George Starostin
National Research University Higher School of Economics, Moscow / Santa Fe Institute, New Mexico; gstarst1@gmail.com

Lexicostatistical studies in Khoisan III/I: Reconstructing a Swadesh wordlist for Proto-Khoe (items 1–25)[1]

This paper is the first part of a comprehensive study whose goal is to revise and expand upon the author's previously published lexicostatistical analysis of the basic lexicon for languages belonging to the Khoe family of South Africa. Covering more than 20 different lects, the paper gives a brief introduction into the relevant issues of the internal genetic classification of Khoe languages, following it up with a detailed etymological analysis of the first 25 items taken from the "ultra-stable" sub-section of the Swadesh wordlist. An interesting preliminary conclusion is that Khoekhoe, one of the two primary branches of the family, seems to be consistently more innovative in its basic lexicon than Kalahari Khoe (Non-Khoekhoe), the other primary branch.

Keywords: Khoisan languages; Khoe languages; lexicostatistical analysis; onomasiological reconstruction; historical semantics.

Introduction

In several previous publications (Starostin 2018, 2021, 2022) I have attempted to take on the matter of reconstructing standard Swadesh wordlists for two of the commonly and uncontroversially accepted phylogenetic units within the so-called Khoisan area, Ju (= "North Khoisan" in the older terminology) and Tuu (= "South Khoisan"), with the ultimate goal of simplifying the challenge to demonstrate their genetic relationship on deeper chronological levels by filtering out as many late-period lexical innovations as possible. It is only logical that the present paper — like Starostin 2021, the first section of a planned multi-part study — should now turn to the third and last major (i.e. consisting of more than one language) taxon within the same linguistic area, namely, Khoe (or "Central Khoisan" within the traditional taxonomy, as the term was used by, e.g., Dorothea Bleek and Joseph Greenberg).

Compared to its two other "Khoisan" neighbors, the Khoe family holds several distinct advantages when it comes to the diachronic stage of analysis of its linguistic data. First, it is relatively large (*Glottolog* 5.1 currently recognizes 13 distinct languages, but the figure could easily be enlarged if more distinctive criteria are applied toward the differentiation of their alleged dialects), and, what is even more important, most of these languages, with the exception of several extinct Khoekhoe lects, are still living, which ensures the slow, but steady publication of new data on their phonology, grammar, and lexicon. This is different from the situation in both the Ju cluster (consisting of, at best, 2–3 different, but still very close, languages) and the Tuu group (where the overall number of known languages may be comparable to the same

[1] The present work was conducted as part of the research project "From antiquity to modernity" (National University Higher School of Economics, 2024). I am also extremely grateful to Chris Collins for providing me with some of the newer materials on Khoe languages, and to all the other specialists who have made their data available in published form or online, making this comparative analysis possible.

numbers for Khoe, but most of them have been extinct for decades, with incomplete and inaccurate records as the only source data for comparative analysis).

On the other hand, serious diachronic study of the Khoe language family also presents its own challenges which require different strategies and solutions from those to which we are accustomed when dealing with Ju or Tuu languages. First, Khoe-speaking populations on the whole show a larger degree of cultural and anthropological diversity than their San neighbors, with a particularly sharp divide between the Khoekhoe and Kalahari Khoe speakers (see e.g. Güldemann 2006); these distinctions, although not having a direct impact on the state of genetic relationship between the different branches, nevertheless inevitably affect their lexicon as well as other layers of the language. Second, the family as a whole demonstrates specific typological features which oppose it to both Ju and Taa, such as a comparatively smaller inventory of click-type phonemes, the category of grammatical gender, a well-developed agglutinative verbal morphology, and a predominantly SOV word order (as opposed to SVO in Ju and Taa) — in other words, it looks somewhat "typologically mixed", with such obviously "Khoisan" features as the presence of clicks and monosyllabic morphemic structure contrasting with the ones listed above (Güldemann 2013). This points to a complex history of the genesis of the Khoe family, as well as some potentially unusual scenarios of its disintegration into the daughter lineages we see today.

It could be argued that more diachronic research has been conducted on Khoe languages to this day than on any other "Khoisan" group — chiefly through the meticulous fieldwork and subsequent historical analysis by Rainer Vossen (1997), though his research had both predecessors (e.g. Baucom 1974, a pioneering study that has largely been made obsolete by Vossen's research) and followers (such as Derek Elderkin and Henry Honken). Despite this, multiple questions remain open about the phonemic and lexical inventory of the reconstructed Proto-Khoe, its internal structure, the timeline of its disintegration into modern lineages, and, of course, the exact nature of its relationship with other linguistic families of South Africa (or perhaps even the African continent as a whole). Being primarily interested in this latter issue (though acknowledging that it can hardly be properly resolved without systematically taking all the others into consideration), I strongly believe in the necessity of first producing a robust qualitative and quantitative basis for any attempts at solving it, in which the quantitative part should be centered around lexicostatistical analysis and the qualitative part should involve etymological analysis of any evidence submitted to formal computational procedures.

The first (and only) comprehensive lexicostastical survey of Khoe data was carried out by Vossen (Vossen et al. 1988a, 1988b) based largely on his own field data; fortunately, most of it was published along with the results and can still be used today. The lexicostatistical comparison, based on a slightly modified version of the Swadesh list, resulted in a phylogenetic classification whose final variant can be found in Vossen 1997 and which, for a long time, has served as a base reference point for learning about the internal structure of the Khoe family (for instance, it is still listed as the chief reference source in *Glottolog 5.1*). A typically "cautious" subclassification of known languages, as presented in Vossen 2013: 10–11, looks as follows:

(a) KHOEKHOE: Namibian Khoekhoe (Nama/Damara), !(G)Ora, Hai‖om;
(b) KXOE: Khwe, ‖Ani, Buga;
(c) NARO;
(d) ‖GANA: ‖Gana, |Gui, ǂHaba;
(e) SHUA: Ts'ixa, Danisi, Deti, Cara, |Xaise;
(f) TSHWA: Cua, Kua, Tsua.

Furthermore, in Vossen 1997: 386 additional topological nodes are suggested:

— clusters (b)–(f) all form part of a primary "Non-Khoekhoe" (today, more commonly called "Kalahari Khoe") branch, opposed to "Khoekhoe";
— clusters (b)-(d) are opposed, as a "Western Non-Khoekhoe" taxonomic entity, to the "Eastern Non-Khoekhoe" grouping of (e)-(f), based on a number of substantial lexical and phonetic distinctions;
— finally, within the Western cluster Vossen proposes to link (c) and (d) together as part of an intermediate "Naro-‖Gana" grouping.

Of these three classification proposals, only the demarcation line between Khoekhoe (KK) and Kalahari Khoe = Non-Khoekhoe (NKK) is today seen as more or less incontestable. Existence of a separate "Eastern" and "Western" Non-Khoekhoe remains on more flimsy ground (particularly "Western"), as does the "Naro-‖Gana" cluster; additionally, based on evidence presented by H. Nakagawa (2011), the poorly documented ǂHaba language is now commonly thought to be most closely related to Naro (within a "Naro-ǂHaba" cluster) rather than |Gui and ‖Gana. A special problem concerns the status of Ts'ixa, which, albeit decisively classified as a Shua language by Vossen, is known to also a share a number of strong isoglosses with the Kxoe cluster; this has even resulted in *Glottolog* adopting a cautious position and listing Ts'ixa as a separate third branch of NKK, positioned as a "bridge" of sorts between the Western and Eastern languages.

Starostin (2013) presented the results of a new lexicostatistical classification for Khoe languages, largely based on Vossen's published data (with the introduction of a few additional sources) and on the application of the "Morris Swadesh / Sergei Starostin" formula (see Starostin 2000 for details). The results, for the most part, matched Vossen's, but without a clear confirmation of the "East–West" dichotomy within NKK, while also offering support for Nakagawa's hypothesis of an intermediate "Naro-ǂHaba" node. The study itself focused, however, not so much on statistics as on the reconstruction of a Swadesh-type proto-wordlist (for 50 lexical items) for Proto-Khoe, to be used as next stage data in a general assessment of the degrees of lexical proximity between various "Khoisan" taxa.

More than ten years later, it seems reasonable to offer a revised and expanded version of the same research for scholarly consideration, expanding the dataset to include a complete 100-item (or, rather, 110-item, as currently used in the Global Lexicostatistical Database[2]) wordlist as well as a number of new data collections that remained unpublished or unavailable to the author back in the early 2010s (additionally, the earlier study was only published in Russian and thus remains inaccessible to various members of the international scholarly community). Just as it was in Starostin 2013, the emphasis will be placed not so much on pure percentages of matches between different languages, but rather on the onomasiological reconstruction of the respective Swadesh items for various intermediate levels and on tracing the lexical history of these items, in particular, on attempts to distinguish the lexical and semantic innovations that took place in between Proto-Khoe and its descendants from archaisms retained by those descendants.

Given the sheer size of the family and the complexity of its genetic and areal history, the revised study will be split into several parts. The present paper gives a brief overview of the data and some condensed notes on the current state of Proto-Khoe reconstruction and on the methodology of analysis, followed by a more specific discussion of the first half of the "ultra-stable" 50-item part of the Swadesh wordlist. The second part will tackle the next 25 items,

[2] The Global Lexicostatistical Database (GLD) is an open repository of high quality lexicostatistical wordlists for a variety of the world's languages; among other things, it also includes curated 110-item wordlists for Khoe languages which serve as the basis for this publication (Starostin 2017, 2020).

along with intermediate lexicostatistical conclusions based on the analysis. Lastly, I plan to expand the study to cover the rest of the 110-item wordlist so as to show whether this expansion of the analysis to the less stable part of the vocabulary results in any important modifications of previous conclusions on Khoe classification and dating.

The data

As is the case with all the families of the "Khoisan" area, there is no direct correlation between the numbers and names of formally distinguished Khoe "languages" (as listed, e.g., in *Glottolog* or any general overviews such as Vossen 2013) and the actual lects that may be represented by Swadesh-type wordlists. Some of the recognized languages still do not have sufficient amounts of published lexical data to qualify for inclusion into any lexicostatistical survey; conversely, others may be represented by several closely related dialects, each of which does present such data, so that it would not make sense to intentionally restrict ourselves to arbitrarily choosing just one of such dialects.

For this study, I rely on the following principal guidelines:

(a) all Khoe lects represented by at least a small (but acceptable for lexicostatistical purposes) amount of published lexical data should be grouped into a number of "mini-clusters" — small bunches of languages and/or dialects that are always uncontroversially and justifiedly grouped together based on overwhelming phonetic, lexical, and grammatical similarities;

(b) these "mini-clusters" should cover all the recognized minor sub-divisions of Khoe and include at least 1–2 lects from each so as to make the classification truly representative;

(c) each "mini-cluster" should be available for lexicostatistical analysis in two forms — as actual attested data from all of its members, and also in the condensed form of an intermediate reconstruction for its supposed ancestral state (provided such a reconstruction is possible). Such an arrangement will permit us to conduct lexicostatistical comparison on different chronological levels, any discrepancies between which may have useful implications for disentangling genetic links from areal ones.

Below I list all the eight "mini-clusters", put together for this study, along with the main sources of lexicographic data on each one (for a larger overview of the pre-2013 literature available on these lects, see Starostin 2013: 392–398).

I. **Khoekhoe** (**KK**): consistently recognized as one of the two primary branches of the Khoe family. Three lects available for lexicostatistical comparison:

[1] Nama (N; primary source on the "standardized" language — Haacke & Eiseb 2002; the older dictionary of Rust 1969 is sometimes used for additional details on specific word usage; Haacke et al. 1997 is a useful go-to source to investigate occasional dialectal discrepancies between the different varieties of Nama, which are quite rare in the basic lexicon);

[2] !Ora (K; the single most comprehensive source is Meinhof 1930, closely followed by the somewhat smaller and phonetically less reliable Wuras 1920; Du Plessis 2018 is a new important compendium which, among other things, includes fresh lexical data collected from the few remaining native speakers of the language);

[3] Xri (Gri, Xiri, Griqua; not included into Starostin 2013 for lack of data, but Haacke & Snyman 2019, a new publication that includes most of Jan Snyman's fieldwork data on this particular lect, makes it possible both to better ascertain the position of Xri relative to both Nama and !Ora and to use its data to strengthen various aspects of Proto-Khoekhoe reconstruction).

II. **Non-Khoekhoe** (**NKK**) (= Kalahari Khoe; we use the term Non-Khoekhoe for technical purposes, so as to avoid confusion between similar abbreviations for Khoekhoe and Kalahari Khoe):

II.1. **Khwe**: a very clearly distinguished cluster, centered around the Khwe language proper. Four lects available for lexicostatistical comparison:

[4] ǁAni, [5] |Ganda, [6] Buga. Main sources of lexical data on all these varieties are still Vossen 1997 and Vossen et al. 1988b (also Vossen 1986 for ǁAni). |Ganda is not generally recognized even as a separate dialect in its own rights in any later sources, but is still included for the sake of thoroughness;

[7] Khwe. Replacing Oswin Köhler's earlier data as the major source on this language (specifically, the ǁXom dialectal variety) is Kilian-Hatz 2003 (one of the most comprehensive published dictionaries on any Non-Khoekhoe languages).

II.2. **Naro-ǂHaba**: as established by H. Nakagawa, these two languages clearly belong together in their own "mini-cluster". The main sources are:

[8] Naro — Visser 2001, although earlier (less phonetically accurate) data from Barnard 1985 are very useful to observe dialectal discrepancies, as are fieldwork results from Vossen 1997;

[9] ǂHaba — systematic published data still only found in Vossen et al. 1988b, Vossen 1997.

II.3. **|Gwi-ǁGana**: these two languages are always recognized as very close to each other.

[10] |Gwi, [11] ǁGana: the main sources are Vossen et al. 1988b, Vossen 1997, with the dictionary Tanaka 1978 for both languages used as an auxiliary source (very poor quality of data transcription, but useful as a source of additional verification for Vossen's data).

II.4. **Shua**: inclusion of (at least) four separate linguistic varieties is uncontested:

[12] Cara, [13] |Xaise, [14] Deti, [15] Danisi. Data on these lects come almost exclusively from Vossen et al. 1988b, Vossen 1997. According to *Glottolog*, they all represent varieties of the same language, but there are clearly at least a few dialectal discrepancies between them, even within the basic lexicon.

II.5. **Ts'ixa**. This language was originally included by Vossen into the Shua cluster, but in light of its "mixed" nature (see above), it seemed prudent for the aims of this particular study to provisionally separate it into its own "mini-cluster".

The main source on data for [16] Ts'ixa remains Vossen et al. 1988 + Vossen 1997; however, more recent publications by Anne-Maria Fehn (most notably Fehn 2014) have also been considered so as to add more depth to the perspective.

II.6. **Tsua**: in Vossen 1997, three distinct varieties belonging to this cluster are given as Cua, Kua, and Tsua, but systematic data is only listed for [17] Kua and [18] Tsua. Additionally:

[19] Tjwao (Tshwao) is a recently described variety of the same cluster, very close to both Kua and Tsua, but lexically somewhat distinct from both. We use the comparative wordlist published in Phiri 2021 as the basis, along with a few other publications on the same variety;

[20] Hie (Hietshware) is another variety known only from the older description by Samuel Dornan (1917). Despite being predictably unreliable by way of transcription, adding a separate wordlist for Hie (clearly belonging to the Tsua cluster) is especially useful as it is one of the few Non-Khoekhoe varieties suitable for lexicostatistical analysis with records from more than 100 years ago.

II.7. **Cua**. Vossen's data (1997) includes a few samples of a lect he calls Cua (the samples are quite scarce compared to his data on Kua and Tsua, insufficient to produce even a Swadesh wordlist); however, they are very notably different from the idiom spoken in Diphuhudu (Kweneng district of Botswana) as described in Collins & Wellstood forthc. In an earlier

sociolinguistic paper (Chebanne & Dlali 2020), A. Chebanne classifies this Cua, along with such other idioms as Tshoa and Cire-Cire, into a separate "Central" cluster of Eastern Kalahari Khoe, in between Shua (Northern) and Kua (Southern).

Even a preliminary analysis of lexical data on Cua (Wellstood 2024) shows that it has very significant discrepancies from the Tsua cluster; additionally, such notable features as, e.g., its consistent preservation of the palatal click (affricativization of PNKK *ǂ- is one of the most important shared innovations within all of Eastern Kalahari Khoe, with the notable partial exception of Ts'ixa), strongly dictate the necessity of putting *this* particular Cua (*not* to be confused with the seriously different Cua in Vossen 1997!) into its own separate "mini-cluster", at least for the purposes of this study.

The methods

As stated above, the goal of the present study is to present a historical (classificational) model of the Khoe family not merely by comparing percentages of pairwise matches between potential etymological cognates, but by investigating the individual histories of all the words on the Swadesh list, distinguishing between retained archaisms and diffused innovations with the aid of historical (phonetic, semantic, distributional, and etymological) analysis. In addition to the necessity of looking at the relevant data from the "micro"-point of view (treating each micro-cluster as its own separate entity), this requires constant appeals to the following areas of knowledge:

1. **Historical phonology**. Our general understanding of the inventory and diachronic evolution of the Proto-Khoe (PK) phonological system mostly comes from Vossen 1997; the reconstruction presented in that monograph, carried out in general agreement with the classic Neogrammarian model, still remains relevant and has so far not been supplanted by any radically newer proposals. A few minor additions were proposed in Starostin 2013, the most significant of which is the idea of distinguishing between PK *e : *ε and *o : *$ɔ$, introducing a possible +/-ATR opposition (based on a seemingly unconditional split of vocalic reflexes in Khoekhoe, where NKK *e, *o sometimes correspond to KK *e, *o and sometimes to *a). These and any other serious differences from Vossen's reconstructions will be commented upon in the data section, whenever necessary.

More recent research on the systems of click consonants in Khoe has occasionally shown results that are not at all reflected in either Vossen's synchronic descriptions of the languages or his historical projections (see, e.g., Nakagawa 1996 on click accompaniments in |Gui and their correspondences in ǁGana, with at least one additional phonological opposition not noted by Vossen); taken together with the occasional notable irregularities or fluctuations in the correspondences postulated in his monograph, this hints at possibilities for further amendments to the click system as it is presented, e.g, in Vossen's summarizing table (Vossen 1997: 319). However, since a systematic revision of PK reconstruction falls beyond the scope of this series of papers, I shall only address such issues occasionally, over the course of discussing specific etymological connections within the data collection itself.

Significant advances have also been seen in recent years in the field of comparative Khoe tonology, mainly due to works by Derek Elderkin (2004, 2008), Henry Honken (2008) and others. Comparative-historical prosody is always a very intricate field, with Khoe languages being no exception to the general rule, so I shall largely abstain here from discussing peculiarities of tonal development, sporadically marking tonal reconstructions for intermediate stages when they are more or less trivial. I generally concur with Elderkin on adopting a simple bitonal model for PK — L and H — but as far as its current importance for Khoe etymology is

concerned at present, I would say that tonal matching in the scope of this study is only truly vital for several cases when a potential etymological match between KK and NKK presents doubts from the semantic, or, more rarely, segmental phonological point of view, in which case prosodic arguments can, at least in theory, strengthen or weaken the etymology.

Overall, if no phonetic commentary accompanies the presented data along with the reconstruction, this means that the correspondences are generally regular and follow the basic model as outlined in Vossen 1997. Irregular discrepancies are usually addressed and pointed out.

2. **Historical semantics**. The presence of both intermediate stage and top level reconstructions in lexicostatistical lists necessitates the application of semantic reconstruction, as one has to select the correct or, at least, the optimal candidate (root morpheme or stem) for the expression of a particular basic, neutral meaning on the respective proto-levels. Construction of such a scenario is sometimes possible merely by a systemic investigation of the corresponding semantic field in the languages concerned, but at other times "external" sources of knowledge — containing, for instance, typological data on polysemies or typical directions of semantic shifts — are quite useful in making an unbiased, empirically justified decision. In such situations, I occasionally consult such resources as CLICS (data on colexifications), DatSemShift (manually curated data on polysemies and semantic shifts around the world), and CODASELF (a large comparative database on basic lexicon reconstruction across Eurasia). Additionally, I also rely on my own expertise in diachronically-oriented research on Khoisan languages (including such other lineages as Ju and Tuu), given that actual Khoisan data are severely underrepresented or even completely absent from the abovementioned resources.

3. **External connections of the Khoe family**. Speaking in genetic terms, the only currently recognized genetic "outlier" in relation to Khoe is the extinct language Kwadi, data on which are scarce, but sufficient to strongly suggest that Kwadi does really share a common linguistic ancestor with Khoe (currently called Khoe-Kwadi in scientific literature), while not being itself an integral part of the Khoe family as such. A fuller etymological and lexicostatistical investigation of Kwadi's relationship with Khoe lies outside the scope of this paper (see Fehn & Rocha 2024 for the current state of affairs), but for now it is perfectly justifiable to occasionally resort to parallels from Kwadi, for instance, in situations where an observed binary split between KK and NKK requires external support to understand which lexeme is more archaic (see, e.g., the case of BONE below).

On the much more complicated issue of a possible genetic connection between Khoe and the "San" groupings it is preferable to remain agnostic, although it is undeniable that areal contacts between certain Khoe languages and various lects belonging to the Ju or Tuu families have indeed taken place. It is important to systematically check the data of individual languages for potential borrowings; however, in those cases when a PNKK or PCK reconstruction shows exceptionally close phonetic similarity to a Ju or Tuu form (see, e.g, NAIL or FIRE below), I refrain from automatically labeling it as a borrowing into the proto-language, since there is always a chance that it could in reality reflect a deeper genetic connection between the families. (In any case, the issue of potential borrowings into Proto-Central Khoisan, as opposed to borrowings into the individual branches after its disintegration, is a separate topic that has no bearing on the question of the internal classification and history of the family).

Note on transcription

For the most part, the transcriptional system adopted in this paper follows the current standards of the Global Lexicostatistical Database (GLD), which, in their turn, are largely based on

standard IPA conventions. The only major changes involve consistent transcription of the palatal glide as *y* (instead of *j*) and modifications to the affricate system (IPA *ts*, *dz* = GLD *c*, *ʒ*; IPA *tʃ*, *dʒ* = GLD *č*, *ǯ*; IPA *tɕ*, *dʑ* = GLD *ɕ*, *ʒ)*. In transcribing clicks, we also adopt Rainer Vossen's convention of representing clicks with voiced effluxes with an underline tilde (*ǀ̲*, *ǂ̲*, etc.), while clicks with nasal effluxes get a superscript tilde (*ǀ̃*, *ǁ̃*, etc.).

Some of the newer sources on Khoe languages use more complicated notation to transcribe both click and non-click phonemes, concentrating more on the actual phonetic values of the sounds rather than their phonological status. In order to provide more comparative transparency and avoid confusion for readers who are not so well versed in the specific intricacies of Khoe phonetics, we have simplified those transcriptions: thus, e.g., the exact same click that is transcribed as *ǂx* (velar fricative efflux) or *ǂχ* (uvular fricative efflux) in different sources — which, quite often, does not even reflect a true dialectal / idiolectal difference in articulation, but rather a transcriptional convention — will be consistently unified to *ǂx* so as to avoid creating the illusion that we may be dealing with different phonemes. Likewise, the velar/uvular ejective affricate, alternately transcribed as *kx*, *kx*, *kx'*, *qχ*, or *qχ'* depending on the source, shall be consistently simplified and unified to *kx*.

Khoe basic lexicon: Items 1-25

1. ASHES

○ KK: ***t^{h}ao-** (N *cáó-*, K *t^{h}ao-*). ◊ Not attested in Xri.
• Khwe: ***ǂ̲oa** (ǁAni *ǂ̲óà*, ǁXom *ǂ̲ōá*). ◊ ǀGanda *ǂ̃òà* and Buga *ǂ̃òá* share an irregular shift to nasal articulation of the click, possibly due to the semantic influence of **ǂ̃u* 'black' and/or **ǂ̃um* 'coal'.
• Naro-ǂHaba: ***t^{h}au** (Naro, ǂHaba *t^{h}áú*).
• ǀGwi-ǁGana: ***ǂ̲oa** (ǀGwi *ǂ̲úà*, ǁGana *ǂ̲óà*).
• Shua: ***ʒoa** (Cara *ʒòà*, ǀXaise *ʒóá*, Danisi *d^{y}úà*). ◊ Deti *dù* 'ashes' = ǁXom *dǔ* 'place with charcoal; medicinal charcoal' (clearly a semantic innovation).
• Ts'ixa: **d^{y}úà**. ◊ Aligns with Shua[3].
• Tsua: ***ʒoa** (Kua *ʒóá*, Tsua *ʒùá*; Tjwao *ʒoa ~ coa*; Hie. *ǯoa:*).
• Cua: **ǂ̲ǒà**. ◊ For semantic nuances cf. also *qámⁿ* 'ashes with no coal', *qáró* 'ashes (with coal)'.
○ NKK: Common NKK ***ǂ̲oa** 'ashes' is persistent, but opposed to Naro-ǂHaba ***t^{h}au** = ǁAni *t^{h}àú* 'flame', ǁXom *t^{h}éú* 'spark; tinder', ǁGana *t^{h}áú* 'fireplace'. The latter root is sufficiently well distributed to be reconstructible to the Proto-NKK stage, but its semantics is vague (perhaps something like 'burning embers' could accommodate all the attested meanings).
○ CK: An intricately twisted situation. KK ***t^{h}ao-** 'ashes' is formally opposed to NKK ***ǂ̲oa** 'ashes'; the latter root, however, is also present as Proto-KK ***ǂoa-** (N *ǂ̲ò̀à-*, K *ǂòà*, X *ǂ̲oa-*) 'clay, mud', with no clear indication of the direction of semantic change. At the same time, KK ***t^{h}ao-**, as has been shown above, also has a Proto-NKK correlate with an ambiguous meaning ('burning embers'?). It should be noted that ***ǂ̲oa** is definitely an areal word that gets around; note !Xóõ (Taa) *ǂ̲òa* 'ashes' (probably a borrowing from ǀGwi-ǁGana) and Ju|'hoan *ǂ̲oa^{h}* 'soap', Angolan !Xũ (Snyman) *ǂ̲òā* 'ashes' (possibly ← KK, hinting that 'clay, mud' may not be the original meaning).

[3] Given the alleged "mixed" nature of Ts'ixa as a language that includes both a Shua (Eastern) and a Khwe (Western) layer, it makes sense to consistently add notes on whether the form in question is closer (lexically or phonetically) to its basic equivalent in either of these two groups.

One thing is clear: all lexical connections for ***tʰao-** point toward semantics of 'burning', while the connections for ***ǂ̰oa** always refer to the function of 'ashes' as a resource ('clay', 'soap'). This may have been the original semantic opposition in Proto-CK, although the likelihood of such a scenario is diminished by the fact that it is not directly attested in any known language. In any case, the situation in Naro-ǂHaba is clearly innovative compared to the rest of NKK (merger of both semantic aspects in a single root) and we are fully justified to postulate a lexicostatistical discrepancy between Proto-KK ***tʰao-** 'ASHES' and Proto-NKK ***ǂ̰oa-** 'ASHES'.

2. BIRD

- KK: ***kxani-** (N *àní-*, K *kxàní-s*). ◊ Not attested in Xri as a separate word, although root shapes *kxani-* and *ani-* are attested in compounds (e.g. *ǂâ-kxani-s* 'butcher bird', *ǁʾau-ani-p* 'heron').
- Khwe: ***ʒa(ː)ra** (‖Ani, Buga *ʒàrá*, |Ganda *ʒárá*, ‖Xom *ǯāárá*). ◊ The bimoraic structure of the first syllable in ‖Xom may be historically significant (see discussion in the NKK section below).
- Naro-ǂHaba: ***ʒa(ˤ)ra** (Naro *càˤrá*, ǂHaba *ʒàrá*). ◊ Pharyngealization in Visser's transcription of the Nara word is not confirmed either by Vossen's (*ʒàrá*) or Barnard's (*càrá* ~ *ʒàrá*) data; it may be a local positional development, but see further considerations in the NKK section below.
- |Gwi-‖Gana: ***ʒera** (|Gwi *ʒérá*, ‖Gana *ʒárá*).
- Shua: ***ʒera** (Cara *ʒàrà*, |Xaise *ʒérà*, Deti *ʒàrà*, Danisi *ʒàrá*).
- Ts'ixa: **ʒìrá**. ◊ Phonetically aligns with Shua (front vowel in first syllable).
- Tsua: ***ʒera** (Kua *ʒérá*, Tsua *ʒérà*; Tjwao *ʒiraː*; Hie. *zera*).
- Cua: **ʒàrā**.
- NKK: Although the root seems to be very stable across the entire group, there are two phonetic problems with R. Vossen's (1997: 503) reconstruction of **dzada* (= **ʒara*); (a) the unpredictable appearance of a front vowel reflex in many languages and (b) the pharyngealized articulation of the vowel in (at least some) Naro dialects. Personally, I find the idea of sporadic fronting after an affricate quite suspicious (it does not happen in any other roots with similar structures, to the best of my knowledge). Also of note is the odd reflex *ǯāárá* in ‖Xom (confirmed in Köhler's older records as *djǎdà* with a contour tone); the bimoraic sequence suggests that the first syllable may have originally been a dipthong, e.g. ***ʒaera**, with either subsequent contraction (→ *ʒera*) or assimilation (→ *ʒaː/ːra*) in daughter languages.

  Since such a structure is formally impossible in basic roots, a form like **ʒaera* could only be understood as a derived extension from a former **ʒae-* — a root which, incidentally, may be found (with additional nasalization) in Naro *cə̀ə̀ˤ* 'to fly'. Internal etymologization of 'bird' as a nominal derivate from 'to fly' (with the addition of a fossilized suffix **-ra* = **-da*, whose presence can also be suspected in several other CK stems) is typologically realistic and could, indeed, account for most of the peculiarities of this root (with the exception of pharyngealization in Naro that remains somewhat mysterious).
- CK: While NKK ***ʒera** (← ***ʒae-ra**?) has no clear parallels in KK, KK ***kxani-** 'bird' naturally corresponds to NKK **kxãĩ* 'vulture' (Vossen 1997: 442). In Starostin 2013: 411, I argued that ***ʒera** must be the original CK root for 'bird', with ***kxani** 'vulture' probably widening its semantics in PNKK; with the newly suggested internal etymologization for ***ʒ/a/era**, however, I am not so certain any more, as it is quite easy to visualize a situation in which (a) ***kxani-** narrows its meaning from 'bird' (general) to 'vulture' (as a kind of 'bird *par excellence*') in PNKK, while at the same time (b) the innovation ***ʒ/a/era** 'flying thing' stands in to express the idea of 'bird' in general. Meanwhile, KK languages would preserve the original situation. With no obvious resolution to this situation, PKK ***kxani** 'BIRD' and PNKK ***ʒ/a/era** 'BIRD' have to be accepted as "technical synonyms" on the PCK level.

3. BLACK

○ KK: ***ǂ̃u** (N *ǂ̃ùù*, K *ǂ̃ū*, X *ǂ̃u:*).
• Khwe: ***ǂ̃u** (‖Ani *ǂ̃nú*, |Ganda, ‖Xom *ǂ̃ú*, Buga *ǂ̃ŭ*).
• Naro-ǂHaba: ***ǂ̃u** (Naro *ǂ̃ŭ*, ǂHaba *ǂ̃nú*).
• |Gwi-‖Gana: ***ǂ̃u** (|Gwi *ǂ̃ú:*, ‖Gana *ǂ̃nú*).
• Shua: ***nʒu** (Cara *yú*, |Xaise *nʒú*, Deti *yŭ*, Danisi *ndú*).
• Ts'ixa: ***ǂ̃ú***. ◊ The entry in Vossen's dataset phonetically aligns with Khwe and other WKK languages vs. Shua (preservation of the palatal click); but cf. *ngyú:* 'black' in Fehn 2014: 81.
• Tsua: ***nʒu** (Kua *ʒú*, Tsua *dú*; Tjwao *nʒu: ~ ncu:*; Hie. *ǯu-ɲe*).
• Cua: ***ǂú***:.
○ NKK: ***ǂ̃u**.
○ CK: ***ǂ̃u**. ◊ Highly stable root. A few languages also use secondary synonyms formed from the marginal noun **du* 'k. of charcoal', e.g. Khoe *dŭ-ʃĩ* 'to be black' (Kilian-Hatz 2003: 38), but this seems to be a very recent development.

4. BLOOD

○ KK: ***|ʔao-** (N */ʔáò-b*, K */ʔau-b*). ◊ Not attested in Xri.
• Khwe: ***|ʔáò** (‖Ani, |Ganda, Buga, ‖Xom */ʔáò*).
• Naro-ǂHaba: ***|ʔáò** (Naro, ǂHaba */ʔáò*).
• |Gwi-‖Gana: ***|ʔáò** (|Gwi, ‖Gana */ʔáò*).
• Shua: ***|ʔáò** (Cara, |Xaise, Deti, Danisi */ʔáò*).
• Ts'ixa: ***|ʔáò***.
• Tsua: ***taka** (Kua, Tsua *tâkà*; Tjwao *ta:ka*; Hie. *t^{h}aka*).
• Cua: ***|ʔáò***.
○ NKK: ***|ʔáò**. ◊ Tsua ***taka** is an obvious innovation, both because of its isolated distribution and decidedly non-Khoe phonetic shape; however, no obvious source of borrowing has been detected so far in the neighboring Bantu languages. Remnant of a more ancient substrate?
○ CK: ***|ʔao**. ◊ Highly stable root, with the exception of the Tsua branch.

5. BONE

○ KK: ***ǂxo-** (N *ǂxòő-*, K *ǂxo-b* [E], Xri *ǂxo:-p ~ /xo:-p*). ◊ Meinhof lists the !Ora word only in the meaning 'stone / pith of fruit', but Engelbrecht's data, as well as the Xri entry, clearly show that the root is reconstructible in the meaning 'bone' for the PKK level.
• Khwe: ***|ʔő̃á̋** (‖Ani, |Ganda, Buga, ‖Xom */ʔő̃á̋*).
• Naro-ǂHaba: ***|ʔő̃á̋** (Naro, ǂHaba */ʔő̃á̋*).
• |Gwi-‖Gana: ***|ʔő̃á̋** (|Gwi, ‖Gana */ʔő̃á̋*).
• Shua: ***|ʔő̃á̋** (Cara, |Xaise, Danisi */ʔő̃á̋*; Deti */ʔò̃á̋*).
• Ts'ixa: ***|ʔő̃á̋***.
• Tsua: ***|ʔõ̃ã** (Kua */ʔò̃á̋*, Tsua */ʔő̃à̃*; Tjwao */ʔũ̀ã̀*; Hie. *n|gwa*).
• Cua: ***|'ù̃ā̃***.
○ NKK: ***|ʔõ̃ã**. ◊ Preserved in all daughter languages.
○ CK: ***|ʔõ̃ã**. ◊ KK ***ǂxo-** and NKK ***|ʔõ̃ã** find themselves in complementary distribution; however, external data from Kwadi (*/ʔũ̀ã̀* 'bone') decisively confirm the NKK entry as more archaic. Speculatively, one might assume that the alternate meaning 'stone / pith of fruit', attested for the word in !Ora, may be primary for KK ***ǂxo-**, but there are no additional arguments to strengthen that scenario.

6. CLAW (= NAIL)

○ KK: ***‖oro-** (N *‖ǀórò-s*, K *‖ōrō-b*, Xri *‖oro-p*).
• Khwe: ***‖a** (‖Ani, ‖Xom *‖à:*, Buga *‖â*).
• Naro-ǂHaba: Naro *‖óˢrò*. ◊ Not attested in ǂHaba.
• |Gwi-‖Gana: —. ◊ Attested only in Tanaka 1978 as |Gwi, ‖Gana *!ore*; it is quite likely that *!* is a mistranscription for *‖*, and the root is the same as in Naro, etc., but more reliable sources are necessary to confirm that.
• Shua: [A] ***‖oro** (Cara *‖órò*, |Xaise, Deti *‖óró*, Danisi *‖áró*); [B] ***‖a** (Cara *‖â*, |Xaise *‖ʰâ*, Danisi *‖à*). ◊ Due to absence of detailed lexicographic descriptions for Shua languages, semantic difference between these quasi-synonymous roots remains unclear (Vossen specifically notes polysemy 'fingernail/claw' for Cara *‖â*, but not for Cara *‖órò*; however, most Khoe languages do not lexically distinguish between these two meanings, so the cue may be misleading).
• Ts'ixa: **‖à**. ◊ Aligns with both Shua and Khwe.
• Tsua: ***‖oro** (Kua *‖órò*; Tjwao *‖xore*; Hie. *ǀara*). ◊ The accompaniment *-x-* in Tjwao may be indicative of a less trivial original segment (cf. the uvular efflux in Cua), but more data are needed to confirm. In Hie. (as in a few other cases), it is probable that the word was erroneously printed with a *|* (dental) click instead of a *‖* (lateral click).
• Cua: **‖qólē**.
○ NKK: [a] ***‖oro**; [b] ***‖a**. ◊ [a] It is tempting to reconstruct the first of the two roots as ***‖qoro**, based on such reflexes as *-q-* in Cua or pharyngealized articulation of the vowel in Naro (where *‖óˢrò* is highly likely ← **‖qoro*). However, although reflexes of potential clicks with uvular effluxes are notoriously unstable in Kalahari Khoe languages (see Vossen xxx), one would probably expect them to be at least slightly more prominent in this case throughout the subgroup. At the same time, it is impossible not to note the phonetic similarity with !Xóõ (Taa) *‖qû-le*, pl. *‖qû-n* 'nail, hoof' (Traill 1994: 115) — see Starostin 2021: 117 for notes on its internal etymology within the Tuu family, which more or less precludes borrowing from a Khoe source. It makes more sense to view such forms as attested in Naro and (especially) Cua as isolated re-borrowings from various dialects of Taa or, at least, contaminations of the original root with any such borrowings. [b] The root ***‖a** can also easily lay claim to Proto-NKK status; at our present state of knowledge, it is impossible to establish their semantic difference.
○ CK: ***‖oro**. ◊ A straightforward and phonetically unproblematic (except for the abovementioned elements of uvular or pharyngealized articulation) isogloss between KK and NKK which suggests that Proto-NKK ***‖a** 'fingernail' (no etymological connections discovered in KK) is most likely an innovation.

7. DIE

○ KK: ***‖ʔo** (N *‖ʔó:*, K *‖ʔo*, Xri *‖ʔo:*). ◊ It is curious to note the possibility of reconstructing an additional synonym **|xai* 'to die', based on comparative data from Nama and Xri; however, the basic meaning of this root is 'to be absent', with the partial shift → 'to die' clearly reflecting a recent tabooistic development.
• Khwe: ***‖ʔó** (‖Ani, |Ganda, Buga, ‖Xom *‖ʔó*).
• Naro-ǂHaba: ***‖ʔó** (Naro, ǂHaba *‖ʔó*).
• |Gwi-‖Gana: ***‖ʔó** (|Gwi, ‖Gana *‖ʔó*).
• Shua: ***‖ʔó** (Cara, |Xaise, Deti, Danisi *‖ʔó*).
• Ts'ixa: **‖ʔó**.

- Tsua: ***‖ʔo ~ *ʔo** (Kua, Tsua *ʔǒ*; Tjwao *‖ʔo:*; Hie. *o:*). ◊ Irregular click loss (also attested in several other Tsua roots with initial **‖ʔ-*), but since Tjwao usually preserves the click in all such cases, probably not reconstructible to the proto-level (at most, as a dialectal phenomenon).
- Cua: **ʔó**:. ◊ Note the same irregular click loss as in Tsua.

○ NKK: ***‖ʔó**.

○ CK: ***‖ʔó**. ◊ A stable and well-preserved lexical item throughout the entire family.

8. DOG

○ KK: ***ari-** (N *ǎrí-*, K *ʔàrì-b*). ◊ Not attested in Xri.

- Khwe: [a] ***ari-** (‖Ani *ʔérì-kù*); [b] ***ʔapa** (|Ganda, Buga *ʔápà*, ‖Xom *ʔápā*). ◊ [a] Based on external comparanda, ‖Ani *ʔérì-kù* is identifiable as the result of assimilation ← **ʔari-*, although this seems to be a relatively unique development for this language, nor is it clear at which particular chronological stage the change took place. [b] The main competing etymon is formally reconstructible as ***ʔapa**, although it should be noted that intervocalic *-p-* is a glaring anomaly for Kalahari Khoe.
- Naro-ǂHaba: ***haˢru-gu** (Naro *hǎˢ-gù*, ǂHaba *hǎˢrú-gù*). ◊ One of several lexical isoglosses that transparently demonstrates the exclusive proximity between Naro and ǂHaba, although the internal development in Naro (contraction) is somewhat unique.
- |Gwi-‖Gana: [a] ***haru-gu** (|Gwi *hárú-gù*); [b] ***ʔaba** (|Gwi *aba* [Ta.], ‖Gana *ʔábà*). ◊ Tanaka (1978: 29) lists the forms *arugu* and *aba* as synonyms for 'dog' in |Gwi. Both forms could technically pretend to be protolanguage synonyms.
- Shua: ***ʔaba** (Cara, |Xaise *ʔábà*, Deti, Danisi *ʔábá*).
- Ts'ixa: **ʔábá**. ◊ Aligns with Shua.
- Tsua: ***ʔaba** (Kua, Tsua *ʔábà*; Tjwao *ʔaba*; Hie. *aba*).
- Cua: **ʔāpā**.

○ NKK: [a] ***(h)aˢri-**; [b] ***ʔaba** (~ ***ʔapa**). ◊ A complex situation. Of the two competing roots, ***ʔaba** is clearly the more widespread; however, it has two different phonetic variants, one of which (***ʔapa**) seriously violates a basic phonotactic rule of Kalahari Khoe (no voiceless *p* in the intervocalic position), which usually suggests a borrowed / substrate origin. Given the presence of phonetically similar etyma in Bantu ('dog' is commonly reconstructed as **-búà* for Proto-Bantu), we may be dealing here with a relatively recent borrowing (or a series of independent borrowings) from one or more Bantu languages, with subsequent areal diffusion. On the other hand, the forms attested in ‖Ani, Naro, ǂHaba, and |Gwi are not identifiable as borrowings *and* have plausible external correlates in KK. The only serious problem here is reconciling ‖Ani *ʔérì-kù* with ***haˢru-** in the other two branches; it is possible that a protoform like ***haˢri-gu** (with right-to-left assimilation in ‖Ani and left-to-right in the other languages) could explain all the reflexes, but only according to a unique scenario.

○ CK: ***(h)aˢri-**. ◊ The original Khoe word for 'DOG' is represented by an isogloss between Proto-Khoekhoe ***ari-** and Proto-NKK ***(h)aˢri-**, where the latter turns out to be unstable in descendant languages but potentially preserving more archaic phonology (with pharyngealized articulation). Fluctuation between variants with *h-* and without it is a problem that either has to do with spontaneous reasons (onomatopoeia?) or rare dissimilative processes in roots with specific structures (such as **HVˢ-*). As for ***ʔaba**, no traces of which can be found in Khoekhoe, it is probably advisable to leave it out of any external comparisons due both to its transparently sound-symbolic ("barking") associations and a high probability of borrowing from a Bantu source.

9. DRINK

○ KK: ***kxa** (N *â*, K *kxā*, Xri *kxa:*). ◊ With regular deletion of ***kx-** in Nama.
• Khwe: ***kxâ** (‖Ani, |Ganda, Buga, ‖Xom *kxâ*).
• Naro-ǂHaba: ***kxâ** (Naro, ǂHaba *kxâ*).
• |Gwi-‖Gana: ***kxâ** (|Gwi, ‖Gana *kxâ*).
• Shua: ***kˀâ** (Cara, |Xaise, Deti, Danisi *kˀâ*).
• Ts'ixa: **kˀâ**.
• Tsua: ***kˀâ** (Kua *kˀâ*; Tjwao *kˀa:*; Hie. *ǂʰa:*). ◊ Hie. *ǂʰa:* definitely belongs here, as this is one of several curious cases of transcribing the Tsua velar ejective phoneme with a palatal click in Dornan's notation (cf. also Hie. *ǂʰae-yo* 'a laugh' = Kua, Tsua *kˀâĩ̃* 'to laugh').
• Cua: **kxáà**.
○ NKK: ***kxâ**.
○ CK: ***kxa**. ◊ A highly stable Khoe verbal root. It is also one of the best known transparent lexical matches between Khoe and Tuu (Starostin 2021: 120), but no certain conclusions can be drawn at present at whether it reflects an ancient borrowing (and if so, in which direction) or should be interpreted as a rare remainder of an ultra-deep genetic connection between the various "Khoisan" lineages.

10. DRY

○ KK: ***|ˀo** (K *|ˀō*, Xri *|ˀu: ~ |ˀo:*). ◊ In N, the old root (*|ō*) has largely shifted to the figurative meaning 'dry = barren' (of cows, etc.); the basic meaning 'dry' (of clothes, etc.) has been replaced by the adjectival derivate *ǂ̃ãã̀-sã́* from the verb *ǂ̃ãã̀* 'to dry up; to wither', which, in its turn, may be an old derivate from the more basic verb *ǂ̃â* 'to pour' (of Common Khoe ancestry).
• Khwe: ***‖xo** (‖Ani, ‖Xom *‖xó*, Buga *‖xŏ*).
• Naro-ǂHaba: [a] ***|ˀo** (Naro *|ˀô*); [b] ***‖xo** (ǂHaba *‖xó*). ◊ The second root is also found in Naro, but in a somewhat restricted function: *‖xó:* 'dry, stiff, hard (wood, porridge, ice)'.
• |Gwi-‖Gana: ***‖xo** (|Gwi, ‖Gana *‖xó*).
• Shua: ***‖xo** (Deti *‖xŏ*, Danisi *‖xó*).
• Ts'ixa: **‖xó**.
• Tsua: [a] ***|ˀo** (Kua, Tsua *|ˀô*; Hie. *|o*); [b] ***‖xo** (Tjwao *‖xo:-ha*). ◊ The two roots have so far not been explicitly encountered within the same lect to determine their precise semantic difference.
• Cua: [a] **|ˀóò** ('dry' = 'pliable'); [b] **‖χó:** ('dry' = 'brittle'). ◊ According to the vocabulary, the first word primarily renders Setswana *swaba* 'dry' = 'withered', whereas the second one rather corresponds to *oma ~ omelela* (basic 'dry', 'dry out', of clothes, etc.).
○ NKK: ***‖xo**. ◊ Existing data point to two roots commonly glossed as 'dry' across NKK languages: **|ˀo* and **‖xo*. Distribution of the first one, however, is limited to Naro (where semantic influence on the part of Khoekhoe cannot be excluded), Tsua, and Cua, with the semantics in Cua leaning more toward the meaning 'dried up, withered'; it is also found in ‖Xom (*|ˀóò*) with the meaning 'to dry out, to boil over'. It seems reasonable to assume that the basic Swadesh meaning 'dry' (of clothes, hair, etc.) is better aligned with **‖xo*.
○ CK: ***‖xo**. ◊ It is important to note that the semantics of **|ˀo* in Nama seems to correlate with the data from Cua, in that the root is more strongly reflective of the meaning 'dried up, withered, (fig.) barren' than 'dry' proper. Meanwhile, the other root is also preserved in Khoekhoe, but only in figurative meanings: N *‖xőó* 'dangerous, serious; strict', K *‖xō* 'evil; unpleasant'. The most reasonable scenario is that PCK **‖xo* 'DRY' acquired this figurative meaning relatively early, which led to the neutral meaning 'dry' finding different ways of

expression (either merging with 'dried, barren' or replaced by new derivates from verbal roots, as in N). The merger of 'dried out, withered' with the basic 'dry' in parts of Khoekhoe (K, Xri) and in Tsua is most likely the result of independent (and quite trivial) semantic shift.

It should also be noted that both **|ʔo* and **‖xo* have external parallels in other "Khoisan" languages, especially in Tuu: for the former, cf. !Xóõ *|ʔòo* 'to be dry', |Xam *|ò:wa* 'dry' (of bones), etc.; for the latter, cf. !Xóõ *‖úaˤ* 'dry', |Xam *‖o*, etc. Some of these forms may be borrowings from Khoe (particularly the reflexes of **|ʔo*, which are not widely distributed across Tuu), but some might reflect deeper connections beyond the scope of this paper.

11. EAR

- ○ KK: ***ǂae** (N *ǂàé̋-s*). ◊ Only Nama preserves the original word for 'ear'; the alternate stem K *ǁ̃ãũ-b*, Xri *ǁ̃ãũ-(m)p* 'ear' is a transparent innovation, derived from the verb **ǁ̃ãũ* 'to hear' (which is in itself probably a KK innovation, see HEAR below); it is interesting that the noun *ǁ̃ãũ-b* is also attested in Nama, but only in the more abstract meaning 'hearsay' (cf. also the fem. correlate *ǁ̃ãũ-s* '/sense of/ hearing').
- • Khwe: ***ǂé** (‖Ani, |Ganda, Buga, ‖Xom *ǂé*).
- • Naro-ǂHaba: ***ǂé** (Naro, ǂHaba *ǂé*).
- • |Gwi-‖Gana: ***ǂê** (|Gwi, ‖Gana *ǂê*).
- • Shua: ***ç̡é** (Cara, |Xaise, Deti, Danisi *ç̡é*).
- • Ts'ixa: **ǂé**. ◊ Phonetically aligns with Khwe (preservation of the palatal click).
- • Tsua: ***ç̡e** (Kua, Tsua *kʸê*; Tjwao, Hie. *če:*).
- • Cua: **ǂéē**.
- ○ NKK: ***ǂé**. ◊ A highly stable etymon, preserved in all languages without exception.
- ○ CK: ***ǂe**. ◊ The isogloss between Proto-NKK and Nama *ǂàé̋-s* (with the regular development **-e:* → **-ae* in KK) identifies the PCK etymon beyond any reasonable doubt. However, it is interesting to observe the beginning of the typologically common semantic shift 'hear' → 'ear' in various KK lects (typically not a "Khoisan" feature).

12. EAT

Note: most CK languages feature a distinct binary differentiation between the sub-meanings 'eat soft food' (vegetables, etc.) and 'eat hard food' (such as meat; incidentally, the second lexeme usually has the same root as the noun 'meat', see below) — an interesting feature of lexical typology that they share with such Afro-Asiatic taxa as Chadic (see, e.g., the distinction in Jungraithmayr & Ibriszimow 1994: 55–56). Due to the strong lexical connection with 'meat', it makes sense to regard the sub-meaning 'eat hard food' as somewhat less "basic", so in this section we shall only be surveying the meaning 'eat soft food', taking it as the rough equivalent of 'eat' in general.

- ○ KK: ***ǂʔũ** (N *ǂʔũ̏ű*, K *ǂʔũ*, Xri *ǂʔû*).
- • Khwe: ***ǂʔũ** (‖Ani, |Ganda, ‖Xom *ǂʔű*, Buga *ǂʔũ̌*).
- • Naro-ǂHaba: ***ǂʔű** (Naro, ǂHaba *ǂʔű*).
- • |Gwi-‖Gana: ***ǂʔũ** (|Gwi *ǂʔű*, ‖Gana *!ʔón* [Ta.]). ◊ Tanaka's orthography involves a common mistranscription of the palatal click in ‖Gana.
- • Shua: ***ʔyũ̌** (Cara, |Xaise, Deti, Danisi *ʔyũ̌*).
- • Ts'ixa: **ǂʔũ̌**. ◊ Phonetically aligns with Khwe.
- • Tsua: ***ʔyũ** (Kua, Tsua *ʔyũ̌*; Tjwao *ʔyũ: ~ ʔɲũ:*; Hie. *ɲo:*). ◊ Nasalization of the (former) click influx in Tjwao and Hie. is secondary, under the influence of the nasal vowel of the root.

- Cua: **túm̀** ‘to eat /soft foods/, to swallow’.

○ NKK: ***ǂʔũ**. ◊ The reconstruction is unambiguous, and the lexeme is generally quite stable; it is curious to note, however, that in Cua the old word may have been replaced by a reflex of the common CK etymon ‘to swallow’ (PCK **tom*, see Vossen 1997: 502).

○ CK: ***ǂʔũ**. ◊ A solid isogloss between all branches of Khoe.

13. EGG

○ KK: ***!ʔubu** (N *!ʔùɯ̀ɯ́-s*, K *!ʔú:bú-b* ‘ostrich egg’, Xri *!ʔubu-p*). ◊ Interestingly, for Xri the form *ʃ̃ara-p* is listed as a synonym for ‘egg’; it probably corresponds to Nama *ʃ̃àrâ-b* ‘umbrella thorn tree’, *ʃ̃àrâ-s* ‘pod of umbrella thorn’ (possibly as a figurative term).

- Khwe: ***ǂʔúbí** (‖Ani, Buga, ‖Xom *ǂʔúbí*).
- Naro-ǂHaba: ***ǂʔúbí** (Naro, ǂHaba *ǂʔúbí*).
- |Gwi-‖Gana: ***ǂʔúbí** (|Gwi, ‖Gana *ǂʔúbí*).
- Shua: ***ʔyubi** (Cara *ʔyìbí*, Deti *ʔyùbì*, Danisi *ʔùbí*).
- Ts’ixa: **kʰābí** ‘ostrich egg’.
- Tsua: ***ʔibi** (Kua, Tsua *ʔíbí*; Tjwao *ʔibi*; Hie. *ibi*).
- Cua: **ǂʔíbī**.

○ NKK: ***ǂʔubi**. ◊ A generally stable bisyllabic stem (with regular click loss in most of the Eastern languages). Does not seem to be attested in Ts’ixa, where the only known equivalent is the word *kʰābí* ‘ostrich egg’ = Cua *kʰăbī* ‘ostrich egg flask’; if this is indeed the basic term for ‘egg’, it is likely a recent generalization of a formerly narrow cultural term.

○ CK: —. ◊ Despite superficial similarity, click influxes do not match between KK and NKK, preventing etymological relationship between ***!ʔubu** and **ǂʔubi**. It cannot be excluded, however, that Proto-KK **!ʔubu* ← **ǂʔubu* as a result of semantic contamination with *!ʉ̃ɯ́* ‘round, spherical’; while this hypothesis is impossible to prove, it slightly increases the probability that the NKK form is more archaic for this concept.

14. EYE

○ KK: ***mũ-** (N *mʉ̃ʉ̃́-s*, K *mũ-b*, Xri *mũ-/m/p*). ◊ Clearly connected with the verbal root **mũ* ‘to see’.

- Khwe: ***ǂxáí** (‖Ani, |Ganda, Buga *ǂxáí*, ‖Xom *ǂxéí*).
- Naro-ǂHaba: ***ǂxáí** (Naro *ǂxéí*, ǂHaba *ǂxáí*).
- |Gwi-‖Gana: ***ǂxáí** (|Gwi *ǂxéí*, ‖Gana *ǂxáí*).
- Shua: ***ɕxai** (Cara *ɕxáí*, |Xaise *ɕáí*, Deti *ɕxàí*, Danisi *ɕxáí*).
- Ts’ixa: **ǂxáí** [Vossen], **ǂxáí ~ ɕxáí** [Fehn]. ◊ Both the click-containing and clickless form are attested among different groups of speakers, according to Fehn.
- Tsua: ***ɕxai** (Kua, Tsua *ɕxàí*; Tjwao *čxai*; Hie. *čai:*).
- Cua: **ǂxàī**.

○ NKK: ***ǂxai**. ◊ Regular affricativization of the click in the Eastern languages. It is interesting to observe the development **-ai* → *-ei* across much of Kalahari Khoe territory; this seems to be triggered by the palatalizing influence of the click, since normally Proto-CK **-ai* stays unchanged in most of these languages.

○ CK: ***ǂxai**. ◊ The common KK term for ‘eye’ has to be considered as a lexical innovation (nominal derivation from the verb ‘to see’) in light of the NKK evidence. Interestingly, the old root seems to have shifted to a verbal function itself in KK: cf. N *ǂxáí*, K *ǂxai* ‘to wake up’ (← ‘to open one’s eyes’?), developing along the same lines as, e.g., ‖Xom *ǂáó* ‘heart’ → ‘to be happy, glad’.

15. FIRE

- o KK: ***|ʔae-** (N *|ʔàẽ́-s*, K *|ʔae-b*, Xri *|ʔae-p*).
- Khwe: ***|ʔe** (‖Ani, Buga *|ʔě*, |Ganda, ‖Xom *|ʔé*).
- Naro-ǂHaba: ***|ʔé** (Naro, ǂHaba *|ʔé*).
- |Gwi-‖Gana: ***|ʔe** (|Gwi *|ʔé*, ‖Gana *|ʔê*).
- Shua: ***|ʔe** (Cara, Danisi *|ʔě*, |Xaise *|ʔé*). ◊ Deti *‖ǔ* 'fire' is a local innovation, apparently a nominalization of the original verb 'to light, kindle /fire/' (cf. ‖Xom *‖ūú*, Naro *‖ǔ*; Cua *‖ǔú* id.).
- Ts'ixa: **|ʔě** [Vossen], **|ʔēé** [Fehn].
- Tsua: ***|ʔe** (Kua, Tsua *|ʔé*; Tjwao *|ʔe:*; Hie. *|e:*).
- Cua: **|ʔéē**.
- o NKK: ***|ʔe**.
- o CK: ***|ʔe**. ◊ The basic designation for 'fire' is well reconstructible on the PCK level and quite stable in daughter languages. It is strikingly similar to Proto-Tuu **|ʔi* 'fire' (Starostin 2021: 124), but since both words are perfectly reconstructible on the respective deepest chronological levels, any possible connections through either borrowing or common inheritance must predate the disintegration of PCK.

16. FOOT

- o KK: ***ǂʔai-b** (N *ǂʔâì-b/s*, K *ǂʔai-b*, Xri *ǂʔai-p*).
- Khwe: ***!̃are** (‖Ani *!̃nàré*, |Ganda *kárì*, Buga *kárè*, ‖Xom *káré*).
- Naro-ǂHaba: ***!̃are** (Naro *!̃àrè*). ◊ Not attested in ǂHaba.
- |Gwi-‖Gana: ***!̃are** (|Gwi *ngare* [Ta.], ‖Gana *ng!àre* [Ta.]). ◊ Only attested in Tanaka's problematic dictionary, but the forms agree reasonably well with external data.
- Shua: ***ʒĩ** (Cara *ʒĩ́*, |Xaise, Deti, Danisi *ʒĩ̂*).
- Ts'ixa: **ʒĩ̂** [Vossen, Fehn]. ◊ Aligns with Shua.
- Tsua: ***kare** (Kua *kárì*, Tsua *káré*; Tjwao *kare(:)*; Hie. *kare:*).
- Cua: **kārè**.
- o NKK: ***!̃are**. ◊ The "majority rule" strongly supports ***!̃are** as the original equivalent for 'foot'. Proto-Shua ***ʒĩ** is also reconstructible for Proto-NKK, based on such parallels as ‖Ani *ʒê* 'toe', Buga *ʒĩ̂* 'big toe; foot' [Vossen], ‖Xom *ǯĩ̌* 'bird's foot, talon'; the semantics of these forms, however, rather speaks in favor of the original meaning 'toe' (possibly with polysemy: '(human) toe / (bird's) talon'), and a semantic shift 'toe' → 'foot' in Proto-Shua seems quite credible. (Cf. also a more remote, but also quite telling parallel in N *cĩ̏ĩ-b* 'claw of ostrich, esp. long big toe').
- o CK: **—**. ◊ The problem of choosing an optimal candidate between PKK ***ǂʔai-b** and PNKK ***!̃are** is exacerbated by the fact that both of these nominal stems have verbal etymologies. PKK ***ǂʔai-b** 'foot' is cognate with such forms as ‖Xom, Naro *ǂʔí*, Cua *ǂʔí:* 'to kick'; PNKK ***!̃are**, on the other hand, is clearly related to N *!̃àrì* 'to ride, to drive', K *!̃ārí-s* 'wheel', etc. In both cases, semantic derivational shifts 'kick' → 'foot' or 'ride / drive' → 'foot' seem typologically more justified than the development of the corresponding verbal meanings out of nominal ones (parallel to such developments as 'see' → 'eye', 'hear' → 'ear', etc., see above). This means that there is a high probability of the original nominal root for 'foot', whatever it was, to have independently become lost in both primary branches of CK. It is also possible that PCK did not distinguish between the sub-meanings 'foot' and 'leg' (the latter is well reconstructible as PCK ***|̃u** (Vossen 1997: 422).

17. HAIR

○ KK: ***|ʔũ̀-b** (N */ʔù:-b*, K */ʔũ̄-b*, Xri */ʔu-mp ~ /ʔo-mp*).
• Khwe: ***|ʔû̃** (‖Ani, |Ganda, Buga */ʔû̃*, ‖Xom */ʔű́*).
• Naro-ǂHaba: ***|ʔû̃** (Naro, ǂHaba */ʔû̃*).
• |Gwi-‖Gana: ***|ʔû̃** (|Gwi, ‖Gana */ʔû̃*).
• Shua: ***|ʔû̃** (Cara */ʔű́*, |Xaise, Deti, Danisi */ʔû̃*).
• Ts'ixa: **|ʔù̃**.
• Tsua: ***|ʔû̃** (Kua, Tsua */ʔû̃*; Tjwao */ʔũ:*; Hie. */ʰo:*).
• Cua: **|ʔũ̄:**.
○ NKK: ***|ʔũ̀**.
○ CK: ***|ʔũ̀**. ◊ One of the most stable and phonetically transparent elements in the wordlist.

18. HAND

○ KK: ***!ʔom-i** (N */!ʔőḿ-mi*, Xri */!ʔum-mi ~ /!ʔom-mi*). ◊ Comparison of Nama data with the !Ora entries from J. Engelbrecht's records (as quoted in Du Plessis 2018: 313) allows to reconstruct the paradigmatic opposition ***!ʔom-i** (neut.) 'hand' : ***!ʔom-s** (fem.) 'fist' for PKK. However, our main source on !Ora (Meinhof 1930) gives the impression that in the dialect described by Meinhof, only 'fist' has been retained; the equivalent for 'hand' there is K */!ùm-ma* (without the glottalic efflux), which has no equivalents in N and is rather comparable with PNKK **ǀ̱oma ~ *ǀ̱uma* 'lower arm' (Vossen 1997: 499). Of course, some semantic or phonetic contamination is possible here.
• Khwe: ***c^{h}àú** (‖Ani *c^{h}àú*, |Ganda, Buga *c^{h}àú*, ‖Xom *ṣèú*).
• Naro-ǂHaba: ***cau** (Naro *càú*, ǂHaba *cáù*).
• |Gwi-‖Gana: ***cáú** (|Gwi, ‖Gana *cáú*).
• Shua: ***c^{h}àú** (Cara *c^{h}àú*, |Xaise, Deti *càú*, Danisi *c^{h}àú*).
• Ts'ixa: **càú** [Vossen], **c^{h}āú** [Fehn].
• Tsua: ***c(h)au** (Kua, Tsua *càú*; Tjwao *c^{h}au*; Hie. *cau*).
• Cua: **c^{h}ăū**.
○ NKK: ***c^{h}au**. ◊ Highly stable; reconstruction with initial **c^{h}-* is well confirmed by multiple reflexes spread across all branches, although the phoneme is rare and the distribution of attested aspirated or unaspirated reflexes tends to fluctuate.
○ CK: —. ◊ At present, it is impossible to make a choice between PKK ***!ʔom-** and PNKK ***c^{h}au** as to which one is a better candidate for PCK status. It may be kept in mind that **c^{h}-* is a generally rare phoneme in Khoe, and most of the etyma that are safely reconstructible with in on the PNKK level have no cognates or weak parallels in PKK (cf. only 4 reconstructions in Vossen 1997: 523 for PNKK and not a single one for PCK), which could theoretically hint at a substrate origin for PNKK ***c^{h}au**. However, this is still an open issue; overall, both candidates are more or less equiprobable for inclusion into any external comparisons.

19. HEAD

○ KK: ***dana-** (N */dàná-s/b*, Xri */dana-p*). ◊ !Ora is substantially different here. Meinhof records K */bi-!ʔã̀-b* or simply */!ʔã̀-b* 'head', where the second root is equivalent to PKK ***!ʔã̀** 'to hear, perceive' (see HEAR below), but the first one is unclear. Of particular interest is the "newer" form */mũ-!ʔã̀-b* 'head', recorded relatively recently from the last living speakers of !Ora (Du Plessis 2018: 286): it has the same structure as */bi-!ʔã̀-b*, but replaces the first root with the far more

transparent *mũ-* ‘to see; eye’ (thus, ‘head’ = ‘[that which] sees and hears’). The only possible internal etymology for *bi-* is to equate it with K *bi:* ‘to suckle’, but an analysis of ‘head’ as = ‘[that which] suckles and hears’ is quite bizarre from a typological perspective. Nevertheless, it is still most likely some sort of descriptive compound formation which can hardly pretend to PKK (much less PCK) status.

- Khwe: ***ǂú** (‖Ani, |Ganda, Buga, ‖Xom *ǂú*).
- Naro-ǂHaba: Naro *ǂú*. ◊ ǂHaba *m̀-kʔá* is an alternate candidate, but it is clearly a compound formation (see notes on NKK).
- |Gwi-‖Gana: ***ma[-ʔm].** |Gwi *mâ* [Ta.], *m̀-ʔm̂* [V.], ‖Gana *má*.
- Shua: ***ma*** (Cara *má*, |Xaise *mâ*, Deti *má*, Danisi *mà*).
- Ts’ixa: ***ǂú ~ ɕú*** [Vossen], ***ǂu*****:** ~ ***kyú*****:** [Fehn]. ◊ Aligns with Khwe.
- Tsua: ***m-ʔa*** (Kua *mâ*, Tsua *ḿ-ʔà*; Tjwao *ma:*; Hie. *hma*).
- Cua: ˑ **máà**.

○ NKK: ***ǂu**. ◊ Application of the simple "majority rule" would likely lead us to accept a form like Vossen’s Proto-East Kalahari Khoe **ma* ‘head’ (Vossen 1997: 458) as the most likely candidate for PNKK, especially since its reflexes are also seen in such "Western" languages as ǂHaba and |Gwi-‖Gana. However, it is precisely such forms as ǂHaba *m̀-kʔá* and |Gwi *m̀-ʔm̂*, as well as Tsua *ḿ-ʔà*, that show the etymon in question is actually a compound, possibly organized along the same lines as K *bi-!ˀã-b* since the second half is easily identifiable with PCK **!ˀã* ‘to feel, perceive; to know’ = ǂHaba *kʔâ* ‘to know’, Tsua *ʔã́* id. We have, therefore to reconstruct bimorphemic **m-!ˀã*, where the first morpheme is itself most likely a contraction from another form (perhaps **mĩ* ‘to say’ or **mũ* ‘to see’?). It is possible that both ***ǂu** and this ***m(V)-!ˀã** were partially synonymous already on the PNKK level, but if so, the former must have been the neutral equivalent and the latter must have represented a descriptive formation with an expressive function; subsequently, in those branches and languages where it was preserved it eventually replaced the old form, as well as underwent various idiosyncratic contractions.

○ CK: ≈ ***ǂu**. ◊ A strong argument in favor of seeing the Khwe-Naro root for ‘head’ as the original Khoe root is its potential cognacy with N *ǂgűúrò* ‘first, initial; prime’ (where *-ro* may easily be a derivational element); the semantic shift ‘head’ → ‘first’ is quite trivial. However, this cannot be viewed as a final solution until a solid etymology has been proposed for KK ***dana-** as well.

20. HEAR

○ KK: ***‖̃ãũ** (N *‖̃ãű*, K *‖̃ãũ*, Xri *‖̃ãũ*).
- Khwe: ***kóḿ** (‖Ani, |Ganda, Buga, ‖Xom *kóḿ*).
- Naro-ǂHaba: ***kúḿ** (Naro, ǂHaba *kúḿ*).
- |Gwi-‖Gana: ***kúḿ** (|Gwi *kúáḿ*, ‖Gana *kúḿ*).
- Shua: ***kóḿ** (Cara, Deti *kóḿ*, Danisi *kúḿ*). ◊ For |Xaise, Vossen only records *‖ám̀* ‘to feel, to hear’ ← PNKK **‖am* ‘to feel’.
- Ts’ixa: ***kúḿ*** [Vossen, Fehn].
- Tsua: ***kúḿ** (Kua *kúḿ*, Tsua *cóḿ*; Hie. *čom*). ◊ For Tjwao, Phiri only records *‖am* ‘to hear’ ← PNKK **‖am* ‘to feel’.
- Cua: ***ɕúḿ***.

○ NKK: ***kum**. ◊ Vossen reconstructs vocalic variation (**kúḿ* ~ **kóḿ*), but overall, judging by the data, the fluctuation seems more indicative of an original high vowel. Note the (possibly areal) displacement of the root by **‖am* ‘to feel’, shared by |Xaise and Tjwao.

○ CK: ***kum**. ◊ The relative archaicity of the NKK equivalent for 'to hear' is indirectly proven by the preservation of this root in K in a slightly shifted meaning: cf. *kóm-sén* [Meinhof], *kum-sin* [Wuras] 'to listen closely/attentively'. It may be speculated that the common KK adverb **koma* 'supposedly; so it is said' (N *kőmá*, K *koma*) could go back to the same root as well (= 'hearsay'). On the other hand, PKK *ǁãũ finds no reliable cognates in the other branch of the family.

21. HEART

○ KK: ***ǂao-b** (N *ǂ̋áó-b*, K *ǂáó-b*, Xri *ǂao-p*).
• Khwe: ***ǂao** (ǁAni, |Ganda *ǂáó*, Buga *ǂáò*, ǁXom *ǂáó*).
• Naro-ǂHaba: ***ǂao** (Naro, ǂHaba *ǂáó*).
• |Gwi-ǁGana: ***ǂao** (|Gwi, ǁGana *ǂáó*).
• Shua: ***ɕao** (Cara *ɕáó*, |Xaise *ɕó:*, Deti *ɕàó*, Danisi *ɕáó*).
• Ts'ixa: **ǂáó** [Vossen], **k^{y}áó** [Fehn].
• Tsua: ***ɕo** (Kua *k^{y}ó*, Tsua *k^{y}ô*; Tjwao *co: ~ čo:*; Hie. *čo:*).
• Cua: **ǂó**:.
○ NKK: ***ǂao**. ◊ A highly stable root. Phonetically, the development **-ao* → Tsua *-o* is strange, but the monophthongization may have been influenced by the initial palatal click.
○ CK: ***ǂao**. ◊ The original root is preserved in all languages without exception.

22. HORN

○ KK: ***ǁ̃ã-b** (N *ǁ̃ãã̀-b*, K *ǁ̃ã-b* [Engelbrecht], Xri *ǁ̃a-mp*).
• Khwe: ***ǁ̃â** (ǁAni, |Ganda, Buga *ǁ̃nâ*, ǁXom *ǁ̃â*).
• Naro-ǂHaba: ***ǁ̃â** (Naro *ǁ̃â*, ǂHaba *ǁ̃nâ*).
• |Gwi-ǁGana: ***ǁ̃â** (|Gwi, ǁGana *ǁ̃nâ*).
• Shua: ***ǁ̃â** (Cara, |Xaise, Deti, Danisi *ǁ̃nâ*).
• Ts'ixa: **ǁ̃nâ**.
• Tsua: ***ǁ̃â** (Kua, Tsua *ǁ̃nâ*; Tjwao *ǁ̃ã:*; Hie. *ŋǁga:*).
• Cua: **ǁ̃ã̂**:.
○ NKK: ***ǁ̃â**.
○ CK: ***ǁ̃ã**. ◊ The original root is preserved in all languages without exception. I prefer to see the nasalized vowel of PKK as primary, with potential dissimilation ***ǁ̃ã-** → ***ǁ̃a-** in PNKK, rather than vice versa, but this decision is not very relevant for our current purposes anyway.

23. I

○ KK: ***[ti=]ta ~ *[ti=]de** (N *tà*, emph. *tìí̋-tà*, *tìí̋-r*, K m. *ti-re*, f. *ti-ta*, Xri *ti-r ~ (i)r*). ◊ Following Vossen 1997, I identify **ti=* as an emphatic particle marking the full form of the personal pronoun. The opposition between masc. ***=ta** and fem. ***=de** is perhaps best preserved in K, although all the other lects seem to generalize one of the variants (*=*ta* in N, *=*de* in Xri).
• Khwe: ***tí** (ǁAni, |Ganda, Buga, ǁXom *tí*).
• Naro-ǂHaba: ***tí(-rá)** (Naro *tí(-rá)*, ǂHaba *tí*).
• |Gwi-ǁGana: ***tí(-re)** (|Gwi *tí-rè*, ǁGana *tê*).
• Shua: ***tá** (Cara, |Xaise, Deti, Danisi *tá*).
• Ts'ixa: **tí**. ◊ Aligns with Khwe.

- Tsua: ***ti(-ra)** (Kua *kʸé(-dì)*, Tsua *kʸé(-dè)*; Tjwao *ti(-ya) ~ či(-ra)*; Hie. *či*).
- Cua: **ɕí(-é)**.

○ NKK: ***tí(-de).** ◊ This system, with the root morpheme ***tí-** and the optional stem extension ***-de**, is found everywhere except for Shua which, upon first glance, might seem to preserve the same morpheme ***ta** as in PKK ***[ti=]ta**. However, it is much more probable that Shua ***tá** simply generalizes the old object form ***tí-a**, well attested in most other branches; cf., e.g. Gwi subj. *tí-rè*, obj. *tì-à*, but Deti subj., obj. *tá* (same form for both functions).

○ CK: ***[ti=]ta ~ *[ti=]de.** ◊ Essentially, this reconstruction projects the situation attested in !Ora (and nowhere else) onto the PCK level, which may seem improbable; however, it is more or less a fact that all the other situations can be explained out of it, whereas the opposite (innovation in !Ora) would require us to set up extremely odd mechanisms of development. The most important innovation of NKK is that already on the proto-level the old emphatic particle ***ti=** seems to have been re-analyzed as the base marker of the 1st p. pronoun itself; conversely, the old root (only the masculine morpheme ***=de**) became optional, in a curious case of "semantic metathesis".

Alternately, it is possible to historically treat the "emphatic particle" ***ti=** as a (slightly modified) reduplication of the actual pronominal stem (since it is unique in this situation, and setting up a special emphatic particle in conjunction with but one morpheme is odd).

24. KILL

○ KK: ***!am** (N *!àḿ*, K *!ām*, Xri *!am*).

- Khwe: ***|kxũ** (‖Ani *|kxǔ*, |Ganda *|kxû*, Buga, ‖Xom *|kxű́*).
- Naro-ǂHaba: ***|kxũ** (Naro *|kxõ:*, ǂHaba *|kxű́*).
- |Gwi-‖Gana: ***|kxű́** (|Gwi, ‖Gana *|kxű́*).
- Shua: ***|kxũ** (Cara, |Xaise *|ʔǔ*, Danisi *|kxǔ*). ◊ Deti *kʰàé* ‘to kill’ is a secondary semantic shift from Shua **kʰàé* ‘to stick, pierce’ ← PNKK **!ʰàé* id. (Vossen 1997: 495).
- Ts’ixa: **|ʔù̃** [Vossen], **|ʔũ̄ű́** [Fehn].
- Tsua: ***|ʔũ** (Kua, Tsua *|ʔű́*; Tjwao *|ʔũ:*; Hie. *ǰo:*).
- Cua: **|kxű́ũ̄**.

○ NKK: ***|kxũ**. ◊ A stable verbal root (preserved everywhere except for Deti).

○ CK: —. ◊ The simple binary opposition between PKK ***!am** and PNKK ***|kxũ** ‘to kill’ is difficult to resolve, but at least for the second root KK offers a solid cognate in the form of N *|ʔù̃ű́* ‘to stop, cease, desist (from smth.)’, semantically acceptable and phonetically impeccable (note also the contour tonal pattern that is recurrent in NKK and correlates well with the tonal pattern observed in N). From a typological perspective, semantic shift from ‘kill’ to ‘stop’ is far less probable than the opposite, making it likely that the original root is preserved in PKK, replaced by a euphemistic term in PNKK (‘to kill’ = ‘to make smbd. cease [existence]’).

Theoretically, PKK **!am* ‘to kill’ could perhaps correspond to PNKK **!áḿ* ‘to throw (away)’ (Vossen 1997: 507); if so, the situation here would not be fundamentally different, as once again we would probably be dealing with a euphemistic semantic shift. However, the actual semantic distance here seems to be wider and the shift less typologically credible than the ‘kill’ / ‘stop, cease’ connection (on the other hand, cf. such derivation as N *ǂ̃òá* ‘to throw, shoot’ → *ǂ̃òá-!áǹ* ‘to kill [by throwing or shooting]’, which may point the way to such a shift). In any case, all of these considerations make both candidates somewhat dubious, and strengthen the idea that the original PKK equivalent for the meaning ‘kill’ may have been eliminated in both primary branches of the family through euphemistic replacements.

25. LEAF

- KK: —. ◊ In both Nama and !Ora, the word for ‘leaf’ is a semantic extension of the basic word for ‘ear’: N *ǂàé̋-s/b*, K *ǁãũ-b* [Wuras]; this reflects a very common African type of polysemy, not at all present, however, in any of the NKK languages. Not attested in Xri.
- Khwe: ***ǀ̃ã*** (‖Ani *ǀ̂ã*, |Ganda *ǀ̌ã*, ‖Xom *ǀ̌ã*). ◊ Reconstructed with a contour tone (probably LH) which in this particular case reflects contraction from a formerly bisyllabic structure (cf. external data). For Buga, Vossen records */kxɛ́ù* ‘leaf’ which may be the same as ‖Xom */kxáó* ‘green’ (cf. ‖Xom */kxáó-ǀ̌ã* ‘green leaf’), although the vocalic correspondences are strange (cf. also ‖Gana */xɛ́ù* below).
- Naro-ǂHaba: —. ◊ The situation in Naro is unclear, as the sources are in conflict with each other. The default source for Naro (Visser) lists *toàˢrà* ‘leaf’ = *tòːárà ~ dòːˢáràˢ* id. (Barnard), a form that is transparently equivalent to Ju|’hoan (North Khoisan) *dòàˢrà* ‘leaf’ (Starostin 2018: 36); both of these are, however, isolated in their respective families and are likely to represent a common areal isogloss of non-genetic origin (some common Naro-Ju substrate?). As a synonym, Barnard also lists *ǀ̃ana* ‘leaf’ = Visser’s *ǀ̃ànā* ‘little branch which has leaves, part of any tree’. This semantic discription, although somewhat vague, suggests that this word — clearly the original term for ‘leaf’ — has shifted to a somewhat more specialized meaning in Naro. Complicating matters even further, Vossen (1988) records Naro *dáˢnàˢ-sá* ‘leaf’ = ǂHaba *dáˢnàˢ-sà* id. If true, this is an important lexical isogloss between the two languages; however, neither Visser nor Barnard confirm such a recording for Naro. In any case, the striking phonetic resemblance between *toàˢrà* and *dáˢnàˢ-* is unexplainable in terms of regular correspondences (if we view it as a parallel between Naro and ǂHaba) and further contributes to the idea of potential substrate influence (independent borrowing into different lects with various phonetic idiosyncrasies).
- |Gwi-‖Gana: ≈ ***ǀ̃ana*** (‖Gana *ǀ̃ánà*). ◊ For |Gwi, Vossen only records *dàˢnã̀* ‘leaf’ (same root as in ǂHaba, probably of non-CK origin); for ‖Gana, he lists the form */xɛ́ù* as an additional synonym — it seems to be the same as Buga */kxɛ́ù*, but the click correspondence is irregular. Meanwhile, Tanaka lists the equivalent *dana* ‘leaf’ for both languages, and his data are supported by Nakagawa, who lists |Gwi *dáˢnà*, ‖Gana *dáˢnà ~ ʒáˢnà* ‘leaf’ (Nakagawa 2006: 240). Because of this, it would seem formally justified to project this form onto the Proto-|Gwi-‖Gana level; however, external data show that it must clearly be an innovation, and that only ‖Gana *ǀ̃ánà* is a reflex of the original PNKK root for the concept. More data, including representative semantic contexts, are necessary to clarify the picture here.
- Shua: ***ǀ̃ana*** (Cara *ǀ̃ànáˀ*, |Xaise *ǀ̃ànà*, Deti *ǀ̃ã̌*, Danisi *ǀ̃ànáˀ*).
- Ts’ixa: ***ǀ̃àná*** [Vossen], ***ǀ̃áná*** [Fehn].
- Tsua: ***ǀ̃ana*** (Tjwao *ǀ̃ana*; Hie. *ǀ̃ana*). ◊ Kua *tʰáná*, Tsua *tʰànáˀ* are notably similar to |Gwi-‖Gana *dáˢnà*, etc., but do not correspond to it regularly, once again, suggesting independent borrowing from some common substrate. Cf. also Cua *tʰănà* ‘to blossom’ as a possible cognate.
- Cua: **qàrū**. ◊ This is clearly related to Ts’ixa *qárù*, Danisi *qàrù* ‘to sprout’ (Vossen 1992: 385), although this root, with the rare initial **q-*, finds no parallels outside of the Eastern NKK area.
- NKK: ***ǀ̃ana***. ◊ Despite all the individual difficulties, this root is well reconstructible for PNKK based on combined evidence from at least Khwe, Shua, and Tsua branches, as well as "residual" evidence from Naro and ‖Gana.
- CK: ≈ ***ǀ̃ana***. ◊ The NKK root is cognate with PKK **|ã-* ‘grass’ (N *ǀ̃ãã̋-*, K */ã-b*); the NKK word for ‘grass’ is **dòá* (Vossen 1997: 445), with no equivalents in KK, so the semantic shift ‘leaf’ → ‘grass’ in PKK is quite likely. It is impossible not to notice the resemblance to Taa (South Khoisan) forms, cf. !Xóõ *ǀ̃āna*, Masarwa *ǀ̃ana* ‘leaf’ (Starostin 2021: 132), but this is most likely

the result of borrowing into Taa from a NKK source (probably |Gwi-‖Gana). A general note should perhaps be made here that, on the whole, the concept of 'leaf' is notoriously unstable in the "Khoisan" area, which makes the relative ease with which it is reconstructed at least for PNKK a local typological oddity; note how easily various languages adopt different strategies of expression after the disintegration of CK (e.g. the common African shift 'leaf' → 'ear' in KK, etc.).

Preliminary observations (by way of conclusion)

In Table 1, I summarize the results of intermediate reconstructions for Proto-Khoekhoe (PKK) and Proto-Non-Khoekhoe (PNKK), as well as for the top level of Proto-Central Khoisan = Proto-Khoe (PCK). Reconstructions in square brackets represent probable lexical / semantic innovations in the respective intermediate protolanguage. Question marks in the PCK column note the impossibility of making an "optimal" choice between PKK and PNKK given our current state of knowledge on Khoe etymology (note that this can mean one of *three* possible options, including a scenario in which the original equivalent was replaced independently in both primary nodes of the family). ≈ marks a questionable situation in which the decision is based upon inconclusive (if valid) semantic arguments.

Additionally, I also list all the semantic connections (both polysemies and historically more or less certain semantic shifts) between the analyzed etyma and other meanings, as such a list might be useful for anybody interested in the general diachronic typology of semantic change; for details on particular connections (names and number of languages, direction of shift, etc.) the actual data lists in the paper should be consulted.

Based on the table, the following points of interest should probably be noted.

1. It is relatively easy to reconstruct 24 out of 25 concepts for both primary branches of Khoe, with the notable exception of 'leaf' in PKK (but not in PNKK, where it is almost surprisingly well-reconstructible, given the general "aversion" toward the concept in Khoisan and other African families).

2. On the other hand, only 12 out of 25 concepts are completely unambiguously reconstructible for PCK based on their preservation in both primary branches (very strictly speaking, this number could even be reduced to 11 if we consider the re-evaluation of **ti*= as a basic pronominal morpheme in PNKK to fall under the category of lexical replacement).

3. In 5 cases, we have a good reason to postulate lexical replacement from PCK to PKK: 'pith/stone' → 'bone', 'dried out / withered / barren' → 'dry', 'to see' → 'eye' (parallel to 'eye' → 'to wake up'), 'head' → 'first, prime', 'hear' → 'hearken, listen attentively'. At the same time, there is not a single case where the collective phonetic, semantic, and distributional properties of the respective lexemes would make us strongly suspect such an innovation in PNKK.

This last conclusion is particularly striking. Although replacements in individual languages, as seen from the data above, do happen from time to time (and will likely get even more frequent when we advance to the overall less stable parts of the Swadesh wordlist), and although there are still 7–8 more items for which we cannot make a definitive choice on the PCK level (so, in pure theory, all of them could be innovations in PNKK), for now the data suggest that Proto-Khoekhoe may have replaced a significantly larger part of its core basic lexicon than Proto-Non-Khoekhoe — either gradually or over the course of some abrupt historical event, such as a language shift. In light of the serious anthropological and cultural distinctions between speakers of Khoekhoe and Kalahari Khoe, the latter seems like a more probable scenario, but more on this issue shall be said in the subsequent parts of the current study.

Table 1. Khoekhoe, Non-Khoekhoe, and Proto-Central Khoisan reconstructions for Swadesh items 1–25.

Word	PCK	PKK	PNKK	Semantic connections
ASHES	**?**	**t^{h}ao-*	**ǂoa-*	‘charcoal’; ‘flame’; ‘fireplace’; ‘clay’
BIRD	**?**	**kxani-*	**ʒ(a)era*	‘vulture’
BLACK	***ǂ̃u**	**ǂ̃u*	**ǂ̃u*	
BLOOD	***ǀˀao**	**ǀˀao-*	**ǀˀáò*	
BONE	***ǀˀõã**	[**ǂxo-*]	**ǀˀõã*	‘stone, pith’ (of fruit)
CLAW (= NAIL)	***ǁoro**	**ǁoro-*	**ǁoro (+*ǁa)*	‘claw’
DIE	***ǁˀo**	**ǁˀo*	**ǁˀó*	
DOG	**?**	**ari-*	**(h)a^{ʕ}ri- (+*ʔaba)*	
DRINK	***kxa**	**kxa*	**kxâ*	
DRY	***ǁxo**	[**ǀˀo*]	**ǁxo*	‘dangerous’, ‘unpleasant’; ‘barren’, ‘dried out’
EAR	***ǂe**	**ǂae*	**ǂé*	‘hear’
EAT	***ǂˀũ**	**ǂˀũ*	**ǂˀũ*	‘swallow’
EGG	**?**	**!ˀubu*	**ǂˀubi*	‘ostrich egg’; (?) ‘round’
EYE	***ǂxai**	[**mũ-*]	**ǂxai*	‘see’; ‘wake up’
FIRE	***ǀˀe**	**ǀˀae-*	**ǀˀe*	‘kindle (fire)’
FOOT	**?**	**ǂˀai-b*	**ǃ̃are*	‘kick’; ‘ride / drive’; ‘toe’, ‘talon’
HAIR	***ǀˀũ**	**ǀˀũ-b*	**ǀˀũ*	
HAND	**?**	**!ˀom-i*	**c^{h}au*	‘fist’; ‘lower arm’
HEAD	≈ ***ǂu**	[**dana-*]	**ǂu*	‘hear, perceive’; ‘first, prime’
HEAR	***kum**	[**ǁ̃ãũ*]	**kum*	‘hearken, listen closely’; ‘feel’
HEART	***ǂao**	**ǂao-b*	**ǂao*	
HORN	***ǁ̃ã**	**ǁ̃ã-b*	**ǁ̃â*	
I	***(ti=)ta ~ *(ti=)de**	**(ti=)ta ~ *(ti=)de*	**tí(-de)*	
KILL	**?**	**!am*	**ǀkxũ*	‘stop, cease’; (?) ‘throw away’
LEAF	≈ ***ǀ̃ana**	—	**ǀ̃ana*	‘ear’; ‘grass’; ‘green’; ‘sprout’; ‘blossom’

References

Barnard, Alan. 1985. *A Nharo Wordlist with Notes on Grammar*. Durban: University of Natal.

Baucom, Kenneth. 1974. Proto-Central Khoisan. In: E. Voeltz (ed.). *Third Annual Conference on African Linguistics, 7–8 April 1972*: 3–37. Bloomington: Indiana University.

Chebanne, Andy, Mawande Dlali. 2020. Cua language of Botswana: Resilience or a longer road to language loss? *South African Journal of African Languages* 2020: 1–8. doi: 10.1080/02572117.2020.1733831

CLICS = Rzymski, Christop, Tiago Tresoldi et al. 2019. *The Database of Cross-Linguistic Colexifications*. Available online at: clics.clld.org.

CODASELF = George Starostin et al. 2017–2025. *Comparative Onomasiological Database for the Ancestral States of Eurasian Linguistic Families*. Available online at: starlingdb.org/new400.

Collins, Chris, Zachary Wellstood. Forthc. *A Grammatical Sketch of Cua*. Berlin / Bruxelles: Peter Lang.

DatSemShift = Anna A. Zaliznyak et al. 2024. *Database of Semantic Shifts in languages of the world (v. 3.0)*. Available online at: datsemshift.ru.

Dornan, Samuel S. 1917. The Tati Bushmen (Masarwas) and their language. *Journal of the Royal Anthropological Society of Great Britain and Ireland* 47: 37–112.

Du Plessis, Ménan. 2018. *Kora: A lost Khoisan language of the Early Cape and the Gariep*. Cape Town: University of South Africa Press.

Elderkin, Edward D. 2004. The starred tones of Central Khoisan. *Afrika und Übersee* 87: 3–77.

Elderkin, Edward D. 2008. Proto-Khoe tones in Western Kalahari. In: Sonja Ermisch (ed.). *Khoisan Languages and Linguistics*: 87–136. Köln: Rüdiger Köppe Verlag.

Fehn, Anne-Maria. 2014. *A Grammar of Ts'ixa (Kalahari Khoe). Ph.D. thesis*. Köln: Institut für Afrikanistik.

Fehn, Anne-Maria, Jorge Rocha. 2024. Lost in translation: A historical-comparative reconstruction of Proto-Khoe-Kwadi based on archival data. *Diachronica* 40(5): 609–665.

Güldemann, Tom. 2006. Structural Isoglosses between Khoekhoe and Tuu: The Cape as a Linguistic Area. In: Y. Matras, A. McMahon, N. Vincent (eds.). *Linguistic areas: Convergence in historical and typological perspective*: 99–134. Basingstoke: Palgrave Macmillan.

Güldemann, Tom. 2013. Typology. In: Rainer Vossen (ed.). *The Khoesan Languages*: 25–37. London / New York: Routledge.

Haacke, Wilfrid, Eliphas Eiseb. 2002. *A Khoekhoegowab Dictionary with an English-Khoekhoegowab Index*. Namibia: Gamsberg Macmillan.

Haacke, Wilfrid, Eliphas Eiseb, Levi Namaseb. 1997. Internal and external relations of Khoekhoe dialects: a preliminary survey. In: Wilfrid Haacke, E. D. Elderkin (eds.). *Namibian languages: reports and papers*: 125–203. Köln: Rüdiger Köppe Verlag.

Haacke, Wilfrid, Jan W. Snyman. 2019. Lexical proximity of a Xri corpus to Khoekhoegowab. *Studies in African Linguistics* 48(2): 267–328.

Honken, Henry. 2008. The split tones in Central Khoesan. In: Sonja Ermisch (ed.). *Khoisan Languages and Linguistics*: 185–224. Köln: Rüdiger Köppe Verlag.

Jungraithmayr, Herrmann, Dymitr Ibriszimow. 1994. *Chadic Lexical Roots. Volume I*. Berlin: Dietrich Reimer Verlag.

Kilian-Hatz, Christa. 2003. *Khwe-English Dictionary*. Köln: Rüdiger Köppe Verlag.

Meinhof, Carl. 1930. *Der Koranadialekt des Hottentotischen*. Berlin: Dietrich Reimer Verlag.

Nakagawa, Hirosi. 1996. A first report on the click accompaniments of |Gui. *Journal of the International Phonetic Association* 26(1): 41–54.

Nakagawa, Hirosi. 2006. *Aspects of the phonetic and phonological structure of the G|ui language. Ph.D. Thesis*. Johannesburg: University of the Witwatersrand.

Nakagawa, Hirosi. 2011. *Genetic affiliations of ǂHaba and Tshila*. Presentation given at: ICHL-40, 2011/07/29, Minpaku, Osaka.

Phiri, Admire. 2021. *The Verbal and Nominal Morpho-syntax of Tjwao: A Grammaticalisation Approach. Ph.D. dissertation*. Cape Town: Stellenbosch University.

Rust, Friedrich Jr. 1969. *Nama Wörterbuch (Krönlein Redivivus)*. Pietermaritzburg: University of Natal Press.

Starostin, George. 2017. *Annotated Swadesh wordlists for the Khoekhoe group (Central Khoisan family)*. Ms. Available online at: starlingdb.org/new100/kkh.pdf.

Starostin, George. 2018. Lexicostatistical Studies in Khoisan I: The Ju-ǂHoan Relationship. *Journal of Language Relationship* 16(2): 19–61. doi: 10.31826/jlr-2018-161-205

Starostin, George. 2020. *Annotated Swadesh wordlists for the Kalahari Khoe group (Central Khoisan family)*. Ms. Available online at: starlingdb.org/new100/kho.pdf.

Starostin, George. 2021. Lexicostatistical Studies in Khoisan II/1: How to make a Swadesh wordlist for Proto-Tuu. *Journal of Language Relationship* 19(1-2): 99–135. doi.org/10.1515/jlr-2021-191-211

Starostin, George. 2022. Lexicostatistical Studies in Khoisan II/2: Towards a more precise phylogeny for the Tuu family. *Journal of Language Relationship* 20(1–2): 25–70. doi: 10.1515/jlr-2022-201-205

Starostin, Georgiy. 2013. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoj klassifikacii. Tom I: Metodologija. Kojsanskije jazyki*. Moskva: Jazyki slav'anskoj kul'tury.

Starostin, Sergei. 2000. Comparative-historical Linguistics and Lexicostatistics. In: Colin Renfrew, April McMahon, Robert Lawrence Trask (eds.). *Time Depth in Historical Linguistics*: 223–259. Cambridge: McDonald Institute for Archaeological Research.

Tanaka, Jiro. 1978. *A San Vocabulary of the Central Kalahari: G‖ana and G|wi Dialects*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Visser, Hessel. 2001. *Naro Dictionary*. Botswana: Printing & Publishing Co.

Vossen, Rainer. 1986. Zur Phonologie der ‖Ani-Sprache. In: Rainer Vossen, Klaus Keuthmann (eds.). *Contemporary studies on Khoisan (Festschrift Oswin R. A. Köhler), v. 2*: 321–345. Hamburg: Helmut Buske Verlag.

Vossen, Rainer. 1992. Q in Khoe: borrowing, substrate or innovation? In: Derek F. Gowlett (ed.). *African linguistic contributions (Festschrift Ernst Westphal)*: 363–388. Hamburg: Helmut Buske Verlag.

Vossen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe Verlag.

Vossen, Rainer, Sabine Neumann, Christina Patriarchi, Margit Rottland, Rainer Spörl, Beate Vagt. 1988a. Khoe Linguistic Relationships Reconsidered. *Botswana Notes and Records* 20: 61–70.

Vossen, Rainer, Sabine Neumann, Christina Patriarchi, Margit Rottland, Rainer Spörl, Beate Vagt. 1988b. Khoe Linguistic Relationships Reconsidered: The Data. In: Rainer Vossen (ed.). *New perspectives on the study of Khoisan. Quellen zur Khoisan-Forschung / Research in Khoisan studies, Bd 7*: 67–108. Köln: Rüdiger Köppe Verlag.

Vossen, Rainer. 2013. Introduction. In: Rainer Vossen (ed.). *The Khoesan Languages*: 1–12. London / New York: Routledge.

Wellstood, Zachary. 2024. *Cua Dictionary*. Ms. (unpublished).

Wuras, Carl Friedrich. 1920. *Vokabular der Korana-Sprache. Herausgegeben und mit kritischen Anmerkungen versehen von Walther Bourquin*. Berlin: Dietrich Reimer Verlag.

Г. С. Старостин. Лексикостатистические исследования по койсанским языкам III/I: реконструкция списка Сводеша для працентрально-койсанского (пракхойского) языка (элементы 1–25)

Статья представляет собой первую часть большого исследования, в рамках которого пересматривается и дополняется лексикостатистический анализ языков центрально-койсанской (кхойской) семьи на территории Южной Африки, опубликованный автором более 10 лет назад. В рамках данной части дается краткое введение, посвящающее читателя в наиболее актуальные текущие проблемы реконструкции и внутренней классификации кхойской семьи, за которым следует детальный этимологический анализ первых 25 элементов из «суперстабильной» части списка М. Сводеша. В качестве предварительного вывода отмечается, что группа кхойкхой, представляющая собой одну из двух первичных ветвей семьи, оказывается существенно более инновативной в плане лексических замен, чем группа калахари-кхой (вторая из двух ветвей).

Ключевые слова: койсанские языки; кхой языки; лексикостатистический анализ; ономасиологическая реконструкция; историческая семантика.