


RESEARCH ARTICLE

Tracing contact and migration in pre-Bantu Southern Africa through lexical borrowing

Anne-Maria Fehn^{1,2} , Bonny E. Sands³, Admire Phiri^{1,2,4}, Maitseo Bolaane⁵,
Gaseitsiwe Masunga⁶, Ezequiel Fabiano^{1,2,7} and Jorge Rocha^{1,2,8}

¹CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Porto, Portugal; ²Biopolis Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal; ³Department of English, Northern Arizona University, Flagstaff, AZ, USA; ⁴Department of Linguistics, University of the Free State, Bloemfontein, South Africa; ⁵San Research Centre, University of Botswana, Gaborone, Botswana; ⁶Okavango Research Institute, University of Botswana, Maun, Botswana; ⁷University of Namibia, Windhoek, Namibia and ⁸Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
Corresponding author: Anne-Maria Fehn; Email: afehn@cibio.up.pt

Abstract

Lexical borrowing may provide valuable clues about the sociohistorical context of language contact. Here we explore patterns of vocabulary transfer between languages from three families (Kx'a, Tuu, Khoe-Kwadi) comprising the linguistic unit commonly referred to as Southern African Khoisan. In our data set, 20% of 1,706 roots are shared between at least two families. By applying a carefully chosen set of linguistic and extralinguistic criteria, we were able to trace the origin of 71% of shared roots, with the remaining 29% constituting good candidates for ancient contact or shared common ancestry of the forager families Kx'a and Tuu. More than half of the shared roots for which an origin could be determined trace back to Khoe-Kwadi and were borrowed into languages of other families within two major confluence zones with different sociohistorical profiles: (i) the Central Kalahari characterized by egalitarian interaction between languages of all three families and (ii) the southern and south-western Kalahari Basin fringes showing unilateral transfer from Khoe-Kwadi-speaking herders into resident forager groups. The findings of this study complement genetic and archaeological research on southern Africa and testify to the value of linguistics in the multidisciplinary inference of contact and migration scenarios.

Keywords: Southern African prehistory; Southern African Khoisan; evolutionary linguistics; quantitative linguistics; lexical borrowing; language contact

Social media summary: Lexical borrowing between click languages of three distinct families provides new insights into the pre-history of southern Africa.

1. Introduction

Multidisciplinary research findings combining evidence from linguistics, archaeology and genetics support the existence of three major population layers in pre-colonial southern Africa (Fehn et al., 2022): (i) an ancient foraging layer presently represented by speakers of Kx'a and Tuu languages; (ii) a Late Stone Age migration of sheep herders from eastern Africa associated with languages of the Khoe-Kwadi family; and (iii) the agriculture-related dispersal of Bantu-speaking Iron Age farmers from West Central Africa.

The pre-Bantu languages of the Kx'a, Tuu and Khoe-Kwadi families all have unusually large phoneme inventories and share cross-linguistically rare click sounds which are also attested in the

Tanzanian languages Hadza and Sandawe. Based on this feature and other superficial commonalities observed during the mass comparison of fragmented data sources, Greenberg (1963) grouped all non-Bantu, non-Cushitic click languages of southern and eastern Africa into a macro-family he dubbed ‘Khoisan’. Although the concept of a ‘Khoisan’ family remains widely used among non-linguists, scholars actively working on the languages in question have failed to back up the genealogical validity of this claim by means of the comparative method or morphological reconstruction (Güldemann, 2008a, 2014a; Sands, 1998; Westphal, 1971). Instead, bottom-up reconstruction involving newly collected data from a variety of languages has led to the establishment and widespread acceptance of three southern African clades (Kx’a, Tuu, Khoe-Kwadi) and two eastern African isolates (Hadza, Sandawe) (Fehn & Rocha, 2023; Güldemann, 2005, 2014a; Heine & Honken, 2010; Vossen, 1997). Although the possibility of a deeper-level relationship between Khoe-Kwadi and Sandawe on the one hand and Kx’a and Tuu on the other has been entertained, no conclusive reconstruction has been attempted for either link (Güldemann, 2014a).

Despite the present inability to show a genealogical connection, it is widely acknowledged that languages of the Kx’a, Tuu and Khoe-Kwadi families share many typological features, including similar sound inventories and syllabic constraints (Güldemann, 2008b; Güldemann & Fehn, 2017; Nakagawa *et al.*, 2023). They are therefore thought to have a history of long-term interaction in a linguistic contact area roughly delimited by the Kalahari Basin, which is here referred to by the term ‘Southern African Khoisan’ (SAK) (Güldemann, 1998, 2014a). Within this area, Kx’a and Tuu are closer to one another than to Khoe-Kwadi.

Previous studies on the role of contact in forming the Kalahari Basin linguistic area mostly concentrated on the transfer of typological features from resident Kx’a and Tuu foragers to incoming Khoe-Kwadi-speaking herders (Güldemann, 2006; Güldemann & Fehn, 2017). Güldemann and Fehn (2017) suggest a gradient of adherence to the SAK typological profile, with Khoe-Kwadi languages spoken in and around the Central Kalahari displaying more influence from Kx’a and Tuu than their relatives which remained at the fringes of the contact area. However, advances in the documentation of individual Khoe-Kwadi languages have shown that some features previously ascribed to relatively recent Kx’a and Tuu influence are actually widespread across the family, indicating the SAK profile may have extended well into eastern Africa, or was obtained at the Proto-Khoe-Kwadi stage, immediately after reaching southern Africa (Fehn & Phiri, 2022; Fehn & Rocha, 2023).

Despite their possibly contact-induced adherence to the Kalahari Basin typological profile, it has been implicitly assumed that Khoe-Kwadi speakers, due to their pastoral subsistence, became dominant in the region before the arrival of Bantu-speaking farmers. However, there is little evidence to support this model. Genetically, Kx’a and Tuu speakers had considerable influence on the genetic profile of modern Khoe-Kwadi speakers, all of whom display varying degrees of ancestry derived from southern African forager groups while retaining only small fragments of their original eastern African genetic heritage (Oliveira *et al.*, 2023; Pickrell *et al.*, 2014). Furthermore, only speakers of the Khoekhoe and Kwadi branches of Khoe-Kwadi have been associated with pastoralism in historical times (Barnard, 1992; Fehn & Rocha, 2023).

The situation within the Kalahari Khoe – the family’s biggest and most diverse branch – is more complex: although the raising of small-stock has been observed among speakers of Kalahari Khoe in the Central Kalahari since the 1960s (Ikeya, 1993; Russell, 2017; Tanaka, 1969), the focus on goats suggests a Bantu origin of herding practices, rather than a retention of an older, sheep-centric, pattern. Furthermore, all Kalahari Khoe speakers, irrespective of their current access to livestock, share a foraging lifestyle and worldview with their Kx’a and Tuu-speaking neighbours (Barnard, 1992). Whereas the present situation testifies to the relative simplicity in which hunter-gatherers may pick up and integrate livestock into their traditional lifeways (Russell, 2017), the striking differences between the subsistence patterns of the Khoekhoe on the one hand and the Kalahari Khoe on the other have led to intense speculation on whether modern Kalahari Khoe speakers are former herders who lost their livestock, or simply Kx’a- and Tuu-speaking hunter-gatherers who changed their language (Fehn *et al.*,

2022; Güldemann, 2008b). In a unifying attempt, it has further been suggested that the ancestors of the Khoe-Kwadi were hunter-herders who led a life combining foraging and livestock-keeping practices, effectively allowing their descendants to switch between both subsistence practices in response to environmental and societal triggers (Fehn et al, 2022; Sadr, 2008).

As lexical borrowing – that is, the horizontal transfer of vocabulary from one language to another – is often associated with cultural innovations or elevated prestige of one language over another (Thomason & Kaufman, 1988), the vocabulary domain seems especially prone to provide insights into social relations between Kx'a, Tuu and Khoe-Kwadi speakers in the pre-Bantu population landscape of southern Africa. However, no large-scale quantitative assessment of lexical borrowing across the entire SAK area has been attempted, even though the cross-family sharing of vocabulary has long been noted and significantly contributed to Greenberg's (1963) original 'Khoisan' proposal (Gerlach, 2016; Güldemann & Loughnane, 2012; Sands, 1998, 2001; Traill & Nakagawa, 2000). More generally, the historical implications of lexical borrowing across large-scale multilingual data sets have been neglected in favour of assessing the borrowability of semantic fields and meanings (e.g. Kelih, 2024; Tadmor et al, 2010; van Hout & Muysken, 1994), as well as on the automatic detection of borrowings in data sets used for genealogical language classification (e.g., (List & Forkel, 2022; Nelson-Sathi et al, 2011). Still, lexical borrowings are more than just an obstacle to uncover the true relationships between languages; they contain part of a language's history, thereby granting us a glimpse into the history of its speakers and the way they interacted with others.

In order to uncover the history behind the complex network of lexical borrowing in the Kalahari Basin, we here apply a bottom-top approach which notably differs from mass comparison based on superficial similarity in that the detection of borrowings is combined with the careful historical-comparative reconstruction of lexical roots for each previously established genealogical unit (Kx'a, Tuu, Khoe-Kwadi), taking into account known and newly detected patterns of sound correspondence. To this end, we devised a set of diverse criteria for establishing the direction of transfer: these include language-internal criteria linked to reconstructability, attestation and phonological complexity which allow us to detect violation of established rules of sound change within each family, as well as additional criteria like the presence of alternative roots, historical parsimony and geography which may help to decide on more complex cases for which phonological criteria remain inconclusive. Rather than focusing on the small number of Kx'a, Tuu and Khoe-Kwadi languages for which dictionaries are available, we base our analysis on a unique data set consisting of 57 doculects including published, archival and newly collected data from both vital and extinct languages which belong to all three families and cover a wide geographic area spanning Angola, Namibia, Botswana, Zimbabwe and South Africa, today known as Southern African Development Community (SADC; Fig. 1, Supplementary Table S1). Using 228 meanings encompassing words from the Swadesh 200 list of basic vocabulary and additional meanings widely attested across southern African languages (Supplementary Table S2), we identify a total of 1,706 lexical roots (Supplementary Table S3), 20% of which are shared between languages from at least two families (Supplementary Table S4). We were able to identify 71% of shared roots as borrowings, that is, roots for which an origin within a given family, as well as a distinct direction of transfer (e.g. from Kx'a to Khoe-Kwadi), could be determined. Of those, 41% originated within the Khoe-Kwadi family, whereas only 20% and 10% originated in Kx'a and Tuu respectively. Roots that originated in Khoe-Kwadi are most prominent in languages of the Tuu family, whose speakers were in contact with Khoekhoe-speaking herders, suggesting elevated prestige of food producers among resident forager groups. The data further yielded 29% of shared roots for which no origin could be determined. We do not treat them as borrowings but consider them unresolved cases that may be indicative of ancient contact, or a distant common ancestor shared by Kx'a and Tuu, in line with previous proposals (Collins & Honken, 2016; Güldemann, 2014a). We note that our findings on borrowing also hold if a deep genealogical relationship between all three families (Kx'a, Tuu, Khoe-Kwadi) is assumed, given that such a relationship would significantly pre-date

the historical movements under consideration, and go much beyond the presently established sound correspondences which form the basis of our analysis.

The findings of this study complement genetic and archaeological research on southern Africa as a contact zone and testify to the value of linguistics in the multidisciplinary inference of contact and migration scenarios.

2. Material and methods

2.1. Assembly of the data set

This study is based on a list of 228 meanings initially assembled for 63 doculects from the Kx'a, Tuu and Khoe-Kwadi families, from study locations in Angola, Namibia, Botswana, Zimbabwe and South Africa (Fig. 1). The wordlist consists of 132 meanings from the Swadesh 200 wordlist of basic vocabulary (Swadesh, 1952) and 96 additional meanings for which we could obtain good data coverage, particularly for under-attested and extinct languages of the !Ui (Tuu) cluster. Swadesh 200 meanings that were either (a) function words (grammatical items like pronouns or adverbs), (b) for which the data coverage was insufficient or (c) which have no generic expression in the data at hand (like 'leg' or 'burn') were removed.

The database combines newly collected data from Kx'a (Northern !Xun: Mashangara, Nkurenkuru; Northwestern !Xun: Mupa) and Khoe-Kwadi (Shua: Deti, Shua, Danisi; Khwe: !Ani, Buga, G'landa, !Xoo, Ts'ixa) with archival and published sources. The name, linguistic affiliation, subsistence pattern, sampling location and bibliographical source for each doculect is provided in Supplementary Table S1.

This study was developed within the TwinLab network of collaborations linking CIBIO/InBIO to ISCED (Instituto Superior de Ciências de Educação)–Huíla, Angola, to the University of Namibia, and to the Okavango Research Institute and the San Research Center of the University of Botswana. All newly collected data were gathered with official research permits issued by the Governments of Namibia and Botswana, and with the support and permission of the Provincial Governments of the Namibe and Kunene provinces (Angola).

2.2. Coding of the linguistic data

The linguistic data were coded as follows: for every meaning, we identified all roots attested in the data set, based on previous works and general knowledge of sound correspondences attested in the SAK languages (see Supplementary Text S1 and Supplementary Table S5 for a complete list of all historical-comparative works consulted). We then coded each language for the presence (1), absence (0), or lack of information (?) for a given root. Synonyms were allowed, that is, a language was permitted to have more than one root for a given meaning. However, a root can appear only once in the data set, even if it is attested with different meanings within the universe of our 228-meaning list. These cases of polysemy are always marked in the data set and filed under the cross-linguistically most widespread meaning of each root. Bantu borrowings and complex expressions (e.g. 'chest bone' for 'rib') were removed. To achieve a better data coverage in the !Ui group of Tuu, the three doculects for #Khomani (Doke, 1936; Köhler, n.d.; Maingard, 1937) and the five doculects for !Xegwi (Bleek, 1956; Köhler, n.d.; Lanham & Hallows, 1956; Westphal, 1953-1971; Ziervogel, 1955) were lumped into a single doculect, respectively. The division between individual sources is maintained in the lexical database (Supplementary Table S2). The final data set which all analyses are based on consists of 1,706 roots, based on 228 meanings, which were coded for a total of 57 doculects from the Kx'a, Tuu and Khoe-Kwadi families (Supplementary Table S3).

2.3. NeighborNet analysis

To create a network graph of our data set (Fig. 2B, Supplementary Figure S1), we used the NeighborNet algorithm implemented in the software SplitsTree 4.17.2 (Huson & Bryant, 2006),

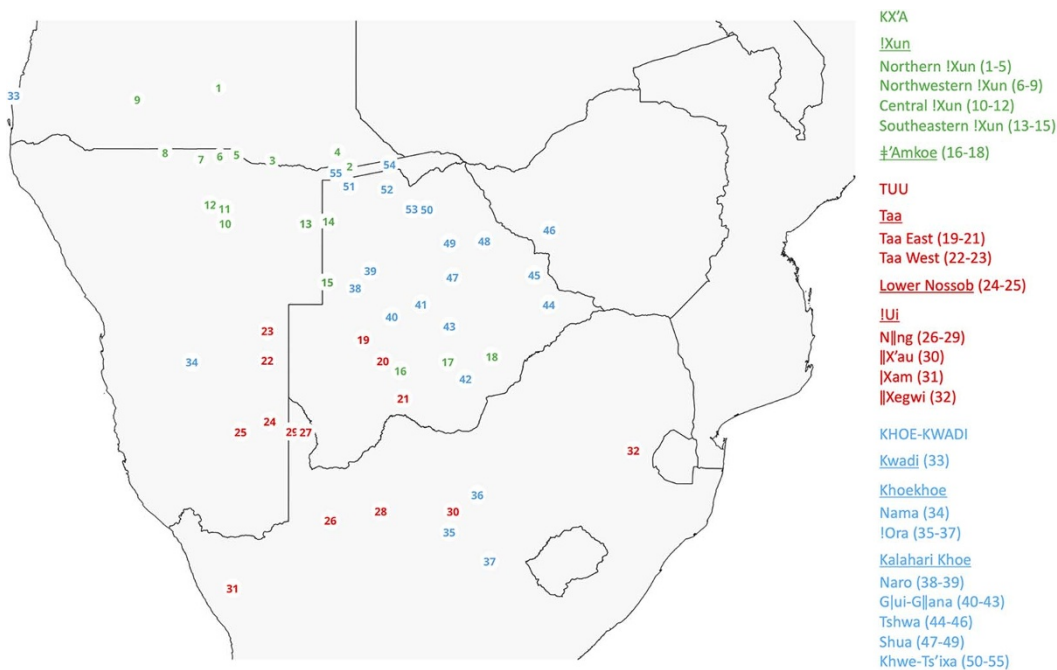


Figure 1. Map showing the approximate geographic origin of each data set used in the present study. Individual data sets are identified by language family and number as referenced in Supplementary Table S1. The legend groups doculects into language families and widely accepted subgroups (cf. Fehn & Rocha, 2023; Güldemann, 2014a; Heine & Honken, 2010; Heine & König, 2015; Vossen, 1997).

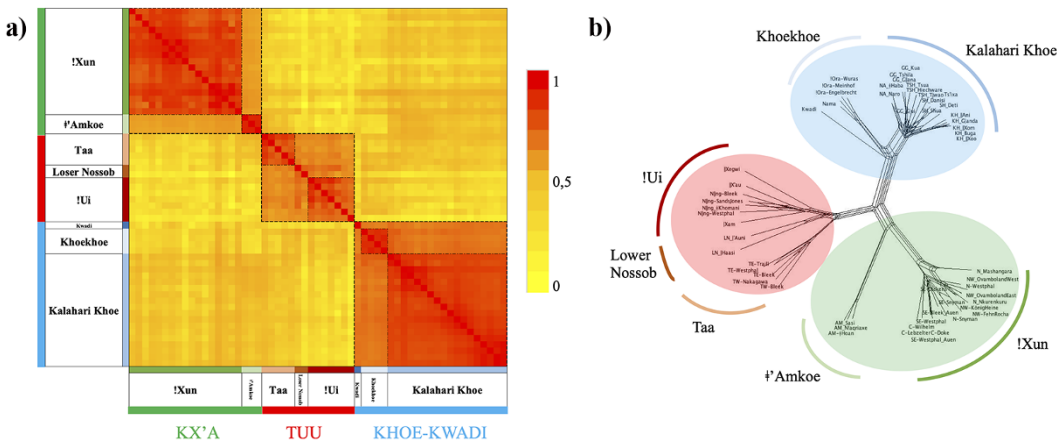


Figure 2. Language relationships as assessed with the GeneContent distance (Gray et al., 2010); both the heatmap based on linguistic similarity (a) and the NeighborNet (b) clearly identify the three families Kx'a (green), Tuu (red) and Khoe-Kwadi (blue), as well as their major subgroups. Lexical sharing between the three families is evident as reticulation in the centre of the network.

applying GeneContent distance and filtering weights at 0.005 as recommended by Gray et al. (2010). We further created a corresponding heatmap based on a similarity matrix $1 - \text{GeneContent distance}$, which was visualized in MS Excel (Fig. 2A).

2.4. Bayesian phylogenetic analysis

We performed a Bayesian phylogenetic analysis for each language family (Kx'a, Tuu, Khoe-Kwadi) by importing the ascertained binary data in the software Beast 2.5 (Bouckaert *et al.*, 2019). In order to obtain a basic phylogeny and posterior support values for individual language clusters, we opted for the Continuous Markov Chain Model (CTMC) included in the Babel package (<https://github.com/rbouckaert/Babel>) under a Yule model, following a strict clock. We acknowledge possible caveats of this simplified model choice (Hoffmann *et al.*, 2021), but as we do not use priors with the aim to date our phylogeny, we consider our analysis satisfactory for the purpose of this work. We ran each analysis for 20 million generations, during which we could obtain adequate values for effective sample sizes (ESS) of all parameters. Consensus trees were subsequently visualized in FigTree v.1.4.4 (Rambaut, 2018).

2.5. Establishing the source for shared roots

To determine the genealogical origin for each root, we defined a set of criteria adapted to the SAK linguistic area that take into account previous guidelines for the identification of borrowings (e.g. Güldemann & Loughnane, 2012; Haspelmath, 2009). We identified five primary criteria (reconstructability, scope of attestation, phonology and morphological analysability) as well as four secondary criteria (presence of lexical alternatives, geography, culture and historical parsimony). To determine the origin and direction of transfer for a given root, we usually employed more than one criterion. Details and examples for each criterion are provided in Supplementary Text S1.

2.6. Analyses of lexical sharing

Analyses of lexical sharing are based on roots shared between doculects from at least two families (Supplementary Table S4), comprising four types of sharing patterns: KX'A + TUU + KHOE-KWADI; KX'A + KHOE-KWADI; TUU + KHOE-KWADI; KX'A + TUU (Fig. 3B). Percentages of sharing were obtained for each doculect, and subsequently averaged across doculects forming a language cluster. Language clusters were established according to posterior probabilities obtained through a Bayesian phylogenetic analysis as shown in Supplementary Figures S2–4.

3. Results

3.1. Distance-based and Bayesian phylogenetic analyses

In a distance-based analysis, our data set reproduces the expected division into distinct families while clearly showing lexical overlap between all three of them (Fig. 2). In both the heatmap (Fig. 2A) and the NeighborNet analysis (Fig. 2B), the Kx'a, Tuu and Khoe-Kwadi families are distinguished, along with all known subgroups within each family: Kx'a is divided into two distinct branches, !Xun and †Amkoe (Heine & Honken, 2010); Tuu has two clearly defined branches, Taa and !Ui, as well as the Lower Nossob language cluster with unclear affiliation within the family (Güldemann, 2005; Starostin, 2021, 2022); Khoe-Kwadi consists of the outlier language Kwadi and the Khoe branch divided into the Khoekhoe and Kalahari Khoe subgroups (Fehn & Rocha, 2023; Güldemann, 2004; Vossen, 1997). The NeighborNet analysis further shows box-like structures in the centre of the network (Fig. 2B), indicating reticulation and therefore lexical sharing between languages of all three families. Reticulation is also visible within the !Xun branch of Kx'a and the Kalahari Khoe subgroup of Khoe-Kwadi, respectively, possibly pointing to family-internal contact between related languages spoken in close geographic proximity. The clear division into major families and subgroups also holds if singletons are removed from the data set, indicating that the differentiation within SAK is driven by broader sharing patterns and not by micro-variation within single doculects (Supplementary Figure S1).

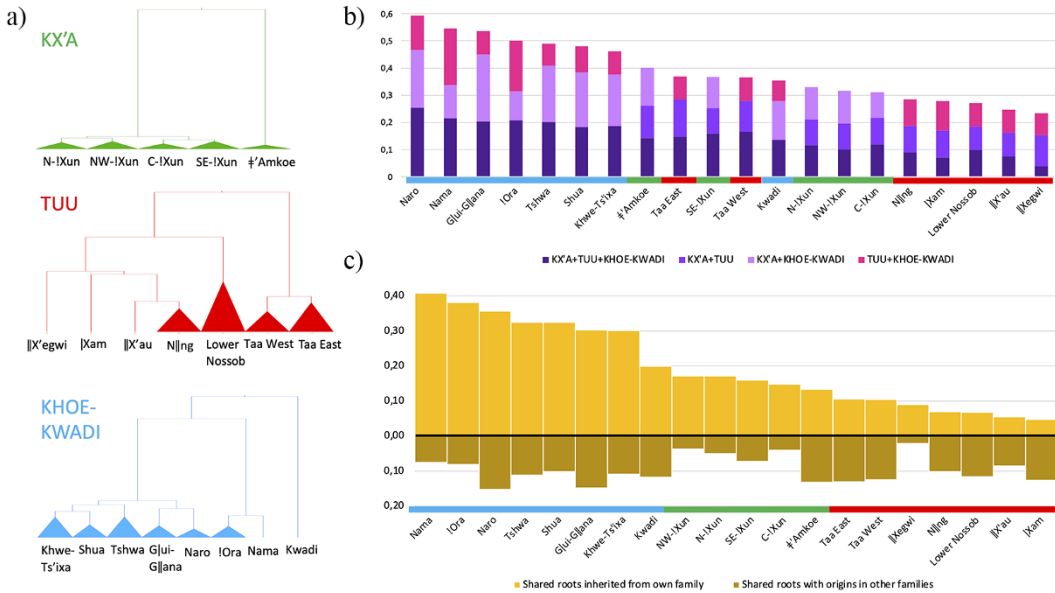


Figure 3. (a) Bayesian phylogenetic trees used to identify major language clusters within each family. (b) Proportion of roots with different sharing patterns in each language cluster. (c) Percentage of roots that were transferred from one's own family to other families (top bars) or that originated in other families; that is, loanwords (bottom bars). All proportions refer to fractions of total roots per language cluster. The horizontal bars below the bar plots indicate the language family: Kx'a (green), Tuu (red) and Khoe-Kwadi (blue).

To assess how the doculects included in our study align with previously identified language clusters within each family, we also performed a Bayesian phylogenetic analysis for Kx'a, Tuu and Khoe-Kwadi (Fig. 3A, Supplementary Figures S2–4). In accordance with other studies, the consensus tree for Kx'a shows a deep split between !Xun and †Amkoe (Heine & Honken, 2010), as well as a relatively shallow differentiation within each branch (Supplementary Figure S2). Our findings further replicate previous research on Kx'a subclassification (Heine & König, 2015; Sands, 2010; Snyman, 1997) in rendering support for a single †Amkoe cluster, made up of three doculects, as well as for four !Xun clusters, namely Northern, North-western, Central and South-eastern !Xun, consisting of at least three doculects each. The Tuu phylogeny clearly delimits a Taa branch divided into a western and an eastern cluster, and an !Ui branch, consisting of the N!ng cluster including N!uu and †Khomani, as well as the isolated doculects !X'au, !Xam and !Xegwi (Güldemann, 2014a; Supplementary Figure S3). The position of Lower Nossob (here grouped with !Ui) is problematic (see also Starostin, 2022): Lower Nossob shares grammatical features with Taa and has been included in a genealogical unit termed Taa-Lower Nossob (Güldemann, 2014b). The lexical affinities between Lower Nossob and !Ui may therefore be a secondary feature, related to family-internal language contact in the region of what is now the Kgalagadi Transfrontier National Park.

In line with a previous Bayesian phylogenetic analysis (Fehn et al., 2022), our tree for Khoe-Kwadi clearly replicates the major division between the isolated Kwadi branch and the Khoe languages, which are further subdivided into Khoekhoe and Kalahari Khoe (Supplementary Figure S4). Within Khoekhoe, the Nama language can be distinguished from the !Ora cluster represented by three closely related doculects. The Kalahari Khoe subgroup consists of five language clusters: Glui-Glana, Naro, Khwe-Ts'ixa, Shua and Tshwa.

Overall, we observe a close resemblance between the topologies inferred by qualitative linguistic studies and the results of our phylogenetic analyses, despite the obvious presence of contact-induced reticulation in the data set (cf. Greenhill et al., 2009).

4. Quantitative assessment of roots shared between at least two families

In the analyses of lexical sharing in SAK that follow, the languages and language clusters identified in the Bayesian study are used as units of comparison (Fig. 3A).

The affiliation of individual doculects to distinct language clusters is provided in Supplementary Table S1. In a first step, we identified all roots that appear in languages of more than one family. These amount to 20% (340/1,706) of the full data set. As much as 27% of shared roots were found in all three families, 21% in Kx'a and Tuu, 23% in Tuu and Khoe-Kwadi and 29% in Kx'a and Khoe-Kwadi. The distribution of shared roots across languages from different families shows that the highest percentage of shared roots is attested in languages of the Khoe branch of Khoe-Kwadi, reaching up to 60% in Naro (Fig. 3B).

We then determined the source family for each root, following the criteria outlined in Supplementary Text S1. We found that 41% of shared roots had originated within the Khoe-Kwadi family, whereas only 20% and 10% had originated within Kx'a and Tuu, respectively. Our criteria did not allow us to identify a definite origin for 29% of shared roots, which will be discussed further below. When ranked by the percentage of shared roots that originated in one's own family, the Khoe-Kwadi languages rank first, displaying sharing rates between 41% (Nama) and 20% (Kwadi); languages of the Kx'a family (17–13%) rank next, followed by the Tuu family (11–5%) (Fig. 3C).

Most languages in our data set share more roots through transfer from their own to other families than through borrowing (Fig. 3C). The highest percentages of borrowed vocabulary are displayed by languages of all three families spoken in the Central Kalahari, namely Naro, Glui-Glana (Khoe-Kwadi, 15% each), †Amkoe (Kx'a, 13%) and Taa East (Tuu, 13%). Relatively high amounts of borrowed vocabulary are also found in languages of the Tuu family spoken along the western Kalahari Basin fringe, namely Taa West (13%), !Xam (13%) and Lower Nossob (12%). The lowest percentages of borrowings are found in languages of the !Xun branch of Kx'a spoken in the northwest (7–4%), as well as in the easternmost Tuu language !Xegwi (2%).

When borrowings in individual languages are broken down by family of origin (Fig. 4A), distinct, family-internal patterns emerge.

In Kx'a (Supplementary Figure S5A), languages of the !Xun branch spoken on the north-western Kalahari Basin fringe display only few borrowings, which are almost exclusively associated with the Khoe-Kwadi family. In contrast, the †Amkoe branch located in the Central Kalahari of Botswana has considerably more loanwords that derive from both Khoe-Kwadi and Tuu, reflecting the linguistic diversity of its neighbours.

In the Tuu family (Supplementary Figure S6A), the Taa branch of the Central and Western Kalahari derives about three-quarters of its borrowed vocabulary from Khoe-Kwadi, and about one-quarter from Kx'a. Further south, languages of the !Ui branch display mostly loanwords of Khoe-Kwadi origin, which are considerably more frequent in the western part of South Africa than in areas further east.

In the Khoe-Kwadi family (Supplementary Figure S7A), loanword frequencies are highest in Naro and Glui-Glana spoken in the Central Kalahari; intermediate in Kwadi, Khwe-Ts'ixa, Shua and Tshwa spoken along the northern Kalahari Basin fringe; and lowest in the Khoekhoe varieties spoken by Nama and !Ora herders in Namibia and South Africa. Although loanwords in Khoe-Kwadi languages predominantly trace back to Kx'a, Tuu contributions are equally present in all three main branches (Kwadi, Khoekhoe, Kalahari Khoe), in line with previous suggestions of both Kx'a and Tuu influence on the ancestral language Proto-Khoe-Kwadi (Fehn & Rocha, 2023).

In order to assess the most recent contact layer for individual languages from different geographic regions, we divided the identified loanwords by their resolution level within the family of origin (Fig. 4B–D), assuming that roots which can be reconstructed to lower-level subgroups would provide additional information on the affiliation of the donor.

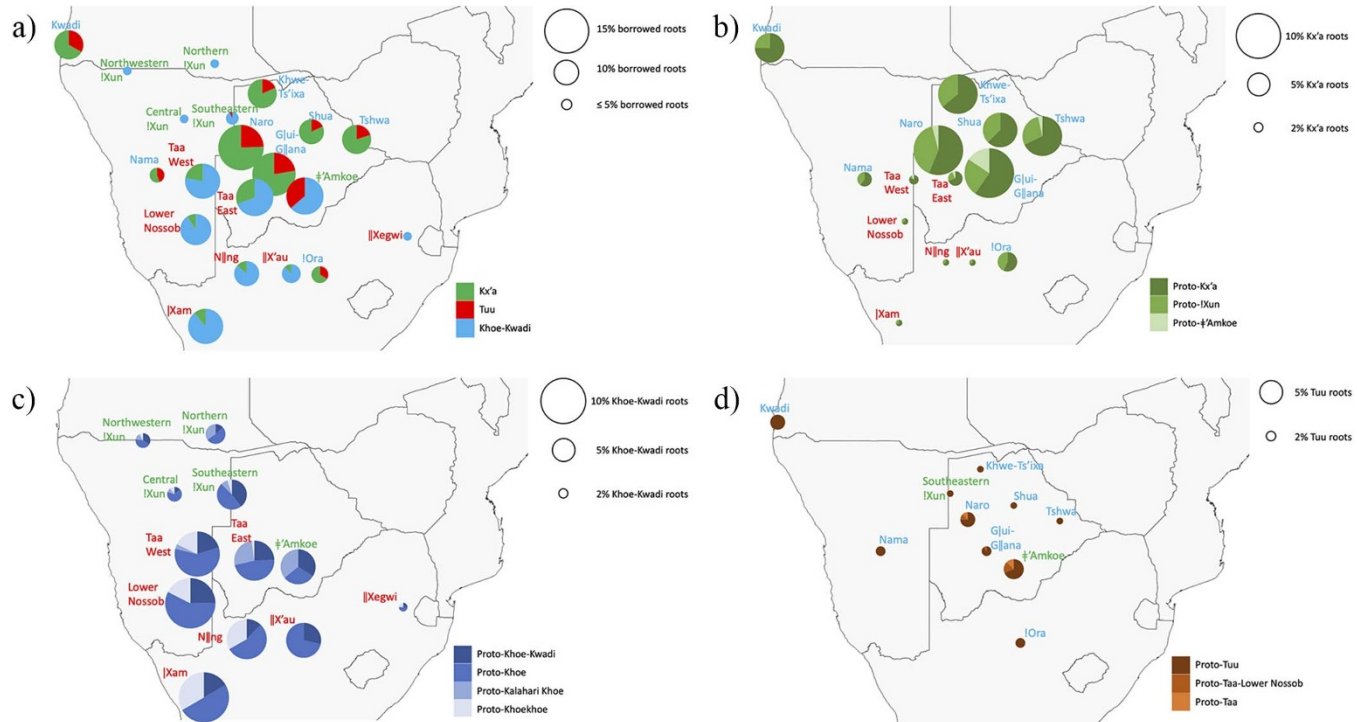


Figure 4. Borrowing profiles of Kx'a, Tuu and Khoe-Kwadi languages. The size of the pie chart corresponds to the percentage of borrowed vocabulary, as indicated for each figure. (a) shows the composition of the borrowed vocabulary in each language; (b) shows Kx'a borrowings in languages from the Tuu and Khoe-Kwadi families, broken down to the highest resolution level that could be obtained for each root (Proto-Kx'a, Proto-!Xun, Proto-ǀAmkoe); (c) shows Khoe-Kwadi borrowings in languages from the Kx'a and Tuu families, broken down to the highest resolution level that could be obtained for each root (Proto-Khoe-Kwadi, Proto-Khoe, Proto-Kalahari Khoe, Proto-Khoekhoe); (d) shows Tuu borrowings in languages from the Kx'a and Khoe-Kwadi families, broken down to the highest resolution level that could be obtained for each root (Proto-Tuu, Proto-Taa-Lower Nossob, Proto-Taa).

Applying this approach, we found that roots with a Kx'a origin (Fig. 4B) were predominantly spread by languages of the !Xun subgroup. Borrowings from !Xun are common in languages of the Khoe-Kwadi family (Supplementary Figure S7B), including those spoken in areas far to the east (Shua, Tshwa) and south (!Ora) of regions where !Xun languages are presently attested. This observation either suggests a different distribution of !Xun in the past, or ancient contact affecting now extinct proto-languages (Proto-Khoe-Kwadi, Proto-Khoe), rather than their living descendants. The latter scenario would support a movement of Kalahari Khoe and Khoekhoe speakers from west to east, at odds with earlier proposals (Barnard, 1992; Elphick, 1985) but in line with the archaeological record (Lander & Russell, 2018) and genetic patterns (Fehn *et al.*, 2022; Oliveira *et al.*, 2023). Kx'a-borrowings tracing back to †Amkoe are mostly found in Glui-Glana, in support of a localized borrowing network in the Central Kalahari (Gerlach, 2016; Güldemann & Loughnane, 2012; Traill & Nakagawa, 2000).

In accordance with the observation that the majority of shared roots (41%) originated within Khoe-Kwadi, we find evidence for a widespread influence of this family on other southern African languages. The distribution of Khoe-Kwadi borrowings (Fig. 4C) shows influence of both the Khoekhoe and Kalahari Khoe subgroups on Kx'a (Supplementary Figure S5B) and Tuu languages (Supplementary Figure S6B) from different geographic regions. The spread of Khoekhoe-associated borrowings shows a striking overlap with the known distribution of Khoekhoe-speaking herders, namely along the Orange River (N!ng), on the west coast of South Africa towards the Cape (!Xam) and across Namibia up to the border with Angola (Lower Nossob, Taa West, !Xun). The intensive borrowing of Khoe-Kwadi lexis among Tuu languages in the southwest reflects the long presence of Khoe-Kwadi herders in this area, as attested by the archaeological evidence for the early spread of livestock along the Atlantic coastline (Lander & Russell, 2018). Further north, only little Khoe-Kwadi influence can be detected in the !Xun subgroup of Kx'a, likely reflecting the relatively recent northward spread of Nama speakers (Barnard, 1992). Contrasting with the widespread influence of Khoekhoe in the west, Kalahari Khoe influence is restricted to Taa East, †Amkoe and South-eastern !Xun, spoken in and around the Central Kalahari, as well as to Northern !Xun, which has a history of contact with speakers of the Khwe-Ts'ixa language cluster.

Compared to Khoe-Kwadi and Kx'a, the contributions of Tuu vocabulary to languages of other families are minor (Fig. 4D). Notable exceptions are contributions to †Amkoe (Supplementary Figure S5C) and Naro (Supplementary Figure S7C), both of which most likely borrowed via their geographically close neighbour Taa within the contact network of the Central Kalahari.

5. Distribution of roots with unidentified origins

As indicated above, our data set contains ~29% (100/340) of shared roots for which no origin could be determined by applying the criteria outlined in Supplementary Text S1; we therefore treated them as unresolved cases of root-sharing, rather than borrowings with a clear origin and direction of transfer. Across the data set, shared roots with unidentified origin are more frequent in languages of the Kx'a and Tuu families than in the Khoe-Kwadi family (Fig. 5A). Moreover, 70% (70/100) of roots with unidentified origins involve sharing between Kx'a and Tuu (Kx'a + Tuu + Khoe-Kwadi, Kx'a + Tuu). Of these, 19% (13/70) can still be explained by contact, as they are shared between neighbouring languages spoken in close geographic proximity, especially within the micro-contact zone of the Central Kalahari (Gerlach, 2016; Güldemann & Loughnane, 2012; Traill & Nakagawa, 2000). Additional 74% (52/70) appear in geographically distant languages or are shared between Proto-Kx'a and Proto-Tuu, thereby suggesting distant common ancestry of the two families, or ancient contact in areas that are presently not occupied by Kx'a and Tuu speakers (Fig. 5B,C). Finally, 7% (5/70) constitute outlier cases that appear in isolated languages across the entire area, but cannot be convincingly ascribed to Kx'a, Tuu or Khoe-Kwadi.

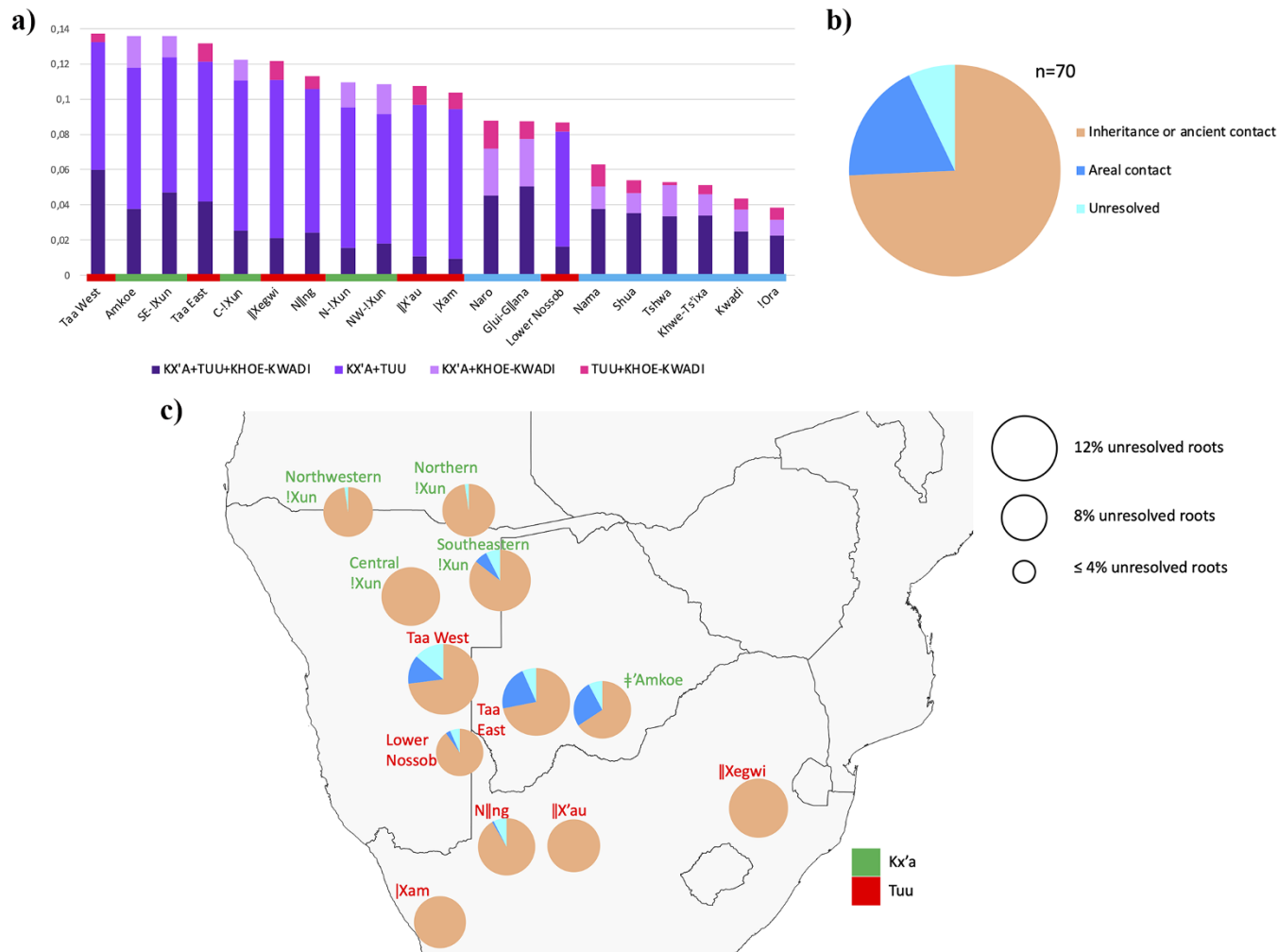


Figure 5. (a) Percentages of roots with unidentified origin in each language cluster, according to their sharing patterns. (b) Pie chart decomposing roots with unidentified origin shared between Kx'a and Tuu, and between Kx'a, Tuu and Khoe-Kwadi ($n=70$), according to whether they are shared: (i) between geographically unconnected units (inheritance or ancient contact); (ii) areally between neighbours (areal contact) or (iii) in an unresolvable setting (unresolved); (c) map showing the geographic distribution of the categories shown in (B) across our Kx'a and Tuu data set. The size of the pie chart corresponds to the percentage of shared roots with unidentified origins in each language cluster, according to the legend.

6. Discussion

The prehistory of southern Africa is characterized by migration and contact between resident forager populations and incoming food producers successively introducing herding, agriculture and metal working (Fehn *et al.*, 2022). Although multiple studies have shed light on the migration routes and ethnolinguistic impact of Bantu-speaking farmers from West Central Africa (Fortes-Lima *et al.*, 2024; Grollemund *et al.*, 2015; Rocha & Fehn, 2016), contact between diverse forager populations as well as the migratory pathways and social influence of early herders remain poorly understood (Fehn *et al.*, 2022; Lander & Russell, 2018; Sadr, 2008; Smith, 2022). Here, we have explored sharing and exchange of vocabulary between languages of the Kx'a, Tuu and Khoe-Kwadi families in order to learn more about contact events and the sociohistorical conditions that characterized pre-Bantu southern Africa.

Kx'a, Tuu and Khoe-Kwadi share click consonants and a particular phonotactic profile (Güldemann, 2008b; Güldemann & Fehn, 2017; Nakagawa *et al.*, 2023). Hence, detecting the origin of a lexical root attested widely across the area poses a challenge to standard methodology from contact linguistics, which relies heavily on the formal differences between borrowed and inherited vocabulary (compare, e.g., Haspelmath & Tadmor, 2009). We therefore opted to use methods from historical-comparative linguistics that place heavy emphasis on the detection of regular versus irregular phonological correspondences, along with a set of linguistic and extralinguistic criteria, to quantify instances of lexical borrowing between the Kx'a, Tuu and Khoe-Kwadi families (Supplementary Text S1).

In our data set comprising linguistic data from 57 doculects, we were able to identify 20% of roots which are attested in more than one language family (Fig. 3B). As this study relied heavily on core vocabulary and therefore did not include semantic fields known for their high borrowability (e.g., terms for local fauna, landscape features), our result constitutes a conservative estimate which may increase if entire dictionaries, rather than just wordlists, are considered. Furthermore, the considerable percentage (29%) of shared roots for which no origin could be assigned include many roots for which an origin in either Kx'a or Tuu is equally plausible (Fig. 5). The close link between Kx'a and Tuu is in line with evidence from archaeology and genetics which hints at a long-term continuity in the population landscape of southern Africa before food production (Sadr, 2015; Schlebusch *et al.*, 2017, 2020). Although it is widely agreed that Kx'a and Tuu form a stronger typological unit with one another than with the Khoe-Kwadi family (Güldemann, 2008b; Güldemann & Fehn, 2017), a deep genealogical relationship or long-term contact in an ancient linguistic area both seem equally plausible explanations for the currently observed similarities (Collins & Honken, 2016; Güldemann, 2014a).

Following an unknown period of exclusive interaction between Kx'a and Tuu languages, the appearance of sheep dating back to around 2000 BP points to an early introduction of herding from eastern Africa, which has since been associated with languages of the Khoe-Kwadi family (Güldemann, 2008b; Lander & Russell, 2018; Smith, 2022). There can be little doubt that the introduction of small-stock pastoralism must have confronted local foragers with a previously unknown way of life. Yet, it has proven difficult to measure the social impact of the Khoe-Kwadi on the basis of archaeological or genetic evidence. Although criteria to distinguish between forager and herder sites have been proposed (Lander & Russell, 2020; Sadr, 2008, 2015), an overturn in material culture following the arrival of the livestock is still disputed. Furthermore, all modern Khoe-Kwadi speakers broadly share the genetic profiles of their Kx'a, Tuu and Bantu-speaking neighbours, making it difficult to assess their demographic impact on the region (Fehn *et al.*, 2022; Montinaro *et al.*, 2017; Oliveira *et al.*, 2023; Pickrell *et al.*, 2012). In addition, the only Khoe-Kwadi speakers for whom a pronounced pastoralist identity and way of life has been documented in historical times are the Khoekhoe of South Africa and Namibia. In contrast, Kalahari Khoe speakers are historically foragers who adopted small-scale pastoralism only recently (Barnard, 1992; Ikeya, 1993; Russell, 2017; Tanaka, 1969), while the

Kwadi of Angola do not preserve an ancient herder profile but mirror the pastoral culture of their Bantu-speaking Kuvale and Himba neighbours (Fehn et al., 2022).

Whereas 41% of all shared roots have a likely Khoe-Kwadi origin and therefore testify to the considerable impact the newcomers had on the Kalahari Basin linguistic area (Fig. 3C), the family-internal division between dominant herding and egalitarian forager populations is also reflected in our results, which suggest differentiated contact scenarios, according to distinct geographic regions.

Along the western and southern Kalahari Basin fringes, Kx'a and Tuu languages receive Khoe-Kwadi roots overwhelmingly from the Khoekhoe subgroup of Khoe, which has been linked to large-scale cattle pastoralism well into historical times (Fig. 4C, Supplementary Figures S5B, S6B). In contrast Khoekhoe only displays residual amounts of Kx'a- and Tuu-derived vocabulary (Figs 4B,D, 6C), indicating that borrowing in this area was a unilateral process.

In the Central Kalahari, on the other hand, languages spoken by Kalahari Khoe foragers (Glui-Glana, Naro) transfer lexical items to Taa East (Tuu) and the Kx'a languages †Amkoe and South-eastern !Xun (Fig. 4D, Supplementary Figures S5B, S6B), while also adopting many loanwords from their neighbours (Fig. 4B,D, Supplementary Figure S7). Therefore, the Central Kalahari forms a micro-contact area characterized by egalitarian patterns of exchange between languages from all three families, probably due to shared livelihoods, intermarriage and multilingualism (Barnard, 1992; Gerlach, 2016; Güldemann & Loughnane, 2012; Traill & Nakagawa, 2000).

As Kx'a and Tuu languages spoken in regions outside the influence zone of the Khoekhoe or the Central Kalahari (i.e. North-western, Northern and Central !Xun, †Xegwi) only display residual amounts of Khoe-Kwadi loanwords (Fig. 4C, Supplementary Figures S5B, S6B), a large-scale impact of the family beyond areas with historically attested contact can likely be excluded.

Beyond characterizing different contact areas, Kx'a and Tuu borrowings into Khoe-Kwadi provide valuable insights into the internal migrations of the Khoe-Kwadi language family inside the Kalahari Basin area, following their initial arrival from eastern Africa. As Khoe-Kwadi languages spoken in regions where no Kx'a and Tuu languages are currently attested (i.e. Kwadi, Khwe-Ts'ixa, Shua, Tshwa) still have 7–9% of Kx'a and 2–4% of Tuu loanwords (Fig. 4B,D, Supplementary Figure S7), we can assume that part of the Kx'a and Tuu lexicon found currently in languages of the Khoe-Kwadi family traces back to historical borrowing events affecting the proto-languages Proto-Khoe, Proto-Kalahari Khoe and Proto-Khoekhoe, which took place in geographic locations where no contact is presently attested. In this frame, we hypothesize that Proto-Khoe-Kwadi must have been spoken in an area which, at the time, was populated by Kx'a and Tuu speakers (see also Fehn & Rocha, 2023; Oliveira et al., 2023), whereas Proto-Khoe and Proto-Kalahari Khoe probably emerged further to the north, in an area dominated by !Xun speakers, in line with findings from other linguistic domains (Güldemann, 2004, 2008b, 2019). In line with the archaeological and genetic evidence (Fehn et al., 2022; Lander & Russell, 2018; Oliveira et al., 2023), the modern distribution of Khoe-Kwadi in southern Africa can thus be interpreted as the result of multiple migratory movements from a proto-population along the western fringe of the Kalahari Basin to the north (Kwadi), south (Khoekhoe) and east (Kalahari Khoe), each with its own genetic and linguistic admixture history (Fig. 6).

Taken together, our results indicate that borrowing patterns mirroring the social relations between the Khoe-Kwadi and their Kx'a- and Tuu-speaking neighbours varied across time and space. Early contacts most likely occurred on the western Kalahari Basin fringe and involved significant borrowing of Kx'a and Tuu roots into Proto-Khoe-Kwadi and Proto-Khoe. While vocabulary related to sheep-herding was indeed transferred to Kx'a and Tuu speakers (most notably *g(w)uu 'sheep' and *TS'áo 'to milk'), our data do not suggest a large-scale demic migration akin to the Bantu expansion, which would have entailed the spread of a distinct ethnolinguistic and genetic package from dominant migrants to resident foragers. We note, however, that a more comprehensive analysis of herding-related vocabulary, including practices like corralling, branding and castrating, might provide further insights into contact and migration patterns associated with the spread of livestock –

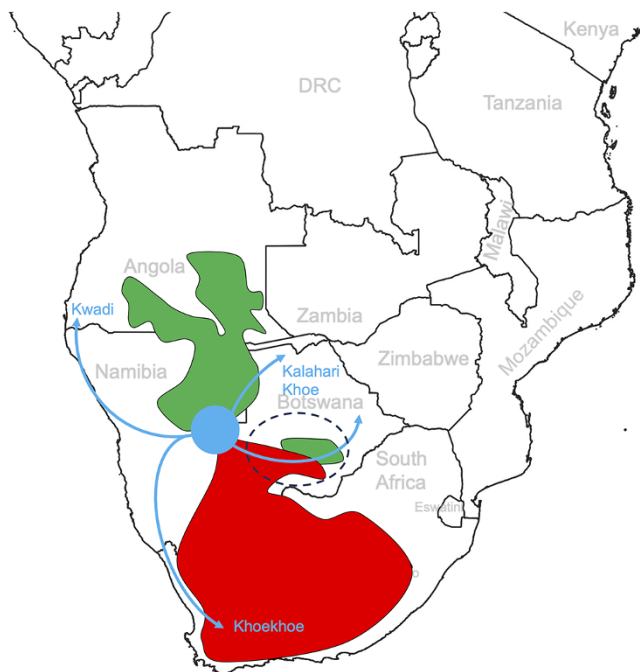


Figure 6. Hypothetical migration routes of Kwadi, Kalahari Khoe and Khoekhoe speakers, out of a core on the western Kalahari Basin fringe where Proto-Khoe-Kwadi and Proto-Khoe emerged (in blue). Historically attested Kx'a- and Tuu-speaking areas are indicated in green and red, respectively. The micro-contact zone of the Central Kalahari where languages of all three families exchange loanwords is encircled by a dashed line.

in particular when considered in the context of the archaeological record (see, e.g., Ehret, 2008; Güldemann, 2008b; Vossen, 2017 for a preliminary discussion).

We alternatively suggest that the first Khoe-Kwadi speakers in southern Africa arrived as small-sized groups of hunter-herders whose lifestyle and technological toolkit broadly resembled resident forager communities (cf. Fehn et al., 2022). A significant cultural impact mirrored in unilateral patterns of lexical borrowing from prestigious into less prestigious languages is still attested in areas historically populated by Khoekhoe herders who, in this framework, constitute a later ethnogenetic development, rather than a remnant group retaining the lifestyle of the ancestral migrants (Sadr, 2008; Fauvelle-Aymar & Sadr, 2008).

7. Conclusion

Our findings suggest a division between multilateral borrowing within egalitarian forager networks involving speakers of Kx'a, Tuu and Kalahari Khoe from the Central Kalahari, as well as unilateral transfer from dominant Khoekhoe-speaking herders into hunter-gatherer groups along the southern and western Kalahari Basin fringes; Khoe-Kwadi languages participate in both contact settings, in line with the historically observed ethnographic diversity characterizing populations speaking languages of this family.

More generally, we demonstrate how the quantitative analysis of lexical borrowing can provide a rich source of information about the contact history between speakers of different language families, especially in regions for which no written records exist (see also Gerstner-Link, 2023; McMahon et al, 2005). In addition, we show the value of linguistic analyses, together with archaeology and genetics, in tracing migration routes and assessing population contact and admixture.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/ehs.2025.10014>

Acknowledgements. This study would not have been possible without the administrative and logistic support of the TwinLabs established between BIOPOLIS-CIBIO and the Universities of Namibia and Botswana, to whom we would like to extend our gratitude. We also thank our colleagues Hiroshi Nakagawa, Chris Collins, Tim Mathes and Andy Chebanne for making their unpublished lexical data and manuscripts available to us, and acknowledge the relevant government agencies, local authorities and institutions that supported our fieldwork in Namibia, Botswana and Angola. We further extend our gratitude to the Teemashani Trust representing the !Anikhwe community at Moheumbo, as well as to Thaddeus Chedau, Mmoloki Mogomotsi, Kelatlhlilwe Gamaxo Moses, Tshiamo Kebuelemang and Tshisimogo Leepang for their unfailing support and friendship during our fieldwork in Namibia and Botswana. Finally, we would like to thank all the ‘Khoisan’ speakers who so generously shared their languages and culture with us. This work is dedicated to them.

Author contributions. J.R., A.M.F., G.M., M.B. and E.F. conceptualized the study and carried out the fieldwork. A.P. assembled and transcribed his previously collected Tjwao data. A.M.F. and B.S. designed the criteria to detect borrowings applied in this paper. A.M.F. assembled the cognate sets and assigned the origin of shared lexical roots. A.M.F. and J.R. carried out the quantitative analyses. A.M.F., J.R. and B.S. wrote the paper. All authors participated in the final editing of the manuscript.

Financial Support. A.M.F. was financed through Junior Researcher Contract CEECIND/02765/2017 granted by the Portuguese Foundation for Science and Technology (FCT), and A.P. by FCT project 2022.06259.PTDC. Lexical data included in this study was collected within FCT projects PTDC/BIA-EVF/2907/2012 (J.R.), PTDC/BIA-GEN/29273/2017 (J.R.) and 2022.06259.PTDC (A.M.F.). This paper was written as part of FCT project 2022.06259.PTDC.

Competing Interest. The authors declare no competing interests.

Data Availability Statement. All linguistic data and cognate judgements used in this paper are provided in Supplementary Tables 2–4.

References

- Barnard, A. (1992). *Hunters and herders of Southern Africa: A comparative ethnography of the Khoisan peoples*. Cambridge University Press. <https://journals.co.za/doi/abs/10.10520/EJC84795>
- Bleek, D. F. (1956). *A Bushman dictionary*. American Oriental Society.
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., & Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>
- Collins, C., & Honken, H. (2016). The plural prefix in Kx'a, !Ui, and Taa. In S. Shah & M. Brenzinger (Eds.), *Proceedings of the 5th international symposium*, July 13–17, 2014, Riezler/Kleinwalsertal. Rüdiger Köppe.
- Doke, C. M. (1936). An outline of !Khomani Bushman phonetics. *Bantu Studies*, X, 433–461.
- Ehret, C. (2008). The early livestock-raisers of Southern Africa. *Southern African Humanities*, 20(1), 7–35. <https://www.sahumanities.org/index.php/sah/article/view/185>
- Elphick, R. (1985). *Khoikhoi and the founding of White South Africa*. Ravan Press.
- Fauvel-Aymar, F. X., & Sadr, K. (2008). Trends and traps in the reconstruction of early herding societies in Southern Africa: Introduction. *Southern African Humanities*, 20, 1–6. <https://www.sahumanities.org/index.php/sah/article/view/186>
- Fehn, A.-M., Amorim, B., & Rocha, J. (2022). The linguistic and genetic landscape of Southern Africa. *Journal of Anthropological Sciences Istituto Italiano Di Antropologia*, 100, 243–265. <https://doi.org/10.4436/jass.10008>
- Fehn, A.-M., & Phiri, A. (2022). Juncture–verb constructions in Northeastern Kalahari Khoe: A comparative perspective. *Stellenbosch Papers in Linguistics Plus*. <https://doi.org/10.5842/65-1-973>
- Fehn, A.-M., & Rocha, J. (2023). Lost in translation: A historical-comparative reconstruction of Proto-Khoe-Kwadi based on archival data. *Diachronica*, 40(5), 609–665. <https://doi.org/10.1075/dia.23022.feh>
- Fortes-Lima, C. A., Burgarella, C., Hammarén, R., Eriksson, A., Vicente, M., Jolly, C., Semo, A., Gunnink, H., Pacchiarotti, S., Mundeke, L., Matonda, I., Muluwa, J. K., Coutros, P., Nyambe, T. S., Cikomola, J. C., Coetzee, V., de Castro, M., Ebbesen, P., Delanghe, J., & Schlebusch, C. M. (2024). The genetic legacy of the expansion of Bantu-speaking peoples in Africa. *Nature*, 625(7995), 540–547. <https://doi.org/10.1038/s41586-023-06770-6>
- Gerlach, L. (2016). *!N!aqriaxe: The phonology of an endangered language of Botswana*. Harrassowitz.
- Gerstner-Link, C. (2023). Multilateral lexical transfer among four Papuan language families: Border, Nimboran, Sentani, and Sko. In M. Klammer & F. Moro (Eds.), *Traces of contact in the lexicon* (pp. 263–306). Brill. https://doi.org/10.1163/9789004529458_009

- Gray, R. D., Bryant, D., & Greenhill, S. J. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1559), 3923–3933. <https://doi.org/10.1098/rstb.2010.0162>
- Greenberg, J. H. (1963). *The languages of Africa*. Indiana University.
- Greenhill, S., Currie, T. E., & Gray, R. D. (2009). Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B*, 276, 2299–2306. <https://doi.org/10.1098/rspb.2008.1944>
- Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., & Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, 112(43), 13296–13301. <https://doi.org/10.1073/pnas.1503793112>
- Güldemann, T. (1998). The Kalahari Basin as an object of areal typology: A first approach. In M. Schladt (Ed.), *Language, identity and conceptualisation among the Khoisan* (pp. 137–169). Rüdiger Köppe.
- Güldemann, T. (2004). Reconstruction through ‘de-construction’: The marking of person, gender, and number in the Khoe family and Kwadi. *Diachronica*, 21(2), 251–306. <https://doi.org/10.1075/dia.21.2.02gul>
- Güldemann, T. (2005). Tuu as a language family. In T. Güldemann (Ed.), *Studies in Tuu (Southern Khoisan)* (p. 23). University of Leipzig Papers on Africa, Languages and Literatures.
- Güldemann, T. (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* (pp. 99–134). Palgrave Macmillan. <https://doi.org/10.1057/9780230287617>
- Güldemann, T. (2008a). Greenberg’s “case” for Khoisan: The morphology evidence. In D. Ibriszimow (Ed.), *Problems of linguistic-historical reconstruction in Africa* (pp. 123–153). Rüdiger Köppe.
- Güldemann, T. (2008b). A linguist’s view: Khoe-Kwadi speakers as the earliest food-producers of Southern Africa. *Southern African Humanities*, 20, 99–132. <https://www.sahumanities.org/index.php/sah/article/view/180>
- Güldemann, T. (2014a). Khoisan linguistic classification today. In T. Güldemann & A.-M. Fehn (Eds.), *Beyond ‘Khoisan’: Historical relations in the Kalahari Basin* (pp. 1–41). John Benjamins. <https://doi.org/10.1075/cilt.330.01gul>
- Güldemann, T. (2014b). The Lower Nossob varieties of Tuu. In T. Güldemann & A.-M. Fehn (Eds.), *Beyond ‘Khoisan’: Historical relations in the Kalahari Basin* (pp. 257–282). John Benjamins. <https://doi.org/10.1075/cilt.330.10gul>
- Güldemann, T. (2019). Person–gender–number marking from proto-Khoe-Kwadi to its descendants: A rejoinder with particular reference to language contact. In R. Vossen, and K. Christa (Eds.), *Patterns of linguistic convergence in Africa (Frankfurter Afrikanistische Blätter 27)* (pp. 53–74). Rüdiger Köppe.
- Güldemann, T., & Fehn, A.-M. (2017). The Kalahari basin area as a ‘Sprachbund’ before the Bantu expansion. In R. Hickey (Ed.), *The Cambridge handbook of areal linguistics* (pp. 500–526). Cambridge University Press. <https://doi.org/10.1017/9781107279872.019>
- Güldemann, T., & Loughnane, R. (2012). Are there “Khoisan” roots in body-part vocabulary? On linguistic inheritance and contact in the Kalahari Basin. *Language Dynamics and Change*, 2(2), 215–258. <https://doi.org/10.1163/22105832-20120205>
- Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In M. Haspelmath & U. Tadmor (Eds.), *Loanwords in the world’s languages: A comparative handbook* (pp. 35–54). Mouton de Gruyter.
- Haspelmath, M., & Tadmor, U. (2009). *Loanwords in the world’s languages*. Walter de Gruyter. M. Haspelmath & U. Tadmor, Eds. <https://doi.org/10.1515/9783110218442>
- Heine, B., & Honken, H. (2010). The Kx’a family. *Journal of Asian and African Studies*, 79, 5–36.
- Heine, B., & König, C. (2015). *The !Xun language. A dialect grammar of Northern Khoisan*. Rüdiger Köppe.
- Hoffmann, K., Bouckaert, R., Greenhill, S. J., & Kühnert, D. (2021). Bayesian phylogenetic analysis of linguistic data using BEAST. *Journal of Language Evolution*, 6(2), 119–135. <https://doi.org/10.1093/jole/lzab005>
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Ikeya, K. (1993). Goat raising among the San in the central Kalahari. *African Study Monographs*, 14(1), 39–52.
- Kelih, E. (2024). Modelling the frequency of loanwords in different semantic fields in core vocabularies (based on WOLD data). *Glottometrics*, 56, 59–77. https://doi.org/10.53482/2024_56_417
- Köhler, O. (n.d.). *Manuscript collection*. Oswin Köhler Archiv, Goethe University.
- Lander, F., & Russell, T. (2018). The archaeological evidence for the appearance of pastoralism and farming in Southern Africa. *PLoS One*, 13(6), e0198941. <https://doi.org/10.1371/journal.pone.0198941>
- Lander, F., & Russell, T. (2020). A Southern African archaeological database of organic containers and materials, 800 cal BC to cal AD 1500: Possible implications for the transition from foraging to livestock-keeping. *PLoS One*, 15(7), e0235226. <https://doi.org/10.1371/journal.pone.0235226>
- Lanham, L. W., & Hallows, D. P. (1956). An outline of the structure of Eastern Bushman. *African Studies*, 15(3), 97–118. <https://doi.org/10.1080/00020185608706989>
- List, J.-M., & Forkel, R. (2022). Automated identification of borrowings in multilingual wordlists. *Open Research Europe*, 1, 79. <https://doi.org/10.12688/openreseurope.13843.3>
- Maingard, L. F. (1937). The !Khomani dialect of Bushman: Morphology and other characteristics. In J. D. R. Jones (Ed.), *Bushman of the southern Kalahari* (pp. 237–275). University of the Witwatersrand Press.

- McMahon, A., Heggarty, P., McMahon, R., & Slaska, N. (2005). Swadesh sublists and the benefits of borrowing: An Andean case study. *Transactions of the Philological Society*, 103(2), 147–170. <https://doi.org/10.1111/j.1467-968X.2005.00148.x>
- Montinaro, F., Busby, G. B. J., Gonzalez-Santos, M., Oosthuizen, O., Oosthuizen, E., Anagnostou, P., Destro-Bisol, G., Pascali, V. L., & Capelli, C. (2017). Complex ancient genetic structure and cultural transitions in Southern African populations. *Genetics*, 205(1), 303–316. <https://doi.org/10.1534/genetics.116.189209>
- Nakagawa, H., Witzlack-Makarevich, A., Auer, D., Fehn, A.-M., Ammann Gerlach, L., Güldemann, T., Job, S., Lionnet, F., Naumann, C., Ono, H., & Pratchett, L. J. (2023). Towards a phonological typology of the Kalahari Basin Area languages. *Linguistic Typology*, 27(2), 509–535. <https://doi.org/10.1515/lingty-2022-0047>
- Nelson-Sathi, S., List, J.-M., Geisler, H., Fangerau, H., Gray, R. D., Martin, W., & Dagan, T. (2011). Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proceedings of the Royal Society B: Biological Sciences*, 278(1713), 1794–1803. <https://doi.org/10.1098/rspb.2010.1917>
- Oliveira, S., Fehn, A.-M., Amorim, B., Stoneking, M., & Rocha, J. (2023). Genome-wide variation in the Angolan Namib Desert reveals unique pre-Bantu ancestry. *Science Advances*, 9(38). <https://doi.org/10.1126/sciadv.adh3822>
- Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Kure, B., Mpoloka, S. W., Nakagawa, H., Naumann, C., Lipson, M., Loh, P. R., Lachance, J., Mountain, J., Bustamante, C. D., Berger, B., Tishkoff, S. A., Henn, B. M., Stoneking, M., & Pakendorf, B. (2012). The genetic prehistory of Southern Africa. *Nature Communications*, 3, 1143. <https://doi.org/10.1038/ncomms2140>
- Pickrell, J. K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., & Reich, D. (2014). Ancient west Eurasian ancestry in southern and Eastern Africa. *Proceedings of the National Academy of Sciences USA*, 111(7), 2632–26327. <https://doi.org/10.1073/pnas.1313787111>
- Rambaut, A. (2018). *Figtree v.1.4.4*. Institute of Evolutionary Biology, University of Edinburgh.
- Rocha, J., & Fehn, A.-M. (2016). *Genetics and demographic history of the Bantu*. eLS. W. Sons, Ed. <https://doi.org/10.1002/9780470015902.a0022892>
- Russell, T. (2017). ‘Where goats connect people’: Cultural diffusion of livestock not food production amongst Southern African hunter-gatherers during the Later Stone Age. *Journal of Social Archaeology*, 17(2), 115–137. <https://doi.org/10.1177/1469605317701596>
- Sadr, K. (2008). Invisible herders? The archaeology of Khoekhoe pastoralists. *Southern African Humanities*, 20, 179–203. <https://journals.co.za/doi/abs/10.10520/EJC84795>
- Sadr, K. (2015). Livestock first reached Southern Africa in two separate events. *PLoS One*, 10(8), 9–11. <https://doi.org/10.1371/journal.pone.0134215>
- Sands, B. (1998). *Eastern and Southern African Khoisan: Evaluating claims of distant linguistic relationships*. Rüdiger Köppe.
- Sands, B. (2001). Borrowing and diffusion as a source of lexical similarities in Khoesan. In A. Bell & P. Washburn (Eds.), *Khoisan: Syntax, phonetics, phonology, and contact* (pp. 200–224). Cornell Working Papers in Linguistics.
- Sands, B. (2010). Juu subgroups based on phonological patterns. In M. Brenzinger & C. König (Eds.), *Khoisan language and linguistics: The Riezler Symposium 2003* (pp. 85–114). Rüdiger Köppe.
- Schlebusch, C. M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A. R., Vicente, M., Steyn, M., Soodyall, H., Lombard, M., & Jakobsson, M. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 358(6363), 652–655. <https://doi.org/10.1126/science.aao6266>
- Schlebusch, C. M., Sjödin, P., Breton, G., Günther, T., Naidoo, T., Hollfelder, N., Sjöstrand, A. E., Xu, J., Gattepaille, L. M., Vicente, M., Scofield, D. G., Malmström, H., de Jongh, M., Lombard, M., Soodyall, H., & Jakobsson, M. (2020). Khoe-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. *Molecular Biology and Evolution*, 37(10), 2944–2954. <https://doi.org/10.1093/molbev/msaa140>
- Smith, A. (2022). *First people: The lost history of the Khoisan*. Jonathan Ball.
- Snyman, J. W. (1997). A preliminary classification of the !Xöü and ǀZulhōasi dialects. In W. Haacke & E. D. Elderkin (Eds.), *Namibian languages: Reports and papers* (pp. 21–106). Rüdiger Köppe.
- Starostin, G. (2021). Lexicostatistical studies in Khoisan II/1: How to make a Swadesh wordlist for Proto-Tuu. *Journal of Language Relationship*, 19(2), 99–135. [https://www.jolr.ru/files/\(302\)jlr2021-19-1-2\(99-135\).pdf](https://www.jolr.ru/files/(302)jlr2021-19-1-2(99-135).pdf)
- Starostin, G. (2022). Lexicostatistical studies in Khoisan II/2: Towards a more precise phylogeny for the Tuu family. *Journal of Language Relationship*, 20(1), 25–70. [https://jolr.ru/files/\(315\)jlr2022-20-1-2\(25-70\).pdf](https://jolr.ru/files/(315)jlr2022-20-1-2(25-70).pdf)
- Swadesh, M. (1952). Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society*, 96, 452–463.
- Tadmor, U., Haspelmath, M., & Taylor, B. (2010). Borrowability and the notion of basic vocabulary. *Diachronica*, 27(2), 226–246. <https://doi.org/10.1075/dia.27.2.04tad>
- Tanaka, J. (1969). The ecology and social structure of central Kalahari Bushmen: A preliminary report. *Kyoto University African Studies*, 3, 1–26.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.
- Trall, A., & Nakagawa, H. (2000). A historical !Xöü-ǀGui contact zone: Linguistic and other relations. In H. Batibo & J. Tsonope (Eds.), *The state of Khoesan languages in Botswana* (pp. 1–17). Basarwa Languages Project.

- van Hout, R., & Muysken, P. (1994). Modeling lexical borrowability. *Language Variation and Change*, 6(1), 39–62. <https://doi.org/10.1017/S0954394500001575>
- Vossen, R. (1997). *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Rüdiger Köppe.
- Vossen, R. (2017). Frühgeschichtliche Betrachtungen zur Wirtschaftsform der Khoe (Zentralkhoisan) im südlichen Afrika. In N. Rupp, C. Beck, G. Franke & K. P. Wendt (Eds.), *Winds of change. Archaeological contributions in honour of Peter Breunig* (pp. 87–94). Dr. Rudolf Habelt.
- Westphal, E. O. J. (1953–1971). *The E.O.J. Westphal papers. Manuscript collection*. University of Cape Town.
- Westphal, E. O. J. (1971). The click languages of southern and Eastern Africa. In T. A. Sebeok (Ed.), *Linguistics in sub-Saharan Africa* (pp. 367–420). Mouton de Gruyter.
- Ziervogel, D. (1955). Notes on the language of the Eastern Transvaal Bushmen. In E. F. Potgieter (Ed.), *The disappearing Bushmen of Lake Chrissie: A preliminary survey* (pp. 33–64). J.L. van Schaik.