

---

# The Role of Dictionaries in the Documentation and Codification of African Languages: The Case of Khoisan\*

Andy Chebanne, *Centre for Advanced Studies of African Society (CASAS), Cape Town, South African and Department of African Languages, University of Botswana, Gaborone, Botswana (chebanne@casas.co.za / chebanne@mopipi.ub.bw)*

---

**Abstract:** Khoisan speech communities are found in Southern Africa, with Botswana having the greatest ethnic and linguistic diversity. Living in small communities, the Khoisan have always found themselves in a situation of being ethnic and linguistic minorities that are socially marginalized. The Khoisan marginalization is historical, resulting in the Khoisan languages not featuring in education and language development. These languages are underresearched and existing publications are divergent, and not accessible to the speech communities themselves. While some anthropologists, linguists, and a few community organizations have undertaken some development of language resources, as yet these efforts have not had the desired effect. The reasons are that the publications used implement impractical linguistic writing conventions or that no supportive programmes exist to encourage more focused and purposeful research and language promotion through literacy classes. This article will review research and publication materials on the Khoisan languages by arguing that there is a need to document and codify these languages to make them accessible to the speech communities. Because dictionary documentation is suggested as the most viable way to accomplish this, existing lexicographical works will be reviewed to demonstrate their strengths and shortcomings. The article will also indicate why and how dictionaries will be critical in this documentation and codification enterprise. Lemmatization procedures will be suggested to facilitate the capture of the linguistic and the indigenous knowledge system of the Khoisan. In this way the Khoisan languages will be vitalized and research contribute to Khoisan language development.

**Keywords:** KHOISAN LANGUAGES, UNDERRESEARCHED LANGUAGES, LANGUAGE DOCUMENTATION AND CODIFICATION, DICTIONARY, LEXICOGRAPHICAL WORKS, LEMMATIZATION PROCEDURES, LINGUISTIC AND INDIGENOUS KNOWLEDGE, LANGUAGE PRESERVATION AND PROMOTION, BOTSWANA

**Opsomming: Die rol van woordeboeke in die dokumentasie en kodifikasie van Afrikatale: Die geval van Khoisan.** Khoisanspraakgemeenskappe word in Suidelike Afrika aangetref met Botswana wat die grootste etniese en linguistiese verskeidenheid het.

---

\* This article was presented as a plenary address at the Fifteenth International Conference of the African Association for Lexicography which took place at the University of Botswana, Gaborone, Botswana, 19–21 July 2010.

Deurdat hulle in klein gemeenskappe lewe, het die Khoisan hulle altyd in 'n situasie bevind van etniese en linguistiese minderhede te wees wat maatskaplik gemarginaliseer is. Die marginalisering van die Khoisan is histories, wat daartoe gelei het dat die Khoisantale nie 'n belangrike plek in opvoeding en taalontwikkeling inneem nie. Hierdie tale is ondernagevors en bestaande publikasies is uiteenlopend, en nie toeganklik vir die spraakgemeenskappe self nie. Terwyl sommige antropoloë, linguïste en 'n paar gemeenskapsorganisasies 'n sekere ontwikkeling van taalhulpmiddels onderneem het, het hierdie pogings tot dusver nog nie die gewenste uitwerking gehad nie. Die redes is dat die publikasies wat gebruik word, onpraktiese linguistiese skryfkonvensies toepas of dat geen ondersteunende programme bestaan om meer gefokuste en doelgerigte navorsing en taalverheffing deur geletterheidsklasse te bevorder nie. Hierdie artikel sal die navorsings- en publikasie materiaal betreffende die Khoisantale beoordeel deur te redeneer dat daar 'n behoefte bestaan om hierdie tale te dokumenteer en te kodifiseer om hulle vir die spraakgemeenskappe toeganklik te maak. Omdat dokumentasie deur woordeboeke aanbeveel word as die haalbaarste manier om dit te bewerkstellig, sal bestaande leksikografiese werke bespreek word om hulle sterktes en gebreke te toon. Die artikel sal ook aanbeveel waarom en hoe woordeboeke krities in hierdie dokumentasie- en kodifikasieonderneming sal wees. Lemmatiseringsprosedures sal aanbeveel word om die vaslegging van die taal- en inheemsekennissels van die Khoisan te bevorder. Op hierdie manier sal die Khoisantale verlewendig word en navorsing bydra tot Khoisantaalontwikkeling.

**Slutelwoorde:** KHOISANTALE, ONDERNAGEVORSTE TALE, TAALDOKUMENTASIE EN -KODIFIKASIE, WOORDEBOEK, LEKSIKOGRAFIESE WERKE, LEMMATISERINGS-PROSEDURES, TAALKUNDIGE EN INHEEMSE KENNIS, TAALINSTANDHOUDING EN -VERHEFFING, BOTSWANA

## 1. Introduction: Who and why the Khoisan?

The linguistic and ethnographic distinction between Khoe and San is not new and dates back to the early 20th century (Schultze 1928; Barnard 1992; Vossen 1986). The word *Khoe* (in historical texts *Khoi*), meaning "person" for most Khoekhoe (in historical texts, Khoikhoi) speakers has become a generally accepted term for the people and their languages, while the word *San*, which is a Khoekhoe word for "gatherers", comes from *Saon* (with a Nama gender common plural) (Barnard 1992). In recent years, however, the Working Group of Indigenous Minorities in Southern Africa (WIMSA) (Le Roux 1999; WIMSA 2000) has championed efforts to provide Khoe and San languages with a common social reference, the term *San* now being used to transcend ethnic and linguistic distinctions. This may be problematic according to sociohistorical analysis of the label (cf. Chebanne 2003). However, since the name San has the marginalized communities' approval, it could be used advisedly to obviate the historical prejudices associated with it or with some of the communities. The term *Khoisan* is now generally accepted as a common reference for the combined Khoe (those speaking Khoekhoe languages) and San (those speaking non-Khoekhoe languages). Whenever Khoe and San languages are used as separate labels they invariably refer to linguistic entities.

The Khoisan as ethnolinguistic group has an internal diversity that has often led to debates on how exactly the classification should be presented (cf. Güldemann and Vossen 2000). The generally accepted classification of the language families for Khoisan are as follows (adapted from Güldemann and Vossen 2000; Vossen and Keuthmann 1986; Vossen 1988, 1998; Chebanne 2003).

- Proto-Khoe-San**
- 1 **Hadza** (language family, Tanzania)
- 2 **Khoe-San**
- 2.1 **Sandawe** (language family, Tanzania)
- 2.2 **Khoe-San** (Southern Africa)
- 2.2.1 **Khoe-Kwadi** (language family, Botswana and Namibia)
- 2.2.1.1 **Khoekhoe:** Nama and Damara (Khoekhoegowab); !Ora; Hai||om
- 2.2.1.2 **Kalahari Central:** ||Gana; |Gui; Naro
- 2.2.1.3 **Kalahari North:** Khwedam (Kxoe, ||Ani, Buga, |Ganda)
- 2.2.1.4 **Kalahari East:** Kua, Tshoa (Tshwaa), Shua, Tciretcire, Ganadi
- 2.2.2 **San** (language family, Botswana, Namibia and South Africa)
- 2.2.2.1 **Southern San:** Taa and Tuu: !Xóǀ (!Aa): †Ama Oam (†Ama Pfam); !Gwan Oam (!Gwaa Pfam); !Kui (|Xam, ||Xegui, and †Komani)
- 2.2.2.2 **Northern San:** !Xü (!Xung (!Kung)), Ju|'hoan (Ju|'hoasi), †Kx'au||'ein; †Hoa.

The above classification shows the genetic diversity of these ancient languages. Most of the communities who speak these languages are found in Botswana. Also, most of these communities are small and therefore prone to undergoing language shift, dearth and death (cf. Batibo 1997). It is therefore vitally important that these languages should not be left to die before they have even been documented. While the diversity that these speech communities represent within a geographical space gives rise to challenges for linguistic description and anthropological analyses, research should be focused on documentation and codification, and presented in forms readily utilizable by the speech communities themselves.

## 2. The state of Khoisan anthropological and linguistic research

In Southern Africa, the speech communities of the Khoe and the San are generally invisible in social domains (such as language planning and education). As such they can be considered marginalized and underresearched. However, most of the modern understanding of the Khoe and San is credited to Schapera (1930) and Westphal (1956) who described the ethnography and the ethnolinguistic typology of the Khoisan. The classification that is retained here is mainly adapted from Güldemann and Vossen (2000), whose research has contributed a greater linguistic understanding of Khoisan and its internal linguistic

typology. Thus, the largest part of the relevant research on Khoisan which provides a fair understanding of these language communities has been essentially generated by anthropologists and linguists (cf. Chebanne 2008; Willet et al. 2002). Historical and linguistic works that have resulted in the refined classification of the Khoe and San languages by Güldemann and Vossen (2000), have been undertaken by Westphal (1962, 1971) and Köhler (1971, 1981), the most notable works being by Vossen (1997). Köhler's (1971, 1981) hypothesis is that the common structure of word roots, the combination of rare consonants (ejectives) of click and glottal types largely demonstrated that Khoe and San languages had an ancient genetic relationship, and that pursuing such a hypothesis would be scientifically creditable and productive. The research by Güldemann and Vossen (2000) seems to have brought this debate to some consensus.

To address the theme of this article, it is necessary to contextualize the discussion, and to delimit it to some linguistic samples that will allow a quick appreciation of the issues. To achieve this, the following speech communities will be briefly considered as far as documentation and codification are concerned.

## 2.1 !Xóǀ

The !Xóǀ language has been the subject of linguistic research for many years (Traill 1986). However, the usage of these publications has been limited to linguists and academics, and has therefore not been accessible to native speakers. !Xóǀ (or !Aa) is a Southern non-Khoe language, belonging to the Taa and Tuu branch, whose closest related, but now extinct languages were once spoken in what is at present South Africa. Incontestably, there is some affinity with the now extinct !Kui (|Xam, ||Xegui, and †Komani) (Güldemann and Vossen 2000; Traill 1986). The speakers of !Xóǀ are scattered over a large area stretching from western Botswana to the eastern Namibian border (Andersson and Janson 1997). It has been reported by Hasselbring (2001) that !Xóǀ is also spoken in some parts of the South African province of the Northern Cape. The !Xóǀ live in small groups without much contact with each other, but share a dialect continuum. Research by Traill (1985) suggests that !Xóǀ has two main varieties, namely †Ama Oam (†Ama Pfam) (Western variety), and !Gwaa Oam (!Gwaa Pfam) (Eastern variety). Most !Xóǀ speakers are bilingual, speaking another language with which they are in contact, namely |Gui (north and east), Nama (south west) and Shekgalagadi (Chebanne and Monaka 2005). However, this bilingualism does not include Setswana (Andersson and Janson 1997), which they encounter only in the administration and the education of their children. Linguistically !Xóǀ is the only language in the Khoe and San family with the most elaborate click phenomenon. This extensive click incidence alongside other complex sounds seems to have impeded its development (Chebanne 2003).

## 2.2 Ju|'hoasi

Ju|'hoasi (Ju|'hoan), also commonly known as Kaukau in Botswana, is spoken mainly in North Western Botswana and Tsumkwe in Namibia (Güldemann and Vossen 2000; Hasselbring 2000). Ju|'hoasi is considered the central variety of Northern Khoisan (!Xū) which extends into southern Angola. It forms a continuum with †Kx'au||ein (the southern branch of Northern Khoisan). However, the Ju|'hoasi spoken in Tsumkwe is not intelligible to those who speak †Kx'au||ein, and should be considered a dialectal variety of Ju|'hoasi in north-west Botswana (Traill 1986; Güldemann and Vossen 2000). The geographical and social conditions form the basis of the differences between the Namibia and Botswana varieties (Chebanne 2008). The language is mostly used at local village and family level, as well as at community meetings (with interpreters). Ju|'hoasi has been studied extensively by Snyman who wrote a grammar guide and dictionary (1970 and 1975 respectively). Another dictionary was compiled by Dickens (1994). Orthography has been in place since 1969 and was updated in 1987 and 1991. There are many native speakers who can now read and write Ju|'hoasi. This is mainly because of the efforts of the Nyae-Nyae Development Foundation as well as sustained linguistic and cultural research, and missionary work (Chebanne 2008).

## 2.3 Naro

Naro speakers are found mainly in Botswana, and a very few are encountered in eastern Namibia. In Botswana, many Naro speakers now live and work on the Ghanzi Farms. However, they have maintained their way of life. Naro has been studied by many scholars including Barnard (1985), Bleek (1928), and Visser (1994, 2001). The Naro spoken in the west near the Namibian border is said to be slightly different, as it has been substantially influenced by Nama. Historically, Naro belongs together with |Gui and ||Gana, but except for some lexical items here and there, there is not much mutual intelligibility among them. Naro shows differences in grammatical and phonological structure, which suggest that either it has retained the historical forms while others have lost them, or it has acquired them after it separated from |Gui and ||Gana. Work on the Naro Bible translation has greatly contributed to literacy development. A mass of texts have been developed, thus creating a good database for linguistic and lexicographical work.

## 2.4 Nama and Damara (Khoekhoegowab)

The languages Nama and Damara are listed under Khoekhoegowab which can be extended to include !Ora. These are mainly Namibian languages. However, in Botswana, they are all identified by their speakers as Nama. In Botswana,

Nama varieties are referred to with Setswana names such as Sekgothu, Sekhikwe and Seqhanakwe (Batibo et al. 2003). Nama has long been codified, its orthography and dictionary dating as far back as 1889. Developed by Kroenlein (Haacke 1999), it has inspired the orthographies of other Khoisan languages. A dictionary has been compiled by Haacke (2003), and there are also grammar guides in Khoekhoegowab (Batibo et al. 2003). Books and texts for primary schools are available, because, according to the Namibian language policy, the first three years of education (Grades 1–3) should be in the mother tongue, this having boosted the resilience of Khoekhoegowab.

## 2.5 |Gui and ||Gana

The |Gui and the ||Gana languages are mutually intelligible, although the speakers see themselves as different (cf. Nakagawa 2006). The |Gui consider ||Gana as "black Khoe", because of the rather darkish skin of the speakers. Though they have adopted Shekgalagari cultural practices, the people are still relatively positive towards their language (Barnard 1982). In spite of having lived alongside the !Xóǀ speakers, there are only lexical borrowings, but no language switch from |Gui to !Xóǀ. However, the speakers use Naro and Shekgalagari as *lingua franca* (Chebanne 2003) which is a threat to the currency of their own language. This seems to work against the maintenance of the |Gui language as only elderly people, 30 years of age and above, still use it effectively among themselves. Like most of these languages, there is no orthography on which to base any literacy material. Only a few descriptive and socio-historical studies, particularly by Tanaka (1978) are available.

## 2.6 Khwedam (Kxoe)

The Khwedam (Kxoe) is situated in Northern Botswana. It is a cluster language made up of the ||Ani, Buga, and |Ganda. In much research by German linguists, they are grouped under the label Kxoe (pronounced Khoe). The north western Khoe mainly occupy the Okavango delta of Botswana, and therefore are sometimes referred to as the flood-plain (or River) Khoe. They are essentially homogenous owing to their geographical location and their shared history in and around the delta, which has influenced their lives for thousands of years. The ||Ani and the Buga are the main Khoe communities in this area. Their languages are mutually intelligible. Substantial descriptive studies by German linguists and anthropologists have been done on the ||Ani under the label Kxoe. The literacy efforts associated with the Penduka Declaration will see ||Ani lumped together with Buga, and so contributing to the Khwedam. This will need careful planning and an awareness of how the languages will benefit Khwedam in terms of literacy development and linguistic studies (WIMSA 2001).

### 3. From research to documentation and codification: What needs to be done?

In one way or another, languages and cultures have communicated with each other long before the existence of language dictionaries. And indeed, dictionaries cannot replace people's capacity to communicate, especially within the framework of the theory of communication. Other mechanisms such as language policies assist in enhancing this communication between languages and cultures. Dictionaries are excellent pedagogical tools; they are indispensable within the language teaching programme. However, community-based dictionaries also have an informal aspect in their use, and therefore need to be designed for non-expert utilization. While dictionary-using skills must be learnt for dictionaries to appropriately respond to the community needs, users must be made to realise the importance of dictionaries. Policies that facilitate communication between languages need to be in place. If there is no recognition of formal literacy in other languages, there will be generalized illiteracy in the communities speaking indigenous languages. In the UNESCO (1995: 57) report there is the following information which should be noted.

Minorities often find it difficult to participate fully in the activities of societies that favour dominant groups. Sometimes this discrimination is embedded in the legal framework that denies these minorities access to education, employment and political representation. More generally, however, the lack of participation is less a matter of official policy than of everyday practice. The challenge consists in first removing discriminatory barriers and then creating the basis for the empowerment of these minorities.

From this report, it becomes clear that the policies that deny the rights of minorities in areas of culture and education, and that exert domination by linguistic and ethnic hegemonies, create a state of discrimination and denial of access to self-realization. To escape from this vicious circle, the UNESCO World Commission on Culture and Development urged the global community in its 1995 report to adopt a vision for the twenty-first century that would uphold the protection and exercise of cultural rights. This vision of cultural diversity seeks to replace prejudice, discrimination and domination with tolerance, fruitful co-existence and equity. Education must be used to enrich Botswana's cultural diversity. All of the country's languages must be taught to a high standard at primary, secondary and tertiary level. But before this could be achieved all these languages must be documented and codified. These observations point out some practical problems that need urgent attention or consideration:

- in the development of indigenous languages, dictionaries will occupy a central role: they should be the basis for codified languages, uniting different dialects, and promoting translation between languages and cultures;

- in the development of dictionaries, especially in the context of minority languages, there is a need to conjugate efforts at the level of dictionary compilation to render them repositories of linguistic and cultural information and instil a sentiment of common utility;
- in the general development of ethnocultural knowledge, and its understanding or communication, there is a need to utilize existing specialized skills: what are the experiences of other dictionary initiatives inside and outside the country?

Therefore there is also a need to specify what these dictionaries of the indigenous languages want to achieve — should they be monolingual, bilingual, multilingual culturally based, grammatically based dictionaries?

- (a) Languages in the culture systems development will be crucial as instruments for reviving the linguistic and cultural resources of any ethnic community.
- (b) The development of dictionaries for the ethnic languages will ensure a positive way towards democratization and the promotion of social equity and harmony, and indeed mutual respect among all ethnic groups, facilitating cultural appreciation and sharing.
- (c) Culture or ethnoculture documentation should underscore the fundamental need to facilitate the speech community to fully access and utilize its linguistic, cultural and ethnic knowledge.

In the following section, a brief look will be taken at the existing dictionaries of the Khoisan languages and the way they contribute to capacity building in linguistic description and ethnocultural knowledge presentation, making them viable projects of both documentation and codification for language promotion and development.

#### **4. Current dictionaries of the Khoisan languages**

The foregoing discussions have shown that some useful research and language promotion activities do exist among some Khoisan communities. However, as has been evident, much of this research is academic and does not have much effect in the Khoisan communities. For languages that are under this risk of dying, better methods of documentation and codification should be put in place. Researchers (cf. Haacke 1998; Kgasla and Tsonope 1988) have also argued that dictionaries afford better documentation and codification because their data could be reutilized for many other language related activities (literacy, grammar, linguistic descriptions). For a dictionary to become a source of documentation, it must be construed and conceptualized as a linguistic and cultural reference book or information database which is readily accessible or

utilizable. Also, and this is important, it should provide all possible information that constitute the linguistic and cultural repertoire of the language it is dealing with. The codification dimension hinges on the aspects of linguistic description and prescription, reducing the dictionary material to a code. This code provides rules for orthography, grammar, words and general language usage. In this regard, a dictionary, whatever its design or purpose, is didactic and its material constitutes describable or analyzable elements, that is, words, which are the informational repertoire of a language. While each of these elements or words is independent, they maintain a linguistic relationship at the grammatical level. It is the lexicographical technique which facilitates the way a dictionary as a source of documentation and codification can be accessed and utilized.

Except for Khoekhoegowab, there is a dearth of Khoisan lexicography in Southern Africa (cf. Haacke and Eiseb 2002). The reasons for this state of affairs are varied, but the main ones are: (a) Khoisan studies do not form a significant part of the research publications used in Botswana; (b) linguistic research tends to be technical and therefore not readily utilizable by many scholars; (c) language policies in Botswana restrict learning and usage of the marginalized languages, and thus limit them to the informal domains; and (d) Khoisan speakers are not literate in their languages. To assess the way linguistic work has contributed in documentation and codification of the Khoisan languages, some dictionary excerpts will illustrate the usefulness and also the shortcomings of some of the lexicographical works. The lemma **child** will be considered for all the languages discussed.

#### 4.1 Traill's !Xóõ–English dictionary

Anthony Traill who devoted his academic life to the linguistic description of !Xóõ, published his !Xóõ–English dictionary in 1984. From its presentation, it becomes clear that it was to serve as a linguistic and ethno-anthropological documentation of the !Xóõ people and their language (Traill 1984: 7).

⊙'āni 2 II/I children, diminutive plural marker of ⊙āa. Also general plural marker of membership in certain human groups (e.g. !Xóṇa-tê ⊙'āni, the !Xóõ people, !āma ⊙'āni western bands; †hūā ⊙'āni, southern bands; !xām-tê ⊙'āni, the first people) and other animated non-humans (e.g. g†xā-bu-tê ⊙'āni, the two mythical young gemsbok; ⊙àje ⊙'āni, young animals). But with most regular plural nouns it signifies diminutive only, so, !qhúū-tê ⊙'āni, is small white people; †áā-tê ⊙'āni, is small Kgalagadi people.

Currently Hiroshi Nakagawa and Andy Chebanne are working on the finalization of the trilingual !Xóõ–English–Setswana dictionary. The grand objective of this trilingual dictionary is to record the !Xóõ language, which took Anthony Traill more than 30 years. For an outsider, !Xóõ is a complex language which requires a keen learner and recorder. Traill did an immense work on this lan-

guage, also significantly contributing to socio-anthropological, and sociolinguistic documentation.

#### 4.2 Dickens' Ju|'hoan–English dictionary

The Ju|'hoan–English dictionary compiled by Patrick Dickens was published posthumously in 1994. As the compiler died in 1989 before completing the dictionary, Anthony Traill who was Dickens' mentor completed it. The key lemma, **child**, used in this discussion for the assessment of the linguistic adequacy of a dictionary is only given as:

**Dáàmà**, n. child

This entry is too brief to capture the versatility of the word *dáàmà* for it can also be used in semantic derivations. By not including these versatile semantic values of *dáàmà*, the dictionary sacrifices much of the linguistic and indigenous knowledge system of Ju|'hoasi.

#### 4.3 Visser's Naro–English dictionary

Hessel Visser's Naro–English dictionary (2001) is grammar and translation focused. However, it contains few other sources of knowledge, especially those based on pure anthropological observation, not prompted by grammar and translation. For the illustrative lemma, **child**, the dictionary presents:

**cóá** [lɔ́á], 'child'

The Naro dictionary only provides the primary value of the word *cóá*. It does not gloss the semantic derivations the word generates. Here again a lexical work moves on a solely lexical level, not going far in recording the way Naro employs lexical units in the derivation of new lexical terms, and new semantic notions.

#### 4.4 Haacke and Eiseb's Khoekhoegowab dictionary

In 2002, Wilfrid Haacke and Eliphaz Eiseb published their dictionary which represents a significant achievement in the documentation and codification of languages represented, namely Nama and Damara (Khoekhoegowab), and Hai|om. It is an elaborate lexicographical work, providing grammatical information as well as linguistic descriptions. Here follows the treatment of the illustrative lemma, **child**.

**Darob**, **dawon** [D] {**darob**, **dawob**} *n.* boy; (juvenile) son, s.a |GOAB; **daró-i** {**daró-i**} *n.* child; **daron** {**dranon**} *n.* children; **daros** {**daros**} *n.* girl; (juvenile) daughter; **darorob** {**darorob**} *n.* baby boy; small boy (in size); **daroro-i** {**daroro-i**} *n.* baby; small child; **daroros** {**daroros**} *n.* baby girl; small girl.

#### 4.5 Nakagawa's word lists of |Gui and ||Gana

Hiroshi Nakagawa has made extensive linguistic descriptions of the |Gui and ||Gana language cluster. In 2006, he completed his thesis on aspects of the phonetic and phonological structure of the |Gui language, which have hugely contributed to the understanding of the click sound system in |Gui in particular and Khoisan in general. Although he has compiled some lexicographical works (a |Gui–Japanese word list and a |Gui–Japanese–English word list), he has not yet published a dictionary.

#### 4.6 Kilian-Hatz's Khwe dictionary

This dictionary of Christa Kilian-Hatz, dealing with Khwedam (||Ani and Buga), was published in 2003. It came complete with a practical orthography, notes on grammar, and appendices and it also contains a supplement of Khwe place names in West Caprivi compiled by Matthias Brenzinger. It is an example of a lexicographical work that contributes to the documentation and codification of the Khoisan languages. For the illustrative lemma, **child**, this dictionary presents the following:

ǃǃǃ *n* (ǃǃǃ ǃǃǃ as a second part of a n–n compound; var. ǃǃǃ; q.v.) 1. child (boy), girl), embryo. 2. Son, daughter. 3. Niece, nephew [i.e., child of brother]; cf. -ǃǃǃ.  
ǃǃǃ-dà<sup>h</sup>màcǃ *n* afterbirth, placenta (*lit.* younger brother/sister of the child);  
ǃǃǃ-khòè dǃvèè *hon.* 2. *m/f.sg* [from older to younger men/women]; cf. ǃǃǃ-can  
ǃǃǃ-<sup>h</sup>ngu *n.* boys' hut  
ǃǃǃ-ò *n.* childhood

The strength of this Khwe dictionary lies in its judicious application of lexicographical techniques by adding to the definitions descriptions and translations. Yet illustrations are kept within limits, not yielding to unnecessary semantic digressions. Indeed, this dictionary has made and will make an important contribution to the efforts for language development in the Khwe community by adopting and justifying the practical orthography which has already been accepted by the speakers (cf. WIMSA 2001). There is therefore no need for a conversion from phonetic to practical as the symbols used constitute the conventional orthography.

Unfortunately, nothing much exists in the other Khoisan languages of Botswana. There are descriptive linguistic and sociolinguistic works on some of these languages (Shua, Tciretcire, Tshoa (Tshwaa) and Kua) but there has as yet been no serious lexicographical undertaking to contribute to documentation and codification. This situation should be viewed with serious concern as most of these languages are spoken in areas where language shift and language attrition are a daily experience.

## 5. Khoisan dictionary and lexicography work: The way forward

The foregoing observations on the lexicographical content of the current dictionaries present issues of critical concern, which are: (a) some of these dictionaries have been designed as research documents for the sole utilization by linguists; (b) generally these dictionaries have not been planned to contribute to the documentation of the indigenous knowledge of the Khoisan communities; (c) lemmas in these dictionaries are spelt according to the IPA which may not be practical for ordinary writing of the languages; and (d) most of these dictionaries are not accompanied by appendices (grammar, orthography) that facilitate using them as resources for language activities. Because these dictionaries are not monolingual, it means that speakers of these languages should be literate in English and be linguists themselves to access them. This is a critical situation. Therefore the next consideration is: how should Khoisan lexicography proceed to remedy these problems? The other crucial question is: what type (monolingual, bilingual, trilingual, etc.) would facilitate the utilization of Khoisan dictionaries?

For Khoisan communities in Botswana, monolingual dictionaries which lemmatize words in one language may be problematic as these languages are not taught in schools. Bilingual dictionaries which use two languages to lemmatize words are viable but it would require that one of the languages should be accessible to the speakers. Multilingual dictionaries which treat more languages can be a benefit if Setswana is included as it provides a common literacy base for many people in Botswana. Encyclopaedic dictionaries which focus not on just mere words but on concepts in the domain of the function and reference of the word would be ideal for Khoisan. However, the work that would go into an encyclopaedic compilation is huge and nor the Khoisan communities nor dedicated linguistics have resources to realize such a lexicographic undertaking.

In the African languages, lexicographic work has generally not benefited from corpus planning and therefore the quality of the dictionaries varies a great deal (cf. Otlogetswe 2007). For Khoisan specifically, dictionary work has been very insignificant (cf. Haacke 1998). The fact that Khoisan languages exist as marginalized minorities, means that within the countries where they are spoken, they are considered insignificant. Consequently, they are underresearched, and do not feature in any literacy programme. In the main, dictionary material development has not directly resulted from lexicographical work undertaken by researchers, but as a secondary consideration to linguistic data analysis by linguistic researchers, anthropologists and missionaries. These are therefore better qualified as research notes or as glossary materials, especially of a bilingual nature. Later these limitations will be critically considered, and the fact that foreign languages, definitely causing cultural and linguistic obstacles, were and are still used in this kind of work.

The foregoing observations show decisively that if these languages are to be preserved, as they are indeed suffering from the serious assailing effects of language shift, documentation through dictionaries should target language as a vehicle of the cultural and linguistic experience of the speakers. The way out of this situation is that dictionary work should be planned to facilitate functional documentation and codification. The question of a practical orthography is important if this documentation and codification should directly benefit Khoisan speech communities. The current inconsistent and eccentric writing conventions used by some researchers and community language promotion groups compound documentation and codification. The Centre for Advanced Studies of African Society (CASAS) published an orthography for Khoe and San languages developed by Namaseb et al. (2008) with a view to harmonize these divergent writing practices in the documentation and codification of Khoisan languages. However, this work cannot immediately eliminate this orthographic problem, but it should be a reminder to all Khoisan linguistic and ethnocultural promoters to be aware of the need to codify these languages for the ease of cross-Khoisan studies.

Generally, Khoisan languages suffer from poor corpora and the use of non-conventional lexicographical presentations. These linguistic and lexicographical structure problems do not facilitate the exploitation of the existing materials for language development and literacy. The following paragraphs will elaborate on what needs to be done by linguistic researchers and language associations.

One interesting development that facilitates capturing this sort of linguistic and ethnocultural information is the software of the Summer Institute of Linguistics FieldWorks. It is designed to record and generate data in dictionary format. Fieldworks proposes language projects which begin with a careful study of the language and culture of an ethnolinguistic group by:

- data capturing of linguistic and cultural realities observed in practice,
- data analysis and ascertaining linguistic and cultural rules and pre-concepts, and
- generating reports, research papers, dictionaries and write-ups of linguistic and cultural phenomena.

This software helps to manage linguistic and cultural data from the collection point up to the time the linguists wish to publish it (<http://www.sil.org/computing/ddp/DDP>). Chebanne (2007: 22) suggested a keyfile design for a toolbox which would integrate semantic and inter-Khoisan glosses to enhance the study of cognates and the analyses of linguistic and cultural material. Also it is important that the use of foreign languages to define Khoisan entries be adopted with circumspection as the speakers themselves are generally illiterate in these foreign languages. However, this observation also points to the very difficulty of working on these languages and the kind of materials that will

facilitate cross-Khoisan studies. The following keyfile design for Toolbox can be used to input multilingual entries. At the level of formatting there is a choice to generate monolingual, bilingual or trilingual dictionaries.

```
\lx  
\ps  
\sem 1  
\sem 2  
\sem 3  
\ge  
\gs  
\etc
```

(where \lx = lemma in a Khoisan language; \ps = part of speech; \sem 1 ... \sem 3 = to semantic domains; \ge = to gloss English; and \gs = to gloss Setswana).

Note: The back slash (\) is the Toolbox code.

This database design would ensure a fairly informative lemmatization of Khoisan language entries and create access to other Khoisan languages and facilitate inter-Khoisan comparative linguistic and cultural studies. Corpus design is not just a lexicographical, but also a linguistic and cultural imperative. For the Khoisan whose ecology, culture, and languages are endangered, there is a need for lexicographical work that will be holistic in data gathering techniques. While grammar and usage are important for any language, it is the other aspects of life that make the human experience complete. The best approach will be to work, in the first instance, on a bilingual dictionary frame, favouring inter-Khoisan bilingualism: !Xóõ–|Gui, Naro–Nama, etc. The second stage may then consider a multilingual word-list for ease of reference. This way, it will be easier to check lexical adequacies, and lemmatization strategies. But this approach, in contrast to a monolingual dictionary, faces serious semanticization problems when cultural vocabulary items such as proverbial expressions are encountered.

## 6. Conclusion

Languages such as those belonging to the Khoisan ethnic communities possess an important indigenous knowledge system. The Khoisan communities' knowledge of the environment is complex, their adaptation to the rough desert country require a people with special practical knowledge and skills. This impressive knowledge is found in their understanding of botany, zoology, geography, medicine, and indeed, life-saving skills. This is evident in their interaction with their environment and their ingenuity in times of hardship. The documentation of all these aspects is critical in the survival of these languages and cultures. Dictionaries are favourable for this kind of undertaking because they can holistically document language and culture and thus form a

codified base. Generally, in a region such as Southern Africa, and particularly in a country such as Botswana, codifying these languages will form part of the historical and cultural patrimony of regional countries. Archaeologists are interested in the rock art, artists in their cultural artefacts, but it behoves linguists (lexicographers, grammarians, philologists) to effectively vitalize these languages, and use all methods that could contribute to the recording, documentation and preservation of Khoisan.

## References

- Andersson, L.G. and T. Janson. 1997. *Languages of Botswana*. Gaborone: Longman Botswana.
- Barnard, A. 1982. Universal Kin Categorization in Four Bushman Societies. *L'Uomo* 5(2): 219-237.
- Barnard, A. 1985. *A Nharo Word-list, with Notes on Grammar*. Occasional Publication No. 3. Durban: Department of African Studies, University of Natal.
- Barnard, A. 1992. *Hunters and Herders of Southern Africa: A Comparative Ethnography of the Khoisan Peoples*. Cambridge: Cambridge University Press.
- Batibo, H.M. 1997. The Fate of Minority Languages in Botswana. Smieja, B. and M. Tasch (Eds.). 1997. *Human Contact through Language and Linguistics*: 243-252. Frankfurt: Peter Lang.
- Batibo, H.M., J. Mathangwane and J. Tsonope. 2003. *The Third Language Teaching in Botswana. A Consultancy Report for the Department of Curriculum and Evaluation, Ministry of Education, Botswana*.
- Bleek, D.F. 1928. *The Naron: A Bushman Tribe of the Central Kalahari*. Cambridge: Cambridge University Press.
- Chebanne, A. 2003. The Khoe and San Languages and People of Botswana: The Current Situation. Prah, K.K. (Ed.). 2003. *Silenced Voices*: 59-102. CASAS Book Series 34. Cape Town: CASAS.
- Chebanne, A. 2007. Technology and Language Preservation: Lexicographical Techniques in the Study of Under-researched Languages of Botswana. *Journal of the Linguistics Association of the Southern African Development Community (SADC) Universities, LASU* 2, July 2007: 6-17.
- Chebanne, A. 2008. A Sociolinguistic Perspective of the Indigenous Communities of Botswana. *African Study Monographs* 29(3): 93-118. Kyoto: Centre for South East Asia and Africa Area Studies.
- Chebanne, A. and K.C. Monaka. 2005. San Relocation: Endangerment through Development in Botswana. Crawhall, Nigel and Nicholas Ostler. 2005. *Creating Outsiders, Endangered Languages, Migration and Marginalization*: 101-105. Cape Town: Foundation for Endangered Languages (FEL).
- Chebanne, A. and M. Nthapelelang. 2000. The Socio-linguistic Survey of the Eastern Khoe in the Boteti and Makgadikgadi Pans Areas of Botswana. Batibo, H.M. and B. Smieja (Eds.). 2000. *The Future of Minority Languages in Botswana*: 79-94. Duisberg Papers on Research in Language and Culture 40. Frankfurt/New York: Peter Lang.
- Chebanne, A. and L. Nyati-Ramahobo. 2003. *Language Use and Language Knowledge: Data from the 2001 Population and Housing Census*. Paper presented at the Dissemination Seminar on Population and Housing Census, Gaborone, 8-11 September 2003.
- Dickens, P. 1994. *English-Ju/'hoan, Ju/'hoan-English Dictionary*. Research in Khoisan Studies 8. Cologne: Rüdiger Köppe.

- Haacke, W.H.G.** 1998. A Khoekhoe Dictionary in the Making: Some Lexicographic Considerations. Schladt, Mathias (Ed.). 1998. *Language, Identity and Conceptualization among the Khoisan*: 35-64. Research in Khoisan Studies 15. Cologne: Rüdiger Köppe.
- Haacke, W.H.G.** 1999. *The Tonology of the Khoekhoe (Nama/Damara)*. Research in Khoisan Studies. Cologne: Rüdiger Köppe.
- Haacke, W.H.G. and E. Eiseb.** 2002. *A Khoekhoegowab Dictionary with an English–Khoekhoegowab Index*. Windhoek: Gamsberg Macmillian.
- Hasselbring, S.** 2000. Where are the Khoesan of Botswana? Batibo, H.M. and B. Smieja. 2000. *Botswana: The Future of Minority Languages*: 13-32. Duisberg Papers on Research in Language and Culture 40. Frankfurt/New York: Peter Lang.
- Hasselbring, S.** 2001. *A Sociolinguistic Survey of the Languages of Botswana*. Sociolinguistic Studies of Botswana Languages Series 2. University of Botswana/University of Tromsø Collaborative Basarwa Research Programme. Gaborone: University of Botswana.
- Kgasa, M.L.A. and J. Tsonope.** 1998. *Thanodi ya Setswana*. Gaborone: Longman Botswana.
- Kilian-Hatz, C.** 2003. *Khwe Dictionary. With a Supplement on Khwe Place Names of West Caprivi by Matthias Brenzinger*. Namibian African Studies 7. Cologne: Rüdiger Köppe.
- Köhler, O.R.A.** 1971. Die Khoe-sprachigen Buschmänner der Kalahari; ihre Verbreitung und Gliederung. *Forschungen zur allgemeinen und regionalen Geographie: Festschrift für Kurt Kayser zur Vollendung des 65. Lebensjahres*: 373-411. Kölner geographische Arbeiten. Hrsg. von Freunden und Mitarbeitern. Wiesbaden: Franz Steiner.
- Köhler, O.R.A.** 1981. Les langues khoisan. Perrot, J. (Ed.). 1981. *Les langues dans le monde ancien et moderne. Volume 1: Les langues de l’Afrique subsaharienne — Pidgins et créoles*: 455-615. Paris: Éditions du C.N.R.S.
- Le Roux, W.** 1999. *Torn Apart. San Children as Change Agents in a Process of Acculturation. A Report on the Educational Situation of San Children in Southern Africa*. Shakawe, Botswana: Kuru Development Trust.
- Nakagawa, H.** 2006. *Aspects of the Phonetic and Phonological Structure of the |Gui Language*. Unpublished Ph.D. Thesis. Johannesburg: University of the Witwatersrand.
- Namaseb, L., W. Haacke, L. Davids, B.K. Kure, A. Araes, D. Ortman, N. Fredericks and M.C. Moeti.** 2008. *The Standard Unified Orthography for Khoe and San Languages of Southern Africa*. CASAS Monograph Series 232. Cape Town: CASAS.
- Otlogetswe, T.J.** 2007. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Schapera, I.** 1930. *The Khoisan Peoples of South Africa: Bushmen and Hottentots*. London: Routledge & Kegan Paul.
- Schultze, L.** 1928. *Zur Kenntnis des Körpers der Hottentotten und Buschmänner. Zoologische und anthropologische Ergebnisse einer Forschungsreise in westlichen und zentralen Südafrika*. Volume 5, Part 3: 147-227.
- Snyman, J.W.** 1970. *An Introduction to the !Xu Language*. Cape Town: A.A. Balkema.
- Snyman, J.W.** 1975. Some Phonetic and Lexical Aspects of Zu|’hoasi. *Bushman and Hottentot Studies* 2: 61-102.
- Traill, A.** 1984. *!Xóǀ Dictionary*. Quellen zur Khoisan-Forschung 15. Cologne: Rüdiger Köppe.
- Traill, A.** 1985. *Phonetic and Phonological Studies of !Xóǀ Bushman*. Quellen zur Khoisan-Forschung 1. Hamburg: Helmut Buske.

- Traill, A.** 1986. Do the Khoi have a Place in the San? New Data on Khoisan Linguistic Relationship. *Sprache und Geschichte im Afrika* 7: 407-430.
- Visser, H.** 2001. *Naro-English, English-Naro Dictionary*. Gantsi: Naro Language Projects.
- Vossen, R.** 1988. Studying the Linguistic and Ethno-History of the Khoe-Speaking (Central Khoisan) Peoples of Botswana. *Botswana Notes and Records* 16: 16-35.
- Vossen, R.** 1997. *Die Khoe-Sprachen. Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Quellen zur Khoisan-Forschung 12. Cologne: Rüdiger Köppe.
- Vossen, R.** 1998. Historical Classification of Khoe (Central Khoisan) Languages of Southern Africa. *African Studies* 57(1): 93-106.
- Vossen, R. and K. Keuthmann (Eds).** 1986. *Contemporary Studies in Khoisan: In Honour of Oswin Kohler: Part 2*. Quellen zur Khoisan-Forschung 5(2). Hamburg: Helmut Buske.
- Westphal, E.O.J.** 1956. The Non-Bantu Languages of Southern Africa. Supplement to Tucker, A.N. and M.A. Bryan. 1956. *The Non-Bantu Languages of North-eastern Africa*: 158-173. London: Oxford University Press for the International African Institute.
- Westphal, E.O.J.** 1962. On Classifying Bushman and Hottentot Languages. *African Language Studies* 3: 30-48.
- Westphal, E.O.J.** 1971. The Click Languages of Southern and Eastern Africa. Sebeok, Thomas A. (Ed.). 1971. *Current Trends in Linguistics, Volume 7*: 367-420. Linguistics in Sub-Saharan Africa. The Hague: Mouton.
- Willet, S., S. Monageng, S. Saugestad and J. Hermans.** 2002. *The Khoe and San: An Annotated Bibliography, Volume 1*. Gaborone: University of Botswana.
- WIMSA & RDU (UB).** 2000. *Education for Remote Area Dwellers in Botswana; Problems and Perspectives. Report on a Three-day Conference on the Education for Remote Area Dwellers in Botswana*. Windhoek: Research and Development Unit (UB) & Workgroup of Indigenous Minorities in Southern Africa.
- WIMSA.** 2001. *The Penduka Declaration on the Standardization of Ju and Khoe Languages*. Windhoek, April 20-22, 2001.
- World Commission on Culture and Development.** 1995. *Our Creative Diversity*. Paris: UNESCO.