

Analyzing Words as a Social Enterprise: Lexicography in Africa with Specific Reference to South Africa

Professor Danie Prinsloo
University of Pretoria
danie.prinsloo@up.ac.za

Abstract

In this article a critical overview of lexicography in Africa is attempted with focus on the gradual change from a Euro-centric approach to dictionary compilation for African languages mainly performed by colonialists and in particular missionaries, to an Afro-centric approach. It will be shown that mother-tongue speakers of the African languages increasingly take responsibility for their own lexicographic destiny in the form of internally-initiated collaborative lexicographic projects and the compilation of dictionaries for African languages by Africans for Africans. It portrays a continent with hundreds if not thousands of under-resourced languages which lack dictionaries of good lexicographic quality and the challenges mainly in respect of financial constraints. It will be argued that lexicography in Africa is severely constrained by the joint detrimental effect of a lack of a dictionary culture and a lack of good dictionaries. A number of positive driving forces in African lexicography are identified but also some forces impacting negatively on African lexicography by restraining the development of a strong Afro-centric approach. Concern is expressed that the negative restraining factors seem to be on the increase, i.e. that African lexicography does not seem to keep up with modern trends in international lexicography. The article concludes with a heart-warming textbook example of a collaborative dictionary compilation project by mother-tongue speakers, the Ju|'hoan Children's Picture Dictionary.

Keywords: *African language lexicography, Afro-centric dictionary compilation, dictionary culture, Euro-centric dictionary compilation, missionaries, National Lexicography Units, under-resourced languages*

1. Introduction

This article reflects research done for an oral presentation at Australex 2015 against the background of the conference theme *Analyzing Words as a Social Enterprise: Celebrating 40 Years of the 1975 Helsinki Declaration on Lexicography*. It gives an African perspective on the progress made in terms of community engagement in lexicography, with special reference to South Africa and the Bantu languages.

The article describes a cooperative approach towards lexicographic activities for Bantu languages which are under-resourced minority languages in Africa. It will be illustrated how lexicographers, experts, practitioners, fieldworkers and community members cooperate in the compilation of dictionaries for these languages. Role-players and contributors include governmental initiatives such as the National Lexicography Units (NLU's), publisher-sponsored dictionary compilation, in-house compilers, freelance entrepreneurs, community voluntary individuals, community expert bodies e.g. the National Language Bodies (NLBs) and Provincial Language Committees (PLCs), and user feedback initiatives to create, extend and enhance dictionaries.

Community engagement is entrenched in the vision and practice that initial externally-driven initiatives be gradually replaced by community engagement taking over all of the activities and responsibilities of external forces. The focus will be on the many challenges faced by dictionary compilation in African languages being under-resourced languages.

A number of initiatives and cooperative initiatives towards lexicographic activities for Bantu languages currently under way will be described.

As a point of departure, a brief overview of the language families of Africa will be given and an indication of where they are spoken on the continent as well as some estimates on the number of African languages. This is followed by estimates of the number of dictionaries for African languages reflecting the lack of dictionaries for the continent of Africa as a whole. One of the core aspects of a discussion on African lexicography is the Euro-centric character of initial African language dictionaries and the urge to change it to an Afro-centric character. This aspect is discussed in detail. In the following section attention is given to governmental projects and institutes followed by a discussion on resource-scarceness of African languages, especially financial constraints portrayed as the root of all problems in African language lexicography, especially the lack of good dictionaries and the absence of a strong dictionary culture. A brief discussion is given of lexicographic problems caused by grammatical complexity and lemmatization problems in African languages. Finally, a case of community engagement in the compilation of an African language children's dictionary is described.

2. African languages

African languages can be divided into six major language families (Wikipedia, 2017):

- Afroasiatic languages are spread throughout Western Asia, North Africa, the Horn of Africa and parts of the Sahel.
- Austronesian languages are spoken in Madagascar.
- Indo-European languages are spoken in South Africa and Namibia (Afrikaans, English, German) and are used as lingua francas in the former colonies of Britain (English), former colonies of France (French), former colonies of Portugal (Portuguese), former colonies of Spain (Spanish) and the current Spanish Moroccan enclaves of Ceuta and Melilla (Spanish).
- Khoe languages are concentrated in the Kalahari Desert of Namibia and Botswana.
- Niger–Congo languages (Bantu and non-Bantu) cover West, Central, Southeast and Southern Africa.
- Nilo-Saharan languages (unity debated) are spoken from Tanzania to Sudan and from Chad to Mali.



Figure 1: The language families of Africa (Wikipedia, 2017)

Estimates on the number of languages spoken in Africa differ substantially but all sources suggest that the number is huge. Mann (1990, p. 1) refers to Alexandre who claims that 800 languages are spoken in Africa, but Alexandre acknowledges the possible validity of other estimates of approximately 1 200. Mann also refers to Mann and Dalby who sets the number of African languages to 2 250. Wikipedia sources such as Epstein and Kole estimate even more.

The total number of languages natively spoken in Africa is variously estimated (depending on the delineation of language vs. dialect) at between 1 250 to 2 100, and by some counts at "over 3 000" (Wikipedia, 2017).

3. African language dictionaries

In the African context "dictionary" is a relative term, since mere word lists are often regarded as dictionaries.

Awak (1990, p. 13) refers to Hendrix who says that over a period of 350 years 2 600 African lexicons have been compiled, but more than half of them cannot be regarded as dictionaries but are merely word lists. A thousand-plus dictionaries for say, 2 000 languages are in stark contrast with a single language such as English for which there were at least 6 700 English dictionaries 30 years ago, Quirk as cited by Awak (1990, p. 8), in the library of Indiana State University at Terre Haute.

More than a decade ago De Schryver (2003, p. 2) found that there were already nearly 200 internet dictionaries for nearly 120 different African languages at the time. However, from a lexicographic perspective most of these sources were merely word lists or lemma lists with only basic information on form and meaning, or paper dictionaries presented online. They do not answer the expectation of being dictionaries with true electronic features, appealing and effective screen presentation and the ultimate: online dictionaries solving lexicographic problems that could not be satisfactorily solved in a paper dictionary. Awak (1990) points out that at the time of the survey, there were more African languages than there were dictionaries for African languages and that the lexical coverage was inadequate:

With such a proliferation of dictionaries, one finds it disheartening to realize that most languages of the world, particularly African languages, lack adequate lexical coverage. (p. 8)

Assigning one dictionary to each African language, the distribution would be such that many languages would not have any dictionaries at all. This indeed is the reality: many languages cannot boast a single dictionary. (p. 14)

4. The Euro-centric character of African language dictionaries

A further feature of early African language lexicography, which is often criticized by researchers such as Awak (1990, p. 17), Gangla (2001, p. 52), Nkomo (2008, p. 48) & Prah (2007, p. 23), is that it was Euro-centric; compiled by Europeans for their own use and not by mother-tongue speakers for mother-tongue speakers. Awak (1990) formulates it as follows:

The history of lexicography in Africa began as a result of European activities: exploration, evangelization and colonialization. The early lexicons, whether compiled by explorers, missionaries or colonial administrators, were 'Euro-centered', produced in Europe for Europeans rather than for African users. ... Even with the emergence of modern linguistics, lexicographic works have been primarily intended for scholarly interest and not for the needs of ordinary Africans. (p. 17)

Gangla (2001, p. 52) shows that even the pictorial illustrations in the *South African Multi-language Dictionary and Phrase Book* (Reynierse, 1996) are more Euro-centric than African, for instance that the rooms in a house include a pantry while the sports played are cricket and rugby.

Prah (2007) takes a less subtle stand:

... the fact that in the present situation, the cultures and languages of the majorities are suppressed and silenced in favour of a dominant Eurocentric high culture, which everybody is willy nilly obliged by force of circumstance to emulate. (p. 23)

A number of scenarios can be distinguished for African languages dictionaries:

- dictionary compilation by foreigners abroad – a true Euro-centric approach
- dictionary compilation by foreigners in Africa, e.g. on missionary posts using Africans as informants – also a Euro-centric approach
- dictionary compilation by non-mother-tongue speakers of African languages who studied the grammar and even learned to speak African languages, working with mother-tongue speakers. – contains Afro-centric elements
- dictionary compilation by Africans guided by foreigners – contains Afro-centric elements
- dictionary compilation by Africans – a true Afro-centric approach.

Gouws (2007, pp. 314, 315) regards dictionaries compiled by “foreigners” primarily to serve their purposes and not that of the members of the local speech community as externally-motivated dictionaries, and in contrast he regards dictionaries compiled mostly by people from within the speech community but sometimes also by other people as internally-motivated dictionaries. Internally-motivated dictionaries are directed at the needs of the speech community and at enhancing the reference possibilities of that language. Gouws (2007) states:

A characteristic feature of the linguistic situation in the postcolonialization Africa is the reality of emerging indigenous languages. This has led to an increasing need for dictionaries in which these emerging languages are treated. In this regard monolingual dictionaries have to be regarded as an ultimate aim but the immediate need is for bilingual dictionaries in which an emerging language is coordinated with an already standardized language. This situation created the opportunity for a drastic swing from externally motivated to internally motivated dictionaries, resulting in a situation which sees the majority of new lexicographic projects in Africa characterized by an Afro-centered approach that deviates from the Euro-centered approach. (p. 315)

The development in African lexicography from the colonial era where a Euro-centric approach prevailed to an Afro-centric approach is a significant achievement and seems to be problem-free as if “all problems” have now been solved. Gouws (2007, p. 314) even observe “an internal drive by mother-tongue speakers of the languages to take responsibility for the compilation of dictionaries”.

Prinsloo (2017) and Prinsloo & Taljard (2017) however sketch a much less favourable situation and argue that African language lexicography not only seriously lacks behind international lexicographic standards but that the gap is increasing. In Prinsloo’s view the number of challenges for the production of good paper and electronic dictionaries for African languages is on the increase. He distinguishes “tailwinds” and “headwinds” impacting on African language lexicography. Amongst the supporting factors count the increased compilation and use of corpora for macrostructural and microstructural quality enhancement, establishment of commercial and governmental projects and institutes, increased research and publications on African language lexicography and the guidance and encouragement offered and facilitated by the African Association for Lexicography (AFRILEX). On the other hand, impacting negatively on African language lexicography are scarcity of resources, lack of a dictionary culture, problematic lexicographic traditions, excessive purism, etc. Some of these issues will be briefly discussed below.

5. Projects and institutes

Projects and institutes established for the compilation of dictionaries in Africa by Africans and for the benefit of users of the African languages have great potential to fulfil the objectives of an Afro-centric dictionary compilation and works of merit can be quoted as proof. Among these initiatives count the *Institute of Kiswahili Research* (TUKI) (University of Dar-es-Salaam, Tanzania), the Allex project (later: *African Languages Research Institute* or ALRI) in Zimbabwe, the National Lexicography Units in South Africa, the isiZulu National Corpus (Khumalo, 2017) and the recently-established South African Centre for Digital Language Resources (SADiLaR). Unfortunately such projects do not always live up to their full or expected potential. So, for example, is the Pan South African Language Board (PanSALB) often criticized for not fulfilling its intended role in the South African lexicographic landscape as stated by Wolvaardt (2017):

How have the flag bearers for South Africa's bold approach to restoring the nation's indigenous languages, become the neglected poor relations of the deeply flawed institution that is the Pan South African Language Board (PanSALB)? How has the national lexicography project, pioneered in the early years of South Africa's democratic transition by some of the country's greatest language activists and academics, been permitted to degenerate into the scattered efforts of a diminishing band of lexicographers? Forced for the last decade into perpetual begging for adequate funding, the National Lexicography Units (NLUs) hover on the verge of extinction. The critical question is, 'does anyone care?' Dishearteningly, indications from government seem to imply that the response is, 'Not really.' (p. 9)

In South Africa, community engagement is entrenched in the vision and practice that initial government funding to the NLUs for all eleven official languages of South Africa be gradually replaced by community engagement taking over all of the activities and responsibilities of the Government. As is clear from Wolvaardt's statement above, this expected level of community engagement did not materialize in the South African situation. African language lexicography should have benefited more from a number of workshops, and numerous publications on problematic aspects of Bantu languages and efforts to build corpora for these languages. The PanSALB also made a sincere effort in the first few years after the establishment of the National Lexicography Units to employ experts in lexicographic planning and corpus-based dictionaries to guide the units. These experts paid numerous visits to the units and gave on-site guidance to the newly employed lexicographers.

Central to Bantu language lexicography is lexicographic debate and decisions in respect of (a) lemmatization approaches, (b) orthography of the language, (c) lexicographic traditions and (d) lemmatization strategies that are unique to the Bantu languages. These aspects will be discussed in more detail below. The Bantu language lexicographer not only has to deal with all of these aspects, but he or she also has to consider the complex interplay within (a) to (d) for each dictionary to be compiled in order to fulfil the needs of the respective target users.

So, it is unfortunate that a major driving force behind what Gouws (2007, p. 315) refers to as "a drastic swing from externally motivated to internally motivated dictionaries, resulting in a situation which sees the majority of new lexicographic projects in Africa characterized by an Afro-centric approach", fails according to Wolvaardt (2017). The situation portrayed for PanSALB brings the issue of the

unavailability or underutilisation of resources as the main challenge to dictionary compilation in Africa to the fore.

6. African languages as under-resourced languages

It can be argued that being under-resourced languages is the root of all problems in African language lexicography. The overview given by the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era (CCURL, 2014) fits the African lexicographic situation like a glove and is quoted verbatim with emphasis inserted:

Under-resourced languages suffer from a chronic lack of available resources (human-, financial-, time- and data-wise), and of the fragmentation of efforts in resource development. This often leads to **small resources** only usable **for limited purposes** or **developed in isolation** without much connection with other resources and initiatives. The benefits of reusability, accessibility and data sustainability are, more often than not, out of the reach of such languages.

Yet, these languages are those that could most profit from emergent **collaborative approaches** and **technologies for language resource development**. Given the **high cost of language resource production**, and given the fact that in many cases it is impossible to avoid the manual construction of resources (e.g. if accurate models are requested or if there is to be reliable evaluation) it is worth considering **the power of social and collaborative media to build resources**, especially for those languages where there are no or limited language resources built by experts yet (CCURL, 2014).

7. Lack of good dictionaries and a dictionary culture

It could be stated outright that dictionaries for African languages are perceived to be of low lexicographic standards. Gouws (1990) and Busane (1990) formulate it as follows:

Lexicographical activities on the various indigenous African languages . . . have resulted in a wide range of dictionaries. Unfortunately, the majority of these dictionaries are the products of limited efforts not reflecting a high standard of lexicographic achievement (Gouws, 1990, p. 55).

It is thus clear that the bulk of dictionaries of major national or regional languages referred to above still leave the African user in the lurch, because of the nature of their presentation and the arrangement of the entries (Busane, 1990, p. 27).

Good dictionaries and a good dictionary culture go hand in hand, and, as will briefly be illustrated in this paragraph, are both lacking in African language lexicography.

Many African languages do not even have a single dictionary and for languages where there is a dictionary available, a one-size-fits-all product has to suffice because a single dictionary has to fulfil the needs of different target users. Mono-lingual dictionaries are scarce and bilingual dictionaries do not

succeed in providing their users with a wealth of information. They tend to give only very basic treatment of lemmas, i.e. translation equivalents. They barely fulfil the most basic text reception (decoding) needs of the users and are not useful for text production (encoding) purposes. Even the very latest series of bilingual dictionaries for African languages compiled by the government-funded NLUs in South Africa, e.g. *The Official Setswana-English dictionary* (Mareme, 2015) reflect very basic treatment, cf. an extract given of the alphabetic stretch “D” in Figure 2.

D		ENGLISH-SETSWANA BILINGUAL DICTIONARY	
daddy (noun)	papa	dealt with (verb)	kampane; kampanwa
daisy (noun)	pela	dean (noun)	modini
dam (noun)	letamo; tamo	dear wife (noun)	ntshadi
damage ¹ (noun)	tshenyegelo; tshenyō; ² (verb)	death (noun)	leso
damaged ¹ (adjective)	sentseng; senyegile; ² (verb)	death sentence (noun)	katlholeloloso
	sentsē	deaths (noun)	dintsho
dampen something (verb)	nyanya	debate ¹ (noun)	kganetsano; kganetsanyo; ngangisano
dance for oneself (verb)	ipinela		ngangisano; ² (verb)
dance (noun)	motantsho		nganga; ngangisana
dance several times especially because of pain or joy (verb)	binabina	debates (noun)	dingangisano; dikganetsano
dance ¹ (noun)	modikwadikwane; moretelediane; ² (verb)	debauchery (noun)	botlhaswa
	bina	debris left by a flood (noun)	dikgogola
dancer (noun)	motantshi	debt (noun)	sekoloto
dancers (noun)	babini	debtor (noun)	moadimi; mokoloti
dances (noun)	mmimo	debts (noun)	diakhaonte; dikoloto
dandy (noun)	seithati	decade (noun)	ngwagasome
danger (noun)	kotsi	decantation (noun)	tshololo
dangerous (adjective)	diphatsa; mabe	decay (noun)	sebody
dangerous animal (noun)	seoka	deceased ¹ (adjective)	tlhokafetseng; ² (noun)
dangerous animals (noun)	dibata		moswi;
dangerous fights (noun)	marumo		baswi (plural)
dangle (verb)	kadiela; kadietsa	deceit (noun)	boferere
dark marks on the skin (noun)	ditlhoko	deceitful (adjective)	ferekana
dark ¹ (adjective)	gontsho; ² (noun)	deceive (verb)	ferekanya; thetsa; tsietsa
	lefifi; mafifi	deceive me (verb)	ntsietsa
darken (verb)	ntshofatsa	deceive oneself (verb)	iphora
darkness (noun)	lefifi	deceived (verb)	tsietsaga
dates (noun)	matlha	deceiver (noun)	moeki
dates of birth (noun)	matsalo	December (noun)	Desemba; Disemba; Morule;
daughter (noun)	morwadi		Sedimonthole
daughter of (noun)	morwadia	decently dressed (verb)	ikatega
daughter-in-law (noun)	ngwetsi	deception (noun)	kaketso; tsietso
daughters (noun)	barwadi	decide for oneself (verb)	itaolela
daughters-in-law (noun)	dingwetsi	deciduous (adjective)	thothhorega
dawn (noun)	mahube; masa	decision (noun)	tshwetso
		decisions (noun)	ditshwetso
		declare (verb)	kakanya

Figure 2: The first section of D in *The Official Setswana-English dictionary*

As for dictionary culture, African language lexicography is severely hampered by the lack of a strong dictionary culture. Atkins (1998), after having studied the African language situation, remarks as follows:

[T]he speakers of African languages have not in their formative years had access to dictionaries of the richness and complexity of those currently available for European languages. They have not had the chance to internalize the structure and objectives of a good dictionary, monolingual, bilingual or trilingual. (p. 3)

Her observations are confirmed by Taljard, Prinsloo & Fricke (2011, p. 103).

Gouws (1990), in reference to African language dictionaries, establishes a clear link between dictionaries and dictionary culture:

With a few exceptions, these dictionaries offer only restricted translation equivalents, aimed at decoding, with no or little attention given to the encoding function of a pedagogical dictionary. ... there is less information to be exploited by the user. This results in a vicious circle, with the dictionary user not realizing what he can expect to find in his dictionary or how to interpret the given entries because the lexicographer does not include all the possible information categories or treat them on an equal basis. (p. 55)

Atkins & Varantola (1998) state:

There are two direct routes to more effective dictionary use: the first is to radically improve the dictionary: the second is to radically improve the users. (p. 83)

Gouws & Prinsloo (2005) also emphasize the urge to improve the dictionaries and at the same time the dictionary users by means of a schematic illustration given in Figure 3

DICTIONARIES				USERS		
▶▶▶▶	▶▶▶▶	▶▶▶▶	▼▼▼	◀◀◀◀	◀◀◀◀	◀◀◀◀
Bad/useless dictionary or no dictionaries available	Dictionary of relatively low lexicographic achievement	Dictionary of relatively high lexicographic achievement	▼▼▼ Perfect dictionary & Ideal User ▲▲▲	Relatively good dictionary using skills	Relatively poor dictionary using skills	Pre dictionary culture environment
▶▶▶▶	▶▶▶▶	▶▶▶▶	▲▲▲	◀◀◀◀	◀◀◀◀	◀◀◀◀

Figure 3: Towards the perfect dictionary and the ideal user (Gouws & Prinsloo 2005, p. 42)

Gouws (2016) also gives guidelines towards a comprehensive dictionary culture in the digital era.

8. Grammatical complexity and lemmatization problems in African languages

Sources such as Prinsloo (2012), Van Wyk (1995), De Schryver & Prinsloo (2000a, 2000b and 2001) and Prinsloo (2011) give detailed discussions of especially lemmatization problems in African languages with specific reference to the Bantu languages. The Bantu languages are characterized by a nominal class system according to which nouns are sub-classified into different noun classes. These classes have a complex concordial and pronominal system, and complex word formation strategies by means of numerous affixes to verbal and nominal stems, Prinsloo (2012, pp. 127, 128). Van Wyk (1995, p. 87) calculates that a single verb in Zulu for example can have up to $18 \times 19 \times 6 \times 2 = 4,104$ combinations. So, for example, for a single verb *-sebenza* 'work', Prinsloo (2012, p. 128) shows that a set of 2,525 derivations were found in a Zulu corpus. Figure 4 lists the 131 derived forms of the verb *-sebenza* that occur 20 times or more than fifty times in the corpus.

UMSEBENZI (5883), EMSEBENZINI (1456), IMISEBENZI (1009), UKUSEBENZA (548), MSEBENZI (470), LOMSEBENZI (446), NOMSEBENZI (309), USEBENZA (299), ESEBENZA (290), NGOMSEBENZI (253), UKUSEBENZISA (244), IZISEBENZI (242), SEBENZISA (219), NGOKUSEBENZISA (199), NGUMSEBENZI (179), ABASEBENZI (174), NEMISEBENZI (173), USEBENZISE (167), SEBENZA (155), EMISEBENZINI (153), USEBENZISA (143), ABASEBENZA (134), KOMSEBENZI (132), ASEBENZE (131), WASEBENZA (127), OSEBENZA (125), NGISEBENZA (124), NOKUSEBENZA (121), SOMSEBENZI (112), NGEMISEBENZI (112), USEBENZE (101), YOMSEBENZI (100), UYASEBENZA (97), ISEBENZA (96), ASEBENZA (96), BASEBENZA (92), SISEBENZA (90), ISEBENZISA (90), AYESEBENZA (88), KULOMSEBENZI (87), KUNGUMSEBENZI (86), ISISEBENZI (81), SISEBENZISE (74), ASEBENZISE (72), SISEBENZISA (69), LOWOMSEBENZI (68), ZOMSEBENZI (63), KOSEBENZA (63), KUSEBENZA (62), ZEMISEBENZI (58), NASEMSEBENZINI (58), WAYESEBENZA (57), NGOKUSEBENZA (56), BASEBENZE (55), YOKUSEBENZA (53), MISEBENZI (53), ISEBENZISE (53), EZISEBENZA (53), ZISEBENZISA (51), LEMISEBENZI (51), BESEBENZA (51), ZISEBENZA (50), SISEBENZE (50), UKUYOSEBENZA (49), KWEMISEBENZI (49), ESEBENZISA (49), ENGASEBENZI (49), ZOKUSEBENZA (45), YOKUSEBENZISA (45), NGISEBENZE (45), NEZISEBENZI (44), BASEBENZISA (44), BASEBENZISE (42), USEBENZILE (41), NGIYASEBENZA (40), ISEBENZE (37), OKUSEBENZA (36), AKASEBENZI (36), YEMISEBENZI (35), UKUSEBENZELA (35), ABASEBENZAYO (35), ABABESEBENZA (35), ESISEBENZA (34), OSEBENZISA (33), NINGASEBENZI (32), EKUSEBENZENI (32), YIMISEBENZI (31), OMSEBENZI (31), NOKUSEBENZISA (31), IYASEBENZA (31), AYOSEBENZA (31), UZOSEBENZA (30), NABASEBENZI (30), LOKUSEBENZISA (30), USEBENZELA (29), LISEBENZA (29), ASEBENZISA (29), USEBENZAPHI (28), LOKUSEBENZA (28), KUBASEBENZI (28), ENGENAMSEBENZI (27), BESEBENZISA (27), WOMSEBENZI (26), KUSEBENZE (26), ZABASEBENZI (25), OWAYESEBENZA (25), KWAKUNGUMSEBENZI (25), KWABASEBENZI (25), EZISEBENZISA (25), BOMSEBENZI (25), LISEBENZE (24), EMISEBENZI (24), BABESEBENZA (24), NGALOMSEBENZI (23), KOKUSEBENZA (23), BAYASEBENZA (23), UYOSEBENZA (22), NALOMSEBENZI (22), EKUSEBENZISENI (22), ANGISEBENZI (22), AKUNAMSEBENZI (22), ZISEBENZISE (21), NGABASEBENZI (21), IMISEBENZANA (21), EZINGENAMSEBENZI (21), ESESEBENZA (21), YASEBENZA (20), UZISEBENZISE (20), ONGASEBENZI (20), NGASEBENZA (20), ANGINAMSEBENZI (20)

Figure 4: Words containing *-sebenz-* in the Zulu corpus (frequencies in brackets)

In order to find the meaning of, for example, *nomsebenzi* (309), *ngomsebenzi* (253), *emsebenzini* (1,456), *somsebenzi* (112), etc. the user has to identify the noun through stripping off the affixes to isolate *-sebenza* in order to look it up and has to “add-up” the meanings of the stripped suffixes again in order to arrive at the actual meanings of these derivations. Bantu words are thus highly complex in terms of especially morpho-phonological structure and acquire much attention for analysis in the process of lemmatization.

To make things worse, two different word lemmatization strategies viz. a stem-based approach versus a word-based approach were traditionally followed. A stem-based approach requires stem identification as explained for *-sebenza* but can be highly problematic to perform for users who do not have a sound knowledge of the grammar of the language. This problem is so severe that there are instances where neither the lexicographer nor the user knows what the stem is, cf. Van Wyk (1995) for a detailed discussion. Prinsloo (2009) says that lexicographers furthermore err in forcing stem lemmatization onto a number of disjunctively written languages for which word lemmatization is by far the better option:

Prinsloo & De Schryver (1999, p. 261) point out that the user is unnecessarily burdened with numerous problems relating to isolating the stem in many problematic instances such as *ngwana* (*mo-ana) ‘child’, *mmušo* (*mo-bušo) ‘government’, *muši* (*mo-uši) ‘smoke’, where the noun stem is no longer synchronically identifiable. In some cases, (such as stems containing the nasal prefix of class nine or aspirated and non-aspirated noun stems), it is simply not possible for either the user or the lexicographer to determine unambiguously what the form of the isolated stem is. Lexicographers for the disjunctively written languages need not follow the stem lemmatisation tradition for the sake of tradition, nor should they assume that stem lemmatisation is more ‘scientific’ than word lemmatization. (p. 158)

9. Community engagement in the compilation of dictionaries: The Ju|’hoan Children’s Picture Dictionary

Community engagement in the compilation of the Ju|’hoan Children’s Picture Dictionary (Jones & Cwi, 2014), is a heart-warming textbook example of how wings could be given to the Afro-centric approach to dictionary compilation. In its self-description the compilation of this dictionary is described as a collaborative project between the Namibian Ju|’hoan from the Tsumkwe region and academics from various fields. The primary aim of this dictionary is stated as to provide Ju|’hoan children with a piece of mother-tongue literature that is locally inspired and that can also be shared with those from the outside world (Jones & Cwi, 2014b).

Ju|’hoansi is a Kx’a (Northern Khoesan) language. It is the language of the Ju|’hoan people who are San. They can be found in Namibia and Botswana. Ju|’hoansi is an endangered language with only 11000 speakers left (Jones & Cwi, 2014a, p. vi).



Figure 5: Location of Tsumeb in Namibia (Google map, 2017)



The children got to work on their ideas for developing their own dictionary.

Figure 6: Children at work on the Ju|'hoan Children's Picture Dictionary (Jones & Cwi, 2014c).

The themes covered in this dictionary are grouped into thematic categories for animals, birds, insects, home and family, hunt, gather and dance. Pictorial illustrations are given for each lemma and translation equivalents are given in Afrikaans and English. Clickable sound icons are provided in the accompanying CD-ROM.



Figure 7: Illustrations in the Ju|'hoan Children's Picture Dictionary (Jones & Cwi, 2014a).

10. Conclusion and the way forward

In this article it was attempted to give a perspective on the development of African lexicography from a Euro-centric to an Afro-centric approach for dictionary compilation for African languages. Enough

evidence exists that this change is in principle well on its way and it could be expected that it would gain more momentum.

It has however also been shown in some detail that although there are many factors positively impacting on African language lexicography, there might be even more factors hampering the rapid and successful transition to Afro-centric dictionary compilation. The main problematic aspect that was singled out is lack of resources, especially financial constraints.

We believe that the way forward depends on increased willingness by mother-tongue speakers of African languages to be engaged in community projects for dictionary compilation and to follow best international practices for the compilation of modern paper and electronic dictionaries.

References

- Atkins, B.T.S. (1998). *Some discussion points arising from Afrilex-Salex '98*. [Unpublished course evaluation document]. University of Pretoria, South Africa.
- Atkins, B.T.S., & Varantola, K. (1998). Monitoring Dictionary Use. In B.T.S. Atkins (Ed.), *Using dictionaries. Studies of dictionary use by language learners and translators* (pp. 21-81). Tübingen, Germany: Max Niemeyer Verlag.
- Awak, M.K. (1990). Historical Background, with Special Reference to Western Africa. In R.R.K. Hartmann (Ed.), *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989: Exeter Linguistic Studies 15* (pp. 8–18). Exeter, England: University of Exeter Press.
- Busane, M. (1990). Lexicography in Central Africa: the user perspective with special reference to Zaïre. In R.R.K. Hartmann (Ed.), *Lexicography in Africa. Progress reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989: Exeter Linguistic Studies 15* (pp. 19–35). Exeter, England: University of Exeter Press.
- CCURL. (2014): Proceedings overview: *Workshop on collaboration and computing for under-resourced languages in the Linked Open Data Era*. Retrieved from <http://www.ilc.cnr.it/ccurl2014/>
- De Schryver, G.M. (2003). Lexicographers' dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2), 143–199.
- De Schryver, G.M., & Prinsloo, D.J. (2000a). Electronic corpora as a basis for the compilation of African-language dictionaries, Part 1: The macrostructure. *South African Journal of African Languages*, 20(4), 291–309.
- De Schryver, G.M., & Prinsloo, D.J. (2000b). Electronic corpora as a basis for the compilation of African-language dictionaries, Part 2: The microstructure. *South African Journal of African Languages*, 20(4), 310–330.
- De Schryver, G.M., & Prinsloo, D.J. (2001). Corpus-based activities versus intuition-based compilations by lexicographers, the sepedi lemma-sign list as a case in point. *Nordic Journal of African Studies* 10(3): 374–398.
- Gangla, L. A. (2001). *Pictorial illustrations in dictionaries*. Retrieved from: <https://repository.up.ac.za/bitstream/handle/2263/22862/Complete.pdf>
- Google map. (2017). Retrieved from: <https://www.google.co.za/maps/place/Tsumeb,+Namibia/>
- Gouws, R.H. (1990). Information categories in dictionaries, with special reference to Southern Africa. In R.R.K. Hartmann (Ed.), *Lexicography in Africa. Progress reports from the Dictionary Research*

- Centre Workshop at Exeter, 24–26 March 1989: *Exeter Linguistic Studies 15* (pp. 52–65). Exeter, England: University of Exeter Press.
- Gouws, R.H. (2007). On the development of bilingual dictionaries in South Africa: aspects of dictionary culture and government policy. *International Journal of Lexicography*, 20(3), 313–327.
- Gouws, R.H. (2016). Op pad na 'n omvattende woordeboekkultuur in die digitale era. *Lexikos*, 26, 103–123.
- Gouws, R.H., & Prinsloo, D.J. (2005). *Principles and practices of South African lexicography*. Stellenbosch, South Africa: African Sun Media.
- Jones, K., & Cwi, T.F. (2014a). *Ju/'hoan children's picture dictionary*. Pietermaritzburg, South Africa: University of KwaZulu-Natal Press.
- Jones, K., & Cwi, T.F. (2014b). *Ju/'hoan children's picture dictionary. Information leaflet*. Pietermaritzburg, South Africa: University of KwaZulu-Natal Press.
- Jones, K., & Cwi, T.F. (2014c). *Ju/'hoan children's picture dictionary. Interactive disc-gallery*. Pietermaritzburg, South Africa: University of KwaZulu-Natal Press.
- Khumalo, L. (2017). The design and implementation of a corpus management system for the isiZulu National Corpus. In S.E. Bosch, & D. Nkomo, (Eds.), *Abstracts of the 22nd International Conference of the African Association for Lexicography. Conference of the language associations of Southern Africa (CLASA)*. Rhodes University, Grahamstown, South Africa, 26–29 June 2017.
- Mann, M. (1990). A linguistic map of Africa. In R.R.K. Hartmann (Ed.), *Lexicography in Africa. Progress reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989: Exeter Linguistic Studies 15* (pp. 1–7). Exeter, England: University of Exeter Press.
- Mareme, G.B. (Ed.) (2015). *Thanodi ya Temepedi Setswana-English*. Setswana National Lexicography Unit. Mmabatho, South Africa: South African Heritage Publishers.
- Nkomo, D. (2008). Indigenous languages for culture, science and technology in Africa: On the contributions of lexicography and terminology development. *Postamble*, 4(1), 46–63.
- Prah, K.K. (2007). *Challenges to the promotion of indigenous languages in South Africa*. Retrieved from <http://www.casas.co.za>
- Prinsloo, D.J. (2009). Current lexicography practice in Bantu with specific reference to the Oxford Northern Sotho school dictionary. *International Journal of Lexicography*, 22(2), 151–178.
- Prinsloo, D.J. (2011). A critical analysis of the lemmatisation of nouns and verbs in isiZulu. *Lexikos*, 21, 169–193.
- Prinsloo, D.J. (2012). Electronic lexicography for lesser-resourced languages: The South African context. In S. Granger, & M. Paquot (Eds.), *Electronic Lexicography* (pp. 119–144). Oxford, England: Oxford University Press.
- Prinsloo, D.J. (2017). Africa's response to the corpus revolution. *11th International Conference of the Asian Association for Lexicography. Asialex 2017*. 10–12 June 2017. Guangzhou, China.
- Prinsloo, D.J., & Taljard, E. (2017). Taking stock of changes in lexicography with specific reference to their impact on African language lexicography. *AFRILEX workshop: Lexicography: a continuously changing discipline* 22 June 2017. Stellenbosch, South Africa.
- Reynierse, C. (Ed.) (1996). *South African multi-language dictionary and phrase book*. Cape Town, South Africa: Reader's Digest Association of South Africa.
- Taljard, E., Prinsloo, D.J., & Fricke, I. (2011). The use of LSP dictionaries in secondary schools – a South African case study. *South African Journal of African Languages*, 31(1), 87–109.

- Van Wyk, E.B. (1995). Linguistic assumptions and lexicographical traditions in the African languages. *Lexikos*, 5, 82–96.
- Wikipedia. (2017). Languages of Africa. Retrieved from: https://en.wikipedia.org/wiki/Languages_of_Africa#/media/File:Map_of_African_language_families.svg
- Wolvaardt, J. (2017). South Africa's National Lexicography Units: Time for a reboot? In S.E. Bosch, & D. Nkomo, (Eds.), *Abstracts of the 22nd International Conference of the African Association for Lexicography. Conference of the language associations of Southern Africa (CLASA)*. Rhodes University, Grahamstown, South Africa, 26–29 June 2017.