Celtiberian phonology are in order. Epigraphic and linguistic considerations in turn lead to a refinement in the delimitation of Hispano-Celtic regions. These attend to dialectal differences and to the emergence and stabilisation of scribal habits. New etymologies for some hitherto uninterpreted or even misinterpreted personal names are put forward. Some of these have the comparative advantage of matching inherited Celtic forms surviving in Insular Celtic and Gaulish. Finally, three new readings are proposed: BVGAN-SONIS, CLOVSOCVM and AISAE, possibly also VXSEISVS.

Keywords: Celtiberian languages; Indo-European reconstruction; Celtic etymology.

1. Introduction¹

The fate of the Indo-European sibilants, and clusters containing sibilants, in Celtiberian (or, broadly speaking, Hispano-Celtic)² has been a bone of contention for thirty years. The discovery that the use of *san* and *sigma* for the two sibilant phonemes was not arbitrary, as a rather vague scholarly consensus had dictated *faute de mieux*, but followed a precise, discernible pattern, shattered the foundations of the discipline (cf. Villar 1995). From this moment on, I shall use Villar's notation <s> (for *san*) — <z> (for *sigma*).³

The subsequent discovery of the existence of «lost» cases in the nominal inflection, specifically an ablative singular (spelt *-az*, *-uz*, *-ez* and *-iz*), caused commotion (and, I have to say,

¹ As always, I want to thank the editor and two anonymous reviewers for their work. N.B.: Texts written in the Latin alphabet are rendered in SMALL CAPITALS; texts in the Iberian script, here labeled «Early Celtiberian», are in *italics*. The sign <χ> used for Celtic reconstructed forms stands for a possibly uvular voiceless fricative sound that goes back to IE /p/ in coda position preceding alveolar obstruents. Abbreviations of special interest are: CCelt. = Common Celtic; HCelt. = Hispano-Celtic. For syllabograms of stop + vowel, the traditional, overarching system will be used, which assumes that the Celtiberian script does not distinguish between voiced and voiceless stops. This may be simply interpreted as an “archigraphemic” transliteration by those who prefer to see internal differences in some texts. I shall make an exception when a whole text (specifically in the area of the Arevaci) can demonstrably make sense by consistent application of the so-called “dual system”.

² Hispano-Celtic territories exhibit interesting differences in vocabulary, phonetics, onomastic subsystems (e.g. numeral-based names), etc. For the time being, however, we are unable to detect non-trivial differences going back to Common Celtic that are suggestive of different Celtic language subfamilies having penetrated into Hispania at different stages. Under the most economic assumption, one or more waves of speakers of Celtic traversed the Pyrénées, but most of the traits separating Hispano-Celtic from the rest of Celtic belong to a more recent period.

³ This option is favoured over <ś> — <s> in *MLH I* (which is counterintuitive) and <s> — <ð> in *MLH IV* and *V* (which benefited from Villar's discoveries but has turned out to be misleading and simply does not do justice to the phonemic contents of at least the second of these signs). Celtiberian texts will, when possible, be cited according to *MLH I* (coins: inscription numbers are preceded by A.) and *MLH IV* (rest of texts: numbers are preceded by K.).

stubborn disbelief in some quarters, which in a few cases endures to the present day). Nonetheless, in the meantime Villar's acclaimed work has been partly called into question and refined to some degree. In my view, some of the criticism is justified and several of his tenets need correction. Others, however, will remain pending issues for the foreseeable future.

In Celtiberian texts written in the Iberian script (from now on referred to as «Early Celtiberian»), the sign <s> reflects the outcome of Indo-European /s/ in initial and final position (and probably in onsets after voiceless obstruents) and of several clusters in medial position. For a number of scholars, it also represents the unchanged, voiceless outcome of /s/ in intervocalic position, mainly on the strength of *soisum* 'of these, of them', gen. pl. masc. of the anaphoric pronoun *so*, ultimately going back to Indo-European **toisōm*.

By contrast, <z> is a cover symbol that represents the outcome of Indo-European /d/ and /d^h/ in intervocalic and final position (its phonemic status is, however, uncertain; see below), and /t/ in final position. In addition, <z> systematically renders an affricate or fricative phoneme going back to *-t(i)̄-* (cf. Prósper 2014, with literature), and alternates with 'historical' <ti>, often in the same names, for unknown reasons, related to dialectal or sociological differences.⁴ Finally, according to Villar (followed by Prósper 2014) <z> also represents the voiced outcome of /s/ in some contexts in medial position:

- 1) *arznas* < **φarsnās* < **p̄r̄snās* 'parts, ground lots' (K.1.1, Botorrita), a match of OIr. *rann* that is strongly suggestive of the fact that the metathesis **φars- > *φras-* is a late, probably Insular Celtic phenomenon⁵;
- 2) *kabizeti* 'may take' < **kabi-s-e-ti* (K.1.1., Botorrita; cf. 3rd p. pl. CABINT in Novallas);
- 3) *uerzaizokum* (K.1.3, Botorrita) is a family name probably going back to **uer-s-* 'better, higher', in OIr. *ferr*. It may be identical to Goth. *waírsiza*, OHG. *wirsiro* 'worse', Skt. *varṣīyas* 'better' (< **uers-is-*) with an intrusive /a/ found in Hispano-Celtic names going back to comparatives and superlatives.⁶ In superlative forms in **-isamo-*, the sibilant may have undergone expressive strengthening, like in Latin (as opposed to weakening in the rest of Italic), contrasting with *-aiz-* in the comparative.⁷
- 4) *uerzoniti*, subjunctive form of a present **uper=sonh₂-ēie-* (K.1.1., Botorrita), cf. Gaul. SONITI (RIG II, L- 101, Lezoux), possibly U. 3rd p. impv. SONITV/*sunitu* (*Tabulae Iguvinae*) 'cause to obtain?'. This example is especially interesting because a root beginning with /d/ is virtually ruled out: since the first phoneme in the second member of a transparent compound usually behaves as if it stood in word-initial position, /d/ is preserved as a stop in every context potentially conducive to lenition: cf. *uertatuz*, *uertatos* < **d^h₁-* 'put' (K.1.1, Botorrita), *ambitinkounei*, *ambitiseti* < **d^hing^h-* 'build' (K.1.1, Botorrita), *taruotureska* (K.23.2, Osma) = TARVODVRESKA. Accordingly, the rejection of the above etymology by LIV: 533, **senh₂-* 'erlangen, erwischen', with fn. 11, and the proposed asso-

⁴ From now on I shall refer to this sequence as *-t̄i-*, since loss of syllabicity is likely to have happened early in Celtic.

⁵ The original idea must be credited to Eichner (1989: 33–34). The alternative reconstruction **ard-nā* has no visible cognates.

⁶ The existence of comparative/intensive grades of adjectives in *-ais-/aiz-* is ascertained by such personal names as TVRAESIVS 'very strong' (EDCS-22800201, Guadalajara; EDCS-03700402, Cáceres, etc.), *vis à vis* superlative TVRAESAMVS (CIL II: 2957, Contrasta, Álava). The family name *akaizokum* (K.0.14) *vis à vis* Gaul. ACISIVS is suggestive of a Late PIE comparative **ōk-is-* in L. *ōcior* 'faster'.

⁷ Nevertheless, while superlatives contain <s> in Early Celtiberian, examples found in the Latin alphabet show a single <s> and not a geminate, as in BERISAMO (EDCS-11701112, Orense, Callaecia Lucensis), TVRAESAMVS (CIL II: 2957, Contrasta, Álava), BLETISAM(ENSES) (CIL II: 859, Ledesma, Salamanca), but the fact that the attested forms are mostly peripheral blurs the picture.

ciation with **d^hen₂-* (LIV: 144, ‘sich in Lauf setzen’, only attested with certainty in Indo-Iranian and Greek), simply beg the question, in that etymological suggestions are restricted in accordance with MLH V’s arbitrary transliteration *uer₂doniti*. Since a prefix *uer-* must have been as transparent in this as in other forms, the resulting form would have been *tuertoniti*. But then, why not *tuersoniti*? One could venture the possibility that compound forms of this verb were far more common than the simplex or that the latter was no longer used; the lenited outcome of /s/, once it became phonemic, could occur in word-initial position and became the default anlaut for this root (note that this was not necessarily a neutralisation context, since there may have been ancient structures containing the sequence *-rt^st-*, like **uper=d^hh₁-tó- > *uersso-*).

- 5) At any rate, we have no independent evidence for an intramorphemic sound change *-rd- > -r₂đ-*. We might have an instance of *-rd- > -r₂đ-* in the name *burzu* (K.1.3, Botorrita) if it is equivalent to BVRDO (for instance in a vast number of attestations of a Gaulish potter’s name), and then not to Celtic **burro-* ‘inflated, strong’ (cf. OIr. *borr*, if from **b^horsó-* in PGerm. **barzaz* ‘breaking through’).
- 6) The existence of a place name *burzau* on Celtiberian coins (A.48), which unequivocally matches the base of the ethnonym *Bursaonenses* (Pliny 3, 24, nowadays *Borja* in Saragossa), is definitive proof of the existence of a sibilant spelt <z> in Early Celtiberian, and this would still apply even if the place name were demonstrably Iberian. However, if this were the case, the foreign sibilant could have been identified with /s̄/ < -t̄i- (see below), which is immaterial to the problem of Celtib. /z/.

The best counterexample to the voicing rule thus far is the anaphoric pronoun *soisum* (gen. pl. < CCelt. **soisūm*, K.1.3, Botorrita). Besides assimilatory devoicing, there is a possible explanation for the anomaly: Celtiberian may have had a gen. sg. **sosio*, which would regularly show strengthening of medial /s/ (as in *makasiam*, K.1.1.). If **sosio* was metathesised into **soiso* by analogy with the plural form, it could in turn have caused a homogenisation of sibilants in the pronominal paradigm.

To recap: the fact that the lenited outcome of /s/ is reflected in writing suggests that it was underlyingly phonemic. The rise and completion of inter- or intra-paradigmatic analogical processes also crucially depends on this. In addition, orthography may have often resorted to phonemic, not morphophonemic spelling, which would for instance explain the differences in *arznas* (< **φarsnā-*) and *masnai* (< **mad-snā* or **mag-snā*; both in K.1.1, Botorrita), assuming these were still perceived as *-snā-*derivatives.

2. New evidence, new scripts... and new signs for old scripts

A recently unearthed Late Celtiberian document, the so-called bronze of Novallas (Saragossa), probably to be dated not later than 50 BC, has unexpectedly brought to light nothing less than a regional, hitherto unknown version of the Latin alphabet. In my view (Prósper 2017), it contains a description of the general design, size and trajectory of a Roman road built by Roman engineers long after Roman rule had been established in most of Hispania, and was written at a time when the Latin language was not yet definitively imposed for official documents, but the Iberian script had already been abandoned.⁸

The main distinctive trait of this alphabet is the use of a «barred <S>» that characterises, for instance, the ablative endings, and in this way successfully distinguishes the outcome of

⁸ Cf. also the recent edition by Beltrán Lloris et al. 2021.

Indo-European *-Vd* (where the final consonant is represented in Early Celtiberian by <z>, and in the new alphabet by <s>) from that of *-V(n)s* (represented in Early Celtiberian by final <s>, and in the new alphabet by <S>). Jordán Cólera (2016) has also identified the «barred <S>» in other Celtiberian texts in the Latin alphabet, e.g. the rock inscription of Peñalba de Villastar, Teruel (K.3.3).

In my view, formerly expressed in Prósper (2016: fn. 145), this regional version of the Latin alphabet originally constituted a unique means of *transcription* of the Iberian script. This new system may have been implemented by indigenous Celtiberian scribes at the Romans' request, probably some time before the Latin alphabet was finally imposed *and* the official documents switched to Latin. It preserved the digraph <EI> for IE /eᵢ/ and <QV> for /kʷ/, at a time when indigenous names already showed <E> and <P> in Latin funerary and votive epigraphy (incidentally coinciding with the Gaulish outcomes, which suggests that there was already some amount of allophonic variation when Early Celtiberian became a written language two centuries earlier). In addition, it introduced a new sign <S> in order to render a phoneme for which no Latin letter was available, and at the same time to reduce ambiguity in the representation of phonemic contrasts.

When we compare the Iberian and this particular Latin script, we find that Celtic intervocalic /d/ is respectively spelt <z> and <D>, but, as observed above, /d/ in final position is spelt <z> and <S>. In Novallas, we find ODAS < **pod-* 'feet'; BEDAM < **bedā-* 'road'; MEDOM < **medo-* 'middle'. Again, we are at a loss as to whether the Early Celtiberian use of <z>, which reflects a process of fricativisation of /d/ in intervocalic position, constitutes a very rare example of allophonic writing. This *could* be explained by accepting that the use of alphabetic as opposed to syllabographic writing drastically reduced ambiguity (since the reader did not have to guess whether the vowel in the syllabogram was mute or not). At any rate, there is at least one case of initial <z> from /d/: *zizeti, zizonti* 'may give' < **dideti, *didonti* (respectively Iniesta and K.1.1., Botorrita).

By contrast, the ablative forms attested in Novallas are VSAMVS < **uχsamūd* 'from above'; TERGAS, and not, as we would have predicted, †VSAMVD and †TERGAD. This allows us to draw an interesting conclusion: the scribes who designed this hybrid script did not resort to the expedient of writing *-AD, -VD* (which would have been much easier to distinguish at first glance from *-AS, -VS* than the innovative *-AS, -VS*) for some important reason. This points to lenition of final *-d* and, quite conceivably, to its attribution to another phoneme somewhere down the line, a point that could not be clarified thus far, since, as observed above, <z> functioned in Early Celtiberian as a cover symbol for more than one fricative. I shall come back to this problem at the end of this work.

3. How to write the indigenous names of old times in Latin epigraphy: what is the use of «crossing» the letter <S>?

In recent years, it has been argued (cf. Simón Cornago & Jordán Cólera 2018) that a limited number of indigenous names found in Latin epigraphy show the intervocalic sequence <SS>, which in most cases goes back to *-t̪i-*.⁹ Their showcased examples thus far are PRESSVS, SEGOS-
SOQVM, NISSICVM and DERCINOASSEDENSIBVS, to which a new instance TELASSICVM has most recently been added (Jordán Cólera & Díaz Ariño 2022). Let us examine their examples in detail.

⁹ They always speak in phonetic terms; when they vaguely speak about letters in different scripts representing [ð], [θ], or the affricates [dʒ], [tʃ], the reader is often lost as to their respective etymological origins and, crucially, the synchronic phonemic contrasts.

In point of fact, the first one looks doubtful. PRESSVS is mostly attested in Hispania (except for one case in Dijon, Gallia Belgica/Germania Superior), though PRISSIA and PRISSO are well attested in Belgica and Germania. The authors have missed a number of related names, like PRESSILLA in a Latin onomastic context and one AEMILIA PRESSA in Catalonia (*CIL* II: 4469, Isona/Aeso, Lérida). In spite of their efforts, this name has no obvious Celtic etymology, and it may well be simply Latin since, to begin with, the four extant Hispanic cases of PRESSVS are inserted in a purely Roman onomastic context.¹⁰ If, however, one sticks to an indigenous interpretation of this form, and, crucially, if at least *one* example contains <SS>, a conceivable preform would be the derivative in *-(i)jō-* of a very archaic root participle **kʷr-ent-* (*LIV* 391 **kʷer-* ‘(ab)schneiden, schnitzen’). *Non liquet*.

SEGOSSOQVM (*CIL* II: 5790, Buenafuente del Sistal, Guadalajara) was identified with *sekonzos* (K.1.3, Botorrita) for the first time by Ballester (2003), who correctly reconstructed a participial adjective **segont-(i)jō-* ‘victorious’ before the present reading <SS> had even been considered. Both reading and etymology are undoubtedly correct. An abbreviated SEGOSS(VS) from Medellín (Cáceres, Lusitania Emeritensis), edited by Saquete Chamizo — Guerra Millán (2015), is most likely to be the same name. Early Celtiberian attests a place name in the ablative case *sekotiaz* on coins (A.77) ‘from Segontia’, as opposed to the individual names *sekonzos* and *sekontios* (K.1.3, Botorrita). Examples of an indigenous personal name SEGONTIVS abound in northern Celtiberia and its periphery (Burgos, Álava). The fact that the place name *Segontia* has become present-day *Sigüenza* (Guadalajara) may prove irrelevant to this problem: it must have been identified with Roman place names of good omen early in the 2nd c. BC, and adapted as **segontijā* or even **segontjā* with ‘Vulgar Latin’ phonetics (anticipating the hiatus resolution presupposed by all the Romance languages). As has been observed in a number of works, CCelt. **-jō-* had undergone loss of syllabicity in Hispano-Celtic or earlier, so that innovative Latin phonetics would roughly coincide with conservative Celtiberian phonetics.¹¹

DERCINOASSEDENSIBVS VICANIS CLVNIENSIVM (Peralejo de los Escuderos, Soria; cf. *ERS*: 133). The etymology of the second member of the compounded place name underlying this *origo* is obviously **ad-sedo-*, probably meaning ‘settlement, see’. The same form is found in Gaulish onomastics, where it is spelt in various ways, as in the personal name ASSEDOMARI (*CIL* III: 3291, Noricum), AÐÐEDOMARI (*EDCS*-11401186, Aquitania), ADSEDI, ADSEDILI (*CIL* III: 4847, Noricum). This means that *-ds-* underwent a late regressive assimilation, and that, consequently, the resulting geminate sibilant preserves dental features. If, conversely, one were to argue that <SS> stands for an affricate, there is no explaining why the orthographic spelling <DS> did not survive, or why an alternative <TS> was not substituted for it following regressive assimilation of voice, as in the provincial «title» ATSERTORI (reflecting *adsertorī* ‘protector’)

¹⁰ They partly misunderstand the arguments put forward in Prósper (2016): while the study focuses on western Celtiberia and the Cantabri, their assertion (fn. 71) «another, separate question is to accept the isogloss **kʷ-* (and **kw-*) > *p-* proposed by B. M^a Prósper, *The Indo-European names* (note 64) 123–198, which would affect the western Celtiberia that the author delimits» is off the mark. The original text reads on p. 119 «Labialization of the voiceless labiovelar is a late, but probably pan-Hispano-Celtic feature outside early Celtiberian (*i. e.*, transmitted in the Iberian script)», and there is no mention of an isogloss covering a specific area of Celtiberia. As for the rest, they have wholesale ignored virtually every other argument about Celtiberian sibilants put forward in the book.

¹¹ There are of course other names with <nz> from **-nt-(i)jō-* not attested in later epigraphy: the personal name *Melmanzos* (K.1.3, Botorrita) goes back to **menmant-(i)jō-* (cf. the divine name MINMANTIS, dat. pl., Périgueux, *CIL* XIII: 940), from **menm̄-tó-* ‘having sense’. Fricativisation of the cluster with loss of the preceding nasal and regressive assimilation of /nm/ are patent in the potter’s name MEMASVS (*CIL* XIII: 12014, 60 a/b, Noricum, Pannonia). The place name *Mazonza* (A.15), on several coins found in Alcañiz (Teruel), can be unproblematically traced back to **madont-(i)jā*.

awarded to the young Lollianus in the interesting *mischsprachlich* text reading [LOL]LI/ANO DVRE/TA SALDAN/ICA ATSE[O]/RI IVVENTVT/IS (written by a Vaccaeus from Saldania and found in León, see Martínez Chico & Prósper 2021 for details).

In sum, *DERCINOASSEDENSIBVS* originally contained a heteromorphemic and heterosyllabic sequence *-d.s-*: the dental and sibilant segments occur across a morpheme boundary, which neatly explains the preservation of dental articulation after regressive assimilation (if we provisionally accept that CCelt. *-ds-* would otherwise become */s:/*). Other instances of heteromorphemic *-d.s-* are mostly roots beginning with *s-* prefixed by *ad-* and roots ending up in *-d-* and followed by a suffix *-s-*. They are rendered as *<s>* in Early Celtiberian, which points to preservation of the cluster *-d.s-*, possibly realised as [t.s], when Celtiberian became a written language: cf. the subjunctive form *robiseti* < **φro-bid-s-e-ti* < **b^hid^(h)-* (K.1.1, Botorríta), and *asekati* < (hyperthematic) **ad-sφek-ā-(i)-ē-ti*¹² ‘may envisage’ < **spek-* (K.1.1, Botorríta).

These subjunctive forms are, of course, not isolated, and as a consequence the above argument does not incur circularity: other sigmatic subjunctives are *kabizeti* ‘may take’ < **kabi-s-e-ti* (K.1.1, Botorríta; cf. 3rd p. pl. *CABINT* in Novallas),¹³ *ambitseti* ‘may surround, build around’ < **diχse-* < **d^hig^h-s-e-* (K.1.1, Botorríta).¹⁴

Subjunctive forms found in the same syntactic slots (subordinate clauses headed by ‘if’ or ‘whoever’) are as follows.

Hyperthematic subjunctives: *uerzoniti* (corresponding to an IE indicative **uper=sonh₂-eie-* (K.1.1., Botorríta); *asekati* < (hyperthematic) **ad-sφek-ā-(i)-ē-ti*¹⁵ ‘may envisage’ < **spek-* (K.1.1, Botorríta). A similarly built form is *kuati* (K.1.1, Botorríta), in my present view corresponding to the indicative *(*s*)*kouh₁-eh₂-ie/o-* ‘takes a look at, oversees’, identical to OHG. *scourwôn* ‘schauen’ and then a Germano-Celtic isogloss; both are denominative to *(*s*)*kouh₁-eh₂*.¹⁶ At any rate, a subjunctive to the indicative zero-grade primary formation found in Skt. *ā-kuvate* ‘to have in view’ cannot be ruled out.

Thematic subjunctives built from athematic present or aorist indicatives: *zizeti*, *zizonti* ‘may give’ < **di-dh₃-e-ti*, **di-dh₃-o-nti* (respectively Iniesta and K.1.1., Botorríta), with generalised zero grade of the root and restoration of the alternation *o/e* regardless of the preceding laryngeal, as opposed to the athematic imperative forms *tatuz* (K.1.1, Botorríta; Iniesta lead) and *tizatuz* (BB.IV, Botorríta), probably < *(*d^(h)i-*)*d^(h)h_{1/3}-tōd* ‘place, pay, give?’.

Another interesting case in point is *auzeti* (K.1.1, Botorríta), a thematic subjunctive endowed with primary endings, corresponding to the indicative pret. 3rd p. sg. *auz* < **auð* < **auð* <

¹² Assuming that the outcome of the contraction of long vowels of different colours is not wholly predictable. Even if the second vowel often predominates, this may not necessarily be the case, as in Greek. In fact, this is why both the 3rd p. sg. of the indicative and the (hyperthematic) subjunctive of Gk. *τιμάω* ‘to honour’ are identical: *τιμάᾱ*. Note, in addition, that *-ā-(i)-ē-* need not have become a diphthong */ai/* or */a:i/*, since the change */e:/* > */i:/* is not as early as usually assumed. By the same token, unless one chooses to reconstruct this conjugation as athematic **-ā-ti*, a thematic indicative *-ā-(i)-e-* would have undergone the same change.

¹³ Note that the *ad hoc* reconstruction of a preform **kam-bid-* in MLH V, 144 is untenable, since the IE prefix **k_m-* is nonexistent.

¹⁴ The question remains undecided whether these are aorist subjunctives or, as their zero root vocalism would seem to suggest, subjunctives built from athematic desideratives. The comparison of the stem **kabis-* with the desiderative **kapis-* in L. *re-cuperāre* ‘recover’ (Nussbaum 2007: 4) opens new prospects.

¹⁵ See fn. 12 above.

¹⁶ By contrast, L. *caveō* and Gk. *κοέω* are usually considered iterative and not denominative formations in spite of the conceivable essive/stative derivation Gk. *-σκός* ‘overseer’ → *-e-ῖo/e-*, *-eh₁-ῖo/e-* ‘be an overseer, watch over, see’. Cf. on the last forms Vine 2006 and Garnier 2010: 441. Assimilatory reduction *-oua-* > *-uua-*, progressively conducive to loss of syllabicity, is otherwise well attested in vast areas of central and south-western Hispano-Celtic (cf. Prósper & Medrano Duque 2022).

**aud-t* (with regular, probably early loss of the ending *-t* in complex codas, which I deem more likely than a prevocalic outcome *-ss*: see below).¹⁷ The 3rd p. pl. *auzares* (K.0.14, Bronze ‘Res’) reflects Celtib. **auð-ares* << **aud-ars* < **h₂eud^h-rs*.¹⁸ The 3rd p. pl. pass. pret. was *auzanto* < **auð-anto* < **h₂eud^h-nto* (K.1.3, Botorrita). All these forms contain a Celtic neo-root **aud-* that goes back to **h₂eud^h-h₁*, possibly meaning ‘to issue’, in which, contrary to most former treatments of these forms, the laryngeal does not play a role any more. The fact that Novallas attests a present participle in the gen. pl. AVDINTVM points to the existence of a suffixed present stem (probably in *-io/e-*). The segment *-d-* was no longer interpreted as the second member of a compound, and hence became a fricative (see Prósper 2016: 188, fn. 152, which rendered my former analysis **h₂eud^h-s-* obsolete; and, for the general concept, Hackstein 2002).

The family name NISSIC[VM] in the construction LVCIV[S] NISSIC[VM] ACCVT[I] F (Almadrones, Guadalajara) has been traced back to **nit̃jo-* ‘inner’ by Prósper (2016: 144, where the spelling <ss> went unnoticed). Simón Cornago & Jordán Cólera (2018: 198) seem to find this etymology acceptable. While a preform **nit̃jo-* was in itself not implausible before the use of <barred <ss>> was ascertained for this name, other alternatives may now be considered, and this form would be more likely to have been reflected as †NISICVM or possibly †NISICVM (see below). Let us explore further options: the compositional scheme «prefix + past part. **d^hh₁-tó-*» may have been productive to some degree in Hispano-Celtic, witness *ensikum* (K.1.3, Botorrita, as if from **h₁en(i)-d^hh₁-tó-*), *eisubos* (Torrijo del Campo, as if from **h₁epi-d^hh₁-tó-* ‘resident’?).¹⁹ In the same vein, we cannot rule out a reconstruction **ni-d^hh₁-tó-* ‘settled’. For conceivable Eastern Gaulish parallels, cf. NISSA (Aquincum, Pannonia Inferior; EDCS-51400615: 16), NISS[-] (Dacia, CIL III: 6268). Unfortunately, in that case we would definitely expect †NISICVM instead of the attested NISSICVM, where <barred <ss>> shows up for some reason. In sum, the last word on this name has not been pronounced yet.

On the other hand, this form could conceivably be traced back to **nenti-jo-*, from Celtic **nenti-*, **ñt̃ej-* ‘fight’. Gaulish nouns in **nanti-*, attested in compounded personal names, have occasionally been identified with OIr. *néit* ‘fight’ (see LEIA-N). For his part, Hamp (1976: 14) has connected these Celtic forms with Goth. *ana-nanþjan* ‘to dare’, from **nont-*, OHG. *gi-nindan*, from **nent-*, OE. *nóþ* ‘courage’ from the zero grade **nunþ-* < **ñt̃-*), and on the strength of this comparison has set up «a good Germanic-Celtic etymon **nent-* ‘to be bold, aggressive’» (on this form see also Irslinger 2002: 226). While the Irish form has been traced back to **nanti-* with the zero grade of the root, it may contain the full grade /e/ equally well, and this would have yielded the attested form, as in OIr. *cét* ‘first’ < **kentV-*, etc.²⁰ Gaulish names in **nant-*, if related, may lead us to the conclusion that different Celtic branches generalised different root-allomorphs of this form. In Celtiberian, /e/ may have been raised because it was preceding a complex cluster *-nt̃-*, or, crucially, because the palatalisation and fricativisation of *-t̃-*, followed by loss of the preceding nasal, had given rise to a long/tense vowel that was phonemically as /i:/. In that case, an underlying **nent-(i)jo-* or **nenti-(i)o-* would have given **nimtso-* and eventually **ni:šo-*.²¹ However, one cannot entirely rule out a derivative of a root participle **ñ-*

¹⁷ Cf. K.2.1: [R]etukenos auz a[---] (Albalate del Obispo, Teruel); K.5.1: Besku auz uetikubos (Caminreal, Teruel); K.0.8: Letontu auz : soz (unknown origin).

¹⁸ The vowel /e/ was in all likelihood inserted in the ending in analogy to the 1st and 2nd p. pl. *-mes*, *-tes*, thus averting the problems of the unstable cluster /rs/; see Prósper 2016: 197.

¹⁹ Cf. for both etymologies Prósper 2014; in both cases, the laryngeal has been lost in compounds early on.

²⁰ For nasals in coda position in Insular Celtic, cf. Schrijver 1993.

²¹ This definitely speaks against my own interpretation of NARISSIT (Coca, Segovia, CIL II: 2728, allegedly a misreading or misspelling for a genitive NARISSI, unfortunately only surviving in a drawing) as **narit-(i)jo-*, and then a

nt- ‘leading’, from **niH-nt-* (for **neiH-* see LIV: 450), which, needless to say, would have undergone phonotactic rearrangement, by which the preconsonantal zero-grade stem **nī-* (e.g. in the past part. **nī-tó-*) would have been generalised, as is probably the case in other languages with productive long vowel stems (cf. L. **amā-* → **amā-nt-*, etc.).

NESSIA (Reznos, Soria, Arevaci) has been prudently read NES+IA by Gimeno Pascual & Ramírez Sánchez (2002: 277). However, the lower part of the letter is preserved. Their proposal to read the missing letter as is considerably less promising than their own alternative <S>. This name may unproblematically be traced back to **ned^h-tó-*, like Skr. *naddhá-* ‘tied’, and then is a match of the second member of the compound *kounesikum*, a family name in K.1.1, B-1 (Botorrita), from **kom-nesso-*, like OIr. *comnessam* ‘neighbour, relative’ (the descriptive account in MLH V, 202: «*kou-n-es-*» is worthless). In the present state of our knowledge, forms like Gaul. NEÐÐAMON are unlikely to belong here.²² But, of course, the alternative possibility cannot be rejected out of hand that NESSIA is a match of NISSICVM, and that scribes simply hesitated as regards the phonemic attribution of contextually raised /e/. In view of the photograph, the use of «barred <S>» cannot be entirely ruled out for this name, since there seems to be a small diagonal stroke under the first <S>.

TELASSICVS (on a funerary inscription from Sisante, Cuenca, first edition by Corell i Vicent 1992: 584) is a family name agreeing in the nominative with the preceding individual name RETVCENVS.²³ The available photograph suggests that the correct reading is TELASSICVS. In my view, it is a close match of the family name *telazokum* (K.1.3, Botorrita; no etymological explanation is offered in BB.III), with trivial differences in the suffix. Interestingly, these forms have not undergone Joseph's Law, according to which PCelt. **-eRa-*, commonly resulting from an Indo-European sequence **-eRH₂-*, yielded CCelt. *-aRa-* through vowel-to-vowel assimilation. Therefore, they do not contain a Common Celtic phoneme /a/ in the second syllable, but either /a:/ or /an/ with loss or graphic omission of /n/ (as we are going to see, loss is more likely to have taken place). Like many other Celtiberian names, TELASSICVS and *telazokum* can be unproblematically traced back to a derivative in *-(i)io-* of a root part. **tel-ant-* < **telh₂-nt-* or **telh₂-ent-* ‘supporting, holding’ (for **telh₂-*, cf. LIV: 622). As observed by Harðarson (1993: 183–184), Gk. *τάλαντα* ‘scales’ goes back to the original root participle **t^lh₂-ént-*. The Celtiberian form has conceivably reintroduced the full grade from the original alternating athematic aorist stem: sg. **telh₂-*, pl. **t^lh₂-*; **t^lh₂-ént-* would have evolved into Celtic **talant-*, not **telant-*. At any rate, the etymology **telh₂-tyo-s* proposed most recently by Jordán Cólera & Díaz Ariño (2022: 163) for their own, tentative reading TELASSICVM is not possible for the above reasons.

Other conceivable cases are:

The place name Τουριασσώ (Ptolemy, *Geogr.* 2, 6, 57), *Turiassonem* (Pliny, *Nat. Hist.* 34, 41, 144), TVRIASSONESIS (CIL XIII: 586, Bordeaux), today Tarazona in northern Saragossa, could re-

cognate of Gaul. NARITVS, NARITI (Prósper 2016: 170). It could conceivably be taken from **narent-(i)io-*. Besides the river *Narenta* we have a personal name NARENSAI ANNVAE (dat., Dalmatia), a NARENS(?VS) (nom., Dalmatia), neither of them necessarily Celtic, and a NERANT[V]S (Este, Venetia et Histria). Since we lack other parallels, this must remain speculative. If we were to read NARISS(I) or even NARISS+I, this could be an old superlative form **nār-isto-* ‘most noble’ (cf. Mlr. *nār* ‘magnanimous’). In that case, the doubling of <S> may be related to the fact that /t/ is, of the voiceless stops, the shortest in duration (see Méndez Dosuna 1985). This phenomenon is also attested, for instance, in Greek.

²² See Prósper 2018a. Early syncope in this context is attested in nearly all Celtic branches, as in OIr. *nessam*, Gaul. NEÐÐAMON ‘closest, nearest,’ from **nezd-isamo-*, as claimed by Cowgill (1970: 132). The Sabellic forms O. *nessimas* ‘nearest’, NESIMVM, U. *nesimeis*, etc., point to an intermediate stage **neds-iz(v)mo-*, with inner-Italic loss of the penultimate vowel.

²³ The editor read NEIVCENVS. The definitive reading was suggested by Abascal Palazón (2015: 238–240; cf. CIL II-13: 952).

flect a participial adjective **turiant-(i)īā*. The indigenous attestations are *Turiazu* (A.51), TVRIASICA CAR (K.27.1, Monte Cildá, Palencia). The geminate <ss> (never attested as “barred <ss>” in Hispania, however) seems to suggest that this is correct, and that the peripheral, northern place/river names *Toranzo* (Cantabria), beach *Torranza* (Asturias), are its cognates. See Prósper 2014: fn. 26.

A set of names E[L]ANDI (EDCS-00380519, Álava), AELANDI (CILCaceres-02: 793, Cáceres), ELANDVS (Turma Salluitana, CIL I²: 709) may be taken either from **elanti* ‘doe’ (cf. Mlr. *elit*, *ailit* ‘doe’, in turn from **(H)el-ŋ-tiH₂*) or from a participle **φelant-* < **pel(H)-*.²⁴ We may, with due caution, add the family name *Elatunako* (K.9.4, Numancia), which derives from a personal name **Elantū*. In my present view, there are no insurmountable obstacles to identifying the above forms with the base of Celtiberian *Elazunos* (K.6.1, Luzaga),²⁵ *Elazuna*, *Elazunos* (K.1.3, Botorrita), if from **(φ)elant(i)īū*, but their uniformly ‘innovative’ phonetics in Early Celtiberian call for prudence. An inscription reading [E]LASVS N[-] / [-] E(L)ANDI(?) F(ILIVS) (Álava, cf. Sáenz de Buruaga & Sáenz de Urturi 1994) could accordingly contain a son’s name derived from that of his father, but the text is mangled and altogether unreliable.

What we can gather from the above examples is that a phoneme rendered <z> in Early Celtiberian, which goes back to *-tj-* (at least when this cluster was preceded by a nasal), occasionally corresponds to <ss> on inscriptions resorting to the «innovative» version of the Latin alphabet for the rendition of Celtiberian names. This is clear at least in *sekonzos*, *SEGOSSOQVM*, *SEGOSS(VS)*; promising examples are *NISSICVM* and *TELASSICVS*, *telazokum*; and this can hypothetically be extended to the families of *Turiassonem*, *Turiazu* and *ELANDVS*, *elazunos*.

To my mind, scribes were ill at ease with using <ss> and <ts> for the underlying fricative or affricate phoneme, and they chose a digraph because it was in fact a long/geminate phoneme. As we will see, examples of a single intervocalic <s> may have passed undetected. This phenomenon is reminiscent of «tau Gallicum», which arose from clusters in Gaulish epigraphy to note a phoneme alien to Latin and going back to Indo-European, and was thereupon borrowed by workshops with bilingual customers. These workshops commissioned texts in Latin that could, for instance, mention Gaulish gods.

The sociological substrate of this practice may have been similar in both regions, and both spellings may be compared in *formal*, genetic terms, that is to say, the Celtiberian digraph may have come about in imitation of Gaulish practice. Still, the functional connection of «barred <ss>» with «tau Gallicum» partly escapes us. The denomination «tau Gallicum» encompasses a plethora of signs and combinations thereof, <DD>, <Ð>, <ÐÐ>, <TS>, <DS>, <S>, <SS>, <SS>, etc., testifying to the scribal uneasiness concerning the articulation of the represented phoneme, which probably varied across periods, regions and workshops. In my present view, there is no proof that «barred <ss>» could represent the outcome of **-tst-*, and, conversely, «tau Gallicum» could not represent the outcome of **-tj-*. Of course, this does not mean that they could not synchronically be used for a phoneme with similar phonetic features in both systems. To make my case clear, I shall put forward some novel interpretations of the Celtiberian names.

While <ss> could have represented an affricate phoneme, a number of reasons cast doubt on this possibility:

First, the simplex <s> in indigenous epigraphy (specifically Novallas) can reflect a fricative phoneme, at least in word-final position, where it has never been an affricate at any stage.

²⁴ It is additionally found in European onomastics, as in the place name *Elantia* > *Elz*, the Gaulish personal names *ELANTIA*, *ELANTIAE* (Reims, CIL XIII: 3320), etc.

²⁵ My reading for former *Elazunom*; see Martínez Chico — Prósper 2021.

Second, forms containing <ss> show loss of the preceding /n/. This seems to suggest that the original cluster had been weakened and become a fricative, *and* that the nasal had been lost as a consequence of the difficult nasal-fricative transition. If this is true, occasional occurrence of <N> may be due to the synchronic association of a name with its living participial base, or may be reflecting secondary nasalisation of the preceding vowel. I consequently start from an evolution $-nti- > -nts- > -(n)_{\text{ss}}-$ (with or without lengthening of the preceding vowel after loss of the nasal segment).

Our best case of Late Celtiberian <NS> going back to $-nti-$ is the recently edited form BVGANSONIS (Soria, Arevaci),²⁶ the father's name of the deceased person in a funerary inscription. On closer inspection, the photograph offered by the editor shows a vertical stroke crossing the base of <S>, and the reading is consequently <S>. I have traced it back to a Celtic root participle **bugant-* < **b^hug-nt-*, a match of the Venetic name FVGANTIVS (CIL V: 8986, Aquileia). Both names probably go back to the root *2*b^heug-* (cf. LIV 84–85 'jmdm. nützen, Nutzen bringen'), systematically found in the zero grade (cf. Lat. *fungor* 'to enjoy, perform' and Skt. *bhunakti* 'to enjoy, use, consume'). See Prósper (2019: 36–39) for details and comparanda. It definitely points to fricativisation prior to complete loss of the nasal.

In Celtiberian texts written in the Iberian script, both <ti> and <z> reflect inherited $-ti-$, though the reasons for the distribution of these spelling variants are unclear: their distribution may have been dialectally conditioned or simply reflect different social strata. After a Celtiberian sequence $-nti-$ evolved into $-nts-$, the affricate became a long/tense dental sibilant /s̄:/, or, conceivably, /θ:/, since interdental fricatives and dental sibilants are auditorily confusable sounds. The nasal was thereupon lost, and, as we shall see in what follows, a new phonemic contrast emerged between intervocalic /s̄:/ and /s̄/ (for a comparable phenomenon in Italic, see Prósper 2020: 69). The same reasoning applies to the phonetics of heteromorphemic $-d/t-s-$. Where possible, the digraph <SS> was pressed into service in order to distinguish the new phoneme from the already existing long/tense alveolar /s:/ (<s>, <SS>) and the lax /s̄/ (<z>, <S>), but the spelling of BVGANSONIS at least seems to have favoured a compromise solution, reflecting either the nasal or a nasalised vowel. Similarly, the divine name VELONSAE (dat., Navarra, IRMN 55; unearthed in Tobalina, Burgos) goes back to **uel(H)-ont-ijā* 'willing/strong'.

4. The reflection of Early Celtiberian <s> in the Latin alphabet

Many (if not all) cases of Early Celtiberian word-medial <s> go back to expressive geminates, resyllabification/fortition ($-s.i- > -s.si-$), clusters of two dental segments ($*-tst- > -ss-$), heteromorphemic clusters of a dental stop + sibilant, or the Celtic cluster $-\chi s-$. Let us consider some indigenous names in Latin epigraphy, specifically those in which we can reconstruct an Indo-European sequence $*-ntst-$.

The family name COSSOVQVM (Sigüenza, Guadalajara; lost) can be seamlessly traced back to an adjectival **kom-d^hh₁-teu-o-l(i)jo-*, directly derived from the action noun **kom-d^heh₁-tu-* 'confluence of paths or watercourses' (cf. Gaul. *Condate* < **kom-d^hh₁-ti*). It survives in the Roman divine name *Cōsus* (which somewhere down the line became thematic) and its derivative *Cōsuālia*, as well as the Lusitanian divine name COSSVE (dat. sg.) and its variants, which go back to **kom-d^hh₁-tu-* with early laryngeal loss and generalisation of the zero grade of the root (see Prósper 1997).

²⁶ Cf. Alfaro Peña 2017.

The individual name LESSO (K.1.2, Botorrita), LESSONI (CIL II: 3852, Valencia) is in my present view a match of Gaulish names attested in the Venetic record: *le.s.sa* (LV: 208, Cadore), *leso* (LV 93: Este) may derive from **splend-tu-* ‘splendor’ (cf. OIr. *lés* ‘light’).

BESSVCA (CIL II: 3097, Cuenca, lost) probably continues an agent/object noun **b^hend^h-tu-* ‘manners, habits, behaviour’ that became CCelt. **b^hessu-*, and whose perfect match is OIr. *bés*. At the present state of our knowledge, however, we must be prudent, since the stone is lost and the only available drawing may have ignored a “barred <ss>”. This is especially true of the divinity to which the stone is devoted. If the reading and segmentation LEIOSS(A)E G(-) BESSVCA PRO FILIO V L R M are correct, LEIOSS(A)E may be concealing an active participle in *-nt-* followed by the relational suffix *-i_o-*, which has parallels in the indigenous theonymy of Hispania and Gaul (see most recently Prósper & Medrano Duque 2022: 25).

BESSVCA, COSSOVQVM and LESSO may consequently be traced back to forms containing **-nt^st-*, and all of them show loss of the nasal in coda position in contact with a fricative. The Hispano-Celtic outcome of Late Indo-European **-t^st-* was probably /s:/ in every context. This is regularly rendered as <s> in Early Celtiberian, whose script does not note geminates, and <SS> in the Latin alphabet.

To recap, the distribution of <SS> and <ss> reveals an interesting pattern: both digraphs can be used for the outcome of *-nt(i)_o-* depending on the area, since not all regions and workshops had access to the orthographic innovation called «barred <S>»; at the present state of our knowledge, only <SS> can reflect /s:/, resulting from two Indo-European dental segments in contact and processes of gemination and fortition. In the first case, the Early Celtiberian spelling is universally <z>; in the second it is <s>.

5. *-t_i-* in non-nasal environments

We may consequently wonder at the cases of simple <s> in (originally) intervocalic position, of which a single instance has been detected on the indigenous rock inscription of Peñalba de Villastar (K.3.3). Here, we read ENIOROSEI and ENIOROSEI, apparently belonging to two different sentences, as we are going to see in what follows.

In Early Celtiberian, both <ti> and <z> occur when *-t_i-* is *not* preceded by a nasal: cf. the place name *tirzoz* < **trit(i)_o-* ‘third’ (A.45), *arzakoz* < **art_i-āko-*, cf. Northern Italian *Arsago Seprio* (A.36), both probably «Vasconian mints», but *Lutiakēi* (K.6.1, Luzaga), *Lutiakos* (A.76), *Titiakos* (A.58), *Teitiakos* (A.57). These four forms might reflect the «Sievers effect» if they were respectively rendering **lux_{t_i}ti_v-*, **di_{x_{t_i}ti_v-}*, **te_{x_{t_i}ti_v-}*, or simply a *synchronic* process of derivation from participles in *-to-* or nouns in *-teh₂* by means of a productive suffix *-iāko-*. For many other cases of this alternation, which are essentially contingent on the chosen etymologies, see Prósper 2014.

No certain instances of the use of a single <s> have been detected thus far in Latin epigraphy, where fricativised *-t_i-* is occasionally spelt <S>, but at least two promising instances have been overlooked.

The family name of an individual called ALBANVS CLOVSOCVM (Sayatón, Guadalajara, 1st C. AD, cf. Abascal, López de los Mozos, 1993), reflects the evolution of an erstwhile **klout-(i)_o-*, still attested in Early Celtiberian as *koloutios* (K.1.3, Botorrita) and later as CLOVTIVS all over Hispania. The photographs provided by *HEp* online edition²⁷ and EDCS-03700465 are suggestive of an alternative reading CLOVSOCVM with a slanting stroke that runs downwards left to

²⁷ Retrieved from http://eda-bea.es/pub/record_card_1.php?page=6&rec=142.

right from the base of <S>, traversing the deeply incised horizontal guideline. It occurs exclusively in this instance of <S> (and not in the Latin forms ALBANVS and LIBERTVS).

Another inscription from the realms of the Arevaci reads: AISAE · VIANNETIQ(VM) CAV[C(ENSI)] / AN(NORVM) · [-] (Sepúlveda, Segovia, *ERSg*, 27). It has passed unseen that the right reading is probably AISAE. It shows a «barred <S>», again in the form of a slanting stroke that runs downwards left to right from the base of the letter. Consequently, this name can be traced back to **ait-(i)jā*, a derivative of **ai-to-* ‘time’ or, perhaps more plausibly, to **axt-(i)jā* (see below).

The rock inscription of Peñalba de Villastar (K.3.3), the only indigenous document in the «Aragonese» alphabet besides Novallas, does not make consistent use of the new resource. For instance, «barred <S>» has been detected in the word or phrase ENI.OROSEI that opens the text, but the next sentence begins with ENIOROSEI, with no distinctive mark on the sibilant and no interpunction.²⁸ No plausible etymology has ever been proposed for this strange-looking form. Other considerations can still be added. If the first «barred <S>» is employed correctly, this form cannot be identified with the coins *Orosiz*, *Orosi* (A.86) any longer. Since *-i-ej* can hardly be a dative form (*-i-*stems have a dative *-ej*), and a locative would need no preposition to express location, the underlying form may have been **eni-(φ)oro(n)t-ġ-eġ*, the thematic locative of an adjective derived from **(φ)or-ont-*. In turn, this would be the present participle of a denominative **por-e/o-* that survives in OHG. *faran* ‘travel’, which would constitute a remarkable Celto-Germanic innovation. Alternatively, it could be the agentive derivative of IE **poro-* ‘causeway, path’ → **poro-t-*, which looks amazingly similar to L. *interpres* ‘go-between’, a compounded agent noun **enter-pore-t-*, from a noun **poro-* (cf. Nussbaum 2016: 290).²⁹

Cases of *-Vt̃V-* spelt <S> in Latin epigraphy are:

SEGISAMA BRASACA (*CIL* II: 4157, Tarraco), which comes from **brāt̃i-ākā* (cf. Prósper 2014, 122). This etymological attribution is now cemented by other traditionally misinterpreted Gaulish and Early Brittonic examples, like DEO MARTI BRACIACAE in Derbyshire, Britannia (cf. Prósper & Medrano Duque 2022). It undoubtedly corresponds to *barazioka* ‘lawful’ (K.6.1, Luzaga, North of Guadalajara), which shows addition of synchronic *-joko-* to a preexistent base **brāt̃(i)jō-* (cf. *tanioka* < **dāno-*, *risatioka* < **φri-sant-* in K.1.3, Botorrita).

VASCASVS (Hontoria, Burgos, *CIRB*: 332), which goes back to **uφo-skāt-(i)jō-* ‘protector’, attested in OIr. *foscad* ‘shelter’, etc. (see Prósper 2016: 135).

SECOVESO (Lara de los Infantes, Burgos, lost; *CIRB*: 358) is a derivative in *-(i)jō-* of **sego-uet-*, preserved in individual names like SEGOVETIS F(ILIA), SEGGVES, SEGVETI F(ILIA), etc.

Occasionally, the spelling <TI> alternates with <S> for the same name:

VSSVEITIO AMMONIS F(ILIVS) (Clunia, Burgos, *CIRB*: 68); VSSEITIO (Clunia, Burgos), VXSEISVS / ELLICO / STENION/TIS F(ILIVS) (Sotodosos, Guadalajara), where the photograph offered by EDCS-21900141 shows a stroke under the second <S> that is compatible with a reading <S>, and possibly also the western form SVNVA VSEITI (Coria, Cáceres, *CIL* II: 785). This name may be traced back to a compound **uχs-ueχt-(i)jō-/l-ū* > **uχsueġt̃jō-/l-ū* from the past participle of **ueġh-* ‘to move, drive, carry’.³⁰

²⁸ See Jordán 2016 for the new reading.

²⁹ We cannot say whether <SS> was ever used in the innovative «Aragonese» alphabet created to write Celtiberian. Since, as contended above, it was partly designed as a transcription of the corresponding forms in the Iberian script, it might never have used geminate letters, but our evidence is too paltry. This would, for instance, account for TIASSO in Peñalba de Villastar, if it stands for †TIASSO and continues a past participle in **-to-*, related to the form TIATVNEI in the same text.

³⁰ See Prósper (2012; 2016: 178–179). The readings VSSEITIO and VXSEISVS have been put forward by González Rodríguez & Gorrochategui Churruca (2011). The comparison with *useizu* in K.1.3 that they propose without further linguistic analysis is uncertain, because these forms are irreducible to a common etymology unless Botorrita is

In these forms, the sequence *-eχt-* has eventually become *-eġt-*, as in Welsh, Alpine Gaulish, Venetic (see Prósper 2018b), South-Western Slavic (see Šefčík 2020) and a number of Romance languages, notably Catalan, Portuguese and Galician. The change is already apparent in Early Celtiberian: cf. **dū-reχtā* ‘issued’ in *taruotureska tureita* (K.23.2)³¹ and DVREITA TARVODVRESCA (Arevaci, Untermann & Villar 1999) ‘issued by the city of Tarvodurum’. By contrast, more peripheral Hispano-Celtic dialects seem to have lost any trace of preconsonantal velar or uvular fricatives, as transpires from DVRETA SALDANICA (León; see Martínez Chico & Prósper 2021) or the widespread AMBATVS (as opposed to Gaul. AMBACTVS, AMBAXTVS).³² In spite of the traditional vision of Common Celtic as a dialect in which /eġ/ had been monophthongised, it must still have existed in Celtiberian, at least as a phonosyntactic variant, when the Iberian script was adopted, around the beginnings of the 3rd c. BC, but it had already become /e:/ before the Latin alphabet was definitively imposed. Accordingly, late attested, indigenous names containing <EI> are likely to have undergone the abovementioned change *-eχt-* > *-eġt-*.

6. Another letter comes to complicate matters: the use of <Z> for the Celtic names of Central Hispania

In a number of names attested in Latin epigraphy, the outcome of *-tġ-* is reflected as <Z>, occasionally alternating with <TI> and <S> for the same name. This letter was used in a reduced area in Central-Southern Spain (the adjacent provinces of Madrid, Toledo and Cuenca).

A name VA(---) ANNESVS (Pedraza, Segovia, ERSg, 170) is obviously the same name as the more peripheral ANNETIAE AROCI F(ILIAE) (León, CIL II: 2689), ANNETIA (Noricum, EDCS-14400122), and the cognomen of ACILIA ANNEZA (Titulcia, Madrid, CIL II-13: 2; see Prósper 2016: 146). This is a derivative of the name **annet-*, attested as ANNETIS (gen., Segovia, ERSg 138). In this case, we exceptionally find both <S> and <Z> for original *-tġ-*.

In addition, a number of forms show a sequence <IZ>, which in my view goes back to a CCelt. cluster *-χt(i)ġ-*. As in the forms VSSVEITIO, VSSEITIO, VXSEISVS and DVREITA/*tureita* (or *dureita*) of the Arevaci reviewed above, a glide is the only trace left of the fricative uvular segment originally preceding *-tġ-*. In the following cases, the fricative outcome of *-tġ-* is rendered <Z>:

ARBAIZA (Caesaro-briga, Toledo, 150–200 AD). If Hispano-Celtic displayed the same tendency for syncope of the prefix **pare-* in intensive compounds as the British and Gaulish-speaking territories, this name could go back to **pare-uχt(i)ġo-* ‘very bad’,³³ containing the past part. **uaxto-* in Mir. *facht* ‘malice’ (cf. EDCP: 405). Onomastic cognates may be ARVATIVS in Germania (CIL XIII: 7577, Wiesbaden), and VAXTVLLA (CIL XIII: 5864, Langres, Belgica/Germania superior). A potter’s name VAXTI is attested (as usual, in the genitive case) in Britannia and Aquitania.³⁴

reflective of a late and dialectal stage of this form. The form *useizu* may simply reflect an erstwhile **uχs-edġo-*. See the parallel of **seg-edġo-* in A.78 *sekeiza*, and the late coin reading SEGEIDA, for the place name *Segeda*/Σεγηδα. The reading *sekaiza* is definitively superseded thanks to the palaeographic observations of Rodríguez Ramos (2002, neatly confirmed by etymological considerations.)

³¹ Or *taruodureska dureita* under acceptance of the dual system.

³² The variant form VXSEISVS has undergone further, predictable changes: hesitation between <SS> and <XS> bears witness to the difficulties that scribes experienced when trying to reflect what may have been a palato-alveolar fricative [ʃ:], and the glide [ɥ] tends to be absorbed by the preceding long sibilant.

³³ See Prósper 2019.

³⁴ There is a reason for the apparent fricativisation *-rɥ-* > *-rβ-* that explains the sequence <RB> in Hispania. Imported forms attested in Lusitania Emeritensis usually show for <V>, probably indicating that this fricativi-

CAECILIA CONTAIZA (Saelices, Cuenca, *CIL* II–13: 426, 2nd c. AD) furnishes an interesting case in point.³⁵ The text reads CAECILIO VICTORI / CAECILI ISARGYRI LIB(ERTO) / CAECILIA CONTAIZA / VIRO ET SIBI ET / SVIS ET EVTIC(H)IAE F(ILIAE).

CONTAIZA may well go back to **kom-tanχt-(i)io-*, containing the Celtic past part. **tanχ-to-* (< **tŋk-tó-*, cf. **tenk-* ‘to be gathered, become fixed’) in OIr. *técht* ‘calm, undisturbed, frozen’. Its adjectival derivative **tanχt-(i)io-* is also attested in OIr. *téchttae* ‘proper, right’, and W. *teithi* ‘characteristics, qualities’ goes back to **tanχt-* (with *i*-affection). A number of Old or Middle Irish prefixed forms can be directly compared with CONTAIZA: cf. *ántéchtta* ‘splendid and fitting’. But the perfect match of CONTAIZA is OIr. *coitechta* ‘(equally) legal, proper, suitable’ (see *eDIL*, s.u. *coitechta*). As a consequence, this form may be regarded as an inherited compound. The nasal has been lost early on before a fricative, probably with ensuing lengthening of the preceding vowel. Context-sensitive raising [an] > [æ̃n], or its conceivable variant [ã] > [æ̃], seem to have never occurred in Celtiberian, since it would have probably been identified with Latin /ɛ/ and would have been spelt <E> or <AE>, not <A>.³⁶

CONTAITA (showing a ligature in which an <I> «longa» is superimposed on <T>; the resulting sign is cross-shaped) appears on a slightly earlier inscription equally unearthed in Saelices, Cuenca (*CIL* II–13: 474, 70–130 AD), reading MATIRA VALERIORVM / L(VCI) ET NEPOTIS LIB(ERTA) H(IC) S(ITA) E(ST) / CAECILII VICTOR ET / CONTAITA [M]ATRI. As the editors contend (see Abascal et al. 2011: 222), it may actually refer to the same person as CAECILIA CONTAIZA, and, in addition, both *may be* instances of the same name. But this is simply impossible under acceptance of the respective readings <T> and <Z>, which cannot be reconciled as variant spellings of a single phoneme: it transpires from hundreds of cases that intervocalic voiceless stops are neither voiced nor fricativised in Celtiberia. In fairness, there are cases in epigraphy in which we may reasonably suspect that postconsonantal <I> has been omitted in writing because of incipient palatalisation. This is hardly ever the case with *-ti-*, however, especially because alternative means to reflect palatalisation were available.

We could speculate with the idea that CONTAITA, after she and her husband, *liberti* of *Caecilius Isargyrus* (himself probably a freedman), were manumitted, not only took her master’s gentilic name, but adapted her cognomen by attaching *-(i)io-* to it. In this way, she would have managed to harmonise it to a common onomastic formula illustrated by VALERIA LATVRINA (Soria), AEMILIA ITVNDIA (León) or TERENTIA NESSIA (Segovia). But this is an altogether unsatisfactory escape route. The dedicants of the older text are the married couple CAECILI VICTOR ET CONTAITA ‘the Caecilius V. and C.’, who probably were already Caecilius’ *liberti* by that time, and therefore there is no reason to believe there was any change in the wife’s name somewhere down the line.

Under a more trivial assumption, both names are in fact the same, but definitely not because <T> and <Z> can alternate freely. In fact, the form the scribe intended to write in the earlier text could be CONTAITIA with a triple ligature <I+T+I> (of which we have no fewer than two examples in the form <V+A+LERIOR+V+M> on the same inscription). The sequence <TI> was in all likelihood rendering an affricate (at least in careless speech or among the lower classes) that

sation was alien to the indigenous variety of this part of Lusitania at least, where [rβ] was automatically interpreted as containing an original /b/. If context-bound fricativisation of [ɥ] in onsets was an ongoing process in western and/or southern Celtiberia, it could be indirectly reflected in the names of migrants hailing from this region, if the sound change had not taken place in their destination.

³⁵ Note that the same form might occur in the *hápax* CONTAECA (*CIL* II–13: 142, Gálvez, Toledo) if the reading <C> in the suffix, which is hardly legible on the photograph, were erroneous (possibly <S>?).

³⁶ The nasal was never reintroduced (but cf. L. *sānctus*, with /n/ in analogy to *sancio*, as opposed to O. *saahtum*, U. *sahta*, *sahata*, *satam-e*); for Lusitanian SINTAMOM < **sanχto-*, cf. Prósper 2021.

would become a fricative and be naturally spelt <Z> in that region in the lapse of barely one generation: CONTAITIA → CONTAIZA. Accordingly, this case provides a precious testimony of the diachronic evolution of the sequence *-tj-* for the same name and referent.³⁷

As contended above for <EI>, <AI> cannot be plausibly taken to reflect an original diphthong: apart from etymological difficulties, an inherited diphthong would have a recognisable tendency to become /ae/. All these names have one thing in common: they contain <I> for Celtic [χ], the outcome of Indo-European labials and (labio-)velars in coda position preceding an obstruent. Celtiberian in the Iberian script is ambiguous in this regard, since it does not note fricatives in coda position except sibilants, for which it had alphabetic signs available. We cannot reject out of hand the possibility that the change *-axt-* > *-ajt-*, at least, has been favoured by the presence of a following *-i-*. By contrast, metathesis (*-Vtj-* > *-Vjt-*) or palatalisation followed by prevocalisation (*-Vtj-* > *-Vt̥i-*) can in my view be safely ruled out, both for etymological reasons and because there are forms containing *-Vtj-* that militate against this idea: VASCASVS, ANNESVS, ANNEZA, BRASACA, SECOVESO, actually show the culmination of the palatalisation process, with no apparent effect on the preceding syllable.

Still another interesting name containing <Z> is attested in the same area:

TVRPIO / AN(NORVM) XXV / H(IC) S(ITVS) E(ST) S(IT) T(IBI) T(ERRA) L(EVIS) / CAIZITA / F(ACIENDVM) C(VRAVIT) (Polán, Toledo, *CIL* II–13: 150, 71–130 AD)

The dedicant's name has been unanimously read as CAIZITA to my knowledge. All I can discern with certainty on the photograph is <CAIZI>, followed by a somewhat longer, deeply carved <I> and a final blurred letter. The *hápax* CAIZITA can be interpreted in the light of the above examples.³⁸ It might be a comparatively late derivative (probably in imitation of other local names) of †CAIZA, in turn from **kaχt-(i)jā*, a derivative of CCelt. **kaχto-*, from a Late Indo-European past part. **kap-tó-* 'caught, grabbed' (cf. L. *captus*, Goth. *hafts*). This participle has survived in OIr. *cacht*, MW. *caeth* 'slave' (*EDPC*: 197) and a Gaulish personal name CAXTOS.³⁹

In sum, intervocalic *-tj-* (including original *-Vχt̥jV-* in (south-)western areas where this sequence has evolved into *-Vjt̥jV-*) can be rendered <TI>, but in some places also <S>, <Z>, and <S>, indicating that it has undergone affrication and eventually become a short/lax, not a long/tense sibilant.

The letter <Z> was used in Latin for the first time on a denarius in 81 BC. As is well known, Greek <ζ> represented a voiced sibilant from Hellenistic times onwards. While <Z> is very frequent in Greek names in Hispania, these are often spelt with <S>. From the 1st century BC on, <Z> was hesitantly used in Latin epigraphy for the rendition of Greek names, to render a voiced sibilant /z/: the *Paradebeispiel* is Ζώσιμος, which occurs as ZOZIMA, ZOSIMVS and

³⁷ Jordán Cólera & Díaz Ariño (2022: 159) propose a reading CONTATIA. While this is formally possible, it can hardly account for CONTAIZA. Their views are eclectic as concerns the relation of writing with the phonetics (phonology is not mentioned) of the sequences <T+I> or <I+T> and <I+Z>, and, in order to make sense of the erratic spellings of this name, allow for phonetic changes, apparently affecting the same form, which have long since been proven incompatible (see below). Their reconstructed **kontatya* [sic] is not further explained.

³⁸ Under the assumption that the segmentation is right, this name could be read CAIZIIA, with <II> = <E> (which is underpinned by the alternative reading TVRIIO of the deceased person, which has direct cognates in TVREO, Villamanta, Madrid, TVREO, Nunomoral, Cáceres, and TVREA, three times in Lusitania).

³⁹ Such names as CAITTA / ANNAE MAX/VMILLAE SER/VA (Palencia, Vaccae); VAICVS C/AITAE LIBE/RTVS (Cáceres, Emeritensis) could conceivably belong here and go back to the unenlarged form **kaχto-*, if the bearer of the name came from the territory of the Arevaci or from Central-Southern Hispania, which is impossible to ascertain. This might also be the case with some names attested in the Alpine areas: MOSICAITO, MOSGAITO, DIOCAITO (Noricum). However, we are in the dark as to whether we are catching the last glimpses of «trans-Alpine» Venetic. Alternative etymologies are offered in Prósper 2016: 30.

SOSVMVS. An indirect parallel is provided by Late Oscan in the Latin alphabet, which introduces <Z> for its own voiced sibilant. This does not necessarily mean that <Z>, in so far as it represented the outcome of *-tj-*, already had a voiced realisation in Early Celtiberian, or that it ever acquired one. The individual names ANNEZA, ARBAIZA, CONTAIZA, definitely point to a (phonemically voiceless) dental fricative sibilant /s̥/.

The dental fricative nature of the phoneme rendered <Z> is now confirmed by an inscription from Villaminaya (Toledo), originally read as NEZVTVS / CAECILI / LVCANI / SERBVS / H(IC) S(ITVS) EST (CIL II–13: 119, 1st c. AD). Velaza (2008: 369–370) corrects NEZVTVS into MEZVTVS (a suggestion certainly borne out by the ductus, the presence of a short unfinished or erased fourth diagonal stroke, and the room between this letter and the adjacent <E>). He relates this form to Indo-European **med^hu-* ‘mead’ (which he, however, translates as ‘sweet’). His phonetic explanation for the use of <Z> as reflective of a Celtiberian intervocalic dental segment is plausible, but his etymological attribution is problematic, and I remain convinced that this is a mere thematicisation of **med-ūt-* ‘ruler’ (in several cases of MEDVTIVS, MEDVTICA, etc.). The variant MEZVTVS vs. «orthographic» MEDVTVS indicates that the intervocalic dental stop had been fricativised.

The personal name LVRAEZI (gen., Madrid, CIL II–13: 23, 50–150 AD) may contain **-aido-*, like ARRAEDO (Soria), ABLAIDACORVM (Asturias), TVRAEDOQV(M) (Ávila), cf. Prósper 2016: 53. The inscription has been read DOMITIA VIC/CI MALVGENIQ(VM) / <F=E>(ILIA) LVRAEZI AVCALI/Q(VM) VXSOR H(IC) S(ITA) E(ST) S(IT) T(IBE) T(ERRA) L(EVIS). Needless to say, it is arbitrary to spell out <E>, which follows the indigenous family name and not the father’s name, as F(ILIA), and the interpunction is not visible on the available photographs. One could therefore speculate with an alternative possibility: the father’s name might have been ELVRAEZI. The advantage of this segmentation is that it allows us to connect the base **(φ)eluro-* with hypothetically Gaulish names like ELVRINI (gen., EDCS-78600062 Germania Superior). In turn, these may reflect the crossing of **pelh₁-u-* ‘much, many’ and **pelh₁-ro-* in OIr. *ilar* ‘big number’ (cf. Widmer 2004: 90–92) and Celtib. ELARCORVM (family name, Belorado, Burgos, ERPB: 39; see Prósper 2016: 100). On balance, given the existence of hypercorrect forms with <AE> for orthographic <E>, we could reckon with an original **(φ)elur-etio-* (cf. ANDAETI for ANDETI in Lusitania, etc.).⁴⁰

The scribe who wrote MEZVTVS can hardly have *intended* to reflect [ð]: this amounts to accepting that he felt the need to accord a specific spelling to allophonic [ð] (as noted above, the phonemic status of a voiced dental fricative would at best be marginal). This leads us to an interesting, albeit merely provisional conclusion on the phonetics of Hispanic Latin in the 1st c. AD: if workshops occasionally resorted to <Z> when faced with [ð] or a similar sound, this may be taken to mean that fricativisation of Latin voiced stops was still at a very incipient stage or had not begun at all, and, as a consequence, they could not ignore the fricative nature of the foreign sound. Alternatively, slight differences in voice or point of articulation of obstruents may have been at play here: taking into account the existence of the spellings <T>, <S>, <Z> and <S> in exactly the same slots, we may provisionally surmise that the scribe who wrote MEZVTVS (and, in view of SERBVS, was not proficient in Latin orthography, either) actually heard a sibilant [z̥] that he was unexperienced enough to parse as /s̥/ and not as /d/.

We have an indirect clue that points to [ð] tending to evolve into a sibilant [z̥] in central/southern Celtiberian, a fact concealed by orthographic spelling in most places: a new inscription from Abánades, Guadalajara (see Barbas et al. 2011) contains two instances of the family name OBIDOQ(VM). The personal name that forms its base looks identical to OBIDDVS, attested once as a cognomen in Castro Verde (Beja, Portugal, Lusitania Pacensis): M(ARCVS) VLPVVS

⁴⁰ L IVLI RVFI / VZENTI (Madrid, CIL II: 4975, lost) may be another instance of <Z> for expected <D>, but this cannot be proven.

OBIDDVS / H(IC) S(ITVS) E(ST).⁴¹ As other scholars have noticed before, OBIDOQ(VM) has a match in MACER OBISOQ(VM) AMBATI F TOLETA(NVS) (Casas de Don Pedro, Badajoz, Lusitania, *CIL* II-13: E 4)⁴². There is no doubt as to the southern Celtiberian, specifically Carpetanian origin of the deceased individual. It is therefore intriguing that his family name was not written with <D>, but with <S>. It is tempting to assume that the western scribe took it to be not an allophone of /d/ (for which <D> would have been used), but of /s/. Again, this speaks in favour of a realisation closer to [z̥]. Both MEZVTVS and OBISOQVM seem to testify either to the incursion of speakers of dialects with dental fricatives into areas where lenition of voiced stops had not begun, or to assibilation of the lenited dental fricatives in a central zone that we cannot possibly delimit.

The sign <Z> has not yet been found in the special, «Aragonese» version of the Latin alphabet. It serves a conscious need to render dental, probably voiceless fricative phonemes. Our extant cases span a vast area of Central Spain comprising the present contiguous provinces of Madrid, Toledo, Ciudad Real and Guadalajara. This time, the solution to the deficiencies of the received alphabet came from the Baetica, not from Celtiberia. When the need was felt to represent a dental fricative sound in Celtiberian names, <Z> began to be used and succeeded in spreading northwards from the workshops of Baetica and the East. The letters <SS>, <S> and <Z> stand in complementary spatial distribution because they provide alternative solutions for similar problems encountered by scribes trained in writing Latin in the buffer zones between central Celtiberia and peripheral, highly Romanised regions, to which foreigners were presumably attracted by intense trade activities.

7. The evolution of Celtiberian sibilants.

What can be deduced from the increasing amount of graphic variation?

The new signs <S>, <SS>, associated with the creation of a new Latin alphabet designed to write Celtiberian, could be pressed into service to distinguish some Celtiberian phonemes for which the Latin alphabet had no specific letters. But this resource could not be expected to reach long beyond the original boundaries of that alphabet. It was diffused from its locus somewhere in Aragon to the south, reaching the south-eastern periphery of Celtiberia, and in view of our available materials, was circumscribed to Cuenca, Segovia, Guadalajara and their north-eastern boundaries, roughly coinciding with present-day Soria (DERCINOASSEDENSIBVS, BVGANSONIS).

In a nutshell, Celtiberian may have had a phonemic contrast of place between two long/tense sibilants:

1. /s:/ both intervocalic (-ss-) and in onsets after voiceless obstruents (-C-s-);
 2. /s̥:/ <-nt(i)j- (phonemic upon loss of /n/);
- There may have been short counterparts for both, namely
3. /z/ (or simply a lax /s/) <-s- in intervocalic position and after /r/;
 4. /z̥/ <-t̥i-, -t, -d.

The Iberian script, with its characteristic orthographic underrepresentation of phonemic contrasts and its lack of geminate letters, respectively used <s> and <z> for the tense phonemes (1., 2.), and <z> for their lax counterparts (3., 4.).

The dental sibilant /s̥:/, /z̥/ may have shown a distinct tendency to become a fricative interdental /θ:/, /θ/, as in the well-known Spanish evolution *sperantia* > *esperantsa* > *esperanşa* >

⁴¹ See Encarnação (2016), correcting his former reading OBIDVS; one cannot jettison the possibility that the scribe has heard [ovid:us] and that this is a case of *nomen pro cognomine*.

⁴² First edited by Roso de Luna (1904: 127). He, however, erroneously read it as OBISOD(ICVM); only a drawing is available.

esperanza. This change tends to maximise the articulatory space between phonemes, especially prone to happen if the system already had an alveolar and a palatal sibilant (whether there actually was an underlying palatal sibilant in Celtiberian is unknown, however).

Some pending issues remain: what exactly defines the contrast of <D> vs. <S> in the «Aragonese» alphabet, corresponding to one single sign <z> in the Iberian script? We have every reason to assume that word-medial [ð] was still an allophone of /d/ around the change of the era; the fact that it does not occur in final position in the new «Aragonese» alphabet, in the slots in which we would have expected it to, is intriguing. For instance, ablative endings were spelt with the Iberian *sigma*, vaguely transliterated as <z> or <ð> by modern scholarship under the assumption that it was a voiced fricative. And yet, in the «Aragonese» alphabet, *sigma* was replaced by a sibilant <s>. This can only mean that its phonetic content was, counter to expectations, quite different from the intervocalic allophone of /d/, rendered <D> in all the known varieties of the Latin alphabet, and that it was consequently assigned to another phoneme, which I have formalised as /s̥/.

The phonemic contrast /s̥/ (<z>/<s>) – /s/ (<s>, <S>) discernible in word-final position presupposes a reorganisation of the phonemic system somewhere down the line. In that context, the contrast stood between two fricatives, not distinguished by voice, but more likely by point of articulation, stridency, and tenseness: an alveolar /s/ contrasted with a dental fricative was realised as [s̥] or [θ], perhaps already an approximant on its way to total effacement. We have to bear in mind that words in the «innovative» alphabet mimic or transcribe the corresponding words in the Iberian script, except that the Iberian script suffered from some underrepresentation of phonemic contrasts between fricatives, now partly repaired in the innovative version of the Latin alphabet by the introduction of «barred <S>».

Therefore, we may suspect that the new alphabet was designed to be phonemically more consistent than its antecedent, but at the same time tended to reproduce *all* the characters present in Early Celtiberian, even if this procedure was misleading as to the actual phonetics when compared with contemporary Latin usage. Thus, it utilised the digraphs <QV> and <EI> for labiovelars and diphthongs that did not exist any longer, but had evolved into /p/ and /e:/ everywhere, as transpires from contemporary onomastics in the Latin alphabet, which show <P> and <E> (see Prósper 2016: 183 and fn. 144). As we are going to see, the use of word-final <S> as opposed to <S> in the «Aragonese» alphabet may be artificially recording distinctions long vanished in the language of most speakers for the sake of transparent and explicit inflectional morphology.

Since <z> functioned in Early Celtiberian as a cover symbol, the hypothetical shift *-t#*, *-d#* > *-θ/-s̥#* may be older than we think. *-t#* and *-d#* are often held to have merged already in Proto-Indo-European. Lipp (2016: 268–269, 287) starts from lax plosives, which resulted from the neutralisation of the phonetic manner of articulation features like tenseness, voice and aspiration in word final position. In this context, Proto-Indo-European displayed unreleased *lenes* ([-tense, -voiced, -aspirated]). In a vast number of languages, where tenseness (and not voice) was distinctive, they were classified in phonemic terms as [-tense, -aspirated], and not as [+tense, -aspirated]. This includes Italic (witness OL. *FECED*), as opposed to Late Vedic, where a final unreleased *lenis* is identified with /t, k, p/ = [+tense, -voiced], because the phonological identification was based on the common feature [-voiced] and the distinctive feature was voice and not tenseness (cf. *uṣarbhut* ‘awakening at dawn’). This cannot be weighed against the Celtiberian evidence, which, for all we know, does not preserve Indo-European word-final stops (and consequently only uses syllabograms in word-final position when there was, or had been till recent times, a final vowel).

If the contrast /d/ – /t/ had been neutralised in final position, we probably have to start from **-d̥#*.⁴³ It was never voiced and was never identified with word-medial [ð] in Hispano-Celtic. Over time, it was lenited but remained a voiceless dental [s̥], parsed as /s̥/ when this new phoneme arose upon fricativisation of the cluster *-t̥i-*. This explains the 3rd p. sg. pret. *tekez* (K.6.1, Luzaga), < **dēke-t* (<< **d^heh₁k-t*).⁴⁴

CCelt. *-d̥#* was probably dropped after (a number of) consonants. See, for instance, the sequence reconstructed as *-d-t#* in the 3rd p. sg. pret. *auz* (see above), *-s-t#* in the 3rd p. sg. pret. *tures* < **dū-reχ-s-t* ‘issued, ordered’ (Torrijo del Campo, Teruel); *tures/dures* (K.0.7, Gortono)⁴⁵ and *-n-t#* in the 3rd p. pl. pret. *tako* /*dakon*/ < **dak-onḍ* < **d^hh₁k-ont* (lead letter of Iniesta, with trivial omission of the nasal in coda position).⁴⁶ Hence, the outcome of CCelt. *-d̥#* in Celtiberian was either *-ø* (after consonants) or a fricative *-s̥* (after vowels). This is unproblematic and comparable to the situation in other Celtic languages: cf. the Gaulish verb forms GABAS ‘took, has taken’ (RIG II: L–55), PRINAS ‘bought, has bought’ (RIG II: L–32), and READDAS ‘gave, has given’ (RIG II: L–78), as well as probably Insular Celtic preterite forms as they looked prior to the rise of the absolute-conjunct system. Predictably, it has parallels outside Celtic, e.g. OL. DEDRON (CIL I: 30, Rome), DEDERO (CIL I: 2659, Albano Laziale), which preserved the inherited form, later replaced by *dederunt* (with spread of the primary ending *-nt* to the perfect), the Vedic sigmatic aorist *dhāk* ‘burnt’ < **d^hēg^{uh}-s-t*, etc.

Somewhere down the line, the primary ending *-(n)ti* dropped its final vowel in Celtic, as shown by the 3rd p. pl. present CABINT (Novallas). The designers of this alphabet unequivocally parsed post-apocope *-t#* as a voiceless stop, in spite of the fact that an early rule *-ntV- > -ndV-* can be set up for Celtiberian (see Prósper 2016). This points either to comparatively early loss of *-i*, or to preservation of the voiceless quality of the stop, favoured by the phonosyntactic alternation *-nt̥i# V-* vs. *-nt(i)# C-*. We may consequently surmise that a 3rd p. sg. present form would have terminated in <T> in documents written in the same alphabet.

As observed in different previous works (the last of which is Martínez Chico & Prósper 2021), postvocalic final dentals were eventually lost, and the exponent of the ablative case, once its distinctive feature ceased to exist, merged with that of the instrumental endings (and with the nominative in the first declension). This can be detected in a number of inscriptions (see Martínez Chico & Prósper 2021), but it is difficult to draw significant conclusions since this is a sociolinguistic, not only a dialectal problem: when loss of the final dental segment was reflected in official documents, we may be pretty certain that the change had been completed.

⁴³ Concurring with Kiparsky (2006) that there are no languages with a synchronic word final voicing rule, as opposed, for instance, to German or Catalan, which have a devoicing rule. This does not mean that allophonic voicing is not common between words closely connected in speech when the second begins with a vowel, but this does not affect phonemic contrasts.

⁴⁴ If this document is written in the dual writing system (which for the time being can only be claimed with any degree of certainty for the *Arevaci*), we have to read <*dēkez*>. See Martínez Chico & Prósper 2021: 176–177. Jordán (2007) transliterates it as <*tegez*>, overtly misapplying the rules established by himself for texts showing the distinction between voiced and voiceless syllabograms and in this way jeopardising the very possibility of finding a meaning for this document.

⁴⁵ The probably allative preposition DV is now attested in the bronze of Novallas.

⁴⁶ See Prósper 2007: 85–87. The preceding form *silabur* ‘money’ is its DO. The occasionally raised objection that the sign <*ta*> is smaller or written in superscript (which would apparently call into question its obvious status as a syllabogram endowed with phonemic content) does not withstand scrutiny. Upon autopsy (December 12th, 2021) I can confirm that <*ta*> and <*ko*> are roughly the same size, but the scribe tried to spare some room by inserting the upper angle of the second sign into the right side of the X-shaped sign <*ta*>. As a consequence, <*ko*> stands slightly underneath the writing line.

A comparative table of all the Hispano-Celtic forms discussed in the paper

Indo-European	Hispano-Celtic	Iberian script	Latin alphabet	Forms in the Latin alphabet
<i>-ntij-</i>	<i>-ntj-</i>	<i><nti></i> <i><nz></i> <i><z></i>	<i><NTI></i>	SEGONTIVS [<i>sekontios, sekonzos</i>]
			<i><NS></i>	VELONSAE
			<i><NS></i>	BVGANSONIS
			<i><SS></i>	SEGOSSOQVM [<i>sekontios, sekonzos</i>] TELAŠSICVS [<i>telazokum</i>] NIŠSICVM, ?NESSIA/NESSIA ?LEIOŠS(A)E/LEIOŠS(A)E
			<i><SS></i>	SEGOŠS(VS) [<i>sekontios, sekonzos</i>]
<i>-tij-</i>	<i>-tj-</i>	<i><ti></i> <i><z></i>	<i><TI></i>	CLOVTIVS [<i>koloutios</i>] ANNETIAE
			<i><S></i>	BRASACA [<i>barazioka</i>] SECOVESO ANNESVS VASCASVS
			<i><S></i>	CLOVSOCVM [<i>koloutios</i>]
			<i><Z></i>	ANNEZA
<i>-(n)Ktij-</i>	<i>-χtj-</i> > <i>-jtj-</i> (Arevaci)	<i><ti></i> <i><z></i>	<i><ITI></i>	VŠSVEITIO, VSSEITIO CONTAH+T+IA
			<i><IS>/<IS></i>	AISAE ?VXSEISVS/VXSEISVS
			<i><IZ></i>	CONTAIZA ARBAIZA CAIZITA
<i>-T=s-</i>	<i>-ts-</i>	<i><S></i>	<i><SS></i>	DERCINOASSEDENSIBVS

Dental stops show a tendency to weakening and eventual loss in final position that often compromises grammatical distinctions: this is a platitude in Greek and Latin studies, especially as far as verbal endings are concerned. To illustrate the process and its repercussion in synchronic phonemic contrasts, we may note its remarkable similarity with present-day Spanish. In some central varieties of Castilian Spanish, specifically in Salamanca, Valladolid and other provinces of «Old Castile», as well as some areas of Madrid, many speakers only distinguish two dental phonemes in codas, including word-final position: /s/ and /θ/. The latter is realised as an approximant, and results from merger of historical /θ/ and /d/ [ð], still respec-

tively written <z> and <d>. ⁴⁷ In some places, by contrast (e.g. some areas of Madrid), final -ð is laxed into a barely perceptible approximant ^{-ð} and progressively lost, so that -s# and -θ# are the only surviving word-final obstruents, at least in some contexts (since the preceding vowel and the following vowel/consonant also play a role). ⁴⁸

In a nutshell, attrition of final stops proceeded in Celtiberian in a typologically frequent way: the only inherited word-final dental was not specified for voice and tended to disappear early on after consonants, but may have been partly restored for grammatical reasons; it was lenited and may have eventually disappeared after vowels; a released, voiceless dental stop, which came to be in final position after apocope of -i, lasted longer but probably suffered the same fate over time.

References

- Abascal, Juan Manuel, José Ramón López de los Mozos. 1993. Inscripciones latinas de Sayatón (Guadalajara) en Territorium de Ercávica. *Wad-Al-Hayara* 20: 357–364.
- Abascal Palazón, Juan Manuel. 2015. *Estudios sobre la tradición manuscrita de la epigrafía hispano-romana*. Madrid: Real Academia de la Historia.
- Abascal Palazón, Juan Manuel, Geza Alföldy, Rosario Cebrián Fernández. 2011. *Segóbriga V: inscripciones romanas 1986–2010*. Madrid: Real Academia de la Historia.
- Alfaro Peña, Eduardo. 2017. Los Casares de San Pedro Manrique: Un viejo *oppidum*, una pequeña ciudad. *Anejos de Segovia Histórica* 2: 51–69.
- Ballester, Xaverio. 2003. Celtibérico SECoNZOS = ¿*Secundus* o SECoNTiOS?. *Veleia* 20: 351–354.
- Barbas Nieto, Ricardo, Emilio Gamoz Pazos, Helena Gimeno Pascual. 2011. Nuevos epígrafes latinos en el alto Tajo. Abánades, Ocentejo y Zaorejas. *Veleia* 28: 161–173.
- BB.III = Francisco Beltrán, Jürgen Untermann, Javier De Hoz. 1996. *El tercer bronce de Botorrita*. Saragossa: Institución Fernando el Católico.
- BB.IV = Francisco Villar, María Antonia Díaz, Manuel Medrano, Carlos Jordán Cólera. 2001. *El IV Bronce de Botorrita (Contrebia Belaisca): Arqueología y lingüística*. Salamanca: Ediciones Universidad de Salamanca.
- Beltrán Lloris, Francisco, Carlos Jordán Cólera, Borja Díaz Ariño, Ignacio Simón Cornago. 2021. The Novallas bronze tablet: An inscription in the Celtiberian language and the Latin alphabet from Spain. *Journal of Roman Archaeology* 34: 713–733.
- CIL = Mommsen, Theodor, et al. (eds.). 1862–. *Corpus Inscriptionum Latinarum*. Berlin: De Gruyter.
- CIRB = Crespo Ortiz de Zárate, Santos, Ángeles Alonso Ávila. 2000. *Corpus de inscripciones romanas de la provincia de Burgos. Fuentes epigráficas para la historia social de Hispania romana*. Valladolid.
- Corell i Vicent, Josep. 1992. Inscripciones romanas de Iniesta y Sisante (Cuenca) y la Iglesia del Cid (Teruel). In: *Estudios de arqueología ibérica y romana: homenaje a Enrique Pla Ballester*: 583–596. Valencia: Museu de Prehistòria de València.
- EDCS = *Epigraphik Datenbank Clauss-Slaby*. Available online at: <http://www.manfredclaus.de>. Accessed on 02/15/2022.
- eDIL = *An Electronic Dictionary of the Irish Language, based on the contributions to a Dictionary of the Irish Language* (Dublin: Royal Irish Academy, 1913–1976). Available online at: <http://www.dil.ie>. Accessed on 02/15/2022.
- EDPC = Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden: Brill.
- Encarnação, José. 2016. Os romanos de Castro Verde (Conventus Pacensis, Lusitania). *Revista Portuguesa de Arqueologia* 19: 195–210.
- Eichner, Heiner. 1989. ‘Damals und heute: Probleme der Erschließung des Altkeltischen zu Zeußens Zeit und in der Gegenwart’. In: Bernard Forssman (ed.). *Erlanger Gedenksfeier für Johann Kaspar Zeuß, Erlanger Forschungen, Reihe A Geisteswissenschaft, Band 49*: 9–56. Erlangen: Universitätsbibliothek.

⁴⁷ The sibilant goes back to L. -s#, and the two obstruents -θ# and -d# were in word-final position after final vowels were apocopated posterior to the western Romance processes of palatalisation (*radice(m)* > *raíz* ‘root’) and intervocalic voicing (*ite* > *id* ‘go!’).

⁴⁸ This is of course a simplified description of a very complex variation that is in flux. See for details Molina Martos 2016, with references.

- ERS = Jimeno, Alfredo. 1980. *Epigrafía romana de la provincia de Soria*. Soria: Diputación de Soria.
- ERSg = Santos Yanguas, Juan, Javier del Hoyo Calleja, Ángel Luis Hoces de la Guardia y Bermejo. 2005. *Epigrafía romana de Segovia y su provincia*. Segovia: Caja Segovia.
- Ezquerro Lebrón, Beatriz, José B. Vicente Redón. 1999. El bronce celtibérico de Torrijo del Campo (Teruel). In: Francisco Beltrán et al. (eds). *Pueblos, lenguas y escrituras en la Hispania prerromana: Actas del VII Coloquio sobre Lenguas y Culturas Paleohispánicas*: 581–594. Salamanca: Ediciones Universidad de Salamanca.
- Garnier, Romain. 2010. *Sur le vocalisme du verbe latin. Étude synchronique et diachronique*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Gimeno Pascual, Helena, Manuel Ramírez Sánchez. 2002. Dos inscripciones inéditas de la provincia de Soria (España). *Zeitschrift für Papyrologie und Epigraphik* 139: 273–278.
- González Rodríguez, M. Cruz, Joaquín Gorrochategui Churruga. 2011. Nuevas lecturas de tres inscripciones de Clunia (Hispania Citerior) con repercusión sobre la expresión de algunas agrupaciones familiares celtibéricas. *Veleia* 28: 271–279.
- Hackstein, Olav. 2002. Uridg. *CH.CC > *C.CC. *Historische Sprachforschung* 115: 1–22.
- Hamp, Eric P. 1976. On some Gaulish names in *-ant-* and Celtic verbal nouns. *Ériu* 27: 1–20.
- Harðarson, Jon A. 1993. *Studien zum urindogermanischen Wurzelaorist und dessen Vertretung im Indoiranischen und Griechischen*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- IRMN = Castillo, Carmen, Joaquín Gómez-Pantoja, M. Dolores Mauleón. 1981. *Inscripciones romanas del Museo de Navarra*. Pamplona: Diputación Foral de Navarra.
- Irslinger, Britta. 2002. *Abstrakta mit Dentalsuffixen im Altirischen*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Jordán Cólera, Carlos. 2007. Estudios sobre el sistema dual de escritura en epigrafía no monetaria celtibérica. *Palaeohispanica* 7: 101–142.
- Jordán Cólera, Carlos. 2016. Sobre el valor de la *s* marcada en la epigrafía celtibérica en alfabeto latino. *Études Celtiques* 41: 75–94.
- Jordán Cólera, Carlos, Borja Díaz Ariño. 2022. Antropónimos de la Hispania céltica escritos con *Z*. *Emerita* 90: 149–167.
- Kiparsky, Paul. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32: 217–236.
- LEIA = Vendryes, Joseph, Édouard Bachellery, Pierre-Yves Lambert. 1959–1996. *Lexique étymologique de l'irlandais ancien*. Dublin: Dublin Institute for Advanced Studies.
- Lipp, Reiner. 2016. Final stops in Indo-European: their phonological classification as a key to the Proto-Indo-European root structure constraint. *Slovo a Slovesnost* 77: 251–299.
- LIV = Rix, Helmut, et al. 2001². *Lexicon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. Wiesbaden: Ludwig Reichert.
- LV = Michel Lejeune. 1974. *Manuel de la langue vénète*. Heidelberg: Winter.
- Martínez Chico, David, Blanca María Prósper. 2021. A new Celtiberian tessera in the Latin alphabet from Virovesca (Briviesca, Burgos) and the typology of tesserae hospitales. *Zeitschrift für Celtische Philologie* 68: 167–196.
- McCone, Kim. 1996. *Towards a Relative Chronology of Ancient and Medieval Celtic Sound Change*. Maynooth: National University of Ireland at Maynooth.
- Méndez Dosuna, Julián V. 1985. La duración de *S* en los grupos *SP*, *ST*, *SK*: A propósito del orden regular de difusión en algunos cambios fonéticos. In: José Luis Melena (ed.). *Symbolae Ludovico Mitxelena Septuagenario Oblatae*: 647–655. Vitoria: Universidad del País Vasco.
- MLH I = Untermann, Jürgen. 1975. *Monumenta Linguarum Hispanicarum I. Die Münzlegenden*. Wiesbaden: Ludwig Reichert.
- MLH IV = Untermann, Jürgen. 1997. *Monumenta Linguarum Hispanicarum IV. Die tartessischen, keltiberischen und lusitanischen Inschriften*. Wiesbaden: Ludwig Reichert.
- MLH V = Wodtko, Dagmar S. 2000. *Monumenta Linguarum Hispanicarum V. Wörterbuch der keltiberischen Inschriften*. Wiesbaden: Ludwig Reichert.
- Molina Martos, Isabel. 2016. Variación de la *-d/* final de palabra en Madrid: ¿prestigio abierto o encubierto?. *Boletín de Filología* 51/2: 347–367.
- Nussbaum, Alan J. 2007. Latin present stems in *-sā-*: A possibly not so minor type. Handout. Kyoto, September 2007, 1–15.
- Nussbaum, Alan J. 2016. Replacing *locus* ‘place’ in Latin *locuplēs*. In: Dieter Gunkel, Joshua Katz, Brent Vine, Michael Weiss (eds.). *Sahasram Ati Srajas. Indo-Iranian and Indo-European Studies in Honor of Stephanie W. Jamison*: 276–295. Ann Arbor / New York: Beech Stave Press.

- Prósper, Blanca María. 1997. La divinidad paleohispana *Cossue/Coso* y el dios itálico *Consus*. *AION (Sezione Linguistica)* 19: 267–302.
- Prósper, Blanca María. 2005. Estudios sobre la fonética y la morfología de la lengua celtibérica. In: Francisco Villar, Blanca M. Prósper. *Vascos, celtas e indoeuropeos. Genes y lenguas*: 153–364. Salamanca: Ediciones Universidad de Salamanca.
- Prósper, Blanca María. 2007. *Estudio lingüístico del plomo celtibérico de Iniesta*. Salamanca: Ediciones Universidad de Salamanca.
- Prósper, Blanca María. 2008. *Estudios sobre el bronce de Botorrita I*. Pisa-Rome: Fabrizio Serra Editore.
- Prósper, Blanca María. 2011. The instrumental plural in the thematic noun inflection of Continental Celtic. *Historische Sprachforschung* 124: 224–241.
- Prósper, Blanca María. 2012. The Hispano-Celtic divinity *Ilurbeda*, gold mining in Western Hispania and the syntactic context of Celtiberian *Arkatobezom* «silver mine». *Die Sprache* 49: 53–83.
- Prósper, Blanca María. 2014. Time for Celtiberian dialectology: Celtiberian syllabic structure and the bronze tablet from Torrijo del Campo (Teruel). *Keltische Forschungen* 6: 115–155.
- Prósper, Blanca María. 2016. *The Indo-European Names of Central Hispania. A Study in Continental Celtic and Latin Word Formation*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Prósper, Blanca María. 2017. Novallas, ¿un testimonio celtibérico de la ingeniería romana en Hispania?. In: Juan Antonio Álvarez-Pedrosa et al. (eds). *Ratna. Homenaje a la profesora Julia Mendoza*: 163–169. Madrid: Guillermo Escolar.
- Prósper, Blanca María. 2018a. The Indo-European personal names of Pannonia, Noricum and Northern Italy: comparative and superlative forms in Celtic, Venetic, and South-Picene. *Voprosy Onomastiki* 15: 108–138.
- Prósper, Blanca María. 2018b. The Venetic agent nouns in *-tōr-* revisited. In: José M. Vallejo, Carlos G. Castellero, Iván Igartua (eds.). *Studia Philologica et Diachronica in Honorem Joaquín Gorrochategui*: 453–471. Vitoria: Publicaciones de la Universidad del País Vasco.
- Prósper, Blanca María. 2019. Celtic and Venetic in contact: the dialectal attribution of the personal names in the Venetic record. *Zeitschrift für Celtische Philologie* 66: 7–52.
- Prósper, Blanca María. 2020. The Sabellic accusative plural endings and the outcome of the Indo-European sibilants in Italic. *Journal of Language Relationship* 18: 41–79.
- Prósper, Blanca María. 2021. Latin *sancītō* vs. Lusitanian *singeieto*. Is the Lusitanian inscription of Arroyo de la Luz I the westernmost *lex sacra*?. In: Marco A. Coronel Ramos, Ricardo Hernández Pérez (eds.). *PRISCORVM INTERPRES. Homenaje al profesor Jaime Siles (= Studia Philologica Valentina, Anejo 2)*: 339–350. Valencia: Ediciones Universidad de Valencia.
- Prósper, Blanca María, Marcos Medrano Duque. 2022. Ancient Gaulish and British divinities: notes on the reconstruction of Celtic phonology and morphology. *Voprosy Onomastiki* 19: 9–47.
- RIG II = Lambert, Pierre-Yves. 2002. *Textes gallo-latins sur instrumentum (Recueil des inscriptions gauloises. Vol. II, 2)*. Paris: C.N.R.S.
- Rodríguez Ramos, Jesús. 2002. *Okelakom, Sekeida, Bolšken. Kalathos* 21: 429–434.
- Roso de Luna, Mario. 1904. Nuevas inscripciones romanas de la región Norbense. *Boletín de la Real Academia de la Historia* 44: 113–137.
- Sáenz de Buruaga, Andoni, Paquita Sáenz de Urturi. 1994. La epigrafía romana de San Román de San Millán. *Veleia* 11: 49–82.
- Saquete Chamizo, José C., Santiago Guerra Millán. 2015. Una inscripción constructiva procedente de Metellinum (provincia Lusitania). *Zeitschrift für Papyrologie und Epigraphik* 196: 303–306.
- Schrijver, Peter. 1993. On the development of vowels before tautosyllabic nasals in Primitive Irish. *Ériu* 44: 33–52.
- Šefčík, Ondřej. 2020. The development of Indo-European obstruent clusters of types plosive + *t/s* into Slavic. *Zeitschrift für Slawistik* 65: 222–248.
- Simón Cornago, Ignacio, Carlos Jordán Cólera. 2018. The Celtiberian *S*. A new sign in (Paleo)Hispanic epigraphy. *Tyche* 33: 183–205.
- Untermann, Jürgen. 1967. Die Endung des Genitiv singularis der o-Stämme im Keltiberischen. In: W. Meid (ed.). *Beiträge zur Indogermanistik und Keltologie, Julius Pokorny zum 80. Geburtstag gewidmet*: 281–288. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Untermann, Jürgen, Francisco Villar. 1999. Las «téseras» de Gadir y Tarvodurum. In: F. Villar et al. (eds). *Pueblos, Lenguas y Escrituras en la Hispania Prerromana. Actas del VII Coloquio sobre Lenguas y Culturas Paleohispánicas*: 719–732. Salamanca: Ediciones Universidad de Salamanca.

- Velaza, Javier. 2008. La onomástica personal en la epigrafía romana de la Meseta Meridional: una aproximación. In: G. Carrasco Serrano (ed.). *La romanización en el territorio de Castilla-La Mancha*: 367–383. Ciudad Real: Ediciones Universidad de Castilla-La Mancha.
- Vicente Redón, Jaime D., Beatriz Ezquerro Lebrón. 2003. La tésera de Lazuro: un nuevo documento celtibérico en La Caridad (Caminreal, Teruel). *Palaeohispanica* 3: 251–269.
- Villar, Francisco. 1993–1995. El instrumental en celtibérico. *Kalathos* 13–14: 325–338.
- Villar, Francisco. 1995. *A New Interpretation of Celtiberian Grammar*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Vine, Brent. 2006. On Thurneysen-Havet's Law in Latin and Italic. *Historische Sprachforschung* 119: 211–249.
- Widmer, Paul. 2004. *Das Korn des weiten Feldes: Interne Derivation, Derivationskette und Flexionsklassenhierarchie. Aspekte der nominalen Wortbildung im Urindogermanischen*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Б. М. Проспер. Сибилянты в испано-кельтских языках: вопросы фонетики, фонологии и орфографии

На основании недавних открытий относительно использования латинского алфавита в официальных документах, написанных на кельтиберском языке в различных регионах Центральной Испании, а также в результате нового прочтения автохтонных имен, обнаруживаемых в латинской (не кельтиберской) эпиграфике, автор выдвигает ряд новых гипотез об устройстве кельтиберской фонологии. Аргументы, основанные на эпиграфическом и лингвистическом анализе, в свою очередь, позволяют точнее очертить границы испано-кельтского языкового ареала, определить внутри него некоторые диалектные различия и описать возникновение и фиксацию отдельных орфографических норм. Приводятся несколько новых этимологизаций для ряда личных имен, до сих пор не имевших этимологии или этимологизировавшихся ошибочно (включая случаи, когда новая этимология позволяет отождествить ту или иную форму с формами, сохраняющимися в галльском или в островных кельтских языках). Наконец, для трех известных лексем предлагаются новые чтения BVGANSONIS, CLOVSOCVM и AISAE (а также, возможно, VXSEISVS).

Ключевые слова: кельтиберские языки; индоевропейская реконструкция; кельтская этимология.

Lexicostatistical studies in Khoisan II/2: Towards a more precise phylogeny for the Tuu family

The paper is the second part of an extensive study aimed at clarifying the internal relationships within the Tuu (Southern Khoisan) linguistic family of South Africa and reconstructing a reasonably accurate Swadesh wordlist for Proto-Tuu. In this section, I first investigate the issue of extensive areal contact between two languages belonging to two different subgroups of Tuu (N|ng and !'Auni), filtering out potential borrowings from the former into the latter which may obscure both etymological judgement and lexicostatistical calculations. Next, lexicostatistical matrices and resulting phylogenetic trees are offered for the entire family, demonstrating that a ternary model, in which Proto-Tuu splits into three more or less equidistant branches (!Ui, Nossob, and Taa), is likely preferable to any possible binary models; arguments for a closer proximity between Nossob and Taa are analyzed and found generally inconclusive. Finally, some remarks are made about the reconstructibility of the Swadesh wordlist for Proto-Tuu and on some of its peculiar properties as compared to attested Tuu languages. The Appendix section contains the entire second half of the Swadesh wordlist (items 51–100) as reconstructed on intermediate levels and on the Proto-Tuu level.

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

Preliminary remarks

In the first part of the present paper, published in an earlier issue of JLR (Starostin 2021), I have given a brief description of the principal goals, methods, and problems (both technical and substantial) of conducting a detailed lexicostatistical survey of all known languages of the Tuu family; the theoretical part was then illustrated with actual comparative material from the first half of the Swadesh wordlist (items 1–50). Subsequently, this second part of the study, in addition to presenting data on the second half of the list (see the Appendix at the end of the paper), will focus on the actual analysis of the data, starting with some simple statistics and their phylogenetic interpretations, and then concentrating on more “fine-grained” manual analysis of the relevant data. The overall plan is as follows:

1) first, I shall directly address the issue of *areal contact* between the surveyed languages, most notably, concentrating on the consequences of N|ng-'Auni bilingualism which may or may not have resulted in incorrect phylogenetic conclusions in the past;

2) next, I shall present the *lexicostatistical matrices* for the full 100-item wordlists for all languages and doculects of the survey, along with their glottochronological interpretations and a brief discussion of their problematic areas;

3) finally, identification of problems stemming from the crude statistical approach will lead to a manual re-analysis of some of the potential matches in terms of areal contacts, shared innovations, or retained archaisms, resulting in an overall grading of potential phylogenetic schemes in accordance with their respective strengths and weaknesses.

Areal contact in Tuu: the issue of isoglosses between N||ng (N|uu) and |'Auni

In theory, any non-accidental lexical similarity between two languages that are in contact with each other today or may have plausibly been in contact with each other some time in the past may be due to lexical borrowing, regardless of whether the languages in question are genetically related. In practical terms, however, we usually evaluate such a possibility based on what could be called the “common sense scenario”. Thus, if such similarities are seen as proportionally more common across basic rather than cultural vocabulary (or, at least, if they are more or less equally spread across both layers), it makes sense to assume that they stem from a common ancestor, and that special, additional arguments are needed to challenge that default assumption. (That any such lexical similarities should also display recurrent phonetic correspondences is also a vital requirement, but it is somewhat irrelevant for this purpose, since systemic borrowings between languages tend to also take place in accordance with strict correspondence patterns). Such additional arguments in favor of areal contact may come from:

phonetics — if potential cognates between languages A and B feature more than one set of recurrent correspondences, this is a clear indication that at least one of the sets takes place in more recent loans from one language to another (e.g. the case of English and French, or modern Chinese “dialects” vs. the literary Chinese language);

distribution — if language A seems to be related to language B, but language B is clearly more tightly connected to languages C and D, which, on their own, show relatively little affinity with language A, the increase in similarity between A and B is almost certainly due to secondary contact (e.g. the case of Armenian and Iranian languages, see Campbell & Poser 2008: 80);

morphology and semantics — if the formal and semantic properties of lexical items in language A consistently match only a small subset of the respective properties of lexical items in language B (e.g. the words in A are attested only as grammatically complex stems rather than simple roots, or only in what looks like secondary / figurative meanings), this can be interpreted as the result of borrowing from B rather than inheritance from a common ancestor.

When this line of reasoning is applied to the lexical data of Tuu languages, it can be seen that, on the whole, it is difficult to suspect most of them of ever having been in intense secondary contact with each other. Although phonetic correspondences between them, as has already been partially shown in the first part of the paper, sometimes give the impression of being fairly erratic (and it is not always clear how much of that is due to actual phonetic change and how much to inaccurate data transcription), it has so far been impossible to properly identify two distinct sub-sets of correspondences for any given pair of Tuu languages. Distribution-wise, as we shall clearly see below from the lexicostatistical matrices, most of the languages support a rather transparent phylogenetic structure with little out of the ordinary. As for morphology and semantics, most of the etymological work done on Tuu so far finds more evidence for morphological variation on the Proto-Tuu level, or between the different branches of Tuu, than for secondary convergence processes based on the borrowing of words in various fossilized morphological patterns from one Tuu language into another.

Thus far, the lion’s share of loans in Tuu basic lexicon has been identified as stemming from non-Tuu languages — most commonly, Khoe (Traill & Nakagawa 2000), but also occasional borrowings from Bantu (especially in ||Xegwi, see Lanham & Hallows 1956a) or even European sources (modern N|uu is particularly liable to be influenced by Afrikaans, see Sands et al. 2007).

One notable exception to this tendency is the complicated relation between |'Auni, one of two known languages from the Nossob subgroup, and the N||ng cluster from the !Ui subgroup. There is no need to go into detail here on the nature of the alleged N||ng-|'Auni bilin-

gualism that resulted in the adoption of many N||ng words into !'Auni itself; the sociolinguistic foundation for these processes was already described in early works such as Bleek 1937, and later extensively commented upon by Güldemann (2014, 2018). What interests us specifically is a procedure that would allow to realistically distinguish between “primary” (inherited) cognates and “secondary” (borrowed) cognates between these two languages; the correct solution of this puzzle would not only be instrumental in more precisely determining the phylogenetic status of the Nossob languages, but would also be of interest to any specialists beset with similar problems in other linguistic areas of the world.

First, let us draw up a complete table (Table 1) of all potential lexicostatistical (i.e. not just related, but fully equivalent in terms of semantics) matches between !'Auni and N|uu; for the latter, I list the old ||Ng!ke forms as recorded by D. Bleek (1956, 2000), which would be likely to serve as the actual source of the borrowing, and their modern day N|uu forms as reflecting more accurate transcription.

Table 1. Lexicostatistical matches between !'Auni, “Old N||ng” (||Ng!ke) and Modern N|uu.

Word	!'Auni	Ng!ke	N uu
'bird'	si= u	wí ~ wi:	q ^h ui-si
'blood'	xauu	xau	xau-ke
'claw /nail/'	ora-sa	uri-si	qoro-si
'come'	sa ~ sé ~ sí	si ~ se ~ seya ~ sa	sa: ~ ca:
'die'	!ã	!a:	!a:
'dog'	!õ:	!wiŋ	! ^h un ~ ! ^h uŋ
'drink'	x'ã: ~ x'ẽ	x'a: ~ x'ã ~ x'ẽ	x'ãi
'ear'	!ui	!we: ~ !we:-ntu ~ !u:-ntu	!ui-si
'eat'	à ~ hà ~ hàa	ã ~ ẽ ~ ẽi	?ã
'eye'	coo / c'axu	cáxu ~ ca:xem	c'axam
'fire'	!i	!i	!i:
'fly'	zé	—	ze: ^f
'give'	a	a ~ a:	?ã:
'hair'	! ^h óo	u ~ ! ^h ú	! ^h u:-ke
'hand'	kx'a ~ kx'an	kx'a	kx'a:
'head'	!ã: / x:uu	!ã ~ !a:	!ã:
'hear'	tu: ~ tu:i	tu ~ tú ~ tu:i	ɸu:
'heart'	!e: ~ !e:	!ai ~ !e	!e:
'horn'	ẽi	ãi	q ^h oe-si
'I'	n ~ ŋ ~ na ~ m	ŋ ~ n	ŋ
'know'	xai ~ x'e-ki	ai	xae
'lie'	tòa	tia ~ kia:	ɸa:
'long'	!á-si	!a:	!ã:
'meat'	θwe ~ θwi	θwai: ~ θwai	θoe
'mouth'	tu ~ t ^h u / !u:	tu ~ tu:	ɸu:
'name'	!ẽ ~ !ẽn	!ẽ	ka=!ĩ
'neck'	!õi / !ú	!ú ~ kú	!qu:
'night'	àu ~ ò	a ~ !a:	a:
'nose'	!õ	!u-tu	!u-ɸu

‘one’	ʃú ~ ʃú-u	we: ~ 'we:	'oe
‘rain’	hà:a	!hà ~ !a:	(ʃqau)
‘road’	!án	(tirau)	!an ~ !aŋ
‘say’	ko / u	ka	ka
‘see’	ǃà: ~ ǃe	ǃa: ~ ǃe ~ ǃi: ~ ǃi	ǃa:
‘sit’	sã ~ são ~ so	so ~ so: ~ so:	sũĩ
‘sleep’	θwōi	θwoiŋ ~ θwoeŋ ~ θóeŋ	θun ~ θuŋ
‘smoke’	áu	(wi:)	o: ^f -ke
‘this’	a	a	a
‘thou’	a	a	a
‘tongue’	ǃãri	ǃẽ	ǃãn ~ ǃãĩ
‘tooth’	ẽĩ	ãĩ: ~ ẽ: ~ ẽĩ	hãĩ
‘tree’	θwa:a ~ θwa:-sa	θo ~ θo: ~ θ ^h o	θo:
‘walk /go/’	a ~ e ~ a(:)	a(:) ~ ai	aʔa
‘water’	k ^h á: ~ k ^h áá ~ k ^h ái	!hà: ~ !hà ~ !à: ^f ~ !a: ~ hà:	!q ^h a:
‘we’ (excl.)	si ~ se ~ ci	si	si
‘we’ (incl.)	i ~ e	i	i
‘woman’	ǃé:	ǃai- ~ ǃai- ~ ǃe:-	ǃe:-ki

Taken in the general context of other Tuu languages, these parallels may be divided into the following subgroups:

(1) “Pan-Tuu” roots that are present in the same basic meaning in all or most Tuu languages covering all three subgroups. Provided that within ǃAuni they do not feature any specific “non-ǃAuni” traits (e.g. an uncommon phonetic shift or morphological add-on that is more typical of N||ng than ǃAuni), there is no compelling reason to treat such forms as borrowings.

These roots are: ‘bird’, ‘come’, ‘die’, ‘dog’, ‘drink’, ‘eat’, ‘fire’, ‘hair’, ‘hand’, ‘heart’, ‘horn’, ‘P’, ‘meat’, ‘name’, ‘nose’, ‘one’, ‘see’, ‘sit’, ‘sleep’, ‘thou’, ‘tongue’, ‘tree’, ‘water’, ‘we’ (both excl. and incl.).

(2) Logically close to this group are “Pan-!Ui + Pan-Nossob” roots which have no lexicostatistical (or even etymological) parallels in Taa, but are found in both ǃAuni and |Haasi. Since we have no evidence of lexical contacts between N||ng and |Haasi, and since, once again, there are no specific phonetic or morphological arguments for their being borrowed into ǃAuni, we can essentially merge them with the first group for our purposes, also considering them most likely inherited in ǃAuni.

These roots are: ‘blood’, ‘hear’, ‘night’, ‘this’.

(3) Cases when ǃAuni and |Haasi contradict each other, and the ǃAuni form is closer to N||ng than the |Haasi one. This is an apriori suspicious situation for which no universal solution is available; each case has to be evaluated on its own. These roots are:

(3a) ǃAuni ||ora-sa ‘nail’ vs. |Haasi k^a=ǃü id. It is not likely that |Haasi preserves the original situation; in fact, it is even possible that Story’s semantic glossing here is inaccurate, since, strictly speaking, the form he quotes has to be literally translated as ‘fingers’ (plural prefix k^a= + ǃü ‘finger’). Meanwhile, even if ǃAuni ||ora-sa is indeed quite similar to N||ng!ke ||uri-si, N|uu ||oro-si, their codas and morphological properties are different enough to reject borrowing as the likeliest solution. Since this is really a “Pan-Tuu” root as well, we prefer to treat the ǃAuni item as inherited;

(3b) ‘give’ — according to Bleek 1937, ǃAuni has a variety of synonyms here, only one of which, the verb *a*, attested in a single textual example (‘give me tobacco’), has reliable parallels

in !Ui (**a* is the main verb of giving in both |Xam and the entire N||ng cluster). |Haasi, on the other hand, has *i*, another monovocalic root which is probably not related. Technically, this could be the result of N||ng influence on !'Auni, so we are justified in marking this item as a potential (uncertain) borrowing;

(3c) !'Auni 'go' *||a ~ ||e ~ ||a(:)* vs. |Haasi *ʃa* id. Etymology of the |Haasi form is unclear; the !'Auni stem, on the other hand, is the same as the “Pan-!Ui” equivalent for the same meaning (**||a-* or, perhaps, **||aʔa*, as in Modern N|uu). Given that the same root is also found in Taa in a more specialized meaning (!Xóõ *||ʔâe* ‘to go out hunting or gathering’), it cannot be considered a lexical innovation in Proto-!Ui, and chances of it being lost and then restored in !'Auni under N||ng influence are low; we are fully justified in treating it as an inherited item.

(4) Cases when reflexes of a “Pan-Tuu” or a “Pan-!Ui + Pan-Nossob” root in !'Auni bear a notable resemblance specifically to N||ng. Again, these cases have to be evaluated individually:

(4a) !'Auni *ʃui* ‘ear’ features the same morphological structure as N|uu *ʃui(-si)* (though, curiously enough, not the same as ||Ng!ke *ʃwe:-ntu*). Even more importantly, the root shape of the word for ‘ear’ in |Haasi is *ʃa-*, which is morphologically comparable with !Xóõ *ʃũ^h-ã* (assuming that *ʃa-* ← **ʃu-a-* with vowel contraction) — another slight indication that the !'Auni form may have been borrowed from N||ng or, at the least, “influenced” by it in some way. Considering that other probable instances of N||ng body part terms finding their way into !'Auni also emerge (see below), we are within our rights to mark this form as a potential (though not certified) borrowing;

(4b) somewhat more difficult is the situation with !'Auni *||ēi* ‘tooth’. This form is explicitly glossed as ‘tooth’ only in an early source (Bleek 1929: 86), but is not found in this meaning in Bleek 1937, where the only attested meaning for it is ‘horn’ (‘horn’ and ‘tooth’ are transcribed as if they were homonyms in quite a few !Ui doculects, but more reliable data from ||Xegwi and Modern N|uu show that they are, in fact, quite different roots). In any case, it very closely matches ||Ng!ke *||ēi* ‘tooth’, whereas |Haasi has *k'i=||ε* ‘teeth’ without the nasal coda. Still, the distance between these forms is not as large as in the case of ‘ear’; moreover, the morphological similarity between !'Auni and N||ng is not exclusive in this instance (e.g., |Xam also has a nasal coda), so we tentatively continue to regard the !'Auni form as inherited.

(5) “Doublets”, when Bleek records two forms for !'Auni, one of which is almost always closer to N||ng than the other. These include:

(5a) ‘eye’ — !'Auni *coo* vs. *c'axu*. The latter is almost identical to ||Ng!ke *cáxu*, †Khomani *c'axu* etc.; the former, on the contrary, is closer to the contracted form *cxɔ*, found in |Haasi. Bleek’s notes give no hints as to which of the two was the most commonly used, “neutral” form; it is permissible to simply disregard *c'axu* in the calculations and take *coo* as the inherited form, with a !'Auni-specific simplification of the cluster **cx-*, which itself appeared as a result of contraction from **c'a-x-* in Proto-Nossob;

(5b) ‘head’ — !'Auni *ʃa:* vs. *x:uu*. The former is unquestionably the Proto-Tuu equivalent for ‘head’, continuing to function as such in every attested doculect of !Ui and Taa. However, in |Haasi it is not encountered at all; instead, we have *(ɪ=)xɔ* ‘head’ (the meaning is clearly confirmed with multiple text examples), cognate with Proto-!Ui **xu* ‘face’ and most likely reflecting the semantic shift ‘face’ → ‘head’. Since the exact same form is encountered in the closely related !'Auni, it would be logical to postulate that shift on the Proto-Nossob level and analyze !'Auni *ʃa:* as a re-borrowing from ||Ng!ke (this scenario is explicitly advocated for by D. Bleek herself). We can add *ʃa:* to the list of potential borrowings;

(5c) ‘mouth’ — !'Auni *ʃu:* vs. *tu ~ t^hu*. The situation here is similar to ‘head’, except that this time, **tu* is specifically Pan-!Ui, not attested in Taa. Again, |Haasi has *ʃa* ‘mouth’, cognate with !'Auni *ʃu:* and, further on, with Proto-Taa **ʃu-* ‘mouth’ (note the exact same vocalic correspon-

dence as in ‘ear’, again hinting that |Haasi *ʃa* may be historically contracted ← **ʃu-a*); |’Auni *tu* subsequently looks like a borrowing from N|ng (also suspected by D. Bleek). Another potential borrowing;

(5d) ‘neck’ — |’Auni *ʃōĩ* vs. *||ú*. Here, the situation is different. The former stem does indeed seem more similar to N|uu *ʃqu*: ‘neck’ than the latter, but this similarity is hindered by differences in the codas, as the |’Auni form seems likely to reflect an original morphological structure like **ʃo-inj* or **ʃo-ni*, whereas neither Modern N|uu *ʃqu*: nor ||Ng!ke *!ú* ‘neck’ show any traces of nasal consonants. It is more likely here that the first form is genetically related to !Ui, and the second simply reflects a different root (it is impossible to tell the semantic difference from Bleek’s records);

(5e) ‘say’ — |’Auni *ko* vs. *|u*. Both verbs are used to introduce direct speech and are found virtually interchangeable with each other in Bleek’s recordings of |’Auni texts. In |Haasi, the only attested form for ‘say’ is *|wa*, obviously cognate with |’Auni *|u*. The verb *ku* ‘say’ is one of the main speech verbs in N|ng, so it is more than likely that *|u* in |’Auni is the inherited term, whereas *ko* is a N|ng verb that used to occasionally substitute it, i.e. a likely borrowing;

(5f) ‘woman’ — |’Auni *|é*: vs. *|e:ki*. Both forms feature the same lexical root, but the morphological shape of the second variant is the same as in N|uu (cf. the ||Ng!ke variants listed by Bleek: *|ai-ti* ~ *|ai-ki* ~ *|ai-ti* ~ *|ai-ki* ~ *|e-ki*), whereas the first variant, showing nasality and fully matching |Haasi *|ĩ* id., is especially typical of the Nossob area (more on this in the Appendix below). Most likely, the second variant is a borrowing, while the first one should be counted as an inherited term. (It is interesting that in Bleek’s texts, *|e:ki* ‘woman’ is encountered far more often than *|é*: — yet there is at least one text in which the two are essentially interchangeable, so we cannot assume that in colloquial |’Auni the newer borrowing had completely displaced the original lexeme).

(6) Cases where |’Auni and N|ng feature exclusive isoglosses vs. everybody else. There is only one of these in the 100-item wordlist, and it is almost certainly a borrowing: |’Auni *||xai* ‘know’ = ||Ng!ke *||ai* id. Meanwhile, |Haasi has *|üma* ‘know’, cognate with !Xóó *||úmā* id. This is a case of “criss-cross” distribution, breaking up a well-established phylogenetic structure; typically, in such cases one match at best reflects inheritance and the other one should be ascribed to contact, and since there is little linguistic evidence for secondary contact between |Haasi and Taa (although geographically this would be possible), it makes more sense to treat the |’Auni form as a potential borrowing.

(7) Various unclear cases:

(7a) ‘fly’ — |’Auni *zé* and Modern N|uu *ze:^f* are clearly the same item, and it is highly likely that they are further connected to Taa forms such as !Xóó *zāĩ^h*, Kacia *žōĩ^f* ~ *žwe^f*. This would seem to be a fine Proto-Tuu candidate, but the problem is that Bleek’s ||Ng!ke has *||óu* ‘to fly’ = |Xam *||au* ~ *||^hau* ~ *||^hóu* id.; meanwhile, the verb **zV^f* ‘fly’ seems to have a general areal distribution, perhaps amplified by its sound-symbolic nature (cf. also |Hoan *zoe^f* ‘fly straight’, Ju|’hoan *zōĩ^f* ‘to swarm /of bees/’, Naro *cāē^f* ‘to fly’ etc.). All of this raises suspicions that N|uu *ze:^f* may be a relatively recent innovation, and the |’Auni form could be easily borrowed from N|uu — although the genetic explanation cannot be fully ruled out either;

(7b) ‘lie’ — |’Auni *tòa* features a distinctly different coda from Proto-!Ui **ta* and Proto-Taa **tu*, but given the usual amount of vocalic variation in verbal stems, this is not enough to deny it cognacy with both of these forms; at the same time, it is notably different from N|ng to be judged as a borrowing. We treat it as inherited from a Common Tuu root **tV-*;

(7c) ‘long’ — |’Auni *|á-si* is a good match with ||Ng!ke *|’a:*, N|uu *|ā:*, as well as ||Xegwi-Z *|ā* id. (although the mismatch in click effluxes with the latter is a little puzzling). Unfortunately, no equivalent for ‘long’ is recorded in |Haasi, so there is no sure way of knowing here if the

word was borrowed or not (addition of *-si* means little, since it is a highly productive adjectival suffix in !'Auni). The word may be marked as a potential, but inconclusive, borrowing due to lack of significant evidence to the contrary;

(7d) ‘rain’ — !'Auni *hàa*, glossed as ‘water, rain’ in Bleek 1937, seems to be just a phonetic variant of *kʰá:* ‘water’ (← Proto-Tuu **!qʰa*); it is not clear if this polysemy may be reliably projected onto the Proto-Tuu or even the Proto-!Ui level (see RAIN in the appendix for more information), but in any case, there is hardly any sufficient reason here to suspect the word of being a borrowing from Nǁng;

(7e) ‘road’ — !'Auni has *!án* here, phonetically identical with Modern Nǀuu *!an ~ !aŋ* which, in turn, is cognate with Xegwi *kaŋ* and can thus go back to Proto-!Ui **!an* ‘path, road’. Since the word ‘road, path’ is not attested for !Haasi, there is no way of telling if it is inherited or borrowed. It may be noted that there are two more words with the same meaning in Bleek 1937: *ùru* and *ʃkʰei*, without any recorded semantic distinctions (and without any external etymologies) — this may be a hint that at least one of them may be the inherited term, while *!án* is really a borrowing, but all of this is inconclusive. Again, for specific purposes the item may be marked as at least a potential borrowing;

(7f) ‘smoke’ — although !'Auni *!áu* is phonetically similar to Nǀuu *oːke*, it is even more so to !Haasi *!au*, implying Proto-Nossob **!au* and a genetic connection to the Nǀuu form (as well as Xam *!á:*). No need to suggest borrowing in this case.

Altogether we have thus identified 28 matches which offer no specific arguments for being treated as loans (groups 1 and 2), 9 matches where the evidence clearly points to inheritance as the most likely reason for similarity, and 9 matches where the evidence is either ambiguous or points out to borrowing as the likeliest scenario (‘give’, ‘ear’, ‘fly’, ‘head’, ‘mouth’, ‘say’, ‘know’, ‘long’, ‘road’). While it may, of course, be possible that some of the judgements presented here are erroneous due to insufficient data, any particular errors would be likely to outbalance each other (i.e. undetected borrowings would be compensated for by falsely assumed borrowings), meaning that the phylogenetic results received from a dataset in which these 9 forms are marked as borrowings will probably be more reliable on the whole than results from a dataset in which they are marked as inherited. Nevertheless, for the sake of a more accurate experiment we shall apply the standard lexicostatistical procedure to both sets, and compare the results.

Tuu phylogeny based on the classic lexicostatistical method

The first step is to construct a standard lexicostatistical matrix for all compared languages. In this, we rely on the well-tested Swadesh / Starostin method (see S. Starostin 2000), which requires preliminary elimination of all borrowed items from the dataset in order to produce a “normalized” matrix and avoid potential errors in the resulting tree structure (as well as glottochronological dates). This correction is particularly essential for situations of “mass borrowings”, which can drastically speed up the rate of lexical replacement; within Tuu (if we discount obvious minor impediments such as the presence of Afrikaans lexemes in Modern Nǀuu, etc.), such a situation is only found between Nǁng and !'Auni. However, as has already been mentioned above, for the sake of additional transparency we shall first construct the matrices according to the “maximalist” principle, i.e. pretending that (perhaps) all of the attested matches between these two languages are due to inheritance, not contact. The resulting matrix (Table 2) is as follows:

Table 2. “Maximalist” lexicostatistical matrix between Tuu languages

	Ng!ke	‡Khomani	N uu	Xegwi	!’Auni	Haasi	!Xóǒ	Kakia	N u en
Xam	74 %	79%	73%	59%	53%	43%	42%	40%	43%
Ng!ke		85%	85%	59%	59%	43%	43%	41%	43%
‡Khomani			95%	74%	69%	53%	57%	56%	60%
N uu				64%	59%	48%	47%	45%	48%
Xegwi					50%	40%	41%	41%	46%
!’Auni						72%	47%	48%	49%
Haasi							43%	41%	45%
Xóǒ								85%	77%
Kakia									78%

This matrix, in accordance with the Starling-NJ method¹, yields the phylogenetic structure of Figure 1 (the latter is also accompanied with glottochronological dates, calculated according to the Swadesh / Starostin method).

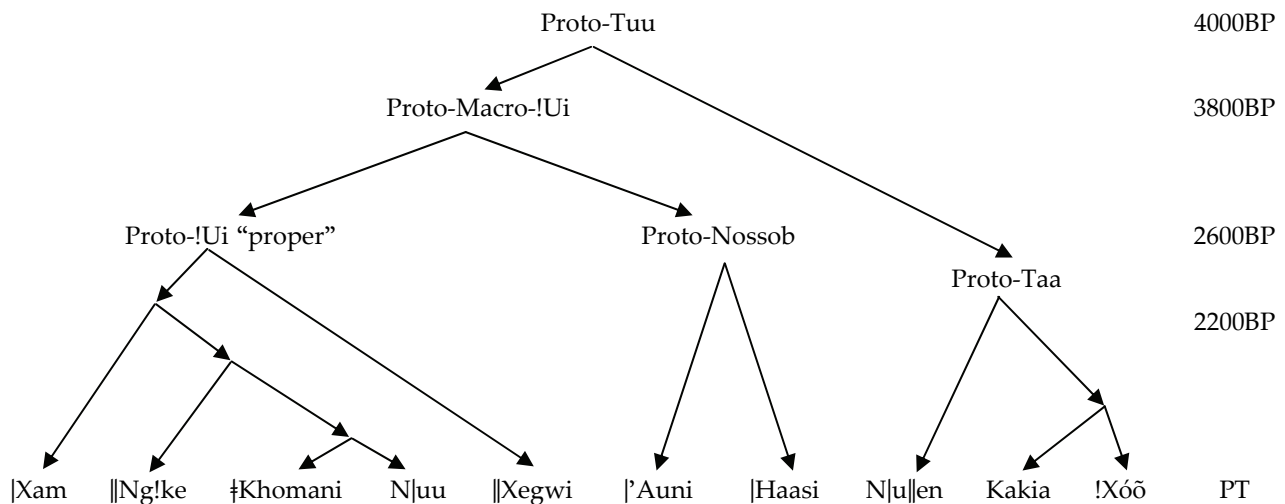


Figure 1. “Maximalist” tree for Tuu languages (= all !’Auni-N||ng matches are counted as inherited)

Before making any comments on the matrix or on the accompanying tree, let us now present the second, “minimalist” scheme, in which the abovementioned 9 highly likely or potential borrowings from N||ng into !’Auni are removed from the calculations altogether, reducing the total number of !’Auni lexemes counted as matches or mis-matches to 68 (Table 3).

This matrix, in accordance with the Starling-NJ method, yields the phylogenetic structure of Figure 2.

It is immediately noticeable that the biggest – and, in fact, the only – change in the phylogenetic structure concerns the position of Nossob languages: on this scheme, they are actually seen as the first branch to split off the common Tuu stem, rapidly followed by Taa, whereas Figure 1 reverses the process, putting Taa as the first branch to diverge, almost immediately followed by Proto-Nossob.

¹ The Starling-NJ method is a simple clustering method for producing phylogenetic trees, introduced by Sergei Starostin and commonly employed in the Moscow school of comparative linguistics; see Kassian 2015 for a detailed description.

Table 3. “Minimalist” lexicostatistical matrix between Tuu languages
 (= 9 !’Auni-N||ng matches are considered to be borrowings and removed from calculations)

	Ng!ke	‡Khomani	N uu	Xegwi	!’Auni	Haasi	!Xóō	Kakia	N u en
Xam	74 %	79%	73%	59%	52%	43%	42%	40%	43%
Ng!ke		85%	85%	59%	55%	43%	43%	41%	43%
‡Khomani			95%	74%	64%	53%	57%	56%	60%
N uu				64%	53%	48%	47%	45%	48%
Xegwi					46%	40%	41%	41%	46%
!’Auni						74%	47%	47%	48%
Haasi							43%	41%	45%
Xóō								85%	77%
Kakia									78%

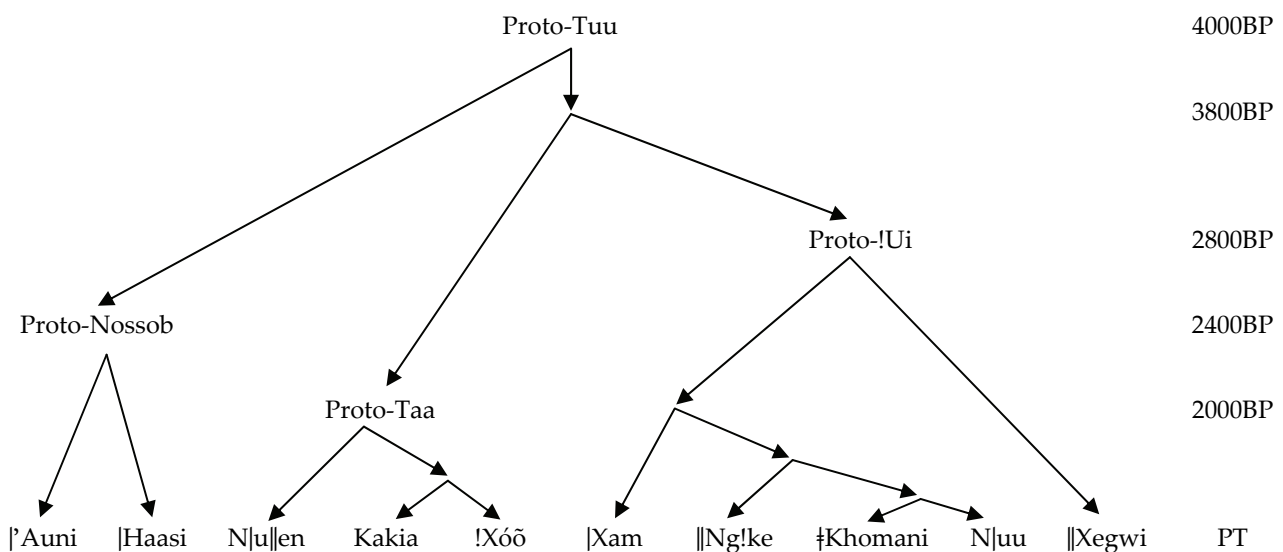


Figure 2. “Minimalist” tree for Tuu languages (= 9 !’Auni-N||ng matches are removed from calculations).

Amusingly, neither of the two schemes agrees with Güldemann’s classification, which would have Proto-!Ui as the first outlier. However, in all fairness, the glottochronological distance between the first two splits of Proto-Tuu on both schemes is so minuscule (approximately 200 years) that it lies well within the borders of statistical error. In such cases, the logical compromise is to postulate a trifurcation, reserving any definitive judgement on the exact chronological order in which it might have taken place, i.e. agree with the primary classification of Tuu as consisting of three more or less equidistant branches (Figure 3).

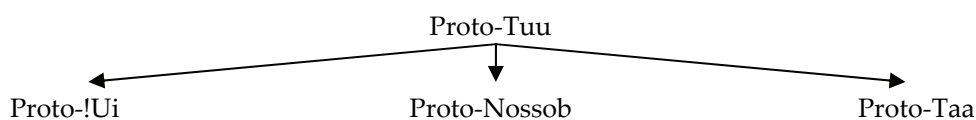


Figure 3. The likeliest evolution scenario for Proto-Tuu.

The veracity of such a scheme may be further confirmed or weakened by (a) running alternate formal methods, e.g. UPGMA or Bayesian phylogenetics, or (b) a manual analysis of

the individual isoglosses between the three branches, based on phonetic, morphological, semantic and distributional peculiarities of the compared items. Given the overall scarceness of the material, it seems far more sensible to me to prefer the latter approach. However, before proceeding to that stage, let us make some additional comments on the internal classification of all three branches.

A. *!Ui*. This is the most deeply divergent branch of all three, but largely due to the inclusion of *||Xegwi*, which comes out as an obvious outlier on the *!Ui* tree (this agrees with Güldemann 2014: 258 listing it as an outlier). The divergence between *|Xam* and the entire *N||ng* cluster is not so deep in comparison (approximately 2000 glotto-years according to the Swadesh-Starostin formula). Interestingly, “Modern *N|uu*” emerges as being much closer to the “*‡Khomani*” doculects as recorded by Meingard and Doke rather than the “*||Ng!ke*” as recorded by Bleek — although it has to be kept in mind that the wordlist for *‡Khomani* is the most deficient one in my collection, and for that reason, match percentages between it and the other Tuu languages are inevitably skewed (cf. 59% between *||Xegwi* and *||Ng!ke* vs. 74% between *||Xegwi* and *‡Khomani* — the only explanation for such a blatant discrepancy is the severe incompleteness of the *‡Khomani* list, creating the illusion of extra proximity).

Due to the impossibility of properly raising even half of the Swadesh list for the remaining attested *!Ui* languages (*||Ku||e*, *||Kxau*, *!Gã!ne*, etc.), their place in this classification remains indeterminate; all that can be said with certainty is that they are probably closer to “Narrow *!Ui*” than to *||Xegwi*, since I have not been able to discover even a single exclusive isogloss between any of them and *||Xegwi*.

B. *Nossob*. As previously recognized by Güldemann, there is no reason to speak of *!’Auni* and *|Haasi* as dialects of a single language; despite all the obvious exclusive isoglosses between the two, their degree of divergence translates to almost 2,300-2,400 glotto-years (this is slightly more than for Turkic or Slavic languages, for instance). The position of “*Xatia*”, or *‡Einkusi* (another small dialect recorded by D. Bleek), in between *!’Auni* and *|Haasi* remains inconclusive due to critical scarceness of data.

C. *Taa*. Given the relative proximity of all attested variants of *Taa*, there is not much that standard 100-item wordlist-based lexicostatistics can say about their internal classification (Naumann 2014 offers a much more thorough perspective on the issue, albeit mainly based on grammatical and phonological, rather than lexical, isoglosses); however, given that at least Bleek’s *N|u||en* (a subdialect of Western *!Xóõ*) differs from Traill’s Lone Tree *!Xóõ* (Eastern) by more than 20% of attested entries, it is clear that we are really dealing with at least several distinct *languages*; in practical terms, it means that cross-checking the lexical data of Lone Tree *!Xóõ* against words with the same meanings in other varieties of *!Xóõ*, whenever available, is a must in any historical studies of Tuu lexicology (as a particularly transparent example, cf. the situation with the 1st person pl. pronoun, where *N|u||en* and *Kakia* agree with *!Ui* and *Nossob* languages, while the situation in *!Xóõ* is innovative).

Notes on the reconstructibility of the Proto-Tuu wordlist

The second most important task of the present study, in addition to constructing a phylogenetic tree for Tuu languages, was to assess the possibility of reconstructing a Swadesh wordlist for Proto-Tuu — a task equally important for the internal historical study of Tuu languages and for solving the problem of their external connections. To somewhat formalize the procedure, as well as reduce the risk of errors resulting from “homoplastic” developments (the same protolanguage item independently shifting its semantic properties to the same Swadesh

meaning in two or more distantly related languages) and make more transparent the processes of lexical replacement and semantic shift in the various Taa subgroups, we have advocated for a strict step-by-step approach, reconstructing first (whenever possible) the respective Proto-!Ui, Proto-Nossob, and Proto-Taa entries, then determining which of these can be genetically related on the basis of established phonetic correspondences (or, at the very least, regular correspondence patterns whose phonological status on the Proto-Tuu level still remains unclear, but whose recurrence can be reliably demonstrated).

Details of the reconstruction process have already been laid out in the first part of the present paper; here, I shall simply restate that a certain item is deemed to be *reconstructible* for Proto-Tuu if its reflexes are found in at least two out of three primary branches (if we accept the trifurcation scheme as the most likely one, this means that !Ui-Nossob, Taa-Nossob, and !Ui-Taa isoglosses all have comparably high chances of reflecting the situation in Proto-Tuu). Also, it is important to state that the intermediate reconstructions can technically be *pseudo-reconstructions* (i.e. a word whose regular reflexes are only found in one !Ui language and in one Taa language can perfectly well reflect the corresponding Swadesh item in Proto-Taa), but only if they do not come into conflict with actual reconstructions (i.e. if there is a “better” candidate for the Swadesh meaning, in terms of its distribution across daughter languages, it is given preference over such “minor” isoglosses). However, to the best of my knowledge, such conflicting situations are extremely rare in Taa languages.

Based on various degrees of their reconstructibility, the entries on Swadesh’s 100-item wordlist can be divided into the following groups.

1. *Not reconstructible even on (most or all of) the intermediate levels*: ‘bark’, ‘belly’, ‘burn’, ‘cloud’, ‘feather’, ‘fish’, ‘good’, ‘green’, ‘leaf’, ‘root’, ‘round’, ‘seed’, ‘swim’, ‘warm’, ‘white’, ‘yellow’.

Some of these words have to be excluded simply based on the lack of corresponding realities (e.g. ‘fish’, ‘swim’) or on their highly specialized representation in the South African environment (e.g. most of the flora-related terms, such as ‘leaf’, ‘root’, ‘seed’). Others simply turn out to be highly unstable, including several adjectives (‘good’, ‘round’, ‘warm’) and, interestingly enough, the majority of color terms included in the Swadesh wordlist — thus, for Proto-!Ui only ‘black’ and ‘red’ can be reconstructed with a certain degree of reliability.

2. *Reconstructible on some or even all of the intermediate levels, but not reconstructible for Proto-Tuu due to lack of cognacy between the individual branches*: ‘black’, ‘cold’, ‘earth’, ‘egg’, ‘foot’, ‘full’, ‘give’, ‘kill’, ‘knee’, ‘man’, ‘mountain’, ‘new’, ‘person’, ‘rain’, ‘red’, ‘sand’, ‘say’, ‘small’, ‘stand’, ‘star’, ‘stone’, ‘that’, ‘two’, ‘who’. It is not always clear which of these words are simply lexicostatistical mismatches (i.e. cognates exist in other branches but underwent semantic shift) and which ones are etymological mismatches (i.e. cognates simply do not exist) — the solution of this issue has to be postponed until the preparation of a full-fledged etymological corpus for Tuu languages.

3. *Reconstructible (or at least “pseudo-reconstructible”, i.e. present in at least one language) for all three subgroups and cognate with each other*: ‘bird’, ‘bite’, ‘claw /nail/’, ‘come’, ‘die’, ‘dog’, ‘drink’, ‘ear’, ‘eat’, ‘fat’, ‘fire’, (?) ‘fly’, ‘hair’, ‘hand’, ‘heart’, ‘horn’, ‘I’, ‘lie’, ‘meat’, ‘name’, ‘nose’, ‘one’, ‘see’, ‘sit’, ‘sleep’, ‘thou’, ‘tongue’, ‘tooth’, ‘tree’, ‘water’, ‘we’, ‘woman’. (‘Fly’ is debatable, see point 7a above where it is argued that the word may be an areal inter-family isogloss with a sound-symbolic stamp on it).

This is the single largest group of them all, providing the strongest evidence for a genetic connection between all three subbranches. It is pleasing to note that no fewer than 23 out of 32 of its constituents belong to the “ultra-stable” 50-item wordlist, suggested in Starostin 2010 and since then used as the basis for deep-level linguistic comparison in my own studies on African languages.

4. *Reconstructible based on isoglosses between !Ui and Taa*: ‘all’, ‘ashes’, (?) ‘big’ (problematic because of phonetic difficulties), ‘breast’, ‘dry’, ‘head’, ‘not’, ‘skin’, ‘sun’; add ‘bone’, ‘liver’, ‘louse’ — three items not attested in Nossob languages at all (the other nine items have etymologically different equivalents in the Nossob branch).

5. *Reconstructible based on isoglosses between !Ui and Nossob*: ‘blood’, ‘eye’, ‘night’, ‘smoke’, ‘this’, ‘walk /go/’; add ‘hear’, ‘long’, ‘neck’, ‘road’ if the Nossob equivalents are inherited rather than borrowed (if we decide to count them as borrowings after all, all four have to be moved to group 2).

6. *Reconstructible based on isoglosses between Taa and Nossob*: ‘know’, ‘many’, ‘moon’, ‘mouth’, ‘tail’, ‘what’. Amusingly, if we discard the four potential borrowings from group 5, the number of exclusive !Ui-Nossob isoglosses on the Swadesh list is precisely the same as the number of Taa-Nossob isoglosses, further confirming the trifurcate division.

Although we are still a long way away from producing reliable Proto-Tuu reconstructions based on well-verified phonetic correspondences for all the constituting segments, certain useful observations on their phonetic features can be made even now. Thus, out of the 54 items reconstructible for Proto-Tuu 38 begin with click consonants (4 with labial *θ-, 14 with dental */-, 3 with alveolar */-, 3 with palatal *ʃ-, 12 with lateral */-, and 2 with the enigmatic “sixth click”) and 16 begin with egressive consonants (6 with alveolar sibilants / affricates, 3 with coronal stops, 2 with velar stops or affricates, 1 with nasals, 4 with initial vowels or glottal stop).

While the statistical results for non-click consonants are relatively predictable and in general agreement with the click-to-non-click rates in such languages as N|uu and !Xóõ, the distribution of different click influxes presents a big surprise, with words beginning with */- and */- being 3 to 4 times more frequent than those beginning with */- or *ʃ-. Even if we ultimately conclude that evidence for the “sixth click” is unconvincing and decide to merge it with either one of those two, the frequency rates will change only slightly. This is markedly different from, e.g., the situation in !Xóõ, where, according to Traill’s dictionary, approximately 330 lexical roots have the dental click /, approximately 430 have the alveolar click !, approximately 450 have the lateral click //, and approximately 350 have the palatal click ʃ (the numbers may be slightly different depending on which items are considered to have the same lexical root and which ones are not).

It is instructive to compare these statistics with the distribution of clicks across 42 common Tuu stems containing clicks which were tentatively reconstructed by T. Güldemann (2005: 24–28) for a general study showing the overall genetic relationship of the Tuu languages. His sample, also rooted primarily in basic lexicon but not restricting itself to Swadesh items, was an attempt to put together the most stable and widely distributed etyma, which would make any resulting statistics run on them worthy of attention. They are as follows: 3 items with labial *θ, 20 items with dental */, 6 items with alveolar */, 4 items with palatal *ʃ, 9 items with lateral */. Again, we see a huge discrepancy when it comes to the dental click, although the figure for the lateral click is slightly lower than in my case (however, I would definitely change Güldemann’s *!ab- ‘leopard’ to a reconstruction with */-, which would bring the distribution slightly closer to the one in my sample).

From a purely theoretical standpoint, this discrepancy, if its purely accidental origins are to be ruled out, can only be explained by one of two possible reasons: (a) undetected processes of click loss or click articulation shift in !Ui and Nossob languages, during which many items with alveolar and palatal clicks underwent significant phonetic change; (b) conversely, mass borrowing of lexical items from other languages into Taa (from Khoe or maybe even other, now extinct, “Khoisan” lineages) which led to an increase of items with formerly rare click influxes. Unfortunately, concrete evidence for both of these speculative hypotheses is so far lack-

ing (there are, of course, plenty of Khoe borrowings in Taa, but nowhere near enough to breach such an enormous gap). Without a doubt, some light on this problem shall be shed in the process of putting together a robust etymological corpus for Tuu, but for now this puzzle remains unresolved.

An additional look at the Nossob language issue: shared features and innovations

It would probably make sense to suggest that where pure statistics fails to properly resolve an issue (such as produce a robust binary-split tree rather than have us accept a trifurcation model), detailed manual analysis of the underlying data may potentially reveal important classificatory arguments that are inevitably lost in the cogs of crude automated algorithms. For instance, it may be important to not merely look at the *quantitative* aspects of !Ui-Nossob-Taa comparison, but also take into account the *quality* of detected matches — it is for this reason that I attempt, whenever possible, in the Appendix sections of both these papers to indicate not just the fact of cognacy between different items, but also the *degree* of cognacy, recognizing that, for instance, a certain Proto-Nossob item may be related to both Proto-!Ui and Proto-Taa, but be phonetically or morphologically closer to one of these branches rather than the other. Likewise, it would make sense to try and assess all the detected binary isoglosses (!Ui-Nossob vs. Taa, Taa-Nossob vs. !Ui) in terms of shared innovations vs. shared archaisms — this is extremely difficult in light of the overall poor state of Tuu etymology, but even a small handful of transparently resolvable cases might be quite helpful.

Below I list, in alphabetical order, all the Swadesh items which satisfy the following conditions: (a) they are reconstructible for Proto-Nossob or are at least found in one of two Nossob languages; (b) they are not suspected, on the whole, of being borrowed into !'Auni from N||ng; (c) they have reliable cognates in either !Ui, or Taa, or both of these branches; (d) they *either* have phonological or morphological features that bring them closer to !Ui or Taa, *or* they can be regarded as shared lexical innovations (rather than retentions) with one of these two branches.

1) BIRD: Proto-Nossob $*si=|u$ has the same desemanticized prefix $*si-$ that is frequently found in numerous dialects of Taa. Additionally, the word has a voiced efflux in common with Proto-Taa $*|u^{(i)}$, rather than the uvular efflux in Proto-!Ui $*|q^hu-$.

2) BITE: Proto-!Ui $*c'i$ and Proto-Nossob $*c'i$ vs. Proto-Taa $*si?(i)$. This could be qualified as a shared phonetic innovation, since it is easier to explain $*si?(i) \rightarrow *s'i \rightarrow *c'i$ than the opposite development in Taa. However, it is not very diagnostic since we also know cases where glottalic articulation of the affricate in !Haasi, as marked by R. Story, is most likely secondary (e.g. $c'i$ 'to come', which has no cognates with glottalized affricates in any other Tuu languages); therefore, Proto-Nossob $*c'i$ could simply reflect $*si-$ or ($si?$), and its extra similarity to Proto-!Ui could be an illusion.

3) DOG: both Proto-!Ui and Proto-Nossob feature a nasal suffix at the end of this stem ($*\dagger^hu-ni \sim *\dagger^hu-i\eta$ vs. $*\dagger^h\text{ }-i\eta$) respectively as opposed to Proto-Taa $*\dagger q^ha-i$ without any traces of nasality. Given the existence of such !Ui variants as ||Xegwi $\text{ } \leftarrow *\dagger^h/u-ai$ (?), it is quite possible that different morphological variants were already present in Proto-Tuu, and it is hard to determine the relative degrees of archaism and innovation in this case.

4) EYE: even if we discard !'Auni $ca-xu$ as a recent re-borrowing from N||ng, !Haasi $cx\text{ } \text{'eye'}$ still transparently shows that Nossob languages have a strong isogloss with !Ui $*c'a-xu$. In the first part of the paper (Starostin 2021: 123) it was already shown how this situation is better explained as a shared innovation for both groups rather than retention of a common archaism.

5) FAT: Nossob $*so-a$ or $*so-\tilde{a}$ is morphologically closer to Taa $*s\tilde{a}f$ than to !Ui $*so-ni \sim *so-i\eta$.

6) FIRE: !Ui and Nossob have */i vs. Taa */ā. It has been previously argued that *i is more likely to reflect the original root vocalism, with Taa */ā ← */i-ā, thus, the similarity between !Ui and Nossob is probably a shared retention rather than innovation.

7) LIE: !Auni tōa is closer to Proto-Taa *tu than than Proto-!Ui *ta (although it is likely that the base root underlying all stem variants is ultimately the same).

8) NAME: this stem's codas in !Ui (*|e ~ *|ē) and Nossob (*|eN) are notably closer to each other than to Taa */ā-ū), but it is not clear which of the two full stems is more archaic.

9) ONE: reflexes of this word, featuring the “sixth click”, are notably closer between Nossob (*|ʰu-ŋ) and Taa (*|ʰu-) than between either of them and !Ui (*|ʰoaʔ-/i). It must be noted that the very fact of the “sixth click” being consistently reflected in Nossob and Taa as palatal |ʰ (opposed to alveolar and lateral reflexes in !Ui languages) is in itself an important phonetic isogloss, more closely binding Nossob and Taa to each other.

10) TOOTH: phonetically and morphologically, !Ui */hā is closer to Nossob */|e-[iŋ] than Taa */|qʰā ~ */|qʰan (no signs of front vocalism in Taa codas).

11) WOMAN: paradigmatically, this stem behaves more similarly in !Ui and Nossob languages than in Taa, although this is likely to be a shared archaism (since Taa sg. *|ā-qāe ‘woman’, lit. ‘person-mother’, is a transparent lexical innovation).

The list turns out to be a little more skewed in the direction of !Ui-Nossob proximity (7 cases vs. 4 cases of closer similarities between Taa and Nossob), but only in two of these cases is it possible at all to make a well-argued (though still far from waterproof) judgement about the direction of innovation, and they are, ironically, equally divided between !Ui-Nossob (‘eye’) and Taa-Nossob (‘one’). The other lexical isoglosses between !Ui and Nossob, on one hand, and Taa and Nossob, on the other (listed under p. 5 and 6 of the previous section), also do not allow to make serious judgements about the directions of lexical replacement. In the end, it turns out that even such fine-grained manual analysis of the evidence does not allow us to properly depart from the trifurcate model of Tuu phylogeny.

One last thing that remains is to critically evaluate the specific arguments given by T. Güldemann in favor of a closer connection between Taa and Nossob and see if they are sufficient to go back to the binary-split classification and admit that the lexicostatistical data simply fail to properly reflect the true historical process. The main argumentation, laid down in Güldemann 2014, consists of two points.

1. Güldemann notes a closer affinity between the overall numeric systems of !Xóõ (|ʰu- ‘one’, |ʰV- ‘two’, |āe ‘three’, |āli ‘many’) and !Auni (|ʰū- ‘one’, |ā ~ |āa ‘two’, |ai- ‘three’, |āni ‘many’). There are, however, multiple problems with this argument. First, |Haasi — !Auni’s closest relative — seems to have only one common element with !Auni in this paradigm (|ʰ- ‘one’). Second, it remains unclear if Bleek’s transcriptional variation |ā ~ |āa can really be interpreted as misheard variants of |ā (as suggested by Güldemann in order to justify the comparison with !Xóõ |ʰV-); note that, as a rule, Bleek transcribes the palatal click in !Auni correctly (see DOG, MOUTH, ONE for examples).

Third, the very “numerals” |ā ~ |āa and |ai- are actually defined by Bleek not as numerals, but as special “particles” that precede the actual numeric stems, as in |ŋ |ā tis |am ‘two huts’, |ŋ |ais |wōna-a ‘three huts’, where the identifiable numerals are |am ‘two’, |wōna ‘three’ (both of them transparent borrowings from Khoe). Güldemann suggests that these morphemes may be behaving like grammaticalized markers (e.g. for dual or plural), to which language speakers are then adding actual numerals borrowed from a different language, but this is hard to prove based on the very limited number of examples in Bleek’s texts.

Furthermore, even if Güldemann is correct in all of his hypotheses here, and the available evidence truly allows us to reconstruct a four-element numeric paradigm for the common an-

cestor of Taa and Nossob, it would still be impossible to present it as a shared innovation between these two groups, rather than an archaic retention from Proto-Tuu; to do this, one would need to demonstrate, at the very least, that Proto-!Ui **!'u* ‘two’ is more archaic in this shape and meaning than Nossob-Taa **fV-* id. (‘three’ is also borrowed from Khoe across all !Ui languages, and ‘many’ is not properly reconstructible). Therefore, the numeric argument is at worst etymologically dubious, and at best inconclusive.

2. The second argument ultimately has more to do with typology than reconstruction: analysis of the small preserved syntactic corpus for |Haasi, as well as a few textual examples from D. Bleek’s records of |’Auni, allows Güldemann to conclude that Nossob languages had an active system of grammatical agreement and lexical gender closely resembling the one described for Taa, but not attested in the !Ui language complex (or, for that matter, anywhere else in the Khoisan-speaking region). Moreover, some, if not all, of the identifiable concord markers (allowing to roughly distribute all the nominal stems of |Haasi into three different classes marked as I/E, A, and U respectively) have direct counterparts in !Xóõ.

While this may be a very strong argument in favor of reconstructing such a system for Proto-Tuu (and assuming, for the moment, that it was lost in Proto-!Ui), it is not necessarily indicative of a specific bond between Taa and Nossob languages as such. One problem is that the actual distribution of nouns across classes in |Haasi rarely matches the same distribution in !Xóõ. Furthermore, the system of markers matches Taa only partially (it is unclear, for instance, if |Haasi *A* can be equated with Taa *Ã*, for which nasalization is essential). Most importantly, perhaps, in !Xóõ class-defining morphemes are frequently seen in conjunction with the roots (e.g. Class 3 nouns often end by themselves in *-e*, which also functions as the agreement marker, etc.), and while comparative data show clearly that some of the codas in |Haasi nouns are old fossilized suffixes, they are rarely the same as in !Xóõ.

Thus, even if a word like ‘dog’ belongs to the same I-Class in |Haasi and !Xóõ, more important is the discrepancy in the actual shape of the stem ‘dog’, which is *ʔq^hã-i* in !Xóõ, but *ʔ^haŋ* in |Haasi (closer in its morphological constitution to the respective forms in !Ui, actually). As has been shown above, cases where Nossob languages show “Taa-like” morphology are statistically more or less the same as cases in which they show “!Ui-like” morphology, which once again brings us back to the trifurcate model.

Conclusions

Perhaps the most important conclusion from this study should be the realization that it is still too early to draw any definitive conclusions on the phonology and lexical constitution of Proto-Tuu, as well as the internal classification of its members — many of the judgements, assumptions, and reconstructions in this paper will yet be subject to further amendment as more data from old sources (such as the slowly emerging notebooks of D. Bleek and other researchers) and new sources (such as contemporary research on surviving Taa dialects) come to light.

Nevertheless, even the limited amount of data taken into consideration here allows to draw up a reasonably realistic picture of the evolution from Proto-Tuu to its descendants. The most important details of that picture are as follows:

(a) Proto-Tuu is, indeed, the deepest and oldest of all known “Khoisan” lineages (with the possible exception of Ju-#Hoan, a.k.a “Kx’a”), glottochronologically dated to approximately 4000BP; this explains many of the difficulties with reconstructing a Proto-Tuu Swadesh list due to cognate loss;

(b) the disintegration of Proto-Tuu must have happened according to a trifurcate pattern, with speakers of Proto-!Ui, Proto-Nossob, and Proto-Taa parting ways at more or less the same time. Even if there was a brief period of “!Ui-Nossob” or “Taa-Nossob” unity, it did not last long enough to produce a statistically significant number of shared innovations;

(c) Proto-Tuu must have had a relatively complex, but not particularly stable, system of nominal morphology, as is seen by the numerous stem variants for the same lexical roots witnessed across (and sometimes even within) the individual branches. Reconstruction of this system is severely hampered by lack of data and inadequate grammatical descriptions for extinct languages;

(d) most curiously, distribution of click phonemes across basic lexicon morphemes reconstructible for Proto-Tuu significantly differs from usually attested distributions of click phonemes across attested languages, such as !Xóǒ. This indirectly hints either at huge, hitherto not comprehensively described or understood areal influence on Tuu languages after the family’s disintegration, or at our inadequate understanding of the historical development of Tuu phonology — or, possibly, both.

It is not likely that anything other than a comprehensive, systemic investigation of the etymological connections between the various Tuu languages, hopefully resulting in a proper comparative dictionary following the classic Neogrammarian principles, will allow us to move beyond these observations, which can only, for now, be used as general guidelines for further research. Whether such an endeavour will be at all possible in the near future, remains to be seen.

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 , z = IPA ts , dz) and palatal (ζ , ζ = IPA $t\zeta$, $d\zeta$) 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. $\text{!Auni } si=|u$ ‘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 51–100).

In this Appendix, I list the results of intermediate and Proto-Tuu reconstructions for the (alphabetically) second half of the Swadesh wordlist (more or less closely following the semantic specifications set out in Kassian et al. 2010). Structure of the entries follows the same guidelines as the first part, reprinted here:

(1) Name of the item, together with a formal notation of the presence / absence of lexicostatistical parallels between the three branches: e.g. [$\text{!Ui} + \text{Taa}$] [$- \text{Nossob}$] 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 reconstructions. 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.).

All data sources remain exactly the same as in the first part of the paper — Bleek & Lloyd 1911, Bleek 1929, Bleek 1956 for |Xam; Bleek 1956, 2000 for ||Ng!ke; Doke 1936, Maingard 1937 for the †Khomani” doculects of N||ng (†Kho-D, †Kho-M respectively); Miller et al. 2009, Collins & Namaseb 2011 and Sands p.c. for Modern N|uu; Ziervogel 1955 and Lanham & Hallowes 1956a, 1956b for ||Xegwi (||Xegwi-Z and ||Xegwi-LH respectively); Bleek 1937, 1956 for |’Auni; Story 1999 for |Haasi; Traill 1994 for Lone Tree !Xóõ; Bleek 1929, 1956 for “Kakia” Taa and N|u||en.

51. MAN [-]

- !Ui: *ʔo ~ *ʔo (||Ng!ke ʔo: ‘male’, †Kho-M, N|uu ʔo, ||Xegwi-Z ʔo, ||Xegwi-LH kwi-ʔo). \diamond Bleek’s ||Ng!ke data as well as ||Xegwi samples collected by Lanham and Hallowes indicate that the primary meaning of this morpheme must have been ‘male’, since the meaning ‘man’ is actually expressed by a compound whose first part (*kwi*) means simply PERSON q.v., and it is also encountered in other compounds denoting male animals, e.g. *širi-ʔo*: ‘male buck’. More problematic is the observed variation in click efflux articulation: the entire N||ng cluster as well as Ziervogel’s ||Xegwi shows *ʔo (zero efflux), whereas Lanham and Hallowes record an ejective lateral affricate which reflects original *ʔo (less indicative is Bleek’s ||Ku||e form t’o ‘man’, since ||Ku||e t’- is also substituted for simple *ʔ- in at least one other case, see MOON below). The “majority rule” suggests regarding *ʔo as the original form, although it is not excluded that the variation actually reflects a separate efflux².

In |Xam, the most common equivalent for ‘man’ is the compound form !wi=gwai, where !wi = PERSON q.v. and gwai means ‘male’. It is unlikely that gwai is related to *ʔo,

² It is interesting to note the existence of phonetically similar roots with the meaning ‘male’ in Proto-Ju (*ʔo, with a voiced retroflex click) and in Khoe (*ʔo ‘male’ → Nama ||gō-b ‘bull; tough person’, Naro ||ô ‘bull, male’): whether they may be used as evidence for more distant relationship with Tuu remains to be seen, but it should be noted that neither of the two shows a glottalized click efflux.

since random click loss is not typical of |Xam, but no other etymology can be suggested at the moment.

Interestingly enough, the plural form ‘men’ is suppletive in most !Ui languages, being formed from the common !Ui root **tu*: |Xam *tu-kən*, ||Ng!ke *tu-kən ~ tu-ŋən*, N|uu *ɕu:-ke*. The only doculect in which *tu* is occasionally found in singular use is Bleek’s ||Ng!ke, cf. *ŋ tu e //ŋ a* “that man is at the hut” (Bleek 1956: 240), but since no other recorded dialect of N||ng reflects such usage, and given that the strong association between **tu* and plurality persists into Nossob and Taa languages as well (see notes below and on PERSON), one should rather suggest either a misglossing on Bleek’s part or a special back-formation in one or more old dialects of N||ng.

- Nossob: (?) **be* (’Auni *be ~ bε*, |Haasi *bi*). ◊ In addition to this form, common for both ’Auni and |Haasi, Bleek also lists *da ~ de* ‘man, person’ for ’Auni, without any clear semantic distinctions (the default word for PERSON is more likely to have been *ʃi ~ ʃe*, see below). Archaic origin for Common Nossob **be* is dubious for phonetic reasons, since labial *b-* is not well reconstructible as an inherited phoneme for Proto-!Ui; it is quite possible that **dV* is the original word for ‘man’, still retained in some functions in ’Auni but already replaced in the most basic usage on the Common Nossob language by an innovation. Additionally, the suppletive plural form ‘men’ in ’Auni is either *tu-ke* or *tu-tu-s(i)*³; the first variant may be borrowed from N||ng, but the second is more likely to be inherited.
- Taa: **λa=aʃ* (!Xóõ *tâ:=à:ʃ*, Kakia *laʃ ~ la:ʃ ~ la*). ◊ For !Xóõ, cf. also Westphal’s data: †Hūa *laʃa*, N|amani *táʃaʃ*, pl. *táʃaʃ-tu* (Westphal 1965: 139). The word is clearly a compound formation, where **λa* = ‘person’ (see notes on PERSON for issues of phonetic reconstruction) and **aʃ* = ‘father’. It is possible that a more archaic root for this meaning is preserved in the suppletive plural, cf. !Xóõ *//xâ:* ‘men’ (†Hūa *íj//taá* in Westphal’s transcription), or in a separate root represented by N|u|len **!ã* ‘man’, pl. *!ã-te* (etymological connection between this form and !Xóõ *//xâ:* is hardly possible). But the compound form is the only isogloss between two distant nodes of the entire branch.
- Tuu-: Formally not reconstructible, but possibly **tu*? ◊ There is no clear evidence for a separate lexical root with the specific meaning of ‘male human being’ on the Proto-Tuu level; most frequently, we see this meaning expressed by various compound formations which do not match each other across different lineages (e.g. **!ui-fo* in !Ui vs. **λa-aʃ* in Taa). That said, given the fact that (a) the root **tu* is reconstructible at least on the Proto-|Xam-N||ng level specifically in the plural meaning ‘men’ and (b) etymologically the same root **tu* is reconstructible for Proto-Taa in the plural meaning ‘people’ (see PERSON), it is a distinct possibility that **tu* may have denoted specifically the ‘male human being’ in Proto-Tuu (either in the plural only or irrespective of number), later becoming generalized to ‘people’ in Proto-Taa. However, the opposite scenario (**tu* as originally ‘people’ is also possible); see further notes on PERSON and WHO.

52. MANY [Nossob + Taa] [- !Ui]

- !Ui: Not reconstructible. ◊ The concept is extremely unstable in !Ui: most languages have their own equivalent, e.g. (a) |Xam */kxʷai:-yal*, (b) ||Ng!ke *ĩai ~ ĩãĩ ~ ĩẽ*, (c) †Kho-M *ɕebe-ɕe* = N|uu *kebe-ke*, (d) ||Xegwi-Z *kʰyũ ~ gyeĩŋ* = ||Xegwi-LH *qʰiŋ* = ||Xegwi-B *//xain*.

³ See Güldemann 2002: 189 on the detailed morphological analysis of this form, which he interprets as a combination of the root *tu* with an agreement marker and a copula.

- Nossob: (a) !'Auni *||áni ~ ||áři*; (b) |Haasi *!ɔ:ɔ-k'a*. ◇ Formally not reconstructible for Proto-Nossob, but at least the !'Auni form has a perfect etymological parallel in Taa. The |Haasi form (also transcribed as *!o:ɔ-k'a* in one example) might, perhaps, have something to do with N|uu *!xo*: 'big' (although this would imply incorrect transcription of the efflux by R. Story).
- Taa: *||a-ri (*!Xóõ ||áli*, *Kakia ||ari ~ ||ári*, N|u||en *||an-te*). ◇ The root seems to be common for all varieties of Taa; Bleek's N|u||en form shows that the stem is probably segmentable (with morphological variants *||a-ri and *||a-n).
- Tuu-: Not properly reconstructible. ◇ The apparent isogloss between Taa *||a-ri and !'Auni *||ani* may very well be etymological, and count as an argument in favor of specific proximity between these two branches; however, lack of any parallels in !Ui does not allow to formally postulate it for the Proto-Tuu level.

53. MEAT [!Ui + Nossob + Taa]

- !Ui: ***θoa-** (*||Ng!ke θwai ~ θwai*, †Kho-M *θwoe*, N|uu *θoe*, *||Xegwi-Z*, *||Xegwi-LH θa:*). ◇ A stable lexical item in !Ui, except for |Xam, where the word was seemingly replaced by *ã*: *~ eη*, a secondary nominalization of the verb 'to eat' (*ã*); the same replacement probably took place in *||Kxau ('a:η)* and *||Ku||e (ða-si)*. Semantic derivation of 'meat' from 'to eat' is not uncommon in the Khoisan area (cf. Kalahari Khoe **k'o-xu* 'meat' ← **k'o* 'to eat /meat, hard food/'), and external data clearly show that **θoa-* is the original lexical item.
- Nossob: ***θoe** (!'Auni *θwe ~ θwi*, |Haasi *θwi*).
- Taa: ***θa-** (*!Xóõ θàye*, *Kakia θwe*, N|u||en *θwe: ~ θwi*).
- Tuu+: ***θ/o/a-**. ◇ An obvious isogloss between all three branches (vocalic reconstruction is uncertain, given the tendency of the labial click to labialize the following vowel in !Ui).

54. MOON [Nossob + Taa] [-!Ui]

- !Ui: ***†oro** (*|Xam !au!arro ~ !au!auru*, *||Ng!ke !orre ~ !urru ~ turro*, †Kho-D *†ɔ̀rɔ̀*, N|uu *†oro*, *||Xegwi-LH λolo*, *||Xegwi-B klolo*). ◇ Diphthongization in |Xam is likely to be secondary (cf. the same situation with NECK, NOT below); also of interest is the unique partial reduplication in this language (unless the first syllable is actually a different word, and the whole form is a compound formation rather than reduplication). Doke records a glottalic efflux for †Khomani, but it is not supported by data from other dialects of the N||ng cluster.
- Nossob: ***!hõĩ** (!'Auni *!õĩ*, |Haasi *!hwi*).
- Taa: ***!q^han** (*!Xóõ !q^hàn*, *Kakia !xan*, N|u||en *!xan*).
- Tuu-: Not properly reconstructible. ◇ !Ui and Taa etyma clearly have different origins and no clear mutual parallels. However, the click onset of the Nossob word for MOON regularly corresponds to Taa (cf. DOG with a very similar correspondence), and coda differences may be explained by morphological variation (e.g. **!q^ho-i* vs. **!q^ho-an* → **!q^han*). The case clearly counts as an etymological and lexicostatistical match.

55. MOUNTAIN [-]

- !Ui: (?) ***!ao**. ◇ In all data sources on |Xam and N||ng, the word for MOUNTAIN (or 'hill') is the same as the word for STONE q.v. Only for *||Xegwi* the situation is different:

||Xegwi-Z *t^han* (only quoted once in the phrase *t^han ʔe čwa* ‘black mountains’), ||Xegwi-LH *||u-|a ~ gu-|a* (a compound formation where the second component is probably HEAD q.v.). Both forms are somewhat problematic and have no etymological parallels in the rest of !Ui.

- Nossob: |’Auni *||’wa*. ◇ In the phrase *|’e |^hui ge ||’wa* “we came down from the hill”. Not attested in |Haasi. Somewhat similar to the form in ||Xegwi-LH, but the click effluxes do not match.
- Taa: *!um (|Xóõ *!ù^hm*, N|u||en *um*). ◇ The only equivalent in Kakia is *ũ:n*, i.e. same word as STONE q.v.; however, |Xóõ and N|u||en clearly have a separate lexical root for ‘mountain / hill’ (another attested meaning in |Xóõ is ‘niche for several sp. of plants, characterised by heavy sand’).
- Tuu-: Not reconstructible.

56. MOUTH [Nossob + Taa] [-!Ui]

- !Ui: *tu (|Xam *tú*, ||Ng!ke *tu*, †Kho-M *tu*, N|uu *ɕu*, ||Xegwi-LH *tu ~ t’u*). ◇ A common and stable !Ui etymon, sometimes with polysemy ‘mouth / hole’ (as in |Xam). Lanham and Hallows note variation between *t-* and *t’-* in ||Xegwi, but glottalic articulation is not confirmed by any of the attested external data.
- Nossob: (a) |’Auni *ʔu*; (b) |Haasi *n=ʔa*. ◇ For |’Auni, Bleek’s later records also show the form *tu ~ t^hu* ‘mouth’, noting explicitly that the form “may be †khomani”. Since the earlier data, collected in 1911 and published in 1929, only gives *ʔu*, Bleek’s note is probably correct. No traces of **tu* ‘mouth’ are seen in |Haasi either; the form *n=ʔa* (where *n=*, as in similar body part terms, is really the pronominal possessive prefix ‘my’) is similar to |’Auni *ʔu*, but the vocalic correspondence is completely irregular and could only be resolved in morphological terms (e.g. **ʔu-a* → *ʔa*), for which no corroborating evidence has been found so far. However, because of external parallels in Taa it may be safely assumed that the |’Auni form is more archaic and may be accepted in the status of a “pseudo-reconstruction”.
- Taa: *†u- (|Xóõ *ʔû-e*, pl. *ʔû-m-sâ*, Kakia *!we*, N|u||en *ʔûê*). ◇ Kakia *!-* regularly reflects Proto-Taa *†- (either as a genuine phonetic development or as a regular mistranscription).
- Tuu-: Formally not reconstructible. ◇ !Ui and Taa roots are not related and have no mutual etymological parallels. However, |’Auni *ʔu* is clearly the same as the Taa form and thus, another important isogloss between the two branches.

57. NAME [|!Ui + Nossob] + Taa]

- !Ui: *|e ~ *|ẽ (|Xam *|ẽ*, ||Ng!ke *|ẽ*, N|uu *ka=|ĩ*, ||Kxau *|e*, ||Xegwi-LH *|e*). ◇ It is unclear if nasalization is an inherent part of the root or a remaining trace of a morphological marker (i.e. *|ẽ ← *|e-/V/N), but at least all forms are clearly related and a common !Ui ancestor is reconstructible. N|uu *ka=* is a prefix of inalienability.
- Nossob: *|eN (|’Auni *|ẽ ~ |ẽn*, |Haasi *a=|aŋa*). ◇ The |Haasi form is also transcribed simply as *a=|ã* in the phrase “what is your name?”; it seems likely that *a=* in the vocabulary form is the 2nd p. pronominal prefix, while the second *-a* is a verbal copula. Vocalic discrepancy between |’Auni *|ẽn* and |Haasi *=|aŋ-* is not easy to resolve in terms of common origin, and makes it worth considering the idea that the |’Auni form may actually have been borrowed from N|uu. On the other hand, there are a couple other cases where |’Auni *e* corresponds to |Haasi *a* (e.g. |’Auni *|’e*: vs. |Haasi *=|a-* ‘heart’), so the contact scenario should not be regarded as the most probable by default.

- Taa: *|ãũ, pl. *|ã (!Xóõ |ãũ, pl. |ã; Kakia |kx'ãũ, N|u|len |ã). ◊ Judging by the paradigm in !Xóõ, the N|u|len form recorded by D. Bleek was actually a plural one. Superficially, the plural form looks simpler than the singular, but its underlying morphophonological shape is actually *|ãũ-a → *|ã with regular contraction (exactly the same situation is observed in the case of NECK and TAIL, see below); -a is a frequent semi-productive plural morpheme in Taa, whereas no singulative marker like -u or -ũ can be postulated with any degree of certainty.
- Tuu+: All listed forms can be judged as cognates due to phonetic identity of the click consonant and basic root structure (*|VN). However, reconstruction of the original quality of the root vowel is problematic; observed discrepancies cannot be explained by phonetic change alone and have to have a morphophonological explanation. The form in !Ui could be analyzed historically as *|ã- (root) + *-i (one of the singulative markers), which would also easily allow to equate it with the form in |Haasi. However, Taa *|ãũ (rather than simply *|ã) should then be symmetrically analyzed as ← *|ã- + *-u, despite fairly little evidence for *-u as a separate class or number marker. Could there have been a regular development *-ãĩ → *-ãũ in Proto-Taa⁴? Unfortunately, there is too little comparative evidence for this sequence to make any strong conclusions. Still, this unresolved discrepancy should not be an obstacle for assuming common ancestry: vocalic correspondences between !Ui and Taa are a priori more complicated than consonantal ones.

58. NECK [!Ui + Nossob (?)] [- Taa]

- !Ui: *|qu (|Xam !au ~ !eau ~ !^hou, |Ng!ke !ú ~ kú, N|uu †qu, |Kxau †'u). ◊ A solid isogloss between |Xam and N|ng (diphthongization in |Xam is probably secondary, as in many similar cases). Of interest is the clickless variant in |Ng!ke, marked by Bleek as an “occasional form”. The |Xegwi equivalent is different: cf. |Xegwi-Z !ele vs. |Xegwi-LH !elen (similar forms, but with incompatible clicks; I suspect a possible mistake in Ziervogel's records).
- Nossob: (a) |'Auni ||u; (b) |'Auni †õĩ. ◊ The word is not attested in |Haasi, whereas for |'Auni Bleek lists two equivalents, with only the former supported by a contextual example (sa ko |kx'esi, ho ha ||ú ||o “bring beads, on my neck put them”).
- Taa: *|kx'ãũ (!Xóõ †kx'ãũ, Kakia !kx'ũm, N|u|len †ũ). ◊ Correspondences between !Xóõ and Bleek-transcribed Taa doculects are generally regular (the efflux -kx'- is often rendered as -k- / = zero/ in her N|u|len records; !- regularly replaces *| in Kakia); a minor problem is word-final -m in Kakia, but there are occasional other instances in which nasalization of labial vowels is rendered as a consonantal segment (e.g. ||ũ ~ ||um STAND, see below).
- Tuu-: Not reconstructible. ◊ It is highly tempting to join !Ui *|qu and Taa *|kx'ãũ in a single etymology, but in order to do that, it would be necessary to at least demonstrate the recurrence of the click efflux correspondence between N|uu -q- and !Xóõ -kx'-, for which no fully convincing examples have been found so far. The two forms in |'Auni also present a challenge, but at least †õĩ is fully compatible with !Ui †qu under the reasonable assumption that *|õĩ ← *|q|u-iŋ (or *|q|u-ni; cf. DOG for a similar case), where the second component is a formerly productive noun suffix.

⁴ For a close-by typological parallel, cf. the curious phonetic variation between -ãĩ ~ -ãũ ~ ã type codas in Ju languages, e.g. Ju|'hoan !ã^h ‘tree’ = !O!Kung !ãũ id. = Ekoka !ã^h id.; Ju|'hoan !ãĩ ‘neck’ = !Kung ||ãĩ ~ ||ãũ id., etc.

59. NEW [-]

- !Ui: Not reconstructible. \diamond This word is fairly well attested only in |Xam, where the most common equivalent for the meaning NEW (with polysemy: ‘new / fresh / raw’) is $\llbracket a:^\xi \eta \rrbracket$, pl. $\llbracket a^\xi \rrbracket a^\xi r̄r̄a$ (with reduplication). Importantly, W. Bleek also records the form $\llbracket we \sim \llbracket wē$ in the same meaning, but textual examples show that it is always used in conjunction with ‘moon’, making it ineligible for lexicostatistical purposes. For ||Ng!ke, D. Bleek lists the adjectival stem $!xe:-k^y a \sim !xe:-t^y a$ ‘new / young’, but for modern N|uu no equivalent has been recorded. The ||Xegwi-LH equivalent, attested in one phrase, is probably $\llbracket i$, with no external parallels.
- Nossob: Not attested in either |’Auni or |Haasi.
- Taa: $*\llbracket qu^\xi \rrbracket$ ($!Xó\tilde{o}$ $\llbracket qu^\xi \rrbracket -V$, Kafia $\llbracket xwe$). \diamond Not attested in N|u|en. The isogloss between $!Xó\tilde{o}$ and Kafia is acceptable, even if Bleek’s $-x-$ for Kafia much more frequently correlates with $!Xó\tilde{o}$ aspirated $-q^h-$ rather than unaspirated uvular $-q-$ (this could theoretically point to an original $*\llbracket q^h u^\xi \rrbracket$ rather than $*\llbracket qu^\xi \rrbracket$).
- Tuu+: Although there are no lexicostatistical matches between any of the three branches, it is important to note that Taa $*\llbracket qu^\xi \rrbracket$ corresponds rather well to |Xam $\llbracket we$ ‘new /of moon/’ (other than the lack of pharyngeal articulation in |Xam); this could imply a narrowing of the original meaning in |Xam and, consequently, a Proto-Tuu origin for the Taa stem (admittedly, this is all highly circumstantial evidence, and the conclusion is liable to change if more data are accumulated).

60. NIGHT [!Ui + Nossob] [- Taa]

- !Ui: $*\llbracket a$ (|Xam $\llbracket a \sim \llbracket a:$, ||Ng!ke $\llbracket a \sim \llbracket a:$, †Kho-D $\llbracket \tilde{a}:$ $\sim \llbracket \tilde{a} \tilde{a}$, N|uu $\llbracket a:$, ||Kxau $\llbracket a:$, ||Ku|e $\llbracket a$, ||Xegwi-LH $\llbracket a:$). \diamond A highly stable and phonetically transparent form, although some fluctuations in click efflux articulation are observable; perhaps the †Kho-D bisyllabic form $\llbracket \tilde{a} \tilde{a}$ may be seen as a clue, although we prefer for now to follow the majority rule in our reconstruction.
- Nossob: $*\llbracket a-$ (|’Auni $\llbracket \tilde{a} u \sim \llbracket \tilde{a}$, |Haasi $\llbracket a-\llbracket a$). \diamond The diphthong in |’Auni finds no confirmation in any other sources; it has to be understood as the result of morphological contraction ($\leftarrow * \llbracket a-u$). The reasons for reduplication in |Haasi are also unclear.
- Taa: $*\llbracket u^\xi \rrbracket$ ($!Xó\tilde{o}$ $\llbracket u^\xi \rrbracket$, pl. $\llbracket u^\xi -m-tê$, Kafia $\llbracket \tilde{d}e^\xi \rrbracket$, N|u|en $\llbracket \tilde{d}e \sim \llbracket we$). \diamond A common and stable Taa root, reliably reconstructible for the proto-level.
- Tuu-: Not reconstructible. \diamond The isogloss between !Ui and Nossob is quite clear, but the Taa root is completely different, and !Ui-Nossob $*\llbracket a$ vs. Taa $*\llbracket u^\xi \rrbracket$ find no mutual etymological parallels.

61. NOSE [!Ui + Nossob + Taa]

- !Ui: $*\llbracket u$ (|Xam $\llbracket \tilde{u}-nu$, ||Ng!ke $\llbracket u-tu$, †Kho-M $\llbracket u-tu$, ||Kxau $\llbracket \tilde{u}-tú$, ||Ku|e $\llbracket u-tu$, N|uu $\llbracket u-\tilde{c}u$, ||Xegwi-Z, ||Xegwi-LH $\llbracket u$). \diamond The pure root is clearly preserved in ||Xegwi; most other languages reflect the extended stem $*\llbracket u-tu$, where $*-tu$ is a special “anatomical” suffix, most likely derived from $*tu$ ‘hole, mouth’. Only |Xam has a different morphological extension, although the form with $*-tu$ is still seen in the plural variant ($\llbracket \tilde{u}-\llbracket \tilde{u}:-tu$ ‘noses’, with root reduplication).
- Nossob: $*\llbracket u$ (|’Auni $\llbracket \tilde{o} \sim \llbracket u:$, |Haasi $\llbracket u$). \diamond In |’Auni, cf. also $\llbracket oi-tu-ke$ ‘nostrils’.
- Taa: $\llbracket u^h-$ ($!Xó\tilde{o}$ $\llbracket u^h-ja$, Kafia $\llbracket u-\tilde{c}a$, N|u|en $\llbracket u-\tilde{s}a$). \diamond As in !Ui, the root avoids being used in pure form (although for Kafia, Bleek claims the possibility of such usage); instead,

the two most commonly encountered morphological variants are $*\tilde{u}^h-na$ and $*\tilde{u}^h-sa$ (both morphemes are nominal suffixes with unclear functions).

- Tuu+: This is one of the most stable and phonetically transparent roots in the entire wordlist, clearly reconstructible as $*\tilde{u}-$ (or $*\tilde{u}^h-$, if breathy articulation of the vowel in !Xóõ is archaic). Of note, however, are the different models of its morphological framing in various branches; only the Nossob languages seem to prefer usage of the pure, unextended root stem.

62. NOT [!Ui + Taa] [- Nossob]

- !Ui: (?) $*\|V-$ ($\|Ng!ke \|u \sim \|j \sim \|e$, $\dagger Kho-M \|o \sim \|e$, $N|uu \|u$, $Seroa \|au$). \diamond Basic expressions of verbal negation in !Ui languages differ highly from language to language. The only one that finds reliable etymological parallels outside of !Ui, and, consequently, the only candidate for the status of a “pseudo-reconstruction”, is the morpheme reconstructible as $*\|V-$ on the basis of $N\|ng$ data. According to Collins & Namaseb 2011: 10, the negation subsystem in $N|uu$ consists of three morphemes: $\|u$ (present tense), $\|am$ (past tense), and $\|ae$ (identificational sentences). It is possible (though not certain) that they may all contain the same root morpheme. The only other language that has a similar negative morpheme is Wuras’ Seroa. There is, however, a possibility that the main negative morpheme in $|Xam$ (recorded as $kx'au$ or $kx'áu-ki$, where $-ki$ is really a copula) is also etymologically related, with irregular click loss ($*\| - \rightarrow kx'-$) possibly due to frequent usage of the auxiliary morpheme.

Other attested morphemes include: (a) the negative verb $\tilde{a} \sim \tilde{o} \sim \tilde{é} \sim \tilde{i}$ in Bleek’s records of $\|Ng!ke$ (e.g. $\eta \tilde{o} \|ai$ ‘I do not know’, $\eta \tilde{i} kien$ ‘I do not sleep’), curiously unattested in any other doculect on $N\|ng$; (b) $\|Xegwi-Z$, $\|Xegwi-LH ?a$ ‘not’, without any etymological parallels.

- Nossob: $*ta$ ($\sim *tu$) ($!Auni kiá \sim tiá \sim tá$, $|Haasi t^y u \sim t^y a$). \diamond Use of a CV-type morpheme with initial $*t-$ (its spelling as $ki-$ $\sim ti-$ in $!Auni$ and as t^y- in $|Haasi$ is consistent with its palatalized articulation in many !Ui languages) is common for both doculects. However, $|Haasi$ shows two differently vocalized variants ($*tu$ and $*ta$) whose rules of distribution are difficult to establish based on Story’s examples. They may or may not have different etymological background; in any case, the only common invariant for both doculects is safely reconstructible as $*ta$.
- Taa: $*\|q^h u-$ ($!Xóõ \|q^h úa$, $Kakia \|wa \sim \|k'a \sim \|a:\acute{a}:\acute{s} \sim \|ai$, $N|u|en \|u$). \diamond In $!Xóõ$, the general negation morpheme $\|q^h úa$ is opposed to the less frequent negative verb $\|à:$ ‘not to be; malfunction’; it is possible that some of the variants in $Kakia$ also reflect not one, but two morphemes (e.g. $Kakia \|wa = !Xóõ \|q^h úa$, $Kakia \|k'a = !Xóõ \|à:$). We provisionally accept the $!Xóõ$ form, with its aspirated uvular efflux, as the most archaic.
- Tuu+: It seems probable that !Ui (more precisely, $N\|ng$ and $Seroa$) $\|V-$ (especially its present tense stem variant $*\|u$) and $!Xóõ \|q^h úa$ are etymologically related, even if click efflux correspondences are not fully regular (Taa $*-q^h-$ typically corresponds to !Ui $*-/k^h-$, see DOG, HAIR), although, on the whole, negation in Tuu languages is quite notoriously unstable, requiring a much more detailed synchronic and historical investigation.

63. ONE [!Ui + [Nossob + Taa]]

- !Ui: $*\mathcal{C}\mathcal{O}\mathcal{Z}-$ ($|Xam !wa:i \sim !w'a:i$, $\|Ng!ke \|we: \sim \|'we:$, $\dagger Kho-M \|oe$, $N|uu \|'oe$, $\|Kxau ?\mathcal{O}e:$, $\|Ku|e \|kx'oa$, $Seroa \|oi$, $\|Xegwi-Z !oa$, $\|Xegwi-LH !wa:$, $\|Xegwi-B \|a:$). \diamond This is an interesting case where, at first, it might seem as if not all the listed forms truly belong together.

However, it may be seen that click influxes generally satisfy the conditions defined for the “sixth click” (* ζ) of Proto-!Ui, namely, an alveolar reflex in |Xam, a lateral release in all the doculects of N||ng, and a (rare) alveolar reflex in ||Xegwi (although note a lateral variant in ||Xegwi-B). A second issue is the odd variation between simple and glottalized variants of the click accompaniment, sometimes within the same dialect cluster (see data on |Xam and N||ng). This, keeping in mind the diphthongial nature of the coda in most dialects, may be interpreted in favor of reconstructing the full !Ui stem as * $\zeta V?V$, more precisely, perhaps as * $\zeta o\eta e$ or even as * $\zeta oa\eta-i$, where *-i is a morphological add-on responsible for diphthongization *-oai \rightarrow -oe in N||ng and ||Kxau. (Although this solution is imperfect in that the function of this *-i remains unclear, its suffixal nature would help explain the discrepancy with ||Xegwi, which could be assumed to still preserve the unexpanded stem * ζoa). The etymological decision to trace all these forms back to a single root is also confirmed by external data (see below) and systemic considerations (not a single !Ui doculect by itself presents evidence for two different inherited roots).

- Nossob: * $\zeta'u\eta$ (!Auni $\zeta'ú \sim \zeta'ú-u$, |Haasi $\zeta'j-k'a$). \diamond If the !Auni and |Haasi forms belong together, this requires assuming that the latter form is either incorrectly transcribed or represents a case of irregular development from an original * $\zeta'V/\eta$, perhaps contracted due to frequent usage in combination with the copula element -k'a. (For an irrefutable example of irregular contraction of numerals in |Haasi, see TWO below). In any case, external evidence from Taa clearly shows that the !Auni form is more archaic in its shape, regardless of whether it is etymologically the same as |Haasi $\zeta'j$ - or if the latter is a separate innovation.
- Taa: (?) * $\zeta'u$ - (!Xóõ $\zeta'úã$, Kakia $k'we$, N|u||en $'oe$). \diamond The variable form of the stem in !Xóõ is $\zeta'u-V$, reflecting $\zeta'u$ - as the original root. Morphological framing of the root in Taa differs from that in Kakia and N|u||en, both of which rather go back to the variant * $\zeta'u-e$ than * $\zeta'u-ã$. The fact that the palatal click in !Xóõ here corresponds strictly to the alveolar click in Kakia and N|u||en (rather than a messy variation between $\zeta-$, ζ' , and ζ' , typical of D. Bleek's records) is perhaps worth noting (see the same situation with TWO below), but it is yet unclear if this can be taken as evidence for reconstructing the “sixth click” on the Proto-Taa level (rather than assuming that it had merged with simple ζ' already before the split of Proto-Taa into its descendants).
- Tuu+: Despite some unclear moments, there is compelling evidence to think that all three branches here show reflexes of a single original root, perhaps to be reconstructed as * $\zeta U?U$ - (which would allow to explain the discrepancy in click efflux correspondences between Nossob and Taa on one hand, and !Ui on the other). Even so, Nossob and Taa forms are clearly closer to one another than !Ui in their phonological reflection of the original root.

64. PERSON [-]

- !Ui: * $\zeta'u-i$ (|Xam $!ú i \sim !ú i-ya$, ||Ng!ke $!wa \sim !wi$, †Kho-D $!wí$, N|uu $\zeta'j=!ui$, ||Xegwi-Z kwi , ||Xegwi-LH kwi). \diamond All languages show reflexes of the exact same proto-stem, sometimes compounded with the stem * $\zeta'a-i\eta$ (\rightarrow N||ng $\zeta'j$) ‘house’. The stem * $\zeta'ui$ itself is morphologically complex, consisting of the original root * $\zeta'u$ (well attested in such |Xam compounds as $!ú 'a$ ‘girl’, $!ú 'a:iti$ ‘woman’, $!u-de$ ‘someone’, etc.) and the common nominal suffix *-i; another possible morphological variant is * $\zeta'u-a$, attested in ||Ng!ke.

A common suppletive plural for this word is reconstructible as $*\#e \sim *\#e'$: |Xam $!e \sim !'e$ (Bleek), $!é \sim !k'é$ (Lloyd), ||Ng!ke $!e \sim !k'e$, N|uu $\#e$, ||Xegwi-Z λe , ||Xegwi-LH $\lambda'e$. The variation observed between click accompaniments is very similar to the one observed in reflexes of HEART q.v. and might go back to an original form like $*\#q'e$, unfortunately, without confirming parallels in Taa.

- Nossob: $*\#e \sim *\#e'$ (|'Auni $\#i \sim \#e$, |Haasi $\#e$). \diamond In Bleek 1937: 218, both |'Auni forms are glossed as 'men, people', but text examples show that they can easily be used as singulatives (e.g. $\#i \text{ ti } \#ú-u$ 'one person'), and a separate plural form $\#i-te$ 'people' is attested as well. For |Haasi, Story lists specifically $\#e$ 'person' and $\#e\varepsilon$ ($=\#e\varepsilon\varepsilon$) 'people'. Comparison with !Ui shows that the original meaning may indeed have been plural, but, unlike !Ui, the Nossob languages have neutralized original suppletion in this paradigm in favor of the plural form.
- Taa: (a) !Xóõ $t\hat{a}$; dial. $l\hat{a}$: 'person'; (b) !Xóõ $t\hat{u}$: 'people', Kakia tu 'person', pl. $tu-ku \sim tu-tu$ 'people', N|u||en tu 'person', pl. $tu-tu$. \diamond The Taa equivalent for sg. 'person' is part of a very small set of morphemes which exhibit the unusual consonantal variation $t \sim l$ across dialects (another well-known example is the plural marker $-te$, dialectally encountered as $-le$). This variation is restricted to only a few morphemes, and no phonetic conditioning for a hypothetical shift $*t \rightarrow l$ has been established; consequently, it is justifiable to see here the reflex of a special rare proto-phoneme, a reasonable phonetic interpretation of which would be a lateral affricate ($*\lambda$). The original paradigm must have been sg. $*\lambda a$, pl. $*tu$ (as in !Xóõ); Kakia and N|u||en show generalization of the plural form, with subsequent formation of new suffixal plurals (the process is structurally identical to that in Nossob languages, though the actual morphemes are different).

It is important to note that Kakia and N|u||en also have additional forms listed by Bleek in the meaning of 'person', notably Kakia da (in such bound formations as $!on-da$ 'old man', $!on-da-ke$ 'old woman') = N|u||en da (in the example $da !oe, du \tilde{l}um$ 'one person, two people'); and N|u||en $\#a$, 'person', pl. $\#a:re$ (not found in text examples). The etymology of these forms and their semantic connection with tu remain unclear, but it cannot be excluded that both da and $\#a$, or at least one of these forms, are actually dialectal reflexes of Proto-Taa $*\lambda a$.

- Tuu-: Not reconstructible. \diamond It is interesting to note that in both Proto-!Ui and Proto-Taa the paradigm for this word was most likely suppletive (!Ui sg. $*!u-i$ vs. pl. $*\#q'e$; Taa sg. $*\lambda a$ vs. pl. $*tu$), yet none of the alternants have any parallels in respective branches. The Nossob languages have their separate agenda here, but at least it is clearly closer to !Ui than to Taa (generalization of the former plural form for singulative usage).

65. RAIN [-]

- !Ui: Not reconstructible. \diamond In Starostin 2013: 378, I have attempted to reconstruct Proto-!Ui $*\#q^{(h)}au$ 'rain' based on presumable common origins for N|uu $\#qau$, on one hand, and |Xam $!wa: \sim !^hwa:$, ||Ng!ke $!^ha \sim !a:$, on the other. This decision now seems hasty to me: although N|uu $\#$ is indeed a regular correspondence for |Xam, ||Ng!ke $!-$, and even the click effluxes may be accommodated, the codas remain incompatible: N|uu $-au$ should regularly correspond to $-au$ in other languages (see $*\#xau$ BLOOD), and there is no evidence for treating $-u$ as a nominal suffix. Moreover, the fact that the earliest attested dialect of N||ng agrees with |Xam, but not with modern N|uu, is indirect evidence for treating $\#qau$ 'rain' as an innovation (of unclear origin).

Similarly, ||Xegwi -LH $\phi^h eu\eta$ ‘rain’, included in the same etymology, should be disqualified for at least three reasons: (a) phonetic — initial consonant should rather be \check{c} - or \check{s} - if it is to reflect Proto-!Ui $*\check{q}^h/-$; (b) semantic — it is actually attested as the verb ‘to rain’ rather than noun ‘rain’; (c) dialectal — for ||Xegwi-B, Bleek lists the equivalent for ‘rain’ as *gaa*, which is seemingly the same word as ||Xegwi-LH *gaʔa* ‘sky’ (← Proto-!Ui $*!aʔa$).

In the end, if there is a minimally likely candidate for RAIN in Proto-!Ui, it should be the same word as WATER q.v., since this is at least a common isogloss between |Xam and some doculects of N||ng. But seeing how ||Xegwi rather neutralizes RAIN and SKY, and given the general lack of stability for this meaning in the Tuu area, it is preferable to leave the slot empty.

- Nossob: |’Auni $||^h\grave{a}:a$. \diamond This is possibly the same word as WATER, even though the two are transcribed differently in Bleek’s data. Not attested in |Haasi (there is a verb $\check{f}i$ ‘to rain’, but these two meanings are often lexically differentiated in Tuu).
- Taa: $*!kx'oe$ (!Xóõ $!kx'ôe$, N|u||en $!xwe$). \diamond The isogloss between !Xóõ and N|u||en seems fairly straightforward, even if this is the only spotted case so far in which !Xóõ $-kx'$ - corresponds to N|u||en $-x-$ ($-kx'$ - and $-k-$ are encountered more frequently in Bleek’s data). The situation becomes more complex, however, if we also take into account the Kafia form $!we-ga-||a$, where $!we$ may be the same morpheme as !Xóõ $!kx'ôe$, $-ga-$ marks possession, and $||a$ is possibly a mistranscription of the word WATER q.v.; if this analysis is correct, it would mean that ‘rain’ was probably not the original meaning of this word and that it could instead have denoted something like ‘raincloud’. This would be consistent with observations on !Ui, where RAIN does not like to behave as a true “semantic primitive” and is typically derived from ‘water’ or ‘sky’. Still, on a purely formal level of analysis this should not prevent us from reconstructing Proto-Taa $*!kx'oe$ ‘rain’, perhaps with polysemy (‘rain / raincloud’).
- Tuu-: Not reconstructible. \diamond Based on analysis of evidence in all three branches, it is reasonable to suggest that there was no separate lexical root for RAIN at all in Proto-Tuu; whether the meaning was linked to a lexeme like WATER, SKY, or CLOUD cannot be properly ascertained.

66. RED [-]

- !Ui: $*ci$ (|Xam $!i: \sim !i:-ya \sim !^hi:-ya$, ||Xegwi-LH $!e$). \diamond A solid isogloss between |Xam and ||Xegwi; preservation of $!-$ in the latter is an argument in favor of the “sixth click” $*c-$. A certain problem concerns the fact that D. Bleek considers the |Xam word to be related to ||Ng!ke $!i$ ‘red ochre used as pigment’, which is a good semantic match, but, according to the model of correspondences laid out in the first part of the paper, should have looked like $*||i$ instead. However, one should also note that ||Ng!ke $!i$ ‘red ochre’ could actually be related not to the |Xam word, but to the word attested as $\check{f}Kho-D \check{f}i$ and N|uu $\check{f}qi$: (Collins & Namaseb 2011: 14), both meaning ‘red’ (although in the sub-dialects of most of the modern speakers of N|uu, the equivalent for ‘red’ is the Khoekhoe borrowing $!kx'aba$). Since phonetic reasons make it impossible to construct an etymology which would unite |Xam, N||ng, and ||Xegwi equivalents at the same time, we prefer to identify two etymologically different roots, $*ci$ ‘red’ and $*\check{f}qi$ ‘red ochre’, the latter represented only within the large N||ng cluster and probably developing the adjectival meaning ‘red’ in some of its daughter dialects.

- Nossob: |Haasi *cxwe-k'a*. ◇ Clearly the same root as |'Auni *coa* 'red colour, ochre', although it is unknown if the |'Auni root also had an adjectival usage (no other words for 'red' are attested in Bleek's data).
- Taa: *|a^h-**ɲa** (!Xóõ *ǎ^hɲa*, Kafia *ǎnyá*, N|u||en *ǎne*). ◇ The root is probably just *|a^h-, since *-*ɲa* is a suffixal component frequently seen in other color terms as well (see BLACK, WHITE). The N|u||en form is only included in Bleek 1929 and is not confirmed in Bleek 1956, but seems close enough to the other two forms to merit inclusion.
- Tuu-: Not reconstructible. ◇ All three branches have their own equivalents, without any clear etymological parallels on the Common Tuu level.

67. ROAD [!Ui + Nossob (?)] [- Taa]

- !Ui: *|aN (N|uu *!an* ~ *!aɲ*, ||Xegwi-Z, ||Xegwi-LH *kaɲ*). ◇ Optimal candidate for Proto-!Ui status is unequivocally pointed out by the phonetically impeccable and semantically precise isogloss between N|uu and ||Xegwi (although the precise phonemic nature of the final nasal remains somewhat unclear). No traces of this root, however, are found in either |Xam or earlier attested forms of N|uu, all of which have their own, etymologically unclear equivalents (|Xam *!xarra* 'path', ||Ng!ke *tirau* 'path').
- Nossob: |'Auni *!án* 'path'. ◇ Attestation of this form allows for the possibility of a Proto-Nossob *|*an*, cognate with Proto-!Ui; however, no equivalents for 'road' or 'path' are attested in |Haasi, and there is nothing to disprove that the |'Auni form may just as well represent a borrowing from one of the N|uu dialects, especially since there is at least one more N|uu-|'Auni isogloss with similar phonetics and semantics: N|uu *||uru-ke* 'path, road, trail' = |'Auni *||uru* 'path'. Finally, Bleek lists yet another |'Auni word with the same meaning: *!k'ei* 'road, path', which, if the other two are N|uu borrowings, could actually represent the inherited Nossob term. The situation cannot be properly resolved at our current level of knowledge.
- Taa: !Xóõ *!ólo*. ◇ This looks like a possibly inherited term, glossed by Traill as 'path'; in the more "cultural" meaning 'road, way' !Xóõ has the Khoe borrowing *dào*, which is also the only known equivalent for this entire semantic cluster in Kafia (*dau*) and N|u||en (also *dau*).
- Tuu-: The isogloss between Proto-!Ui and |'Auni is undeniable, but could reflect either common heritage or areal contact. Given the overall lack of stability of this concept (easily replaceable by borrowings from Khoe or words with unknown etymology), a reconstruction at the Proto-!Ui-Nossob level does not seem particularly trustworthy.

68. ROOT [-]

- !Ui: Not reconstructible. ◇ Given the specific botanic circumstances of the Tuu-speaking area, this concept is notably unstable, and it is never clear just how specific or generic attested terms are; cf. |Xam *!au* 'root' in Bleek 1929: 71, but the same word glossed as 'wild onion' in Bleek 1956: 414. For N|uu *!^habe-si* 'root', B. Sands states (p.c.) that "only one of the Eastern N|uu speakers knows this word", and suggests *!ao-si* ~ *!ãũ-si* 'root of shepherd's tree (*Boscia albitrunca*)', sometimes extended to denote 'root' in general as a more suitable term. For most of the other languages, the exact meaning 'root' is not attested at all.
- Nossob: |'Auni *!au-si* is glossed as 'small roots' in Bleek 1937: 219. ◇ This is clearly the same word as N|uu *!ao-si* ~ *!ãũ-si* 'root of shepherd's tree', mentioned above, and,

given the exact same morphological structure in both cases, is probably a recent borrowing into !'Auni from N|uu.

- Taa: Not reconstructible. ◊ Cf. !Xóõ !k'á-i, pl. !k'á-ba-tê 'woody root'; N|u|en *lau-te* 'root' (a plural form). The precise semantics of both terms remains unclear.
- Tuu-: Not reconstructible, although it must be noted that N|u|en *lau-te* (assuming that the transcription with *l-* is either erroneous for *ʃ-* or reflects a regular development from it) is formally compatible with N|uu *ʃao-si ~ ʃãũ-si*. Still, due to the overall difficulties with this concept, it is probably better to exclude it from analysis.

69. ROUND [-]

- !Ui: Not reconstructible. ◊ This concept is only attested for |Xam (*kuérre-kuérre ~ kwórre-kwórre*, a reduplicated verbal stem applicable to round objects such as 'sun' or 'egg'); there is also ||Ng!ke *kakeriŋ* 'round' in Bleek 1929: 71, which may or may not be related to the |Xam item, but finds no confirmation in any other of the many later sources on the N|uu cluster.
- Nossob: Not attested for either !'Auni or |Haasi.
- Taa: Not reconstructible. ◊ Cf. !Xóõ *ʃnúʔm* /é: 'round shaped, tubular (e.g. a branch, rod), as opposed to flat-shaped'; another synonym with almost the exact same meaning is *ʃnãʔō* /é:.
- Tuu-: Not reconstructible, mainly due to lack of attestation in daughter languages; a rather problematic concept for the Khoisan area as a whole.

70. SAND (= EARTH) [-]

- There is not a single reliable case in which a Tuu word for 'sand' would be lexically different from the corresponding word for 'earth'; it may be safely assumed that such a differentiation did not exist on the Proto-Tuu level, either. See EARTH in the first part of the paper for more details.

71. SAY [-]

- !Ui: ***ku** ~ ***ka** (||Ng!ke *ka*, †Kho-M *ka ~ ku ~ k'u ~ kwa ~ kɔʃ ~ ku:-wa*, N|uu *ka ~ ku*, ||Kxau *ku*, ||Xegwi-Z *kũ*). ◊ Variations in root vocalism are attested for this verbal root in both earlier (Maingard 1937) and newer (Collins & Namaseb 2011) sources on N|uu, seemingly without any explanation; forms such as Maingard's *ku:-wa* hint that *ka* may historically represent a contraction from **ku-a*, but this is not completely certain, though, most likely, both variants do reflect the same root (||Xegwi-Z *kũ* also supports the idea of the labial vowel as original). In |Xam, *ka* is found glossed as 'to wish, intend, think, say' and seems to refer more frequently to mental than verbal activity; the basic equivalent for 'say' (introducing direct speech, etc.) is the morphologically complex verb *ʃákkən ~ ʃákka ~ ʃákən ~ ʃáka ~ ʃággən*, which is perhaps related to N|uu *ʃ^hoa* 'to speak' and, in any case, seems to be innovative in the meaning 'to say'.
- Nossob: ***|u** (!'Auni *|u*, |Haasi *|wa*). ◊ For !'Auni, Bleek records two verbs that introduce direct speech: *|u* and *ko*, without any clear differentiation. It may be suggested, based on the |Haasi parallel, that *|u* is the inherited term, whereas *ko* is borrowed from N|uu *ku*.
- Taa: ***tV-** (?). ◊ The situation here is as follows. For !Xóõ, Traill records *téʔē* 'say', *tám* 'say it, mean', and *tána* 'talk, speak'; all of these three forms could theoretically be related, going back to a single root *tV-* with various suffixal extensions, but this is im-

[Xam records with a different direction of assimilation). Semantic glossing is not always accurate across various sources, but the contrast between stative and non-stative has at least been expressly documented for N|uu (Collins & Namaseb 2011: 20). Ziervogel's transcription *šoge* for ||Xegwi most likely reflects the same stem as *šogaʔane* 'sit and wait form me' in Lanham & Hallows 1956: 116, analyzed as *šo* 'sit' + *gaʔa* 'wait' + *ne* 'me'; Bleek simply lists *šo:* as the primary equivalent for ||Xegwi.

There is also a common !Ui suppletive plural stem for this verb: |Xam *!ʰau:* ~ *!ʰáú-wa* 'to sit' = ||Ng!ke *!ãũ* = N|uu *!qʰãũ* id. (← Proto-!Ui **!qʰau-*).

- Nossob: |ʼAuni *sã* ~ *são* ~ *sõ* ~ *so* 'to sit; to sit down, put down, set, stay'. ◊ Not clear if this item, easily relatable to !Ui **so-*, is etymologically the same as |Haasi *c'i* 'to sit'; the main problem here would not be so much the ejective affricate as the completely unexpected vocalism. In any case, the |ʼAuni forms, due to their own specific vocalic properties, are probably authentic rather than borrowed from N|uu (regarding variants such as *sã*, the correspondence between |ʼAuni *a* and !Ui *o* is quite current; variants such as *sõ* ~ *so* may either preserve the original vocalism in specific grammatical or phonetic contexts, or could indeed be influenced by N|uu).

Both |ʼAuni and |Haasi have additional synonyms for this concept, e.g. |ʼAuni *!ãũ* 'to sit, to squat'; |Haasi *!xi-k'i* 'to sit'. The former is clearly the same item as Proto-!Ui **!qʰau-* 'to sit' (of many); whether this is an inherited Nossob term or a borrowing from N|uu is impossible to determine, both solutions are more or less equiprobable. For the |Haasi term, no diagnostic contexts are available to ascertain its true meaning or usage.

- Taa: **cʰu* (!Xóõ *cʰú:*, Kakia *ču*, N|u||en *šu* ~ *ču*). ◊ The precise nature of the affricate remains to be ascertained; cf. the stunning number of variants for this root listed in Maingard 1958 (*c'o* ~ *č'o* ~ *čə* ~ *č'ou* ~ *su* ~ *šu* ~ *č'ho*). Unlike !Ui, there seems to be no separate morphological variant for the non-stative verb 'to sit down'. There is, however, also a suppletive plural stem: !Xóõ *!á:* = N|u||en *!a:* 'to sit' (of many), reconstructible as Proto-Taa **!ʼa-*.
- Tuu+: **cʰo* is reliably reconstructible, despite some phonetic uncertainties (aspiration, etc.), as the common Tuu morpheme for the meaning 'sit' based on data from all three branches. The suppletive plural stems (Proto-!Ui **!qʰau-* and Proto-Taa **!ʼa-*), despite a matching click influx, are on the whole unreconcilable with each other; note that, because of |ʼAuni *!ãũ*, Nossob and !Ui are closer to each other in this respect than to Tuu.

75. SKIN [!Ui + Taa] [- Nossob]

- !Ui: **Tun* (|Xam *t:ũŋ*, ||Ng!ke *tũ* ~ *twã* ~ *diõ*, †Kho-M *ʒo*, N|uu *zũ:*, ||Xegwi-LH *tun* ~ *tũ:*). ◊ Of note in this case is the unexpected voiced articulation of the initial consonant in modern N|uu, as well as at least some of the older dialects (palatalization **t-/*d-* → **t̪-/*d̪-* in them, on the other hand, is quite regular). A few other cases of such fluctuations are known, but this is the most transparent one; it is possible that the voicing reflects influence of some additional feature that cannot be recovered from available data. Cf. the situation in Taa for a potential clue.
- Nossob: **||ʼU:* |ʼAuni *!ũ:*, |Haasi *||ɔ*. ◊ Although Bleek's and Story's data contradict each other on the precise nature of the click efflux, this is hardly a valid reason to deny the common origin of both items. If the word is related to |Xam *||ɔ* 'outer skin', 'shed skin' with its voiced efflux, this makes the reconstruction **||U* more probable than the one with glottalization.
- Taa: **tuʼm* (!Xóõ *tùʼm*, Kakia *t'üm*, N|u||en *t'üm*). ◊ It is likely that glottalized *t'*, attested in Bleek's transcriptions for Kakia and N|u||en, is a perception error for *t-* + pharyn-

gealized vowel, as in Traill's transcription of !Xóõ. Alternately, it is possible that glotalization was in fact primary, and that !Xóõ underwent a sporadic (or even regular) change from *t^hum to *tu^hm.

- Tuu+: !Ui *Tuŋ and Taa *tu^hm (*t^hum?) are clearly relatable; coda correspondences here are the same as in BREAST and LIVER, implying either different models of suffixation or a phonetic scenario (*-ŋ → *-m in Taa as a regular development?). It also seems as if the strange voicing in N|uu and the fluctuation between t^hu- and tu^h- in Taa may be correlated, but until more examples of such correlation are found, it is difficult to jump to conclusions. In any case, this is yet another instance where !Ui and Taa agree vs. the Nossob languages, which have innovated a different term (perhaps semantically extended from more specialized usage).

76. SLEEP [!Ui + Nossob + Taa]

- !Ui: ***Ōu-** ~ ***Ōi-** (|Xam θ_uoin, ||Ng!ke θwoin ~ θwoeŋ ~ θóeŋ, †Kho-M θ'wō ~ θ'wonna, N|uu θun ~ θuŋ, ||Kxau θan, ||Xegwi-Z θi). ◇ The only thing certain in this reconstruction is the initial consonant; root vocalism may have been *i, as seen in Ziervogel's "future tense" stem for ||Xegwi (he also lists θi-ŋe as "present tense" and θi-ŋa as "past tense" stems), with assimilation to the labial click everywhere else, but this is not obvious. Most of the stems end in nasal consonants, but, again, ||Xegwi θi shows that they may all be suffixal in origin.
- Nossob: ***ŌV-i-** (|'Auni θwōi, |Haasi θwa-ai).
- Taa: ***ŌV-n-** (!Xóõ θân, Kafia θwōi ~ θwoin, N|u|en θwoin). ◇ Cf. also !Xóõ θûm 'sleep (n.)'.
- Tuu+: This is a very stable root, but variation in the coda across all known languages is too extensive to allow for an unequivocal reconstruction other than *ŌV-. It is certain that at least some of the original morphological variants ended in a nasal (*ŌV/-n), but whether it was really an integral part of the root or a suffixal extension is hard to determine.

77. SMALL [-]

- !Ui: ***ʔeni** (|Xam ʔeŋŋi ~ ʔenniŋ, ||Ng!ke ʔĩ, N|uu ʔĩ, ||Xegwi-Z ʔine, ||Xegwi-LH ʔ'ini). ◇ This adjectival stem is preserved in its bisyllabic shape in |Xam and ||Xegwi (where the development of palatal click into a lateral affricate is regular; LH ʔ'ini, with glottalic articulation, is probably a more accurate transcription than Ziervogel's ʔine), but contracts to ʔĩ in N|uu. Of note is the explicit marking of the palatal click articulation in |Xam, since typically Proto-!Ui *ʔ is marked in both W. Bleek's and L. Lloyd's records as !; this may have something to do with the "expressive" nature of the word.

Additionally, in the meaning 'small' some sources list reflexes of a common !Ui morpheme that starts with a labial click, e.g. ||Ng!ke θwain-ki, †Kho-M θ'kō, †Kho-D θónē, ||Xegwi-LH θa-ri. This corresponds to |Xam -θwa ~ θuá, a diminutive morpheme usually found in conjunction with words denoting 'children' or 'young of birds / animals', and to modern N|uu θũ with more or less similar usage. It is quite possible that this is the original equivalent for the unbound adjective 'small', but that already in Proto-!Ui its usage had become more restricted, whereas *ʔeni had already acquired more productive functions.

- Nossob: (a) |'Auni ʔai; (b) |Haasi n^yái-si. ◇ The |'Auni form is not very reliable, as it is attested only in Bleek 1929: 76 (an early source); in any case, it can hardly be the same as |Haasi n^yái-si, which is fairly unique in itself (beginning with a palatal nasal). Unclear situation on the whole.

- Taa: (a) !Xóõ /ʔúí (suppletive pl.: /qʼán-tá); (b) Kákia /óna; (c) N|u|l|en /ari. ◊ The concept is clearly unstable in the Taa branch as well; all three doculects show different equivalents. Note the presence of the diminutive formant *θà:* in !Xóõ, e.g. *θàye-θà:* ‘animal’ (= ‘meat-small’).
- Tuu-: While the diminutive morpheme **θV-* is clearly archaic and reconstructible at the Proto-Tuu level, the same cannot be said about the unbound adjective ‘small’, expressed by different equivalents at the level of each subgroup (and even notoriously unstable within most of them).

78. SMOKE [!Ui + Nossob] [- Taa]

- !Ui: **||oʰ* (|Xam *||ó:*, †Kho-D *||ʔəʔə-ké*, N|uu *||o:ʰ-ke*). ◊ Although the |Xam form is attested only scantily, the isogloss between it and the N|uu cluster allows to make a reliable reconstruction at least on the “narrow !Ui” level. The only other well-attested form is ||Xegwi-LH *kʰaʔa-zi*, whose origins are hard to ascertain (according to Lanham & Hallows, the suffix *-zi* frequently marks verbal derivatives or borrowings from Bantu languages, but *kʰaʔa* is hardly identifiable as either a verbal stem or a Bantu borrowing).
- Nossob: **||au* (|ʼAuni *||áú*, |Haasi *||au*). ◊ The |ʼAuni form is listed as a noun in Bleek’s dictionary; for |Haasi, Story indicates that *||au* is found in both nominal and verbal usage, although the only textual example is within the context ‘I smoke’.
- Taa: (a) !Xóõ *ckʰáye*; (b) Kákia *||álu*. ◊ Not attested in N|u|l|en. The form in !Xóõ is notoriously similar to Proto-Khoe **cʰán(i)* ‘smoke’ (Vossen 1997: 476), especially since it is possible that the former was phonetically realized as **ckʰán-*; however, direct borrowing from a relatively recent Khoe source is excluded for phonetic reasons.
- Tuu+: Comparison between Proto-!Ui **||oʰ* and Proto-Nossob **||au* ‘smoke’ is distributionally and semantically solid; phonetics-wise, the correlation between **-o* and **-au* may raise questions, but is not completely unprecedented (at least Nossob **-au* vs. !Ui **-oe* is recurrent, cf. |Xam *||xwe*: vs. |ʼAuni *||xau* ‘cold’, |Xam *||kʰoe* vs. |ʼAuni *||kʰau* ‘back’). The parallel may be accepted as a lexicostatistical match between the two branches, whereas the situation in Taa is different.

79. STAND [-]

- !Ui: (a) |Xam *!ʰe:* ~ *!ʰé:*; (b) ||Ng!ke *||á*, ||Kxau *||a*; (c) N|uu *||á:* ~ *||áʰa*; (d) ||Xegwi-Z *!ʰoʔo-ge*. ◊ The situation here is complicated not just because of the relative instability of the concept, but also due to insufficiently accurate semantic glosses in most available sources. Thus, while forms in group (b) are clearly the same verb as |Xam *||á* ‘to be (located), stay’, available examples do not make it clear if ||Ng!ke truly extends the usage of this verb to contexts with the meaning ‘stand’ (in a vertical position), or if it is actually the same as in |Xam. In light of this, no particular reconstruction can be reliably offered for Proto-!Ui.
- Nossob: (a) |ʼAuni *!ʰá:*; (b) |Haasi *||wa*. ◊ Despite some phonetic similarity, the two forms are hardly reconcilable with each other (lack of glottal release and extra labialization in |Haasi would be unexplainable in a common etymology). It should be noted that Collins & Namaseb also record an alternate variant *!ʰana* ‘to stand’ for N|uu; if so, the |ʼAuni form could be suspected of having been borrowed from N|uu (with secondary contraction).
- Taa: **||ʰũ* (!Xóõ *||ʰũ:*, Kákia *||ũ ~ ||ō ~ ||ʰũ ~ ||um*, N|u|l|en *||ũ ~ ||ʰu ~ †ʰu*).
- Tuu-: Not reconstructible. ◊ The concept seems to be fairly stable in the Taa cluster, but not in !Ui or Nossob languages.

80. STAR [-]

- !Ui: (a) ***||kx'oa[̣]-** (||Ng!ke *||wai[̣]-sa* ~ *||kx'we:-sa*, †Kho-D *||wāi-ḡē* pl., N|uu *||kx'oe[̣]-si*, ||Ku||e *||ante* pl., ||Kxau *||'an-si*); (b) |Xam *||ua[̣]-ttən*, ||Xegwi-Z *||ou-ni* pl. ◊ The most commonly encountered root for 'star' in !Ui is spread throughout the N|uu cluster and is further confirmed by entries for ||Ku||e and ||Kxau; the common invariant in all these forms is the root ***||kx'oa-** (probably with vowel pharyngealization, expressly attested in Bleek's data on ||Ng!ke as well as modern N|uu), which is found in conjunction with different suffixes (N|uu ***||kx'oa[̣]-i** → modern *||kx'oe[̣]*; ||Ku||e + ||Kxau ***||kx'oa-n**) and additional markers of singularity or plurality.

To this root is opposed the entry in |Xam *||ua[̣]-ttən* (cf. the reduplicated plural: *||ua[̣]-||ua[̣]-ttən*), which finds a distributionally surprising parallel in ||Xegwi-Z (*||ou-ni*). The latter form is somewhat suspicious as to its morphological constitution, and finds no support in alternate sources for ||Xegwi (Bleek has an etymologically unclear *kale* 'star' in its place); also of note is the seemingly full homonymy of the root morpheme in |Xam with *||ua[̣]-* 'cloud' (although any semantic connection between 'star' and 'cloud' would be decidedly non-trivial). Still, phonetically and semantically the match between |Xam and ||Xegwi is impeccable, allowing to reconstruct ***||ua[̣]-** ~ ***||au[̣]-** (vocalism metathesis in ||Xegwi?) as an alternate candidate.

- Nossob: |'Auni *!'^ha*. ◊ Not attested in |Haasi.
- Taa: ***||ona** (!Xóō *||ōna*, Kafia *||wana-te* ~ *||wana-te* pl., N|u||en *||'ana-te* pl.). ◊ This could originally have been a plural form, given the abundance of sg. ***CVn** / pl. ***CVn-a** paradigms in !Xóō (with subsequent formation of a new, more productive plural by means of the formant *-te*).
- Tuu-: It is extremely tempting to compare Proto-!Ui ***||kx'oa[̣]-** and Proto-Taa ***||ona**, but the match would clearly be problematic due to the non-recurrent nature of the efflux correspondence. Until additional supporting evidence surfaces, we prefer to keep these items separate, and assume that no common etymon can be reliably reconstructed for Proto-Tuu.

81. STONE [-]

- !Ui: ***!ao** (|Xam *!au* ~ *!óu*, ||Ng!ke *!au* ~ *!áu*, N|uu *!ao*, ||Kxau *!ao*, ||Xegwi-LH *ḡ'eo*). ◊ Inclusion of the ||Xegwi form is somewhat problematic: *ḡ'eo* should normally go back to ***k'ao** ← ***!'ao**, yet there are no signs of a glottalized efflux in either |Xam or N|uu. On the other hand, glottalization is indirectly supported by the curiously shaped ||Ku||e form *dʔɔ* 'rock', recorded by D. Bleek (*d-* is the normal ||Ku||e reflex for ***!-**, cf. *dōa* 'tortoise' = |Xam *||óě* id., *dɔaxu* 'sky' = |Xam *||waxu* id., *dwene* 'three' = |Xam *||wanna* id., etc.); one may wonder if this does not reflect an original ***!aʔo**, contracted to ***!'ao** in some of the daughter languages. More problematic is that ||Xegwi also has variants of the word 'rock, stone' with initial affricates or fricatives: cf. ||Xegwi-LH *ḡwe*, ||Xegwi-Z *ḡeu*, ||Xegwi-B *ḡe*, *ḡu*. Lanham & Hallows mention that "this is not the common Bushman word for 'stone', and it was obtained from one group of informants only", and propose borrowing from Sotho *li:=ḡwe*. This explanation is not unquestionable, but the alternate solution (suggesting an irregular dialectal development ***!-** → *ḡ-?*) is certainly less preferable.
- Nossob: (a) |'Auni *||kx'ɔ*; (b) |Haasi *!ðè*. ◊ The former item is only found in the early source of Bleek 1929 and is quite dubious. The form in |Haasi is at least well supported by text examples.

numerous other cases of “epenthetic *-w-*” found in this language), but fluctuation between diphthongic *-ai* and monophthongic *-i* has not been explained. However, **!h-* is reliably reconstructible based on data from |Xam, N|uu, and ||Xegwi (could the odd pharyngealization in Doke’s *!āi^f* be a mistake for aspiration?).

- Nossob: (a) |’Auni *ʔwi*; (b) |Haasi *i=|á:-a*. ◊ |Haasi *i=* is probably a possessive prefix. External data (see Taa below) show that the |’Auni form is likely an innovation.
- Taa: **|āũ*, pl. **|ā* (!Xóõ *|āũ*, pl. *|ā:*, Kákia *|āũ ~ |āũ*, N|u|en *|āũ*). ◊ The !Xóõ form is homonymous with NAME (see above) and shares exactly the same grammatical characteristics; the original root may have been **|au-* or simply **|a-* (the external parallel in |Haasi favors the latter choice).
- Tuu+: The undeniable isogloss between Taa and |Haasi *=|a:-* suggests the reconstruction of **|a-* as the original root for TAIL, which Proto-!Ui replaces with an innovation of unknown origin.

85. THAT [-]

- !Ui: (a) |Xam *|e: ~ |e ~ |e:-á*; (b) ||Ng|ke *á*; (c) ||Ng|ke *||e=á ~ ||ŋ-á*, N|uu *||a: ~ kea*; (d) ||Xegwi-Z *!e=ta*, ||Xegwi-LH *!e=na ~ !e=la ~ !e=ta*. ◊ Descriptions of deictic pronoun systems for most !Ui languages are highly inadequate, and textual examples are almost always ambiguous. On the whole, at least for “Narrow !Ui” (without ||Xegwi) it would make sense to reconstruct **a* as a “general” deictic stem, most commonly used to denote proximal deixis (see THIS), whereas distal deixis must have been denoted by using it as a base for various spatial particles — such as *|e* in |Xam and *||e* in N|uu; the latter, as far as modern N|uu is concerned, undergoes irregular phonetic development in both its prepositional (*||e-a* → *ke-a* with click loss) and its postpositional form (*||e-a* → *||a:* with vocalic contraction), cf. *!o: ||a:* ‘that man’ vs. *kea !o:* id. in Collins & Namaseb 2011: 35–36. However, it is unclear which of these variants — if any — is more archaic than the other; moreover, concern can be raised over their similarity with semantically identic morphemes in Khoe (*||a-* ‘that’, **|e-* ‘this’, see Vossen 1997: 377), possibly implying areal interference. The situation is also clearly different in ||Xegwi, where the “general” deictic stem *!e=* is combined with different morphemes (*=na*, *=la*, *=ta*) to form different (and semantically ambiguous) deictic pronouns. Keeping all this in mind, it is perhaps best to refrain at the moment from attempting to identify the principal morpheme(s) responsible for denoting the idea of distal deixis in Proto-!Ui.
- Nossob: (a) |’Auni *ha ~ he ~ hi*; (b) |Haasi *cɔ:-a*. ◊ Fluctuation in |’Auni may be due to contraction with various nominal class markers. Nothing is properly reconstructible for Proto-Nossob, since the attested morphemes are clearly different (moreover, textual evidence to support accurate semantic glossing is pretty much non-existent).
- Taa: (?) **tV-* (!Xóõ *tV?V*, Kákia *ta-le*, N|u|en *ti*). ◊ All contexts for Kákia and N|u|en are highly dubious; as for !Xóõ, *tV?V*, like **a* in !Ui, is more of a “generic” deictic stem than specifically ‘this’ or ‘that’ — in order to form distal deixis stems, it is usually extended with different additional morphemes (nominal stem *tV?V=BV?V*; adjectival or verbal stem *tV(?V:)-yà kV* ‘there, that /proximate/’; adjectival or verbal stem *tV?V:-sà kV* ‘there, that /remote/’).
- Tuu-: All data clearly show that expression of distal deixis widely fluctuates even within the small subgroups, let alone in between them. Nothing is properly reconstructible.

86. THIS [!Ui + Nossob] [- Taa]

- !Ui: ***a** (|Xam $a \sim a$; ||Ng!ke a , N|uu a). ◇ This is the “generic” deictic pronoun which can, on its own, express proximate deixis at least in |Xam and in N|uu. In |Xam, this monovocalic pronoun sometimes changes to e : (most likely, reflecting contraction with a nominal class marker), but in N|uu, it seems to be the only equivalent for THIS. Whether it has anything to do with $?e=$ in ||Xegwi-Z $?e=la$ ‘this’ (= ||Xegwi-LH $?i=la \sim ?e=la$ ‘this, that’) is unclear.
- Nossob: (a) !’Auni a ; (b) |Haasi $g^y a-\eta$. ◇ The !’Auni form is clearly the same as the !Ui pronoun. Story’s $g^y a-\eta$ ‘this’ for |Haasi is, however, quite mysterious (in any case, truly diagnostic contexts with an adjectival ‘this’ are not attested in Story’s manuscript).
- Taa: ***tV-** (!Xóõ $tV?V: \sim tV:?V \sim tán?n$, Kakia ti). ◇ The simple stem in !Xóõ is also used as the basis for other deictic pronouns (see THAT).
- Tuu+: The isogloss between |Xam, N|uu, and !’Auni speaks in favor of * a as the likeliest of all deictic pronominal stems to go back to the Proto-Tuu level. There is, however, a sharp divide in this respect between !Ui and Nossob, on one hand, and Taa, on the other, where the principal “general” deictic stem is * $tV-$, without any parallels in the other two branches.

87. THOU [!Ui + Nossob + Taa]

- !Ui: ***a** (|Xam $a \sim a-á$, ||Ng!ke a , †Kho-M a , N|uu a , ||Xegwi-Z $?a \sim ?a-\eta$, ||Xegwi-LH $?a-?e \sim a-?e$). ◇ The basic root shape, stripped of all additional markers (such as emphatic particles, etc.), is always * a .
- Nossob: ***a** (!’Auni a , |Haasi $g^y \hat{a}:=a$). ◇ The |Haasi form is listed with a prefixal emphatic particle.
- Taa: ***a** (!Xóõ \bar{a}^h , Kakia a , N|u||en $a \sim a-a$). ◇ Breathiness of the vowel in !Xóõ may be an original feature, in which case the reconstruction has to be amended to * a^h .
- Tuu+: ***a** is clearly reconstructible as the common, ubiquitously preserved root morpheme for the 2nd p. sg. pronoun in Proto-Tuu.

88. TONGUE [!Ui + Nossob + Taa]

- !Ui: *!’**ani** (|Xam $’énni \sim ’éřri$, ||Ng!ke $’ē$, pl. $’e:n-yən$, †Kho-M $’an$, ||Kxau $’anan-si$, N|uu $’ān \sim ’āi$, ||Xegwi-B $’ē$). ◇ The original bisyllabic stem shape is arguably best preserved in |Xam ($’énni$, with vocalic assimilation) and in ||Kxau (where the strange form $’anan-si$ looks like a secondary singulative from a plural form, i.e. $\leftarrow *’ani-Vn_{PL-SiSG}$). In most other languages intervocalic $-n-$ is lenited and reduced to nasalization of the vowel.
- Nossob: !’Auni $’āri$. ◇ Not attested in |Haasi, but cf. Xatia $|a: \sim |ā^f a$ (Bleek 1956: 268; marking of pharyngealization is curious, but worth taking into consideration because of the Taa parallel). The !’Auni form looks inherited rather than borrowed from N|uu.
- Taa: *!’**na̱n** (!Xóõ $?|nā^h n$, pl. $?|nā^h n-a \sim ?|nā^h$, Kakia $’a:n$, N|u||en $’a:ni$). ◇ The N|u||en form is slightly suspicious due to lack of nasality in the click efflux; perhaps this is really SVIa (Krönlein’s N|usan, a dialect of |Xam) rather than SVI (N|u||en)?
- Tuu+: All listed forms clearly belong together, although it is hard to say if *!’- or *!’ n - has to be reconstructed for Proto-Tuu (the first scenario would imply that nasality in Taa is secondary, probably through the influence of the nasal coda) due to lack of additional data.

89. TOOTH [!Ui + Nossob] + Taa]

- !Ui: *||^hai(-N) (|Xam ||^hē:i, pl. ||^he||^hēī, ||Ng!ke ||āī: ~ ||ē: ~ ||ēī, pl. ||eŋən ~ ||ēīŋ ~ ||ē||ē, †Kho-M ||ēī ~ ||ēī-si, N|uu ||^hāī, ||Ku||e k'e pl., ||Xegwi-Z, ||Xegwi-LH ||^hi, pl. ||^hi-ŋ). ◇ The click onset is safely reconstructed as *||^h- based on the joint evidence of |Xam, modern N|uu, and ||Xegwi. The coda presents more difficulties, with a unique correspondence series; however, ||Xegwi offers a clue, allowing to assume *||^hai- (→ ||Xegwi ||^hi in a regular manner) as the original singulative root and *||^hai-ŋ as the old plural form, which became generalized as the singular in |Xam and N|uu and from which more innovative plural forms were later formed by various productive means.
- Nossob: (?) *||e- (|'Auni ||ēī, |Haasi k'i=||ε). ◇ The |'Auni form, attested only in the early source Bleek 1929: 86, raises some doubts (it looks too suspiciously close to N|uu to be reliably recognized as inherited), but the |Haasi form is undeniably archaic, reflecting the original stem without nasality (just as in ||Xegwi; k'i= is the productive prefix of plurality).
- Taa: *||q^haN (!Xóō ||q^hā:, Kakia ||xū, pl. ||xa:ni, N|u||en ||'an-te pl.). ◇ Nasality is always a part of the stem here, but (a) it is not clear if it is more appropriate to reconstruct *-ā or *-an, (b) it is highly probable that it was a class marker anyway (the word belongs to class 2 in !Xóō, whose regular concord marker is -ā).
- Tuu+: *||q^ha- is the likeliest reconstruction of the original root shape underlying all the attested reflexes in all three branches. Of note is that !Ui (at least ||Xegwi) and Nossob (at least |Haasi) agree in reflecting the morphological shape *||q^ha-i, as opposed to Taa *||q^ha-N.

90. TREE [!Ui + Nossob + Taa]

- !Ui: *||θo (|Xam θ^ho, ||Ng!ke θo ~ θo: ~ θ^ho, †Kho-M θo 'wood', N|uu θo: 'wood', ||Kxau θo:, ||Xegwi-Z θo ~ θ^ho ~ θ^hoŋ, ||Xegwi-LH θò:-zì 'tree', θo: 'wood'). ◇ Attested variation between click effluxes is quite flabbergasting here; keeping in mind that the correlated Taa parallel begins with *||θn-, it makes sense to suggest that here, too, the original efflux was more complex than the simple velar release attested, e.g., in modern N|uu, but uniqueness of the correspondence series makes it difficult to propose anything with certainty. From a semantic / lexicostatistical perspective, it is important to note that in some languages, the stem is only glossed with the meaning 'wood', most notably modern N|uu, where, according to most sources, the common equivalent for 'growing tree' (or, specifically, for 'shepherd's bush /Boscia albitrunca') is †^hi: ~ †q^hi:. However, both internal and external data show that there is no reason not to project the common 'tree / wood' polysemy onto the Proto-!Ui level.
- Nossob: (a) *||θo- (|'Auni θwa:a ~ θwa:-sa 'wood, stick, tree', |Haasi θöi 'wood, stick'); (b) |Haasi †^hai 'tree'. ◇ The situation in |Haasi seems to be more or less the same as in N|uu, either reflecting a contact scenario or the result of independent ("homoplastic") development. Interestingly, for |'Auni Bleek does not report any similar dichotomy between 'wood' and 'tree', despite it being in far more obvious contact with N|uu than |Haasi.
- Taa: *||θ̃a- (!Xóō ?θnàye, pl. ?θnā:, Kakia θ̃oe: ~ θ̃oi, N|u||en θ'a:). ◇ !Xóō, most likely, preserves the original preglottalized nasal efflux, simplified (or mistranscribed) in the other two doculects.
- Tuu+: The original root for 'tree / wood' is probably to be reconstructed as *||θ̃a- or *||θ̃o-; the complexity of the click efflux would account for the variety of reflexes (sometimes real and sometimes the results of transcriptional inaccuracy) in all daughter languages other than !Xóō.

91. TWO [-]

- !Ui: *!uʔ- (|Xam !ú: ~ !'u:, ||Ng!ke !u ~ !'ú, †Kho-M !'u, N|uu !'u:, ||Kxau !'u:, ||Ku||e !'u, ||Xegwi-Z k^hyũ:, ||Xegwi-LH k'u: ~ ǿ'u:, ||Xegwi-B ||u ~ ||'u). ◊ Most of the attested reflexes would speak in favor of simply reconstructing *!u 'two' for Proto-!Ui. However, it is impossible to discount the recurrent fluctuation between simple velar and glottalized click effluxes in |Xam, ||Ng!ke, and possibly ||Xegwi as well; such correlated fluctuations are a rarity in old records and almost certainly indicate more complexity within the protoform. Provisionally, this is accounted for by the reconstruction *!uʔ- (perhaps *!uʔuʔ?) with glottalization defined on the vowel rather than on the initial click; a contracted variant of this stem could easily result in a common development to *!u in the majority of attested dialects. Note that this is almost the same situation as in the case of the numeral ONE q.v.; this would lead to suggest that, perhaps, *-ʔV- might have been some special morpheme employed in the formation of numerals (*both* numerals, as it seems likely that Proto-!Ui lacked separate lexemes for numerals higher than 'one' and 'two').
- Nossob: (a) |'Auni |am; (b) |Haasi s=||a:-ma:. ◊ The |'Auni form is a transparent borrowing from a Khoekhoe source. The |Haasi form is morphologically complex; initial s= is correctly identified by Güldemann (2002: 193) as a contracted form of the copulative element si- (as encountered in |'Auni si |am, etc.), but his proposal to identify ||a:ma(:) as a single stem may be questioned. In any case, it makes sense to suggest that |Haasi preserves the original Common Nossob stem for 'two', although it finds no etymological parallels in either !Ui or Taa.
- Taa: (?) *ǃum (|Xóõ ǃúm, Kákia ǃum ~ ǃum, N|u||en ǃum). ◊ Correspondences here are almost the same as for ONE q.v., meaning that palatal *ǃ- is the best, but not the only, bet for reconstruction (lack of transcriptional variants with ǃ- for Bleek's two doculects is puzzling).
- Tuu-: All three branches show separate equivalents for this numeral. Despite some phonetic similarity, Proto-!Ui *!uʔ- and Proto-Taa *ǃum can hardly belong together (this would be possible if !- in most !Wi reflexes could be traced back to *ǃ-, but modern N|uu and ||Xegwi unambiguously speak in favor of original *!-).

92. WALK (= GO)⁵ [!Ui + Nossob] [- Taa]

- !Ui: *||'a (|Xam ||'a(:) ~ ||'aŋ ~ ||'é, ||Ng!ke ||'a ~ ||'a: ~ ||'ai, †Kho-M ||a ~ ||'a, N|uu ||'aʔa, ||Kxau ||'a ~ ||'a: ~ ||'aŋ ~ ||'a-i, ||Xegwi-Z ||a ~ ka ~ ga). ◊ Lack of glottalic articulation in ||Xegwi-Z is somewhat puzzling, but the presence of variants ka, ga with click loss show that the word, in general, seems to be subject to irregular phonetic developments. Note that in various sources, the meaning 'go' is sometimes expressed by a different verb: |Xam tàì^f ~ tàè^f, †Kho-M tāī, N|uu (W) ʒa:^fn ~ ǿa:n, (E) ʒāī^f, ||Xegwi-Z t'āʔā-ne, ||Xegwi-LH t'aʔa ~ t'aʔan, reflecting Proto-!Ui *ta^f- with different suffixes. General analysis, however, shows that the probable meaning for this verb in all !Ui languages is really 'walk' rather than 'go'.
- Nossob: (a) |'Auni ||'a ~ ||'e ~ ||a ~ ||aa; (b) |Haasi ǃa. ◊ The |Haasi form must be an innovation, since the |'Auni form clearly belongs with !Ui. Cf. also |'Auni tāī ~ tai ~ taāī

⁵ In accordance with the traditional practice of the Moscow School of comparative linguistics and the currently accepted standards in the Global Lexicostatistical Database, Swadesh's concept of 'walk' is replaced with 'go' (due to the latter's typically higher stability across the world's languages).

‘to walk, to go’ = |Haasi *tʰá-ai* ‘to go’; external data suggest that this is really ‘to walk’ rather than ‘to go’.

- Taa: ***sa** (!Xóõ *sâ*; Kákia *ša*, N|u|en *sa* ~ *ša*).
- Tuu+: |’Auni clearly aligns itself with !Ui here (and there are no significant arguments to assume that all of the attested forms are borrowed from N|uu), as opposed to Taa, where a possible etymological parallel to !Ui-Nossob **ʰa-* could be the !Xóõ verb *ʰa-* (*ʰâe*) with the specific meaning ‘to go out hunting and/or gathering’. It is interesting to note that while Taa (or at least !Xóõ) seems to have a lexicalized opposition of **si* ‘to come’ vs. **sa* ‘to go’, !Ui and Nossob languages show free variation between these two stems, both in the meaning ‘come’; not clear if this is coincidence or correlation.

93. WARM (HOT) [-]

- !Ui: (a) |Xam *káʰo* ~ *k:au:* ~ *k:auʰ-k:áúʰ* ‘warm’; (b) †Kho-M *hã-i* ‘hot (of sun, etc.)’, †Kho-D *há:ʔi* ‘warm’, N|uu *ha:* ~ *haʰ-i* ‘warm, hot (of weather)’; (c) ||Ng!ke *ʰoná* ~ *ʰonà* ‘hot (of sun)’; (d) ||Xegwi-LH *kʰuru* ‘warm’. ◊ Nothing is properly reconstructible here on the Proto-!Ui level, not just because every language (if not every dialect) seems to have a different equivalent, but also because semantic accuracy of the glossing usually leaves a lot to be desired, with the meanings ‘warm’ and ‘hot’ hopelessly entangled with each other.
- Nossob: Not attested in either |’Auni or |Haasi.
- Taa: (a) !Xóõ *kúbi* ‘be hot, warm (e.g. sand, food, water)’; (b) Kákia *θwi* ‘hot (of sun)’; (c) N|u|en *ʰkʰu*. ◊ Not reconstructible for Proto-Taa; same problems as with !Ui.
- Tuu-: The concepts of ‘hot’ and ‘warm’ are poorly documented and generally unstable, which requires us to exclude them from comparison.

94. WATER [!Ui + Nossob + Taa]

- !Ui: ***qʰa** (|Xam *!wa:* ~ *!wá* ~ *!wã*, ||Ng!ke *ʰa:* ~ *ʰa* ~ *!à:* ~ *!a:* ~ *ʰa:*, †Kho-M *ʰa*, †Kho-D *ʰà*, N|uu *!qʰa:*, ||Xegwi-Z *kʰa:*, ||Xegwi-LH *qʰa:*, ||Xegwi-B *ʰa:*). ◊ Most of the transcriptional variation in older sources probably reflects attempts to transcribe the initial click *!qʰ-*, explicitly attested in modern N|uu and evolving with perfect regularity to *qʰ-* in ||Xegwi-LH (Ziervogel mistakenly transcribes the initial uvular as velar, whereas D. Bleek perceives it as a lateral click).
- Nossob: ***kʰa** (***qʰa**?) (|’Auni *kʰá:* ~ *kʰáá* ~ *kʰái*, |Haasi *kà*). ◊ It is important to mention another variant for |’Auni: *ʰà:a*, glossed as ‘water, rain’ (see RAIN). Since there is no serious evidence to suggest the existence of two etymologically different roots for these concepts in |’Auni, Bleek’s spelling with a lateral click, just as it does in the case of ||Xegwi, may actually reflect an initial *qʰ-*, in which case the Proto-Nossob reconstruction should be amended to **qʰa* (and a subset of uvular phonemes should be assumed for the protolanguage). Note also the peculiar transcription of the |Haasi form as *kà*, with a rare vowel that A. Traill identifies as a “clear low front vowel” (Story 1999: 15); one might wonder if this is in any way related to the supposedly uvular articulation of the preceding consonant.
- Taa: ***qʰa** (!Xóõ *!qʰà:*, Kákia *ʰá* ~ *ʰa:* ~ *!xa:*, N|u|en *ʰa*).
- Tuu+: All the forms are obviously related, and matching data from !Xóõ and modern N|uu allow to reconstruct Proto-Tuu ***qʰa** ‘water’ beyond reasonable doubt. Click loss in Nossob is quite exceptional in this case, but it must be noted that relatively few

items with */- in Proto-!Ui and Proto-Taa have reliably identifiable cognates in Nossob languages, and it cannot be stated with certainty that click loss is not regular here (particularly in conjunction with the uvular efflux).

95. WE [!Ui + Nossob + Taa]

- !Ui: (a) ***si** excl. (|Xam *s:i* ~ *s:i-s:i*, ||Ng!ke *si*, †Kho-M *si* ~ *sa*, N|uu *si*, ||Kxau *si*, ||Ku||e *si*); (b) ***i** incl. (|Xam *i*, ||Ng!ke *i*, †Kho-M *i*, N|uu *i*, ||Kxau *ʔi*, ||Ku||e *i*; ||Xegwi-Z *ʔi*, ||Xegwi-LH *ʔi-ʔe*). ◇ All “Narrow !Ui” languages show a clear-cut dichotomy between exclusive **si* and inclusive **i*; however, all attested doculects of ||Xegwi only feature *i* as the default 1st p. pl. pronoun with no regard to clusivity. External data confirm that the situation in ||Xegwi has to be treated as innovative.
- Nossob: (a) ***si** excl. (|’Auni *si* ~ *se* ~ *ci*, |Haasi *ci*); (b) ***i** incl. (|’Auni *i* ~ *e*, |Haasi *i*). ◇ For |’Auni, Bleek explicitly states the same exclusive / inclusive dichotomy as for !Ui languages. In Story’s |Haasi manuscript, the difference between *ci* and *i* is never explained, but both forms are encountered in different contexts (cf. *ci* à *k’i=θwi*: ‘we eat meat’ vs. *i* *c’au* *k’ε* ‘we milk them’), and it is highly likely that the situation here was exactly the same as in |’Auni.
- Taa: (a) ***si** excl. (Kakia *ši* ~ *šia* ~ *ša* ~ *si*, N|u||en *si* ~ *si-sa*; cf. !Xóõ *īsî* ‘we’); (b) ***i** incl. (Kakia *i*, N|u||en *i*; !Xóõ *ī^h* ‘we’ gen.). ◇ This is a rare case when data from older, generally less reliable sources come across as more important than data from Traill’s well-curated description of !Xóõ: Bleek’s Kakia and N|u||en show the same dichotomy between exclusive and inclusive pronouns as !Ui and Nossob languages, whereas Traill’s !Xóõ (the “Lone Tree” variety) shows no signs of it; instead, we find two forms, *ī^h* and *īsî*, listed as synonymous. Of these, *īsî* almost looks like a collocation of **i* + **si*, though one might reasonably doubt the chances of such an odd formation (a general ‘we’ consisting of ‘we incl.’ + ‘we excl.’?). In any case, bisyllabic *īsî* finds no parallels outside of Lone Tree !Xóõ and must be regarded as a likely innovation. Note that Maingard (1958: 106), in his own description of !Xóõ, finds exactly the same dichotomy as in the Bleek-described varieties of Taa (i.e. *i* ‘we incl.’ vs. *si* ~ *ši* ‘we excl.’).
- Tuu+: All three branches rather unequivocally suggest the necessity of reconstructing ***si** ‘we excl.’ vs. ***i** ‘we incl.’ for Proto-Tuu. This opposition is neutralized in ||Xegwi (generalizing the variant *i* for both purposes) and in one or more dialects of Taa, but remains stable everywhere else. Note that it is hard not to suspect a potential link between **si* ‘we (excl.)’ and the morphemic contrast between prefixal **sa*= ‘we (incl.)’ and **si*= ‘we (excl.)’ in Proto-Khoekhoe (Vossen 1997: 234); however, grammatical contrast between exclusive and inclusive markers is only typical of the Khoekhoe branch of Khoe, and is not formally reconstructible to the Proto-Khoe level, which would rather speak in favor of old Tuu influence on Khoekhoe than vice versa (provided this is not just a case of accidental resemblance).

96. WHAT [Nossob + Taa] [- !Ui]

- !Ui: (?) ***de**. ◇ It is not clear if **de* can be reliably reconstructed on the Proto-!Ui level specifically in the pronominal function of ‘what?’ rather than as just a general interrogative morpheme. It is in this latter function that it is explicitly encountered in |Xam: *c’a-de* ‘what?’, where *c’a* = ‘thing’, while *de* on its own is also encountered in other interrogative functions, e.g. ‘where?’, etc.

In N|uu, the situation is as follows: (a) for ||Ng!ke, Bleek lists the complex forms $d^y i$ - $si \sim gi$ - $si \sim k' i$ - $si \sim |i$ - si , all of which are hard to reconcile with each other, but at least the first one definitely goes back to $*di$ - $si \leftarrow *de$ - si with assimilation, further cognate with the general interrogative $d^y e$ ‘where?’ = |Xam de ; (b) for †Khomani, Maingard (1937: 247) only lists zi - si ‘what?’ = Bleek’s $d^y i$ - si ; (c) for modern N|uu, Collins & Namaseb (2011: 63) list two forms, ϕui (Sands quotes this as $\phi \tilde{u}i$ with nasalization) and zi - si , as synonymous. The former would seem to be an innovation and may have developed out of ϕu : ‘who?’ (see below) with additional suffixation.

Notably, for ||Kxau Meinhof (1929: 169) lists $d\epsilon$: $\sim den$ as the default equivalents for ‘what’, while the question ‘where?’ is actually expressed by the combination of this morpheme with other words, e.g. ||xa $d\epsilon$: ‘where?’. This is perhaps the strongest, if still not entirely sufficient, argument for reconstructing ‘what?’ as the original meaning for $*de$.

More problematic is the situation in ||Xegwi, where the only known form is ||Xegwi-LH $t^h \tilde{i}$: ‘what?’. Even if the final vowel is assumed to be a suffixal extension (perhaps the same $*-i$ or $*-i\eta$ as in N|uu $\phi \tilde{u}-i$?), phonetic realisation of the initial consonant as t^h rather than d - is surprising. On the other hand, we really do not know the regular reflex of Proto-!Ui $*d$ - in this language, so it is permissible to tentatively accept this stem as a genuine cognate.

- Nossob: |Haasi $^h a \sim |i$. \diamond No information on |’Auni. Vocalic fluctuation in |Haasi is not explained, but may be of the same nature as in Taa (see below).
- Taa: !Xóõ /V ... ϵ^h . \diamond A combination of the general interrogative particle /V and the 3rd person singular / Class 3 harmonic pronoun ϵ^h . For Kafia and N|u||en, no reliable data are available.
- Tuu+: On a purely formal basis, the isogloss between |Haasi $^h a \sim |i$ ‘what?’ and !Xóõ /V ... ϵ^h id. allows to reconstruct $*/V$ ‘what?’ as the optimal candidate for Proto-Tuu. It must, however, be kept in mind that in !Xóõ, this is a general interrogative particle rather than the pronoun ‘who?’ proper; admittedly, the same concern may be raised over the status of Proto-!Ui $*de$. On the whole, the subsystem of !Ui interrogatives is clearly unstable and easily lends itself to various models of restructuring.

97. WHITE [-]

- !Ui: (?) $*!ui$ (|Xam $!ú i-t\grave{a}n \sim !ú i-ta$; ||Kxau $!ui$ ‘white /of horse/’). \diamond This color term is highly unstable; most languages have their own equivalents, sometimes transparently borrowed (†Kho-M $!uri$ - ya , N|uu $!uri$ - $a \leftarrow$ Khoekhoe $!uri$ ‘white’), sometimes without any etymological connections (||Xegwi-Z $\acute{s}a$, ||Xegwi-LH $\acute{s}a$). On an interesting note, ||Ng!ke $!o:wa$ ‘white’ = |Xam $!o:wa \sim !k'o:wa$, found only in W. Bleek’s records and glossed as ‘pale’ or ‘red’ (Bleek 1956: 321, 339); if the latter glossing is not completely fortuitous, this might be the same word as Proto-Khoekhoe $*!k'aba$ ‘red’, thus, yet another areal borrowing. The only word which looks potentially archaic is |Xam $!ú i-t\grave{a}n \sim !ú i-ta$ ‘white’, further corroborated by its discovery in Meinhof’s ||Kxau records; a rather weak link, but formally acceptable.
- Nossob: |Haasi $\tilde{u}a$. \diamond Not attested in |’Auni.
- Taa: (a) !Xóõ $\tilde{u}i$ - na ‘white’, $k\hat{a}=\tilde{u}i$ - $s\grave{a}$ ‘whiteness’; (b) Kafia $||xw\acute{a}$; (c) N|u||en $!ari$. \diamond The N|u||en item is clearly the same as !Xóõ $!a^h ri$ ‘white’ in Maingard 1958: 102, phonetically and semantically glossed as $!a^h li$ ‘whitish and shiny (silver, light grey, gold)’ in Traill 1994: 75. The form in Kafia has no parallels. The bare root in !Xóõ is $\tilde{u}i$ - ($-na$ is the same

adjectival suffix that is also seen in color terms such as BLACK and RED q.v.), but its reconstructibility for Proto-Taa is uncertain without reliable external cognates.

- Tuu: Unclear. It is highly tempting to trace Proto-!Ui (more accurately, |Xam-||Kxau) **!ui* and !Xóõ *ĩú-* to the same source, but the click effluxes contradict each other; there are no signs of nasality in !Ui and no ways to explain its secondary origin in Taa. Given the overall lack of stability for this concept, it is perhaps best not to stretch the evidence here and leave all the slots empty.

98. WHO [-]

- !Ui: (?) **tu* (||Ng!ke *tú-e*, †Kho-M *ǂ^hu-xai*, N|uu *ɕu*, ||Kxau *tu*, ||Xegwi-Z *to:*, ||Xegwi-LH *towa*). ◊ In most sources on N|uu, the interrogative pronoun *tu* ~ *ɕu* is always encountered only in strict conjunction with the general interrogative morpheme *xai* ~ *xae*, which typically accompanies other types of questions as well. This probably means that, from a historical perspective, *tu* can hardly be judged as an interrogative stem, and, in fact, its phonetic equivalence with Proto-!Ui **tu* ‘men’ (see notes on MAN above) cannot be a coincidence, especially considering that the same derivational model is also found in other Khoisan languages (e.g. Ju!’hoan *ha-žoe* ‘who?’ ← *ha* ‘interr. morpheme’ + *žu* ‘person’, etc.). On the other hand, the likelihood of **tu* being used at least as a significant *part* of the interrogative formation in Proto-!Ui increases with the addition of ||Xegwi (Z) *to:*, (LH) *towa* (probably ← **tu-wa* or **tu-a* with a second component which could also go back to a general interrogative particle). In ||Kxau, according to Meinhof, ‘who?’ is simply *tu*; theoretically, this could also be a contraction or morphological simplification from an earlier complex construction in which *tu* was only the first part.

In stark contrast with this, the main equivalent for ‘who?’ in |Xam is *!u-de*, where *!u* = ‘person’ (see above) and *de* is either ‘what?’ or a general interrogative morpheme. This is a rather transparent derivation from the point of view of |Xam proper, and since it finds no correlations in other !Ui languages, it would be logical to regard it as a recent innovation.

- Nossob: (?) !’Auni *sa*, |Haasi *ci*. ◊ For !’Auni, the only actually attested form is *sa-ka* ‘whose?’ (Bleek 1937: 197), where *-ka* is the productive possessive marker. It is not even clear if the !’Auni and |Haasi forms are related, although the correspondence *s-* : *c-* is quite regular; whether !’Auni *sa-ka* is assimilated from **si-ka*, or |Haasi *ci* is the result of contraction (**sa-i* ?) remains a matter of pure speculation. In any case, both languages present evidence for some sort of **sV*-type morpheme as the principal carrier of the required meaning ‘who?’.
- Taa: !Xóõ /V ... *è^h*. ◊ According to Traill’s description, there is no difference in !Xóõ between the animate ‘who?’ and the inanimate ‘what?’. Separate forms for ‘who?’ are not attested at all in Bleek’s published records on Kakia and N|u|en; it is, therefore, impossible to make any judgements on Proto-Taa.
- Tuu-: Nothing is reconstructible; all three branches have different morphemic strategies of expressing the required meaning, and all of them are just as volatile as in the case of ‘what?’.

99. WOMAN [!Ui + Nossob] + Taa]

- !Ui: **!a-* ~ **|a-* (|Xam sg. *!a:i-ti*, pl. *!á:-gə̀n*, ||Ng!ke sg. *!ai-ti* ~ *!ai-ki* ~ *!ai-ti* ~ *!ai-ki* ~ *!e:-ki*, pl. *!a-gə̀n* ~ *!a:-gə̀n*, †Kho-M *!ai-ǂe* ~ *!ei-ǂi* ~ *!ei-ki*, †Kho-D *!ēiǂī*, N|uu sg. *!e:-ki*, pl. *!a:-ke*, ||Ku||e *!a:-ti*, ||Kxau sg. *!a-ti* ~ *!a-u*, pl. *!a:-kn*, ||Xegwi-Z *!a-zi*, ||Xegwi-B *!a:-ze*). ◊ The plural stem of

this item is actually easier to reconstruct than the singular, since most languages unambiguously agree upon **|a-* as the root morpheme, followed by a nominal marker (**|a-kən* or **|a-ke*). The singular stem, however, shows far more variety. Perhaps the most archaic situation is preserved in |Xam, where the singular form may be analyzed as *|ʼa-* (main lexical root, also seen by itself in such constructions as *!wi |ʼa* ‘girl’, lit. ‘person-woman’) + *-i* (a frequent nominal marker) + *-ti* (a rare singulative marker). If so, the paradigm has to be treated as originally suppletive, with **|ʼa-* ‘woman /sg./’ and **|a-* ‘women /pl./’ reflecting etymologically different roots whose phonetic similarity is accidental (or, if they are related, reflects some ultra-archaic phonetic alternation going beyond Proto-!Ui).

Such a clear contrast, however, is seen exclusively in |Xam, where transcriptions made by both W. Bleek and L. Lloyd consistently show *|a-* for the plural form and *|ʼai-* for singular. Already in D. Bleek’s transcriptions of ||Ng!ke we see variation between forms like *|ʼai-ti* ~ *|ʼai-ki*, on one hand (which closely match |Xam), and *|ai-ti* ~ *|ai-ki* ~ *|e:-ki*, on the other, with a voiced velar efflux instead of glottalization; the same variation is seen in Maingard’s transcriptions of †Khomani, whereas in modern N|uu the only attested variant is *|e:-ki*, with the same simple velar efflux as in the plural form (however, the old contrast is still seen in the vocalism, with sg. *|e:-ki* going back to earlier **|ai-ti*, while pl. *|a:-ke* preserves the original pl. stem **|a-*). Other !Ui languages seem to behave randomly — thus, ||Ku|e *|ʼa:-ti* is recorded with a glottalized efflux, while Meinhof’s ||Kxau *|a-ti* also shows simple velar release.

For now, the most economical explanation for this confusion is to treat the |Xam situation as original and consider all the singular forms with *|-* rather than *|ʼ-* in daughter languages as reformed by analogy with the suppletive plural. This still does not explain, e.g., the voiced release in such ||Ng!ke forms as *|ai-ki*, but the alternate solution — to provisionally set up a special click efflux in Proto-!Ui which would be reflected as this kind of alternation — would need plenty of additional evidence to be properly corroborated.

- Nossob: sg. **|ẽ*, pl. **|ã* (|ʼAuni sg. *|ẽ*, pl. *|an*, |Haasi sg. *|ĩ*, pl. *kʼa=|ã*). ◊ In light of |ʼAuni pl. *|an*, the reconstruction of the pl. form should perhaps be amended to **|a-n*, but it cannot be excluded that this transcription merely inaccurately conveys vowel nasalization. The paradigm is formally suppletive, but at least the click effluxes are always the same this time around. Comparison with !Ui evidence suggests that **|ẽ ← *|a-i(N)*. Note also the synonymous form *|e:-ki* ‘wife, woman’ in |ʼAuni — this is, in all likelihood, a borrowing from N|uu, resulting in yet another etymological “doublet” in |ʼAuni.
- Taa: sg. **|λâ-qáe*, pl. **|a^f-N* (!Xóó sg. *tâ:=qáe*, pl. *|ã^f*, Kafia sg. *la=kai* ~ *la=kái*, pl. *la=ke* ~ *la=ké*, N|u|en sg. *|an*, pl. *|ãĩ*). ◊ The original situation is probably best preserved in !Xóó, which features a suppletive paradigm: a special lexical root for the meaning is found in the plural (*|ã^f*), while sg. ‘woman’ is expressed with a compound formation (*tâ:* ‘person’ + *qáe* ‘mother / female’). Exactly the same situation is observed, for instance, in the †Hūa dialect as described in Westphal 1965: 139: sg. *la-qʼaé* vs. pl. *ɲ|ãʔa*. Other dialects of Taa, however, prefer to generalize the paradigm one way or the other — thus, Kafia reforms the plural form based on the singular, similar to (though not quite the same as) N|amani sg. *ta-qʼaé*, pl. *ta-qʼaé-tu* (also quoted from Westphal 1965: 139); in N|u|en, on the other hand, it is the plural stem that seemingly becomes generalized⁶.

⁶ Additionally, cf. more recent material collected by Christfried Naumann (2011: 21): West !Xóó sg. *sí=|ã^fn* ‘woman’, pl. *sí=|ã^fã* vs. !Ama sg. *táá=qáe* ‘woman’, pl. *sí=|ã^fã* ‘women’ (with the stem further enlarged with the frequent Taa nominal prefix *si=*).

- Tuu+: Oddly enough, the plural stem of this paradigm is more reliably reconstructible for Proto-Tuu than the singular — !Ui */a-, Nossob */ã and Taa */a^ɛ-N all belong together, although the absolute lack of pharyngealization in !Ui is somewhat puzzling. The singular stem, on the other hand, is either transparently innovative (Taa *λâ-qâe ‘mother-person’), derivable from the plural via morphology (Nossob */ē ← */ã-i), or questionable (does Proto-!Ui */ai-/ti/ historically contain the same root as */a-, or is this true suppletivism at work?). Importantly, Nossob languages clearly align themselves closer to !Ui than to Taa in this case; the alternation *e ~ i* in sg. vs. *a* in pl. perfectly correlates with the alternation *ai* sg. vs. *a* pl. in !Ui, reflecting either a shared !Ui-Nossob innovation or an archaic morphological model, lost in all Taa idioms.

100. YELLOW [-]

- This word is excluded from comparison, since it is very rarely attested, and most of the actual attestations are semantically ambiguous (e.g. ||Ng!ke |ã^hla; |Haasi !a; !Xóõ has at least three possible equivalents — †q^húí ‘yellow, Naples yellow, yellow ochre, very light blue-grey, light green-grey’, †ái^{sh} ‘yellow, chrome yellow, cadmium yellow, gold coloured’, †á?u ‘yellow, chrome orange, terra cotta’ — none of which have reliable external parallels).

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.
- Campbell, Lyle, William J. Poser. 2008. *Language Classification: History and Method*. Cambridge University Press.
- Collins, Chris, Levi Namaseb. 2011. *A Grammatical Sketch of N|uuki with Stories*. Köln: Rüdiger Köppe Verlag.
- Doke, Clement M. 1936. An outline of †Khomani Bushman phonetics. *Bantu studies* 10: 433–461.
- Güldemann, Tom. 2002. Using older Khoisan sources: quantifier expressions in Lower Nosop varieties of Tuu. *South African Journal of African Languages* 2002(3): 187–196.
- Güldemann, Tom. 2005. 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. 2018. Historical linguistics and genealogical language classification in Africa. In: Tom Güldemann (ed.). *African Languages and Linguistics*: 58–444. Berlin: DeGruyter Mouton.
- Kassian, Alexei. 2015. Towards a Formal Genealogical Classification of the Lezgian Languages (North Caucasus): Testing Various Phylogenetic Methods on Lexical Data. *PLoS ONE* 10(2): e0116950.
- Kassian, Alexei, George Starostin, Anna Dybo, Vasiliy Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.
- 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 †Khomani 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.
- Maingard, Louis F. 1958. Three Bushman Languages. Part II: The Third Bushman Language. *African studies (Johannesburg)* 17: 100–115.
- 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 N|uu clicks. *Journal of the International Phonetic Association* 39(2): 129–161.
- Naumann, Christfried. 2011. *A preliminary classification of Taa dialects*. Presentation given at: 20th International Conference on Historical Linguistics (ICHLXX).
- 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, Amanda Miller, Johanna Brugman. 2007. The Lexicon in Language Attrition: The Case of N|uu. 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. 2015. *Annotated Swadesh wordlists for the !Wi group (Peripheral Khoisan family)*. Ms. Available online at: starling.rinet.ru/new100/kwi.pdf.
- Starostin, George. 2021. Lexicostatistical Studies in Khoisan II/1: How to make a Swadesh wordlist for Proto-Tuu. *Journal of Language Relationship* 19/1: 99–135.
- 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. Kojanskije jazyki*. Moscow: Jazyki slav'anskoj kul'tury.
- Starostin, Sergei. 2000. Comparative-historical linguistics and lexicostatistics. In: Colin Renfrew, April McMahon, Larry Trask (eds.). *Time depth in historical linguistics. Vol. 1*: 223–265. Cambridge: The McDonald Institute for Archaeological Research.
- Story, Robert. 1999. *K'u|'asi manuscript (MS collections of the Ki|'azi dialect of Bushman, 1937)*. Ed. by Anthony Traill. *Khoisan Forum Working Paper No. 13*. Köln: University of Cologne.
- Traill, Anthony. 1994. *!Xóõ dictionary. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 9*. Köln: Rüdiger Köppe Verlag.
- Traill, Anthony, Hiroshi Nakagawa. 2000. A historical !Xóõ-|Gui contact zone: linguistic and other relations. In: Herman M. Batibo, Joe Tsonope (eds.). *The state of Khoesan languages in Botswana*: 1–17. Gaborone: Basarwa Languages Project.
- Vossen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe Verlag.
- Westphal, Ernst O. J. 1965. Linguistic research in SWA and Angola. In: *Die ethnischen Gruppen Südwestafrikas (Wissenschaftliche Forschung in Südwestafrika, Bd 3)*: 125–144. Windhoek: Südwestafrikanische Wissenschaftliche Gesellschaft.
- 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/2: к уточнению филогенетической классификации языков семьи ту

Статья представляет собой вторую часть обширного исследования, основные цели которого — уточнение внутренней классификации южноафриканской языковой семьи ту (= южнокойсанской) и реконструкция максимально достоверного 100-словного списка Сводеша для пра-ту языкового состояния. Первый раздел статьи посвящен проблеме

ареальных контактов между двумя языками, относящимися к разным ветвям ту (nɪŋɪ и ʔauni), и отсеиванию вероятных заимствований из первого во второй для повышения точности результатов лексикостатистических подсчетов. За этим следуют собственно лексикостатистические матрицы и основанные на них схемы классификации для всей семьи, наглядно свидетельствующие в пользу исходно тернарного (а не бинарного) распада семьи на три ветви (!ви, носсоб и та); аргументы, ранее приводимые в поддержку более близких связей ветвей носсоб и та, признаются недостаточными. Статья также содержит ряд наблюдений над результатами реконструкции стословника для пра-ту состояния и отмечает некоторые любопытные особенности, отличающие лексикон пра-ту от его современных потомков. В Приложении приводятся подробные комментарии относительно возможности реконструкции на промежуточных уровнях (пра-!ви, пра-носсоб, пра-та) и на уровне пра-ту конкретных элементов из второй половины списка Сводеша.

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

Studies in Yukaghir etymology I

This paper presents a number of additions and corrections to the corpus of etymologies published in Irina Nikolaeva's "A Historical Dictionary of Yukaghir" (De Gruyter, 2006). The focus of the paper is on internal Yukaghir etymology rather than on search for loanwords or long-range cognates.

Keywords: Yukaghir languages; etymology; historical linguistics; protolanguage reconstruction.

0. Introduction

The appearance of "A Historical Dictionary of Yukaghir" (HDY) by Irina Nikolaeva represented a significant breakthrough in the study of Yukaghir etymology and in the reconstruction of Proto-Yukaghir. However, despite the comprehensiveness of this dictionary, much remains to be done in the field of Yukaghir etymology.

First, the etymological analysis in HDY is not always as deep as one would want it to be. For example, the Proto-Yukaghir reconstruction **eluč-* (HDY: 157) is based on the Tundra Yukaghir form *elut'ora-* 'to pass by', given under the label "TK" (Tundra Yukaghir form from Krejnovič 1958 or Krejnovič 1982). Unfortunately, HDY never provides the exact reference to the page (or sometimes even to the work) where the cited form is taken from. In this case, the word in question is attested in the phrase <НЭМЭН ИЛИЈЕНОЛЛЭЛК ЭЛУТ'ОРАҢАН ТИТҚАНЭ> (Krejnovič 1982: 278), translated as 'Какой бы то ни было ветер пусть минует вас' (Krejnovič 1982: 283), i.e. 'let any wind pass you by'. The form *elut'orayan* contains the negative prefix *el-* and the jussive ending *-yan*. The same form *el-ut'ora-*, now with the correctly segmented negative prefix, is given once again under the Proto-Yukaghir reconstruction **u:čə-* (HDY: 440). The verb stem *ut'ora-* is identical¹ to T *uučuoore-* 'drop in on somebody on one's way', documented by Kurilov (2001). This verb is derived from T *uu-* 'to go somewhere; to float downstream', which is cited in HDY under the Proto-Yukaghir reconstruction **uj-* (HDY: 441-442).

Second, the approach to Proto-Yukaghir reconstruction taken in HDY is not based on the Neogrammarian notion of *Ausnahmslosigkeit der Lautgesetze*. Frequent references to the influence of neighboring consonants on the development of vowels are made *ad hoc*, without assuming that the changes in question are in any way regular. As was already noted by Elena Maslova in her review of HDY, "it remains unclear in which phonetic contexts the hypothesized sound changes are supposed to have happened. At best, Nikolaeva observes some tendencies (like the presence of labials in the immediate context as a correlate of *PY *o* > T *a*), but they are not supposed to account for all relevant instances of changes and retentions" (Maslova 2008: 258).

Third, the philological treatment of the Old Yukaghir material is rather superficial (Knüppel 2013). Old Yukaghir forms are given without their original glosses after their modern cognates, with the implication that they have the same meaning. For example, MU <andschúb> is given after K *ažu:bə* 'conversation' (HDY: 103), which might mean that the MU form means 'conversation' as well. But on p. 73 the same unglossed MU form is compared to K *ažu:* 'word, language'. Finally, in the Swadesh list of Old Yukaghir varieties, MU <andschúb> is given un-

¹ Vowel length is rendered fairly inconsistently in Krejnovič's works.

der ‘tongue’ (Nikolaeva 2008: 326). Schiefner (1871: 375) gives E (= Ermitage-Manuscript, one of the copies of MU wordlist) <andschub> ‘Zunge’, with the note “wohl richtiger *Sprache*”, so apparently the gloss ‘tongue’ must be correct.

The recent years saw the appearance of a number of works on Yukaghir etymology (Aikio 2014; Blažek forthcoming; Fortescue & Vajda 2022: 45–142; Häkkinen 2012; Knüppel 2010, 2014; Napol’skikh 2020; Piispanen 2018, 2019a, 2019b, 2020a, 2020b, 2020c, 2022). However, almost all of these works are dedicated to external etymology, i.e. search for loanwords from other language families in Yukaghir languages or attempts to find evidence supporting long-range relationship between Yukaghir and other families, mainly Uralic. These are legitimate pursuits, but their success depends on the thoroughness of internal analysis of Yukaghir data. I am convinced that using underanalyzed data in external etymologizing is the main source of errors in both loanword research and long-range comparison. I hope to demonstrate in the present paper that the resources of Yukaghir-internal etymology are far from exhausted.

Proto-Yukaghir forms in this paper are reconstructed following my revised version of Nikolaeva’s Yukaghir reconstruction (Zhivlov 2022). Unless noted otherwise, Kolyma Yukaghir forms are cited after Prokop’eva & Prokop’eva (2021) and Tundra Yukaghir forms are cited after Kurilov (2001).

1. K *čadile* ~ *čad’ile* ‘path’

This word is a cognate of T *čarile* ‘path; summer path; line, notch’. The Proto-Yukaghir form can be reconstructed as **čadilə*, with the regular shift *-d- > -r- in Tundra Yukaghir. In view of the Tundra cognate, the Kolyma variant with -d’- must be secondary. In HDY (121) the Kolyma word is treated as isolated, while the Tundra word is conflated with T *čaril’e* ‘wrinkle’ (a related word with a different suffix) and apparently unrelated T *sorile* ‘wrinkle’ in the following way: “T *čaril’e*, *sorile*, *čarile* wrinkle; summer path; line, notch” (HDY: 414). The resulting mix is compared to K *šorile* ‘picture; pattern; color’ and K *šere-* ‘to embroider’ (tr.) with the remark that “[t]he correspondence K -ö-/e- ~ T -a- is irregular” (ibid.). The reconstructed Proto-Yukaghir **sörö-/sere-* is then compared to Tungusic **sere-* ‘to embroider’. The Kolyma verb *šere-* ‘to embroider’ (tr.) indeed must be a Tungusic loanword, but there is no reason to connect it with other words above, given the divergent semantics and unclear vowel correspondences.

2. T *čald’e(ŋ)* ‘hand’

This word is attested in two Old Yukaghir wordlists²: MU <tolónscha>, MK <tolónscha, tolondschén-> (HDY: 434). The change *o > a and the syncope in trisyllabic stems are recurrent

² The position of these two wordlists with respect to modern Kolyma and Tundra Yukaghir languages is unclear. While most Old Yukaghir wordlists (except, of course, sources on Chuvan and Omok) reflect idioms close to modern Kolyma Yukaghir, both MU and MK combine a number of words otherwise attested only in Tundra Yukaghir with characteristically Kolyma lexemes. This situation can be explained in two ways: either these lists reflect special lects intermediate between modern Kolyma and Tundra Yukaghir (Nikolaeva 2008), or a peculiar mix of Kolyma and Tundra words results from borrowing. The second option can be supported with the following observation: while Kolyma Yukaghir, Tundra Yukaghir, Omok, and Chuvan all have sufficiently distinct systems of numerals, reflecting their long independent history, the numeric systems of MU and MK are practically identical with the Kolyma numeral system. Anyway, these wordlists are a unique source of Old Tundra Yukaghir lexicon, even if Tundra-like forms in them are borrowings.

changes in Tundra Yukaghir as compared to 19th century wordlists (although their exact conditions are yet to be determined), cf. T *al'γa(η)* ‘fish’, MU <ólloga>, MK <oljogá> (HDY: 325). However, the correspondence of Old Yukaghir *t-* to modern Tundra Yukaghir *č-* is irregular. Nikolaeva reconstructs **tolon-/*čolon-*, noting that “[e]ither the MU and MK records are erroneous and the root should be reconstructed with the initial affricate (**čolo-*), or in T the first consonant was assimilated to the third syllable affricate: **tolon-č'ə > *tolod'ə > čal'd'ə*”. The second option can be definitely confirmed when we compare T *im-dald'al-* ‘to be five’. The first part of this compound is also found in Kolyma Yukaghir *in'γańboi*³ <I'n-gan·boi> ‘five’ (Jochelson 1905: 113) and in Chuvan: ChM *emgangan* <ЕМГАНГОНЬ> ‘five’, *emganbagu* <ЕМГАНБАГУ> — the first part of numerals from six to nine, ChB *imo-qánbo-tudole* ‘six’ (HDY: 293). The second part of the Kolyma and Chuvan words for ‘five’ is K *qańbe* ‘palm’, ChM *qańba* <ХАНЬБА> ‘hand’ — the semantic equivalent of T *čald'e(η)* ‘hand’. Another compound preserving the dental stop in the root in question is T *čama-dald'e* ‘tsar’, literally “big hand”⁴. Cf. words with original initial *č*, which changes to *d'* in compounding: T *čajle* ‘day; light’, T *kind'e-d'ajle* ‘moonlight’ (T *kind'e* ‘moon’). Thus, we can reconstruct **tolońžə* as the original form of T *čald'e(η)*.

Now, **tolońžə* looks like a lexicalized active participle (Schmalz 2013: 145; Nikolaeva 2020: 83) from the verbal stem **tolo-* or **tolon-*. The only formally suitable candidate is the verbal root **tol-* (HDY: 433): T *tolie-* ‘to support, to prop up’ (tr.), T *tolej-* ‘to support oneself (by leaning on something)’ (tr.), T *tolii(η)* ‘khorei (a pole for driving reindeer); staff’. While I do not know any typological parallels for the derivation of the word for ‘hand’ from the verb ‘to support, to prop up’, such a semantic connection at least does not look impossible.

3. T *enuníe-* ‘to be of poor quality, thin (of deerskin)’ (intr.)

This verb, together with its derivative T *enuníije* ‘part of the skin, unsuitable (because of its thinness) for the manufacture of clothing and other household items’, is treated in HDY (162) as etymologically isolated. The etymology of these words is quite transparent: they are derived from T *nuníe-* ‘to be thick (of deerskin)’ (intr.), T *nuníije* ‘thick part of the skin’ with the negative prefix *e-*. The problem is that this prefix is non-productive and its very existence is ignored in HDY. The regular negative prefix (or proclitic) in Yukaghir is K *el-*, T *el-* (HDY: 155-156). Other examples of the negative prefix *e-* include K *ejuuke* ‘near’, T *ejuoke* ‘near’ (cf. K *juuke* ‘far’, T *jöke* ‘far’, T *juuke* ‘far, far away’), K *ejedulben* ‘invisible, unclean spirit, devil’ (cf. K *jed-/jen-* ‘to be seen’). One more case, which additionally involves a regressive assimilation of the prefix vowel, is T *araw* ‘naked’, T *araawje* ‘bald spot (e.g., on a deerskin)’, T *arawjaa* ‘bald’, T *arawre-* ‘to become cloudless; to become bald’, T *arawíe-* ‘to have a bare surface without vegetation (of the tundra); to have a plucked, shabby spot’, K *aruóie-* ‘to be naked; to be bald’, K *arouje ~ arojo* ‘lenok’ (cf. T *sawa(η)* ‘hide, skin’; the change of root-initial *s* to *r* is regular in compounding). According to Nikolaeva, “[i]n forms such as *a-ro:jə, a-raw* etc. the consonant *-l-* has fallen out: **əl-sawa(-) > al-sawa(-) > al-rawa(-) > araw, aro:-*” (HDY: 156). This scenario is quite unlikely: since the form without the expected *-l-* is found before different consonants, there is no way to explain the absence of *-l-* by a regular sound change. For example, the regular reflex

³ In modern K *ńayanbuo-* ‘to be five’ *in-* was replaced by the reciprocal prefix *ńa-*.

⁴ This etymology of the Tundra word for ‘tsar’ was proposed by Jochelson (1926: 39). In HDY (138) this word is analyzed as *čamad-ald'e* ‘tsar’ [lit. ‘big chief’]. This analysis is worse than Jochelson’s because it is based on K *and'e ~ and'e ~ anid'e* ‘boss; leader’, KJ *alid'e, anid'e* ‘princeling, head, chieftain’ (HDY: 108), a word not attested in Tundra Yukaghir, which has different words for ‘chief, boss’: *lem'l'e* and *moojče*. I thank an anonymous reviewer for this observation.

of the combination *-l- + -j-* is *-ll'-*, cf. K *jaquj-* ‘to be sufficient’ (intr.), K *ell'aquj-* ‘to lack’ (intr.) (see also Maslova 2003: 43). Therefore, as was pointed out by Krejnovič (1982: 99–100), forms like T *ejuoke* must contain the negative prefix *e-*. All attested cases with this prefix look like archaic fossilized formations. Apparently, at an earlier stage the negative prefix had two allomorphs, **e-* before consonants and **el-* before vowels. Later the allomorph **el-* ousted **e-* in all cases save some fossilized forms. The consonant *-l-* in **el-* was possibly epenthetic, since epenthetic *-l-* is used in Yukaghir morphophonology in order to avoid vowel clusters (Maslova 2003: 56–57).

Finally, T *nunńe-* ‘to be thick (of deerskin)’ (intr.) is itself analyzable, as *-ńe-* is a productive suffix forming denominal verbs with the meaning ‘to have X’ (Krejnovič 1982: 55). The verb in question must have been derived from the nominal root **nun-* ‘thickness’, not attested elsewhere.

4. K *mejnu-* ‘to take’

The new Kolyma Yukaghir dictionary (Prokop'eva & Prokop'eva 2021) points out that this verb is an imperfective form of K *mid'-/min-* ‘to take’. This relationship remained unnoticed in HDY, where K *mejnu-* is given under a separate root **mej-*⁵ (HDY: 262), distinct from PY **menž-/minž-* > K *mid'-/min-* ‘to take’, T *med'-/meń-* ‘to take’ (HDY: 264). The verb *mejnu-* contains the imperfective suffix *-nu-* (Maslova 2003: 182–188). If we added this suffix to the stem of the verb ‘to take’, we could expect ***minnu-* or ***mennu-*⁶, but such a form is not attested in Kolyma Yukaghir. In fact, K *mejnu-* is one of a number of verbal stems where we find *-jn-* instead of expected *-nn-* as a result of a previously unnoticed Proto-Yukaghir morphophonological rule, whereby **-nn-* becomes **-jn-*.

One such case is K *tojnu·m* ‘drive (a herd)’ (Maslova 2003: 555), derived with the imperfective suffix *-nu-* from K *tono·m* ‘drive deer together’ (ibid.). The Kolyma Yukaghir dictionary (Prokop'eva & Prokop'eva 2021) lists only K *tono-* ‘to drive’.

Another case is K *ojnu-* ‘to take away, to steal (many times)’, derived with the same imperfective suffix *-nu-* from K *ono-* ‘to take away, to steal’ (both forms given after HDY: 333). In the Kolyma Yukaghir dictionary of Prokop'eva & Prokop'eva (2021) the forms from this root are given as *ono-* ~ *ojno-* ‘to take away’ (tr.).

Finally, K *pojne-* ‘to be white’ can be compared with K *pod'oγo-* ‘to shine, to glitter’ (intr.), T *pod'aya-* ‘to shine’ (intr.) from the Proto-Yukaghir root **pońž(ə)-* (HDY: 358; see Zhivlov 2022: 50 for this reconstruction). The combination of this root with the qualitative suffix *-ne-* (Maslova 2003: 560) must have given **ponne-*. Instead, we find K *pojne-*. Here an additional complication is the Tundra Yukaghir cognate *pojine-* ‘to be whitish, white; to be faded, whitened (from the spring sun, air)’ (intr.). This form reflects the generalization of the epenthetic vowel *-i-* before the qualitative suffix *-ne-*. This change must have happened when the relationship between root allomorphs *pod'-* and *poj-* ceased to be perceived as regular. As a result, other derivatives were formed from the allomorph *poj-*, e.g., T *pojičeń-* ‘to be faded, whitish (of skin brightened in the spring sun, air)’ (intr.). Thus, synchronically in Tundra Yukaghir *poj-* became a separate root, whereas historically it was an allomorph of *pod'-*.

The rule **-nn-* > *-jn-* is no longer productive in modern Yukaghir languages: cf. K *monnu-*, imperfective of K *mon-* ‘to speak’ and T *mennu-*, imperfective of T *med'-/meń-* ‘to take’.

⁵ HDY also lists Tundra Yukaghir *mei-* ‘to take’ under the label “TJ” (Tundra Yukaghir form from Jochelson 1900). I was unable to find this form in Jochelson (1900) and I suspect that it is a ghost-word.

⁶ Given the irregular correspondence between K and T vowels, it is not clear which vowel to expect here.

The proposed etymology of K *mejnu*- ‘to take’ is incompatible with the hypothesis that Proto-Yukaghir **mej*- ‘to take’ was borrowed from pre-Proto-Samoyed **mɛ(x/j)*- or Proto-Samoyed **mɛ*- ‘to take’ (Aikio 2014: 71).

5. T *ńall'e(ŋ)* ‘sin’

This word⁷ is viewed as etymologically isolated in HDY (287), but a plausible etymology can in fact be proposed. The Proto-Yukaghir root **ńań*- ‘to be sinful’ (HDY: 288) has a wide range of derivatives in Yukaghir languages: T *ńańi*- ‘to consider smth. sinful; to treat smth. with superstitious fear’ (tr.); T *ńańyii*- ‘to consider smth. sinful’ (tr.); T *ńańuu*- ‘to be sinful (of an act)’ (intr.); K *ńańiś*- ‘to be sinful’ (intr.); K *ńand'e* ‘sin’, T *ńand'e* ‘a person with whom one is in a relationship of "mutual avoidance" – sons-in-law, daughters-in-law, brothers, sisters; mistake, oversight’; K *ńančuon* ‘sin’; K *ńańulben* ‘sinner; devil’. T *ńall'e(ŋ)* ‘sin’ is derived from this root with the non-productive nominal suffix *-l'e* (Kurilov 1994: 91-93). The development **ńań-l'ə* > T *ńall'e(ŋ)* is regular. In Kolyma Yukaghir, “nasals /n/, /n'/ alternate with the lateral /l/ when the next syllable starts with /l/ or /l'” (Maslova 2003: 42, see also HDY: 35). The same assimilation is attested in Tundra Yukaghir (Schmalz 2013: 57). Therefore, there is no need to reconstruct a separate root **ńall'ə*.

6. K *ńuu* ‘name’

This word is attested in Kolyma Yukaghir and in Chuvan: K *ńuu* ‘name’, K <*ni'u*> ‘name’ (Jochelson 1926: 324); S <*nɨw*>, R <*niw*>, B <*neve*> ‘Name’ (Schiefner 1871: 383); ME <*niiv*> (HDY: 312); W <*nim*> (Witsen 1705: 687); ChM *nywa* <Ныва> ‘name’ (Matjuškin 1841: 117).

All Old Kolyma data, except Witsen’s <*nim*> and B <*neve*>, point to the form *niw*, which subsequently yielded modern Kolyma *ńuu*. The older form was still preserved in Jochelson’s times: Jochelson’s transcription <*ni'u*> must be read as *niw*, since Jochelson lists <*iu*> among the diphthongs (1926: 315), and the stress mark after *i* makes clear that it is a falling diphthong *iw*. As for B <*neve*>, both the quality of the first vowel and the presence of the second vowel are aberrant.

Chuvan *nywa* <Ныва> has a final vowel, which is not unexpected, given that many Chuvan words have a final vowel where their Kolyma and Tundra cognates have none. Consider the following cases⁸: K *anil* ‘fish’ ~ ChM *anyla* <Аныла> ‘fish’, K *čuul* ‘meat’, T *čuul* ‘meat’ ~ ChM *čulo* <Чуло> ‘meat’, K *ludul* ‘iron’ ~ ChM *lundylo* <Лундыло> ‘iron’, K *met* ‘I’, T *met* ‘I’ ~ ChM *mota* <Мота> ‘I’, K *tet* ‘thou’, T *tet* ‘thou’ ~ ChM *tota* <Тота> ‘thou’. It is not yet clear whether we deal here with the disappearance of a Proto-Yukaghir final vowel in Kolyma and Tundra languages, or rather with a process of epithesis in Chuvan.

The most problematic form is W <*nim*>, attested in the first recorded specimen of a Yukaghir language – the Yukaghir translation of the Lord’s prayer, published in Nicolaes Witsen’s book⁹ “Noord en Oost Tartarye” (Witsen 1705: 687). According to Nikolaeva, the language of this text “was very close, if not identical, to Kolyma Yukaghir” (HDY: 12). Chrono-

⁷ See Nemirovskij (2017) for a detailed discussion of its meaning.

⁸ I list here modern Kolyma forms, but their Old Kolyma counterparts, when these are attested, also lack final vowel.

⁹ The book was first published in 1692. I cite it after the second edition (1705).

logically, this form precedes all other attestations of the word in question, and the consonant <m> makes it especially close to Proto-Uralic *nimi ‘name’ and its Proto-Samoyed reflex *nim, with which the Yukaghir word is traditionally compared (UEW: 305; Aikio 2014: 72). It is usually assumed that /w/ in later forms is a secondary development from original /m/. Nikolaeva suggests two alternative scenarios: “The following phonetic developments can be reconstructed: *nime > niwe > niw > *ńu*: or *nime > nim > niw > *ńu*.” (HDY: 312). The first scenario assumes the development *m* > *w* in intervocalic position, while in the second scenario this development happens word-finally¹⁰. One obvious problem with these scenarios is that the development in question must have happened independently in Kolyma Yukaghir and Chuvan, since both Kolyma and Chuvan have /w/ in the word for ‘name’. The main problem, however, is the irregular nature of the postulated development, since Proto-Yukaghir *m is regularly preserved in daughter languages, both word-medially and word-finally.

Consider the following examples with intervocalic PY *-m-, preserved both in Kolyma and Chuvan: K *ayime* ‘friend’ ~ ChM *agamo* <Агамо> ‘friend’, K *amun* ‘bone’ ~ ChM *amun* <Амунъ> ‘bone’, K *emej* ‘mother’ ~ ChM *eme* <Еме> ‘mother’, K *emil* ‘night; darkness’ ~ ChM *emilo* <Емило> ‘night’, K *nume* ‘house’ ~ ChM *naumagi* <Наумаги> ‘jurt’, K *omós-* ‘to be good’ ~ ChM *omóč* <Омочъ> ‘well’, K *unume* ‘ears’ ~ ChM *inama* <Инама> ‘ears’. These examples show that intervocalic *-m- is regularly preserved both in Kolyma Yukaghir and in Chuvan, so the first scenario above is impossible.

Word-final *-m is reconstructed in two morphemes: PY *-m > K -m, T -m 3rd singular ending of the transitive neutral paradigm, and PY *-m > K -m, T -m 1st singular interrogative ending (HDY: 81; Nikolaeva 2020: 64, 71). Nikolaeva (2020: 66) postulates *-mә as an internal reconstruction of the former ending, based on the comparison with other endings of the transitive paradigm, which contain the original transitive marker *-mә. 3rd singular ending of the transitive paradigm would then consist of the transitive marker *-mә and zero 3rd singular ending. While this internal reconstruction is reasonable, there is no evidence that the word-final vowel was still present in this ending in the last common ancestor of Kolyma and Tundra Yukaghir. Thus, the second scenario above cannot be correct: there was no regular change of word-final *-m to -w in Kolyma Yukaghir. The only remaining possibility is that the form <nim> in Witsen’s text results from some kind of error, which is not surprising given the history of the text. “Witsen himself did not visit the Yukaghir land. He received the Yukaghir materials by post sometime after his visit to Moscow in 1664-1665, probably through his cousin Andrea Vinius, who later headed the Siberian Chancellery in Moscow... It is essentially a word-to-word translation and in some respects preserves the Russian syntax, for example, by postposing a possessive pronoun, which is not typical of Yukaghir. The text is written in Dutch orthography” (HDY: 11). Thus, the text of the Yukaghir Lord’s prayer was probably originally written in Cyrillic and later retranscribed to Dutch orthography. Errors in transcription could have crept in at any stage. If <m> in Witsen’s text results from such an error, the original form of the word in question can be reconstructed as *niw or (less likely) *niwe¹¹.

This form has a previously unnoticed internal Yukaghir etymology: it can be derived from the Proto-Yukaghir verbal root *ńee- ~ *ńi- ~ *ńii- ~ *ni- ‘to call, to name’ (HDY: 292). This root is attested in the following forms: K *ńie-* ‘to call; to invite’ (tr.) (Prokop’eva & Prokop’eva 2021), ‘to call (by name)’ (Nikolaeva & Shalugin 2002), T *ńie-* ‘to call, to invite; to send for; to ask; to ask for a girl’s hand in marriage’ (tr.); K *ńite-* ‘to give a name; to call by name’ (tr.); K *ńiide-* ‘to say;

¹⁰ The Proto-Yukaghir reconstruction *ńu:, given as a headword in HDY, apparently results from an oversight, as it blatantly contradicts both Old Yukaghir data and the scenarios cited above.

¹¹ Or even *nib, given that word-internally, *w and *b were not opposed in Proto-Yukaghir (Zhivlov 2022: 44).

to call by name; to announce; to discuss; to utter’ (tr.); T *nide-* ‘to deliver a speech, to pronounce a word; to say someone’s name’ (tr.), T *nide-* ‘to call by name; to pronounce a word’ (tr.). The relationship between the four attested allomorphs **ńee-*, **ńii-*, **ńi-*, and **ni-* is not yet clear.

Semantically, the connection is unproblematic. Formally, the word **niw* ‘name’ contains the allomorph **ni-*. The suffix can be identified with the suffix *-Vw > K *-u-*, T *-u-* (Nikolaeva 2020: 88), which forms deverbal nouns denoting the result of action (Kurilov 1994: 56-61).

The reconstruction **niw* and the proposed internal Yukaghir etymology make the comparison with Proto-Uralic rather unlikely. Proto-Uralic **nimi* ‘name’ can be compared with the verbal Proto-Yukaghir root **ńee-* ~ **ni-* ‘to call, to name’ only if Proto-Uralic *-m- corresponds to Proto-Yukaghir zero. The hypothesis of borrowing from Proto-Samoyed can be definitely rejected.

7. K *šašqul* ~ *šarqul* ‘claws; fingers; paw; hand’, T *sisqa* ‘finger; paw’

Nikolaeva reconstructs PY **syrq-* and compares this protoform with Proto-Tungusic **siru* ‘span between the big finger and index finger’, noting that “[i]n Yukaghir *-q-* can be a derivational suffix” (HDY: 422–423). Given that the Tundra form and one of the Kolyma forms point to **c* (**s* in Nikolaeva’s reconstruction) as a second consonant, the Proto-Yukaghir form must rather be reconstructed as **cickə*. The Kolyma form has added a suffix *-(u)l*, and the *-r-* in one of the Kolyma variants apparently results from dissimilation. The PY noun **cickə* can be derived from the root **cica-* with the suffix *-*rkə*.

The verbal root **cica-*¹² is attested in derivatives such as K *šašayaj-* ‘to be torn’ (intr.), T *sisayaj-* ‘to be torn; to be split up’ (intr.); K *šašand’i-* ‘to tear’ (intr.); T *sisad’i-* ‘to split, to crack, to tear’ (intr.). The suffix *-*rkə* derives from qualitative verbs names of objects or concepts possessing the quality in question (Kurilov 1994: 43-49; Schmalz 2013: 106), e.g., T *jatarqa* ‘a straight part of something (e.g., of a road)’, T *jatayaj-* ‘to become straight’. This suffix may cause an optional syncope of the root-final vowel, such that we have parallel derivatives from the same root with and without syncope, e.g., K *pońqo* ‘light; candle’ and K *pod’orqo* ‘day’, T *pod’arqa* ‘whiteness, brilliance’ from the root of K *pod’oyo-* ‘to shine, to glitter’, T *pod’aya-* ‘to shine’. When syncope takes place, the resulting cluster is simplified so that in clusters of three or more consonants, only the first and the last consonants are preserved, thus accounting for the lack of *-r-* in derivatives like K *pońqo* or T *sisqa* (see Zhivlov 2022: 49-51). Indeed, side by side with T *sisqa* ‘finger; paw’ a parallel derivative without syncope is attested: T *sisarqa* ‘crack’. It is possible that in this and similar cases, the derivative with syncope is diachronically older.

Finally, the semantic side of this etymology has a partial parallel in such cases as Tundra Nenets *tarka* ‘shoot, branch, offshoot, fork’, *ńudah tarka* ‘finger’ (Tereščenko 1965: 634) or Yakut *tarbax* ‘finger’ from Proto-Turkic **darmaq* ‘branch, fork’ (Tenišev 2001: 255-256).

8. K *ugurče* ‘kamus ski (ski lined with fur)’, T *ugurče(η)* ‘foot; paw; sledge stanchion; footwear’

These words look like participles in *-*čə* (Nikolaeva 2020: 83) from a verbal stem. The deriving stem can be identified as **egur-* ‘to walk’, cf. K *egurte-* ~ *eyurte-* ‘to wander, to walk’ and T *egurie-* ‘to start walking, regain the ability to walk; to start visiting someone’. The same stem with syncopated internal vowel is found in K *ejre-* ‘to walk’, T *ewre-* ‘to go somewhere and return; to

¹² Reconstructed in HDY (423) as **sya-*.

walk, wander; to drop in, to visit; to live, to exist', both from PY *egr̥-. The vowel assimilation *e-u > u-u is regular in Kolyma Yukaghir, cf. K *unmut* 'horn, antler' ~ T *enmur* 'horn, antler', K *unuŋ* 'river' ~ T *enu(ŋ)* 'river'. The absence of this assimilation in K *egurte-* may be due to the analogical influence of related forms like K *ejre-*. The vowel *u* instead of expected *e* in the first syllable of T *ugurče(ŋ)* is unexpected (but see below). However, the Proto-Yukaghir reconstruction *egurč̥ə is supported by the Old Yukaghir forms R <ägurtera> 'Espe' and B <agurtshira> 'id.' (Schiefner 1871: 376), which correspond to modern Kolyma *ugurče-raa* 'poplar', literally "ski tree". Another interesting Old Yukaghir form is MU <egirtscha> (HDY: 320), whose exact gloss is unfortunately unclear because of the way Old Yukaghir data are presented in HDY.

HDY places K *ugurče* 'kamus ski' and T *ugurče(ŋ)* 'foot; paw; sledge stanchion; footwear' in the same entry with K *ugurčie* 'grayling (Thymallus)', T *ugurčie* 'id.'. This comparison does not look plausible from the semantic point of view. Another etymology can be proposed for the name of grayling: it can be derived from K *ugur* 'fish backbone; roughness, knob'. The semantic motivation here is provided by the characteristic feature of the grayling — its large and conspicuous dorsal fin. The denominal suffix *-čie* is unproductive, but cf. T *jengurčie* 'mottled reindeer' from T *jengur* 'something multicolored, variegated'.

K *ugur* 'fish backbone; roughness, knob' itself is listed as an isolated word in HDY (320), but a Tundra Yukaghir cognate can be proposed: T *egur* 'withers' (also listed as an isolated word in HDY: 151). The vowel correspondence is regular (see above) and the Proto-Yukaghir form can be reconstructed as *e(ŋ)gur¹³. The vowel *u* in the first syllable of Tundra Yukaghir words *ugurčie* and *ugurče(ŋ)* may be taken as evidence that Tundra Yukaghir also had an assimilation *e-u > u-u, but with a more limited scope than in Kolyma Yukaghir. Perhaps this assimilation applied only to trisyllabic forms.

Finally, Proto-Yukaghir *e(ŋ)gur seems to be related to the following words: T *egil* 'back side of something', T *egiil* '1) back of the head; 2) back part of a hat; 3) butt of an axe', T *egii* 'a fairly wide space extending immediately behind the hills that stretch from south to north not far from the river, on its eastern side; the east side (from the point of view of someone who is located west of the river); backwards'. These words are given in HDY under the reconstruction *eŋk- (HDY: 161). The original meaning of the Proto-Yukaghir root *e(ŋ)g- was apparently 'back side'.

9. T *waruluu* 'root'

This word is compared in HDY with K *ožuu* 'thin root, used for sewing'. The comparison looks impeccable: both T *war-* and K *ož-* regularly go back to PY *wonč-, and the meaning of the compared words is nearly identical (the basic Kolyma word for 'root' is *larqul*). Morphologically, however, this comparison is far from perfect. There is no suffix *-luu* or *-uluu* in Tundra Yukaghir. The only suitable suffix is *-uu*, but this is a deverbal suffix forming nouns denoting the result of action (Kurilov 1994: 56-61). The deriving stem then should be a verbal stem *warul-. Although such a stem does not exist in pure form, it forms a base of further verbal derivatives T *warulmu-* 'to become strong (e.g., of rope, thread)' (intr.) and T *warulwe-* 'to become strong (e.g., of rope, thread); to get better' (intr.). The derivation of T *waruluu* from *warul- was already suggested by Kurilov (1994: 57), but this etymology was not mentioned in HDY.

The stem *warul- is itself derived from T *war-* 'to be firm, strong' (intr.), related to K *adi* 'firm, strong' < Proto-Yukaghir *wad- (HDY: 449-450). Thus, the comparison of T *waruluu*

¹³ Without Old Yukaghir evidence we cannot know whether to reconstruct *g or *ŋg.

‘root’ with K *ožu* ‘thin root, used for sewing’ is erroneous. The latter word so far looks etymologically isolated, but the presence of a deverbal suffix *-uu* suggests that here, too the deriving root must be verbal. The traditional comparison of these Yukaghir words with Proto-Uralic **wanča(w)* ‘root’ (whether as Uralo-Yukaghir cognates or as Samoyed loans in Yukaghir) is definitely wrong with respect to T *waruluu* and rather unlikely with respect to K *ožu*.

Language Abbreviations

B	Old Kolyma Yukaghir materials of Billings (Schiefner 1871)
ChB	Chuvan materials of Boensing (HDY)
ChM	Chuvan materials of Matjuškin (Matjuškin 1841)
K	Kolyma Yukaghir (Prokop'eva & Prokop'eva 2021)
ME	Old Kolyma Yukaghir materials of Merk (HDY)
MK	Old Kolyma Yukaghir materials of Mueller/Lindenau (HDY)
MU	Old Ust'-Jansk Yukaghir materials of Mueller/Lindenau (HDY)
PY	Proto-Yukaghir
R	Old Kolyma Yukaghir materials of Rajskej (Schiefner 1871)
S	Old Kolyma Yukaghir materials of Suvorov (Schiefner 1871)
T	Tundra Yukaghir (Kurilov 2001)
W	Old Kolyma Yukaghir translation of Lord's prayer (Witsen 1705)

References

- Aikio, Ante [Luobbal Sámmol Sámmol Ánte]. 2014. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62: 7–76.
- Blažek, Václav. Forthcoming. Yukaghir-Uralic comparison: Kinship and social terminology. In: Aleksandr Anikin, Valentin Gusev, Anna Urmanchieva (eds.). *Siberica et Uralica: In memoriam Eugen Helimski*.
- Fortescue, Michael, Edward Vajda. 2022. *Mid-Holocene Language Connections between Asia and North America*. (Brill's *Studies in the Indigenous Languages of the Americas* 17.) Leiden / Boston: Brill.
- Häkkinen, Jaakko. 2012. Early contacts between Uralic and Yukaghir. In: Tiina Hyytiäinen, Lotta Jalava, Janne Saarikivi, and Erika Sandman (eds.). *Per Urales ad Orientem: Iter polyphonicum multilingue. Festschrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012 (Mémoires de la Société Finno-Ougrienne 264)*: 227–250. Helsinki: Société Finno-Ougrienne.
- HDY = Nikolaeva, Irina. 2006. *A Historical Dictionary of Yukaghir*. (Trends in Linguistics, Documentation 25.) Berlin: Mouton de Gruyter.
- Jochelson, Waldemar. 1900. *Materialy po izučeníju jukagirskago jazyka i fol'klora*. I. (Trudy jakutskoj èkspedicii, snarjažennoj na sredstva I. M. Sibirjakova. Otděl III. Tom IX. Čast' III.) Sankt-Peterburg: Imperatorskaja Akademija nauk.
- Jochelson, Waldemar. 1905. Essay on the Grammar of the Yukaghir Language. *Annals of the New York Academy of Sciences* XVI/5, II: 97–152.
- Jochelson, Waldemar. 1926. *The Yukaghir and the Yukaghirized Tungus*. (Mémorial of the American Museum of Natural History IX.) New York: G. E. Stechert.
- Knüppel, Michael. 2010. Zu jukagirisch *tokit* ~ *tokkił* „Pfeil”. *Acta Orientalia* 71: 115–124.
- Knüppel, Michael. 2013. Noch einmal zum historischen Wörterbuch des Jukagirischen. *Rocznik Orientalistyczny* LXVI (2): 109–116.
- Knüppel, Michael. 2014. Jukagirisch-tungusische Lehnbeziehungen. *Studia Etymologica Cracoviensia* 19: 83–93.
- Krejnovič, Eruxim A. 1958. *Jukagirskij jazyk*. Moskva: Izdatel'stvo Akademii nauk SSSR.
- Krejnovič, Eruxim A. 1982. *Issledovanija i materialy po jukagirskom jazyku*. Leningrad: Nauka.
- Kurilov, Gavril N. 1994. *Obrazovanie imennyx slov v jukagirskom jazyke*. Yakutsk: Yakutskij naučnyj centr SO RAN.
- Kurilov, Gavril N. 2001. *Jukagirsko-russkij slovar'*. Novosibirsk: Nauka.
- Maslova, Elena. 2003. *A Grammar of Kolyma Yukaghir*. (Mouton Grammar Library 27.) Berlin: Mouton de Gruyter.

- Maslova, Elena. 2008. [Review of] Irina Nikolaeva. A historical dictionary of Yukaghir. *Studies in Language* 32(1): 254–264.
- Matjuškin, Fedor F. 1841. Sobranie slov Čuvanskago i Omokskago jazykov, sostavlennoe Mičmanom Matjuškinym. In: *Priboavlenija k Putešestviju po severnym beregam Sibiri i po Ledovitomu morju, soveršennomu v 1820, 1821, 1822, 1823 i 1824 g. ekspedicieju, sostojavšeju pod načal'stvom flota lejtenanta Ferdinandona fon-Vrangelja*: 115–125. Sankt-Peterburg: Imperatorskaja Akademija nauk.
- Napol'skikh, Vladimir. 2020. Jukagirisch und Uralisch: Was bleibt übrig? In: Hans-Hermann Bartens, Lars-Gunnar Larsson, Katja Mattsson, Judit Molnár, Tiina Savolainen (eds.). *Kiel joug om šild. Festschrift zum 65. Geburtstag von Eberhard Winkler (Veröffentlichungen der Societas Uralo-Altica 94)*: 269–289. Wiesbaden.
- Nemirovskij, Aleksandr A. 2017. Konceptija n'all'è i voinskij ètos jukagirov (po severojukagirskomu èposu ob Èdilvèe). In: A. N. Alekseev (ed.). *Fol'klor paleoaziatskix narodov: materialy II Meždunarodnoj naučnoj konferencii, g. Jakutsk, 21-25 nojabrja 2016 g.*: 320–328. Jakutsk: RIO media-xoldinga.
- Nikolaeva, Irina. 2008. Chuvan and Omok languages? In: Alexander Lubotsky, Jos Schaecken, Jeroen Wiedenhof (eds.). *Evidence and Counter-Evidence. Essays in Honor of Frederik Kortlandt. Volume 2: General Linguistics (Studies in Slavic and General Linguistics)*: 313–336. Amsterdam: Rodopi.
- Nikolaeva, Irina. 2020. *Yukaghir Morphology in a Historical and Comparative Perspective. (LINCOM Studies in Asian Linguistics 92.)* München: LINCOM GmbH.
- Nikolaeva, Irina A., Vasilij G. Šalugin. *Slovar' jukagirsko-russkij i russko-jukagirskij (Verxnekolymskij dialekt)*. Sankt-Peterburg: Drofa.
- Piispanen, Peter Sauli. 2018. Additional Turkic and Tungusic borrowings into Yukaghir. *Turkic Languages* 22: 108–138.
- Piispanen, Peter Sauli. 2019a. Additional Turkic and Tungusic borrowings into Yukaghir II. *Journal of Old Turkic Studies* 3/1: 54–82.
- Piispanen, Peter Sauli. 2019b. Additional Turkic and Tungusic borrowings into Yukaghir III. *Journal of Old Turkic Studies* 3/2: 321–371.
- Piispanen, Peter Sauli. 2020a. Additional Turkic and Tungusic borrowings into Yukaghir IV. *Journal of Old Turkic Studies* 4/1: 152–187.
- Piispanen, Peter Sauli. 2020b. Additional Turkic and Tungusic borrowings into Yukaghir V. *Journal of Old Turkic Studies* 4/2: 469–498.
- Piispanen, Peter Sauli. 2020c. Some new Late Proto-Yukaghir reconstructions with added thoughts and considerations on various etymologies. *Ural-Altai Studies / Урал-алтайские исследования* 38: 81–91.
- Piispanen, Peter Sauli. 2022. Additional Turkic and Tungusic borrowings into Yukaghir VI. *Journal of Old Turkic Studies* 6/1: 85–110.
- Prokop'eva, Praskov'ja E., Aleksandra E. Prokop'eva. 2021. *Jukagirsko-russkij slovar' (jazyk lesnyx jukagirov)*. Novosibirsk: Nauka.
- Schiefner, Anton. 1871. Beiträge zur Kenntniss der jukagirischen Sprache. *Bulletin de l'Académie Impériale des Sciences de Saint-Petersbourg* XVI: 373–399.
- Schmalz, Mark. 2013. *Aspects of the Grammar of Tundra Yukaghir*. PhD dissertation, University of Amsterdam.
- Tenišev, Èdxjam R. (ed.). 2001. *Sravnitel'no-istoričeskaja grammatika tjurkskix jazykov. Leksika*. Moskva: Nauka.
- Tereščenko Natalija M. 1965. *Nenecko-russkij slovar'*. Moskva: Sovetskaja ènciklopedija.
- UEW = Rédei, Károly. 1988–1991. *Uralisches Etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.
- Witsen, Nicolaes. 1705. *Noord en Oost Tartarye; Tweede Deel: Behelzende de Landschappen Georgia, Mengrelia, Cirkassia, Crim, Astakkia, Altin, Tingoesia, Siberia, en Samojedia*. Amsterdam: François Halma.
- Zhivlov, Mikhail. 2022. Pre-Proto-Yukaghir Consonant Clusters. *International Journal of Eurasian Linguistics* 4: 41–59.

М. А. Живлов. Исследования в области юкагирской этимологии I

В настоящей статье предлагается ряд дополнений и уточнений к корпусу этимологий, опубликованному И. А. Николаевой в «A Historical Dictionary of Yukaghir» (De Gruyter, 2006). Основной упор сделан не столько на поиске заимствований или когнатов в рамках дальнего сравнения, сколько на собственно внутриякагирских этимологиях.

Ключевые слова: юкагирские языки; этимология; сравнительно-историческое языкознание; праязыковая реконструкция.