
A Critical Analysis of the Lemmatisation of Nouns and Verbs in isiZulu

D.J. Prinsloo, *Department of African Languages, University of Pretoria,
Pretoria, South-Africa (danie.prinsloo@up.ac.za)*

Abstract: This article is a critical evaluation of lemmatisation strategies for nouns and verbs in isiZulu with specific attention to the problem of stem identification. The presumed target users of dictionaries compiled according to these lemmatisation strategies are non-mother tongue learners of isiZulu. The advantages versus disadvantages of lemmatising verbal and nominal stems, verbal and nominal stems without suffixes, and nominal words will be considered mainly in terms of the entire paradigm containing the verbal root *-sebenz-* from an isiZulu corpus. The conclusion reached is that word lemmatisation is preferred over both stem lemmatisation and lemmatisation of stems without suffixes. It will be argued that the problem of stem identification can only be solved in electronic dictionaries and the electronic dictionary *isiZulu.net* will be analysed in this regard.

Keywords: CONJUNCTIVE ORTHOGRAPHY, DISJUNCTIVE ORTHOGRAPHY, ELECTRONIC DICTIONARIES, LEFT-EXPANDED ARTICLE STRUCTURE, LEMMATISATION, STEM IDENTIFICATION, STEM LEMMATISATION, USER PERSPECTIVE, WORD LEMMATISATION

Opsomming: 'n Kritiese evaluering van die lemmatiseringstrategieë vir naamwoorde en werkwoorde in isiZulu. Hierdie artikel is 'n kritiese evaluering van die lemmatiseringstrategieë vir naamwoorde en werkwoorde in isiZulu met spesifieke aandag aan die probleem van stamidentifikasie. Die veronderstelde teikengebruikers van woordeboeke wat volgens dié lemmatiseringstrategieë saamgestel is word omskryf as nie-moedertaal leerders van isiZulu. Die sterk- versus swakpunte van stamlemmatisering vir werkwoorde en naamwoorde, werkwoord- en naamwoordstamme sonder suffikse en woordlemmatisering vir naamwoorde sal oorweeg word hoofsaaklik in terme van die volledige paradigma 'n willekeurig gekose wortel, naamlik *-sebenz-* in 'n isiZulu korpus. Daar word tot die gevolgtrekking gekom dat woordlemmatisering bo stamlemmatisering verkies word. Daar sal aangevoer word dat die probleem van stamidentifikasie slegs in elektroniese woordeboeke opgelos kan word en die elektroniese woordeboek *isiZulu.net* sal vir dié doeleindes ontleed word.

Slutelwoorde: DISJUNKTIEWE SKRYFWYSE, ELEKTRONIESE WOORDEBOEKE, GEBRUIKERSPERSPEKTIEF, KONJUNKTIEWE SKRYFWYSE, LEMMATISERING, LINKS-UITGEBREIDE ARTIKELSTRUKTUUR, STAMIDENTIFISERING, STAMLEMMATISERING, WOORDLEMMATISERING

Introduction

The publication of the first dictionary for isiZulu using a word strategy, instead of the traditional device of stem lemmatisation, reopens the debate on stem versus word lemmatisation in African languages. In particular, the question is whether the problem of stem identification — which proved to be the major stumbling block for learners to find lemmas in isiZulu dictionaries — has been solved. To date, most publications on lemmatisation in the African languages contrast disjunctively written languages (e.g., Sepedi, Setswana and Sesotho) with those with a conjunctive orthography (e.g., isiZulu, Siswati and isiXhosa) in order to indicate the advantages and disadvantages of stem as opposed to word lemmatisation. The main argument has been that stem lemmatisation is an accepted, or even the best strategy for conjunctively written languages, but that word lemmatisation is a better option for disjunctively written languages. The principal reason for this is that stem lemmatisation introduces unnecessary problems for the user of a dictionary of a disjunctively written language, especially with regard to the identification of nominal stems. The stem tradition, nevertheless, supported by certain assumptions, such as being the more scientific option gained such momentum that a number of stem dictionaries were compiled for the Sotho languages as well. Word lemmatisation for conjunctively written languages was considered by Van Wyk (1995) and preliminary experiments were conducted at some of the National Lexicography Units in South Africa on the feasibility and possible advantages of word lemmatisation for conjunctively written languages. However, it was only in 2010 with the publication of the *Oxford Bilingual School Dictionary: Zulu and English (OZSD)* that the almost sacred stem tradition of lemmatisation for an Nguni language was broken using word lemmatisation for an isiZulu dictionary.

The focus of this article differs from earlier research in the sense that first, the issue of stem identification takes centre stage, and secondly, that the advantages and disadvantages of stem versus word lemmatisation are not described in terms of conjunctively versus disjunctively written languages but rather in terms of the advantages and shortcomings of these approaches for the conjunctively written Nguni languages, isiZulu being a case in point. Thirdly, although a selection of examples is offered, the analysis of examples focuses on a paradigm of approximately 2 500 occurrences of different words containing the root *-sebenz-* 'work' in the *Pretoria isiZulu Corpus (PZC)*.

Thus the main aim of this article is to critically evaluate lemmatisation strategies for nouns and verbs in isiZulu with specific attention to the problem of stem identification. The prime objective is to evaluate lemmatisation strategies rather than isiZulu dictionaries *per se*. For critical reviews of the two prominent isiZulu dictionaries *Isichazamazwi sesiZulu (ISZ)* and the *OZSD*, see Masubelele (2007) and Prinsloo (2010), respectively. It should be borne in mind, however, that the choice of lemmatisation strategy may depend on the type of dictionary being compiled and the probable linguistic ability of its intended users.

A consolidation of the most prominent views on stem versus word lemmatisation, which lie scattered over a number of publications, is also attempted. Finally, the success or potential of electronic dictionaries to solve stem identification problems that cannot be solved in printed dictionaries, irrespective of the lemmatisation strategy, is evaluated.

Word forms of *-sebenz-* in the PZC and a brief explanation of key terms and concepts

One of the objectives of this article is to study the success of the different lemmatisation strategies on an entire paradigm for a randomly selected word and not only, as has traditionally been done in similar discussions on lemmatisation, by quoting examples in a haphazard way as they fit the author's viewpoint. By taking an entire paradigm of real language use of a word and its derivations as input to the study, strengths and especially weaknesses in the different lemmatisation strategies, which may have been overlooked by an idiosyncratic selection of examples, come to the fore. The paradigm of the verb root *-sebenz-* has been selected on the basis of its high frequency as a verb (*-sebenza* plus verbal prefixes 4 907, *ukusebenza* 548 times in the PZC); its frequent occurrence with suffixes, e.g., *-sebenzisa* (3 373); and also for the high frequency of occurrences of nominal derivations (deverbatives) of *-sebenza*, e.g., *umsebenzi* (5 883), *emsebenzini* (1 456), *imisebenzi* (1 009), *isisebenzi* (81) and *abasebenzi* (174). See the appendix for a list of the forms occurring five times and more in the PZC. The PZC is a raw corpus of approximately six million tokens.

Detailed discussions of the morphological system of isiZulu can be found in grammar books, such as Doke (1945) and in the mini-grammars of dictionaries, such as the *Zulu-English Dictionary* (ZED) and the OZSD.

In their *Zulu-English Dictionary* (ZED), Doke and Vilakazi (1948: xxiv-xxv) define *stem* as "that part of a word depleted of all prefixal inflexions" and *root* as "the irreducible element of a word; the primitive radical form without prefix, suffix or other inflexion, and not admitting of analysis".

In *conjunctively written languages*, such as the *Nguni languages* isiZulu, isiNdebele, Siswati and isiXhosa, most word forms (tokens) contain verbal or nominal roots with affixes (prefixes or suffixes, or both) and they are written as one orthographic word. Consider the examples in (1):

(1)

sebenza 'to work' (verbal root *-sebenz-* + verbal ending *-a*)

sebenzisa 'make use' (verbal root *-sebenz-* + causative suffix *-is-* + verbal ending *-a*)

basebenza 'they work' (class prefix class 2 *ba-* + verbal root *-sebenz-* + verbal ending *-a*)

basebenzisa 'they use' (class prefix class 2 *ba-* + verbal root *-sebenz-*, + causative suffix *-is-* plus verbal ending *-a*)

isineke 'patience' (class prefix class 7 *isi-* + nominal stem *-neke*)

indlovukazi 'queen' (class prefix class 9 *in-* + nominal stem *-dlovu* + augmentative/feminine suffix *-kazi*)

nendlovukazi 'with/and the queen' (conjunctive *na-* + class prefix class 9 *in-* + nominal stem *-dlovu* + augmentative/feminine suffix *-kazi*)

As a prerequisite to subsequent discussion, a brief but more corpus-based analysis is given with the focus on the chosen paradigm of *-sebenz-*. The nominal forms of *-sebenz-* occur with prefixes and suffixes and the combination of such affixes, e.g. *emsebenzini* (1 456) (*e+umsebenz(i)+ini*) 'at work' <*umsebenzi* 'work', *nomsebenzi* (309) (*na+umsebenzi*) 'and the work', *ngomsebenzi* (253) (*nga+umsebenzi*) 'with work', *imisebenzana* (21) (*imisebenzi+ana*) 'little tasks' and *imisebenzikazi* (1) (*imisebenzi+kazi*) 'big jobs'. Affixes and combinations of affixes occurring with the verbal forms of *-sebenza* include the infinitive form *ukusebenza* (548) 'to work, working', subject concords, e.g., *usebenza* (299) 'he/she works', causative suffix, *-sebenzisa* (219) (*-sebenza+isa*) 'use, utilize', infinitive and future *ukusebenzelana* (8) 'working for each other' *ukuyosebenza* (49) (*uku+yo+sebenza*) 'to go and work', combination of concords and auxiliary *wawusebenza* (10) 'it was working'.

The conjunctive way of writing consequently results in very long words; the average word length of isiZulu words (tokens) in the PZC is 6.93 characters, cf. (2a). In *disjunctively written languages*, such as Sepedi, Setswana, Sesotho, Tshivenda and Xitsonga, nouns, verbs, concords, etc. are written as separate orthographic words, e.g., as in (2b). By contrast, the average word length for Sepedi words in the *Pretoria Sepedi Corpus* (PSC) is a mere 3.88 characters.

(2)

- a. Angifuni ukusebenza (isiZulu) 2 linguistic words, 2 orthographic words
 b. Ga ke rate go šoma (Sepedi) 2 linguistic words, 5 orthographic words
 'I do not want to work'

A popular definition of *lemmatisation* is the selection of a canonical form to represent a specific paradigm. A clear though simplified example is that *walk* is chosen as lemma to represent the paradigm *walk, walks, walked, walking*.

Stem lemmatisation refers to the selection of the verbal stem *-sebenza* from verbal forms, such as *ukusebenza* 'to work, working', *usebenza* 'he/she works', *wawusebenza* 'it was working', for instance, as the canonical form for lemmatisation purposes. For nominal forms, the nominal stem *-sebenzi* is selected for *umsebenzi* 'work, worker', *emsebenzini* 'at work', *nomsebenzi* 'and the work', *ngomsebenzi* 'with work', etc. In terms of stem lemmatisation a distinction is drawn between stem lemmatisation and left-expanded stem lemmatisation as opposed to word lemmatisation.

Word lemmatisation would have the lemma *umsebenzi* for the paradigm *umsebenzi, emsebenzini, nomsebenzi, and ngomsebenzi*.

The user perspective

As echoed in many publications, e.g., Hartmann (1989), Gouws and Prinsloo (2005a) and (2005b), contemporary lexicography is dominated by a user-driven approach. Consequently, all arguments in this article depart from the users' needs, and/or their reference skills and ability to find lemmas in isiZulu dictionaries. The target users in mind for this discussion on lemmatisation strategies in isiZulu dictionaries are learners of isiZulu with text production needs as well as the need for text reception of the prescribed books for isiZulu readers.

It may be stated at the outset that the inability of users to identify nominal and verbal stems can impede successful word searches or even result in the total failure to look up words in isiZulu dictionaries successfully. The situation is aggravated by the reality in Africa that users generally lack a dictionary culture and dictionary using skills (cf. Gouws and Prinsloo 2005a: 42).

The stem versus the word tradition in lemmatisation

Bennett (1986) as quoted by De Schryver (2010: 163) rightfully points to the complexity of nouns and verbs in African languages and asserts that stem identification can be problematic.

There has been debate as to the proper arrangement of the Bantu lexicon, and the question is far from settled. The inflection of nominals and verbals by means of prefixes, and the complex and productive derivational system, both characteristic of Bantu languages, pose difficulties [...] If items are alphabetized by prefix [...] a verb will be listed far from its nominal derivations, however transparent these may be. [...] A competing school arranges the lexicon by stem or root; this usefully groups related items, and saves on cross-referencing. Unfortunately, in such a system the user must be able to identify the stem, which given the sometimes complex morphophonemics of Bantu languages may not be easy. Bennett (1986: 3-4)

Van Wyk (1995) puts the issue of stem versus word lemmatisation in perspective in relation to disjunctively versus conjunctively written languages. Van Wyk (1995: 82) notes that two lexical traditions exist in the African languages in South Africa, i.e., the *word* tradition and the *stem* tradition:

According to the word tradition, lemmas are based on complete written words, and there is a one-to-one correspondence between written words and lemmas. According to the stem tradition, lemmas are based on the stems of written words without their prefixes.

Subsequent publications dealing with problematic aspects of lemmatisation in African languages, such as Prinsloo (1994), Prinsloo and Gouws (1996), Prinsloo and De Schryver (1999) and Prinsloo (2009) have departed from Van Wyk's pioneering analysis. The fundamental issues raised in Van Wyk's 1995 study

are not discussed in any detail here — only aspects that are relevant for this article are briefly outlined as a basis for the discussion that follows.

Van Wyk (1995) begins by dismissing the claim that the stem tradition is in any way superior to the word tradition. He states:

[...] many lexicographers have come to the erroneous conclusion that only the stem tradition is linguistically justified. Ziervogel [...], for example, claims that it is scientifically sound, and Ziervogel and Mokgokong [...] state categorically that it is the only scientific method. (Van Wyk 1995: 84).

Then, he refutes the claim that stem lemmatisation is more economical, in terms of dictionary space, than word lemmatisation. Thirdly, he highlights the misconception that verbal affixation and nominal affixation are equally productive and therefore necessitate stem lemmatisation for nouns. Finally, he highlights the problems in respect of stem lemmatisation, especially in the case of some class 9 nouns where neither the lexicographer nor the user can identify the stem. For example, stem identification is very problematic in cases such as *intaba* 'mountain', *intombi* 'girl', *inkosi* 'king' and *inkabi* 'ox'. The uninitiated user would not know that the stem form of *intaba* is *-ntaba*, but for *intombi* it is *-thombi*, for *inkosi*, *-khosi* and for *inkabi*, *-nkabi* in order to look these terms up. Mpungose (1998: 65) agrees by saying that the process of lemmatising nouns in classes 9 and 10 is problematic and he refers to the traditional method as lemmatising the "lexical noun by etymological noun-stem". Mtuze (1992: 17), in reference to nominals of Class 9 and Class 10 in isiXhosa, bluntly states:

You never knew how these nominals were lemmatised [...] In some cases, you had to struggle trying to look up words such as *ingulube* (the wild pig) as the entry could either be under *g* or under *n*.

There is, fortunately, no dispute regarding stem versus word lemmatisation in the case of verbs. Followers of both traditions agree that verbs should be lemmatised on their stems. Sources, such as Prinsloo (2009) debating the issue, consequently focus on nominal stem identification as the problematic area. However, it should be emphasized that the fact that both traditions agree on stem lemmatisation for verbs does not alleviate the problem of stem identification. It is argued here that stem identification for verbs in Nguni languages is as problematic for verbs as it is for nouns — the identification of *-sebenza* from the numerous verbal forms, or *-sebenzi* from the nominal forms in the paradigm of *-sebenz-* is proof thereof. From the 31 orthographic forms in the appendix occurring more than 100 times in the PCZ, 15 are verbs and 16 are nouns. The challenge to identify the stem is exactly the same for nouns and verbs.

The advantages and disadvantages of lemmatising stems

This strategy entails the lemmatisation of nominal and verbal stems with or without suffixes. For verbs, all suffixes can be removed rendering the verbal

root, e.g., *-sebenz-* plus verbal ending *-a*, *-sebenza* as lemma and also for words such as *ukusebenza* (548), *usebenza* (299) and *esebenza* (290) from the appendix. Verb stems can also be lemmatised with their suffixes, e.g., *-sebenzisa* for the paradigm *ukusebenzisa* (244), *-sebenzisa* (219) and *ngokusebenzisa* (199). For nouns, the result of affixal stripping will render the nominal stem *-sebenzi* as lemma for *ngomsebenzi* (253), *izisebenzi* (242), *ngumsebenzi* (179). Consider verbal stem without suffix(es), verbal stem with suffix(es), and noun stem lemmatisation in ZED in (3).

(3) ZED

sebenza (3.2.9) v. [> perf. *-sebenzile*; pass. *setshenzwa*; neut. *sebenzeka*; ap. *sebenzela*; rec. *sebenzana*; caus. *sebenzisa*; int. *sebenzisa*; dim. *sebesebenza*; inbenzo; isisebenzi; umsebenzi; umsebenzi.]
 1. Work, labour; work at. *ukusebenza emasimini* (to work in the fields). *Usebenza imithi* (He is a druggist). *Usebenza izimoto* (He is a motor-mechanic). *Woke usebenze* (You will have a tough job).
 2. Administer medical treatment to. *ukusebenza umuntu ogulayo* (to treat a sick person).
-sebenzi (*isisebenzi*, 2.2-4.4.3.9, *izisebenzi*) n. [< *sebenza*.]
 Servant, employee; workman.
-sebenzi (*umsebenzi*, 2.2-4.4.3.9, *abasebenzi*) n. [< *sebenza*.]
 Workman, labourer.

-sebenzi (*umsebenzi*, 2.2-4.4.3-8.9, *imisebenzi*) n. [< *sebenza*.]
 Work, job, occupation. *Ngifuna umsebenzi* (I want work).
sebenzisa (3.3.2.9) v. [caus. < *sebenza*. > perf. *-sebenzisile*; pass. *setshenziswa*; neut. *sebenziseka*; ap. *sebenzisa*; rec. *sebenzisa*.]
 1. Cause to work, help to work, work with.
 2. Use, utilize. *ukusebenzisa umshanelo* (to use a broom).
-sebenzo (*insebenzo*, 2.4.3.9, *izinsebenzo*) n. [< *sebenza*.]
 Wages, payment for work. *Bathola insebenzo yabo ngesonto* (They get weekly wages).

An advantage of stem lemmatisation is that it is the undisputed option for the lemmatisation of verbs, not only for conjunctively written languages but also for the disjunctively written ones. Van Wyk (1995: 85) states clearly that "except for the notational device of a hyphen [...] the entry for 'see' will [...] be found as **bona** in word dictionaries and as **-bona** in stem dictionaries".

For nouns, stem dictionaries normally provide the stem and the possible nominal prefixes in brackets, e.g., *-sebenzi* (um-, aba-, imi-) and the user can rightfully conclude that the forms are *umsebenzi*, *abasebenzi* and *imisebenzi*. The lexicographer must of course make sure that the possible combinations suggested by this notation, i.e., *umsebenzi*, *abasebenzi* and *imisebenzi* are correct. An example where it is not the case is the lemmatisation of *inkosi/amakhosi* 'king(s)', in *Woordeboek Afrikaans-Zoeloe, Zoeloe-Afrikaans* (WAZ) in (4).

(4) WAZ

-khosi, (in-, ama-), b; 1. koning, regent, hoofman

The user's conclusion is: **inkhosi* 'king', *amakhosi* 'kings', of which the plural form is correct but the singular form is incorrect — it must be *inkosi*. This is a very serious mistake, since the dictionary should never guide the user to such incorrect conclusions.

Van Wyk (1995), however, has fundamental problems with stem lemmatisation for nouns. In his view, the first disadvantage stems from the misconception that nominal and verbal stems are equally productive in combining with prefixes. Van Wyk (1995: 87) quite correctly shows that verbs stems may, for

instance, productively combine with all the subject concords, object concords, negative morphemes and modal morphemes, $18 \times 19 \times 6 \times 2$ which comes to 4 104 possible combinations. Noun stems can only be used with a small number of class prefixes. In the case of nominalizations of *-sebenz-*, nouns occur in classes 1, 2, 3, 4, 7 and 8 as in (5a). The starred forms in (5a) indicate ungrammatical combinations in terms of the class prefix paradigm for isiZulu. For other nouns, the number of possible combinations can be even less, as indicated in (5b).

(5)

- a. Class 1 and 3: *umsebenzi* 'worker, work', Class 2: *abasebenzi* 'workers', Class 4: *imisebenzi* 'work, jobs', Class 5: **i(li)sebenzi*, Class 6: **amasebenzi*, Class 7: *isisebenzi*, 'employee' Class 8 *izisebenzi* 'employees', Class 9: **insebenzi*, Class 10 **izinsebenzi*, Class 11: **u(lu)sebenzi*, Class 14: **ubusebenzi*
- b.
- ntu (umu-, aba-, isi-, ubu-, u(lu)-): *umuntu* 'a human', *abantu* 'humans', *isintu* 'Bantu culture', *ubuntu* 'humaneness', *untu* 'common people'
 - khosi (in- (inkosi), ama-): *inkosi* 'a king', *amakhosi* 'kings'
 - khaya (i(li)-, ama-, um-): *ikhaya* 'home', *amakhaya* 'homes', *umkhaya* 'members of the family'
 - tho (isi-, izi-, in-, izin- (u(lu)-): *isitho* 'limb', *izitho* 'limbs', *into* 'thing', *izinto* 'things', *utho* 'something'
 - daba (in-, izin-, u(lu)-): *indaba* 'story', *izindaba* 'stories', *udaba* 'a serious affair'

Thus Van Wyk (1995) concludes that there is no linguistic justification for treating nouns and verbs in the same way in terms of stem lemmatisation. Van Wyk's criticism is valid if the view is restricted to the consideration of concords in terms of verb stems and class prefixes in relation to noun stems. If, however, the complex orthographic forms of nouns and verbs are considered, e.g., as for *-sebenz-* in the appendix, then noun stems and verbs stems are on a par in terms of productive combination with affixes, such as the conjunctives *na*, in *nomsebenzi* (*na+umsebenzi*) 'and the work', *nga* in *ngomsebenzi* (*nga+umsebenzi*) 'with work' and the possessive *sa* in *somsebenzi* (*sa+umsebenzi*) 'of the work' or with the combinations of affixes. The distinction between verbal stem identification and nominal stem identification therefore effectively falls away for the target users. The question could therefore be asked: if the user has to deal with affixation on such a massive scale anyway in his/her effort to find the lemmas for nouns and verbs in the dictionary, why not depart from the stem in all instances for nouns and verbs?

Identifying the stem remains the underlying challenge for the discussion on the following two lemmatisation strategies: stems lemmatised with their prefixes, and word lemmatisation.

An issue of special relevance for a critical analysis of stem and word lemmatisation in the Nguni languages is the lemmatisation of infinitives. In all of

the lemmatisation strategies for isiZulu, i.e., stem, left-expanded and even word lemmatisation, verbs are by default lemmatised as stems. Linguists agree that the infinitive has characteristics of both nouns and verbs. Infinitives such as *ukuhamba* 'to walk, a/the walking', *ukusebenza* 'work, a/the working' and *ukukhuluma* 'speak, a/the speaking' are nouns (of class 15) and verbs at the same time. In traditional grammars the infinitive is therefore positioned and formally described within the two major categories of *verb* and *noun*. Consider (6a) in contrast to (6b) where the verbal versus nominal meanings of the infinitive are foregrounded.

(6)

- a. Angithandi ukuhamba ngezinyawo 'I do not like *to walk* on foot'
- b. Ukuhamba kuyakhathaza '*Walking* is exhausting'

Unlike the other noun classes, the stem of the infinitive noun is not a nominal, but a verbal stem and, unlike verbs, infinitives contain a noun class prefix (class prefix of Class 15). In dictionaries following a stem lemmatisation strategy, such as the ZED, all infinitives are lemmatised under their stem forms, e.g., the lemmas *-hamba*, *-sebenza* and *-khuluma*. No effort towards the lemmatisation of the nominal forms *ukuhamba*, *ukusebenza* or *ukukhuluma* is made in the ZED and no effort towards treating nominal meanings in the articles of *-hamba*, *-sebenza* and *-khuluma* has been made. Ironically, the advocates of stem lemmatisation are forced to lemmatise full words with *uku-*, e.g., in the case of the ZED for *ukuthi* (conjunctive) 'so that', *ukufa* (interjection) 'how magnificent!', *ukuba* (conjunctive) 'that (after verbs of knowing, etc.)', *ukuphela* 'only that', etc. These words belong to different parts of speech and the lemmatisation strategy could be justified. The problem lies with the fact that users are conditioned to ignore infinitive prefixes in the process of determining the lemma, i.e., not to consider the *uku-* and therefore they will look up *-thi*, *-fa*, *-ba* and *-phela* and indeed find such lemmas with treatment in the ZED without any cross-reference to the lemmatised full forms *ukuthi*, *ukufa*, *ukuba* and *ukuphela*.

Ukuthi, as a conjunctive or connective, is the most frequently used word in isiZulu and the lack of a cross-reference could simply mean that the users will not find the meaning of *ukuthi* 'so that, in order that' representing 90% of its use. IsiZulu dictionaries, such as the *Compact Zulu Dictionary* (CZD) and the *English and Zulu Dictionary* (EZD) do not handle this issue satisfactorily by either not lemmatising the conjunctive or not giving a cross-reference.

(7)

<p>a. CZD -thi (v) (ukuthi), to say.</p>	<p>b. EZD -thi defect. V. 1. say. 2. mean, intend; ... ukuthi conj. 1. (foll. by indic.) that. 2. (foll. by subjunct.) so that, in order that.</p>
--	---

The strengths and shortcomings of stems lemmatised with their prefixes

This is the lemmatisation strategy employed by dictionaries, such as the ISZ, the *Concise SiSwati Dictionary* (CSD), the *Dictionary of the Tebele & Shuna languages* (DTS) (the latter as quoted by Gauton (forthcoming)), and even for a disjunctively written language, Sesotho, in the *Southern Sotho-English Dictionary* (SSED) where stems are lemmatised but the full form of the word is given. In the ISZ, for verbs, it entails lemmatising the stem and presenting it in boldface but adding the infinitive prefix *uku-* in italics, e.g., *ukusebenza* will be lemmatised as *ukusebenza* in the alphabetical stretch S of the dictionary. Likewise, the noun *umsebenzi* will be lemmatised as *umsebenzi* under S by giving its full form with the stem in bold and the class prefix in italics as in (8).

(8)

ikusebénza (sz) [-an-; -el-; -ek-; -is-; -w-] - 1. ukwenza okuthile ngenhloso yokuthola umkhiqizo noma umphumela othile. *Ubaba usebenza eGoli.* 2. ukwelapha. *Inyanga yabe isebenza isiguli.* 3. isimo sokufakazela ukuphila kokuthile. *Ungalibali ukufonela uZitha, phela seluyasebenza ucingo lwakhe.*

isisebénzi (bz) - umuntu oqashiwe owenza umsebenzi othile; um(u)sebenzi; um(u)qashwa.

im(i)sebénzi (bz) - 1. okwenziwa ngenhloso yokuthola umkhiqizo othile noma umphumela othile. *Umsebenzi usihlasele kulezi zinsuku.* 2. idili elenziwa ekhaya njengokubuyisa osashona.

im(i)sebénzi (bz) - oqashelwe ukwenza umsebenzi othile; isisebenzi; um(u)qashwa.

ikusebénzisa (sz) [-an-; -el-; -ek-; -is-; -w-] - ukwenza okuthile kube kukhona okukusizayo ukwenza lokho okwenzayo; ukusebenza ngokuthile. *Ukusebenzisa isikele ukugunda intshebe yakhe yikhona okufanelekile.*

Adding the prefix has certain advantages, e.g., the reassuring factor (in cases of a 1-1 match), e.g., the user wants to look up *isidaka* and looks for *-daka* and finds all the different full nominal forms including *isidaka*. So (s)he knows that the process of information retrieval has been successful. The ISZ goes even further by implicitly giving morphological information about the prefix. Examples, such as *ú(lu)sinsi†* 'hair growing low on the forehead' *í(lí)fasíkoti** 'apron' *í(lí)bhoklóló°* 'a brave, confident male', contain additional information, i.e., that the full form of the prefixes are, respectively, *ulu-* and *ili-*. Presenting the full word also enables the indication of tone. The symbols '†', '*' and '°' following the lemma indicate that no plural form exists for *ú(lu)sinsi*; that *í(lí)fasíkoti* is a loan word; and that *í(lí)bhoklóló* is a neologism. Apart from giving the infinitive class prefix with verbal stems, verbal suffixes are given in brackets following the stem, i.e., not as separate lemmas, e.g., *úkúlobola* ... [-an-, -el-, -ek-, -is-, -w-]. Indicating a number of frequently used verbal suffixes in this way does enhance the comment on form but does not contribute to the comment on semantics. For the user, it means that stems containing these suffixes have to be

looked up under the basic stem, and the meanings conveyed by these suffixes then have to be added on. It also gives insufficient guidance in cases where sound changes occur as a result of affixation. As Masubelele (2007: 460) rightfully remarks, "variants of words which are the result of phonological processes, such as the passive construction have been omitted, e.g., *úkúlobola* which changes to *úkúlotsholwa* in the passive". The same holds true for the passive form of *-sebenza* where the inexperienced user is unlikely to link the passive form *-setshenzwa* with *-sebenza+w*. It would be better to lemmatise and treat derivations containing these suffixes as in (3).

Gouws and Prinsloo (2005b: 29) refer to the lemmatisation strategy where stems are lemmatised, but the full form of the words are given as left-expanded article structures.

[...] a left-expanded procedure [...] can [...] accommodate the prefixal element in a slot preceding the stem. This phasing out of the prefixal element to the article-initial position does not, however, change the status of the lemma sign as guiding element of the article because the lemmatization is still done according to an initial-alphabetical ordering in which the stem is the alphabetical point of reference.

Consider the following examples from CSD in (9).

(9) a. CSD

(kú)-hamba v. walk, go, travel, move,
 ↓
 leave. (For going to ... cf. -ya).
 (kú)-hambela v. t. 1. travel for or on
 ↓
 behalf of. 2. visit. 2. press a claim.
 (kú)-hambelana v. 1. go with each other,
 ↓
 travel together. 2. be on good terms
 si-hambi / ti- n. visitor, tourist stranger.
 um-hambi / bá- n. traveller.
 (kú)-hambisa v. t. 1. drive, cause or
 ↓
 help to move, send off. 2. bid fare-
 well to; send greetings to. 3. purge.
 (kú)-hambisana v. associate, accompany
 lu-hambo / ti- n. journey

b.

<i>kuhamba</i> v. walk, go ...
<i>kuhambela</i> v.t. ...
<i>kuhambelana</i> v. ...
<i>sihambi</i> n. ...
<i>umhambi</i> n. ...
<i>kuhambisa</i> v.t. ...
<i>kuhambisana</i> v. ...
<i>luhambo</i> n. ...

The verb stems *-hamba*, *-hambela*, *-hambelana*, *-hambisa* and *-hambisana* are lemmatised with the infinitive class prefix *ku-*. Full nouns, i.e., *sihambi*, *umhambi* and *luhambo* are given, but they are lemmatised on their stem forms *-hambi* and *-hambo*. As in the case of the verbs, the alphabetization is done on H, the first letter of the stem as indicated by the arrows in (9a). Gouws and Prinsloo (2005b) suggest that the first letters of the stems should be vertically aligned to visually strengthen the alphabetical alignment on H in (9b). In the ISZ, vertical alignment on the first letter of the stem has also not been done as in (8), but indentation of the amount of space equal to approximately three characters and the contrast between italics for prefixes and boldface for lemmas provide for a user-friendly layout and alleviate the need for vertical alignment.

Left-expanded article structures can in principle be extended to go beyond

left-expansion of class prefixes to other types of prefixes and prefixal combinations, such as conjunctives and concords. The lexicographer could for example decide to lemmatise words that occur with a high frequency in the corpus, such as *nokusebenza* (121) 'and to work', *ngokusebenza* (56) 'by working', *wayesebenza* (57) 'he/she/it was working' *ukuyosebenza* (49) 'to go and work' in (10).

(10)

nokusebenza

ngokusebenza

wayesebenza

ukuyosebenza

On the one hand, the lemmatisation strategy in the ISZ and the CSD shows characteristics of stem lemmatisation in the sense that alphabetical ordering runs on the first letter of the stem, thus ignoring the different nominal and verbal prefixes. On the other hand, it resembles a word dictionary since full words, i.e., the full infinitive form of, e.g., the verb, *kuhamba* 'to go' and the full form of the noun *sihambi* 'visitor/tourist' are lemmatised. An even closer resemblance to word dictionaries is found in the ISZ's layout, e.g., by not putting the infinite prefix in brackets or separating the prefix from the stem by means of a hyphen as was done in the CSD but using the normal orthography. Masubelele (2007: 459) quotes the following paradigm: *í(lí)daka* 'dry cattle dung', *ísídaka* 'black soil', *ú(lú)daka* 'mud', *úkú)daka* 'to become drunk' and *úm(u)daka* 'heavy, brown bracelet, bestowed as royal honour'. Advocates of the word tradition for lemmatisation will be quick to point out that no gain in terms of space saving is achieved in such cases. Giving the prefix has certain advantages. First, the reassuring factor should not be underestimated. It is of special value in the case of those class 9 nouns quoted by Van Wyk (1995: 90) where lexicographer and user have difficulty in identifying the stem form. Secondly, utilizing the opportunity to give additional morphological and tonological information is a positive aspect, provided that the user is familiar with the tonal markers. The convention used to indicate the full form of the prefix, however, carries the risk of misinterpretation that, e.g., both *usinsi* and **ulusinsi*; *ifasikoti* and **ilifasikoti*; and *ibhoklolo* and **ilibhoklolo* are grammatical, because this convention normally suggests the part in brackets as being an alternative. Thirdly, the symbols '+' for indicating that the noun does not have a plural form or '‡' in the case of certain plurals not having a singular form is a positive. This convention has, however, to be weighed against the convention '(x/y)' as, for example, used in the OZSD where both singular and plural class numbers are given and related and the appropriate class to which the word belongs is indicated by the boldface as in (14).

Masubelele (2007: 459) regards the fact that no plural indication is given in the ISZ as a problem, "since only a singular noun prefix is given with each stem, this might be problematic, especially to users who are not mother-tongue

speakers, because they might not know what the plural form of the specific word is". The question, however, is whether the user who wants to find the meaning of *umsebenzi* 'worker' will be interested to know what the plural form is? The problem is rather the amount of knowledge required from the user to look up singular or plural forms in dictionaries employing a left-expanded strategy and whether this strategy contributes in any way to resolving the problem of stem identification. The answer is *no*, although this lemmatisation strategy provides for user-friendly elements, such as additional information and reassurance, stem identification still has to be done.

Returning to the issue of infinitives: in dictionaries following a left-expanded lemmatisation strategy, the lemmas will (for example) be *ukuhamba*, *ukusebenza* and *ukukhuluma*, respectively. In the case of *ukufa* 'die, death; how magnificent!' and *ukuthi* 'say, a/the saying; so that' all three semantic distinctions in each case are accounted for and accommodated together in, say, two *subsequent* lemmas as in (11) and (12).

<p>(11) <i>ukufa</i> ... inf 1. to die ... 2. death .. <i>ukufa</i> ... interjection ... how magnificent!</p>	<p>(12) <i>ukuthi</i> ... conjunctive ... <i>ukuthi</i> ... inf 1. to say ... 2. saying ..</p>
---	--

This layout represents an improvement to stem lemmatisation as argued in terms of (7) above, because nominal, verbal and other meanings are presented together.

Advantages and disadvantages of word lemmatisation

The title of De Schryver's (2010) text: *Revolutionizing Bantu Lexicography — A Zulu Case Study*, suggests that word lemmatisation has fundamentally transformed the lexicography of the African languages. Word lemmatisation for nouns, in word dictionaries where both singular and plural forms are lemmatised means that the full singular and plural forms will be lemmatised with alphabetical sorting on the first letter of the word, e.g., *umsebenzi* 'work, worker', *abasebenzi* 'workers' and *isisebenzi* 'employee', can be looked up directly under *u*, *a* and *i*, respectively.

(13) OZSD

umsebenzi *** *noun* 3/4 (pl. **imisebenzi**)
 C-SEBENZA ■ work; job • Kunzima ukuthola
umsebenzi kule mihla. *It is very difficult to get a job these days.*
 ◇ **umsebenzi wokuqondisa**
 ■ comprehension test
abasebenzi ** *pl. noun* 7/2 See sg.
 ISISEBENZI

isisebenzi *noun* 7/2, 7/8 (pl. **abasebenzi, izisebenzi**) C-SEBENZA ■ worker; servant
 • Okhulumela **abasebenzi**, uMnu. Khulekani Mzimela. *The one who speaks on behalf of the workers is Mr Khulekani Mzimela.*

The OZSD goes beyond the lemmatisation of the basic nouns *umsebenzi*, *abasebenzi*, *imisebenzi*, *isisebenzi* and *izisebenzi*, cf. (5a) above, and offers articles for derived forms, such as *ekusebenzeni* 'in the working', *ekusebenziseni* 'in using', *ekusetshenzisweni*, *emisebenzini* 'at work' and *emsebenzini* 'at work'.

Word lemmatisation also solves the difficulties mentioned by Mtuze (1992), Van Wyk (1995) and Mpungose (1998); and for those words where it is difficult to identify the stem, e.g., *intaba*, *intombi*, *inkosi* and *inkabi*. In a dictionary where full nouns are lemmatised, the problem is avoided by the lemmatisation of these forms exactly as they are. So, if the user is given the nominal form, access to the lemma is straightforward and easy.

A typical argument against the lemmatisation of the full forms of nouns as echoed by Van Wyk (1995: 95) is that the alphabetical stretches, especially U, A, I into which nouns fall, will be very large, because nouns in classes 1, 3, 11 and 14 begin with *u-*, classes 2 and 6 with *a-*, classes 4, 5, 7, 8, 9 and 10 with *i-*. Van Wyk's estimate for isiZulu is U: 18%, I:20% and A:5%. In the OZSD, *i-* takes up 62 pages, representing 23.5%, i.e., almost a quarter of the dictionary, *u-* 40 (15.2%) pages and *a-* 14 pages (5.3). However, users are unlikely to find this at all disturbing as can be judged by looking at a typical example taken from the alphabetical stretch *isi-* in (14). In the *Collins COBUILD English Dictionary* (COBUILD), the alphabetical stretch CON- is almost 30 pages long and to the best of our knowledge no complaints have been voiced in this regard.

(14) OZSD

isiZulu–isiNgesi 113	isisebenzi → isitayela
<p>rallying cry "The Power is Ours" means that the power must be with the people.</p> <p>isisebenzi noun 7/2, 7/8 (pl. abasebenzi, izisebenzi) C -SEBENZA ■ worker; servant • Okhulumela abasebenzi, uMnu. Khulekani Mzimela. <i>The one who speaks on behalf of the workers is Mr Khulekani Mzimela.</i></p> <p>isisekelo noun 7/8 (pl. izisekelo) C -SEKELA ■ support; foundation; basis • Uphenyoluthole ukuthi izinsolo eMhlali azinaso isisekelo. <i>The investigation found that the complaints in Mhlali are without foundation.</i></p> <p>◊ gotha imbokodo nesisekelo ■ destroy completely • UShaka wayeseqothe imbokodwe nesisekelo kuzo zonke izizwe ayakhelene nazo. <i>Shaka completely destroyed all the surrounding tribes.</i></p> <p>◊ umthetho-sisekelo ■ constitution</p> <p>isishayamthetho noun 7/8 (pl. izishayamthetho) C -SHAYA + UMTHETHO ■ legislature • Isishayamthetho sesifundazwe sikhethelwa iminyaka emihlanu. <i>A provincial legislature is elected for a term of five years.</i></p> <p>isihiyagalolunye noun 7/- ■ nine</p>	<p>on a programme to lose weight, fibre can prevent you becoming hungry quickly.</p> <p>isisinga noun 7/8 (pl. izisinga) ■ rope (used for tying a calf, goat, ...)</p> <p>◊ inkonyane isethuka isisinga ■ not yet used to • Inkonyane yafike yethuka isisinga kuNompumelelo nxa efika esikoleni esikhulu saKwaMagwaza ngakho wafike waqaphelisisa konke okwenzekayo wakubhala encwadini yakhe. <i>Nompumelelo was not yet used to it when he came to the high school in Magwaza; that's why he first took careful note of everything that happened and wrote it down in his book.</i></p> <p>isisombululo noun 7/8 (pl. izisombululo) ■ solution • Unqamule uhambo lwakhe lwaseNdiya ngesonto ukuze ahlngane nezikhulu zalo mnyango emizamweni yokuthola isisombululo kule nkinga. <i>She shortened her visit to India by a week, in order to meet with the senior officials of the department in an effort to find a solution to the problem.</i></p> <p>isisu ** noun 7/8 (pl. izisu) 1 ■ stomach; belly; abdomen • Waze wafika emanzini,</p>

A second argument against the lemmatisation of the full forms of nouns pertains to the lemmatisation of plural forms of nouns, first, in terms of the additional space in the dictionary taken up by these lemmas and, secondly, that it results in overuse of the mediostructure (cross-referencing system), because all such lemmas function as cross-references to the singular forms. It cannot be denied that lemmatising plural forms takes up a great deal of additional space. However, in terms of the reassuring aspect mentioned above as well as the amount of information carried by these skeleton dictionary articles as in (15), their inclusion in the macrostructure could be justified. First, the user is reassured that (s)he is dealing with the correct lemma; secondly, information on the frequency of use is indicated (by means of e.g. *, ** and ***); thirdly, noun class information is provided; and finally a cross-reference is given to the singular form where full treatment is offered.

(15) OZSD

imiphakathi * <i>pl. noun 3/4</i> See sg. UMPHAKATHI	iziphuzo <i>pl. noun 7/8</i> See sg. ISIPHUZO
imiphefumulo <i>pl. noun 3/4</i> See sg. UMPHEFUMULO	iziqalo <i>pl. noun 7/8</i> See sg. ISIQALO
imiphumela ** <i>pl. noun 3/4</i> See sg. UMPHUMELA	iziqephu <i>pl. noun 7/8</i> See sg. ISIQEPHU
imiqhudelwano <i>pl. noun 3/4</i> See sg. UMQHUDELWANO	iziqu * <i>pl. noun 7/8</i> See sg. ISISU
imiqondo * <i>pl. noun 3/4</i> See sg. UMQHONDO	iziqubulo <i>pl. noun 7/8</i> See sg. ISIQUBULO
imisakazo <i>pl. noun 3/4</i> See sg. UMSAKAZO	izisebenzi <i>pl. noun 7/8</i> See sg. ISISEBENZI
...	izisho <i>pl. noun 7/8</i> See sg. ISISHO
imisebenzi *** <i>pl. noun 3/4</i> See sg. UMSEBENZI	izisu <i>pl. noun 7/8</i> See sg. ISISU
	izitatimende <i>pl. noun 7/8</i> See sg. ISITATIMENDE
	iziteshi <i>pl. noun 7/8</i> See sg. ISITESHI
	izitha * <i>pl. noun 7/8</i> See sg. ISITHA

Consider now the presumed or likely dictionary needs of a learner of isiZulu who wants to use the dictionary for text production. First, a typical situation in class is considered where the learner is instructed to find the meaning of a number of isiZulu words, say nouns, or to write an essay on *abasebenzi* 'workers'. In the latter case, the users simply take the dictionary, look for the lemma *abasebenzi* and the worst that can happen is of (s)he having to follow up a cross-reference to the singular form *umsebenzi* where appropriate treatment is offered. No problematic stem identification is required as in the case of stem or left-expanded lemmatisation discussed in the previous paragraphs. However, a substantial part of learners' needs is to find the meanings of words used in their prescribed books, especially isiZulu literary works, such as novels, poetry and prose.

This means they are from the outset confronted by the full/complex orthographic forms which can be more than 2 000, e.g., in the case of *-sebenza*. In

order to find the meaning of *nomsebenzi* (309), *ngomsebenzi* (253), *emsebenzini* (1 456), *somsebenzi* (112), etc. (s)he has to *identify the noun*. This has to be done principally by stripping off affixes. Even more problematic is where (s)he has to *add* characters to the word in order to reconstruct the full noun in order to look it up, e.g., *umsebenzi* for *msebenzi* (470). Adding or stripping affixes in order to find the *word form* for the word search is as challenging to the user as stem identification. This unfortunately means that the problem of stem identification is simply replaced by the challenge of identifying word forms. Advocates of the stem tradition could argue that, if identification of the lemma entails the selection of a section of the complex orthographic word anyway, why not then also cut the noun prefixes, which brings one back to stem lemmatisation? It is not possible to lemmatise the entire paradigms of all isiZulu words in printed dictionaries. Prinsloo (2010) tries to make a case for selection on the basis of frequency in this regard by saying that the lexicographer should ensure that the frequently used forms are included. The OZSD indeed lemmatises quite a number of frequently used derivations of *-sebenz-*, i.e., *abasebenzi*, *ekusebenzeni*, *ekusebenziseni*, *ekusetshenzisweni*, *emisebenzini*, *emsebenzini*, *imisebenzi*, *isisebenzi*, *izisebenzi*, *-sebenzela*, *-sebenzisa*, *-sebenzisana*, *-setshenziswa*, *-setshenzwa*, *ukusebenza*, *ukusetshenziswa*, *umsebenzi*. This is useful, but for the learner reading an isiZulu novel, the low frequency words also need to be decoded for him/her to understand the specific utterance.

Returning to the infinitive, dictionaries following word lemmatisation will, by default, also at least have the lemmas *-hamba*, *-sebenza* and *-khuluma*, honouring the non-disputed stem lemmatisation approach for verbs, but will lemmatise infinitives as nouns according to the default word lemmatisation strategy for nouns, i.e., on the first letter of full forms. The OZSD accordingly lemmatises the full forms of a number of frequently used infinitive nouns, such as *ukudla* 'food', *ukuhamba* 'departure' and *ukukhuluma* 'a/the talking' in the alphabetical stretch *uku-* and treats them appropriately for their nominal meanings. These infinitive *nouns*, however, stand in contrast with the infinitive verbs *ukudla* 'to eat', *ukuhamba* 'to walk/go' and *ukukhuluma* 'to speak' in isiZulu. *Ukudla*, *ukuhamba* and *ukukhuluma* have, therefore, also been lemmatised in the OZSD under their stem forms *-dla*, *-hamba* and *-khuluma* as well with applicable treatment for their verbal meanings. However, as argued above, dictionary users become used to looking up infinitive verbs under their stem forms. When looking up *ukudla*, *ukuhamba* and *ukukhuluma*, the user is unlikely to consider the possibility that (s)he should also check under *uku-* for the possible existence of an infinitive noun with the same stem. As in the case of lemmatising stems, a cross-reference in this case is imperative from the articles of the verb stems to the full nouns in such cases. In many instances the nominal and verbal meanings are closely related, e.g., *ukukhuluma*: 'to talk, a/the talking', *ukuhamba* 'to travel, a/the travelling', *ukusebenza* 'to work; a/the working', but in cases, such as the infinitive noun *ukujula* 'depth', the infinitive verb stem, i.e., *-jula*, means 'consider carefully'. A cross-reference from *-jula* to *ukujula* is imperative to

avoid misguiding the user. Inserting such cross-references would of course require additional space in the dictionary. In addition, users should be alerted in the user's guide to the dictionary to check for possible nominal forms when looking up infinitives under their verbal stems and vice versa.

There should be no doubt that word lemmatisation contributes substantially to reducing the problems that stem lemmatisation causes to users. However, the problem of stem identification or word identification is still not solved, and probably never will be solved in printed dictionaries and that moves the focus to electronic dictionaries.

Electronic dictionaries for isiZulu — the final frontiers?

In the early nineties, the electronic era was met with great enthusiasm and expectations expressed in relation to electronic dictionaries and their enormous potential to supersede printed or paper dictionaries in imaginative ways. As the title *Lexicographers' Dreams in the Electronic-Dictionary Age* of De Schryver (2003) suggests, early publications on EDs were dreams about the potential of the new medium and the expected revolution it would bring along, such as antiquating the paper dictionary in a decade or two. These publications list dozens of advantages of EDs, such as accessibility, user-friendliness and especially the availability of space and processing speed. Many of these issues are discussed in detail by Dodd (1989), Bolinger (1990), Atkins (1996), Nesi (1999), Geeraerts (2000), Harley (2000) and Prinsloo (2001), to name but a few. Meijs (1990) even predicted the end of the paper dictionary by 2000. Prinsloo (2005) believes that the potential of electronic dictionaries lies in the utilization of what he calls true electronic features, such as pop-up access, bringing together of related items, new routes to the data, less dependency on alphabetical order, fuzzy spelling, intelligent extrapolation of characters keyed in and audible pronunciation. For the purposes of this article, the question is what the status of currently available isiZulu dictionaries is in terms of lemmatisation and solving the issue of stem or word identification that dominate the discussion in the previous sections of this article.

In principle, catering for all of the approximately 2 500 occurrences of *-sebenz-* in the PZC is not a problem in electronic dictionaries, given the almost unlimited available space and the speed of information retrieval, cf. Prinsloo (2001) and De Schryver (2003). The question, however, is whether this goal has been achieved.

A number of electronic dictionaries and word lists are available for isiZulu, such as the *Webster's Online Dictionary*, *Freelang.net* and the *Dicts.info*. However, the most sophisticated online dictionary is the *isiZulu.net*. The major stongpoints of this dictionary are that it is extensive; there is no need for stem search; and it automatically gives a morphological analysis of the stem plus affixes. The *isiZulu.net* offers some promising features in solving the most problematic cases discussed in terms of Van Wyk (1995)

above where stem identification is problematic. To illustrate: it offers two access routes to *impilo*, i.e., *impilo* and *mpilo*, and both *intombi* and *ntombi* for *intombi*. Plural forms of these nouns can also be directly looked up by typing their full forms, *izimpilo* and *izintombi*, respectively. In addition, this dictionary is useful in cases where the learner finds it difficult to isolate stems/words.

From the examples given in (16), it is clear that the stem identification problem has at last been resolved. The inexperienced learner can simply type in the word or part of it and is (re)routed to the appropriate lemma. Moreover, the quality of the treatment is good.

(16)

The image shows two side-by-side screenshots of the isiZulu.net dictionary interface. Both screenshots have a search bar containing the word 'umsebenzi' and a search button. The left screenshot shows the search results for 'umsebenzi', listing 'Direct Matches' for 'umsebenzi/abasebenzi' (n, 1/2 (-sebenzi) [umse'be:nzi] [abase'be:nzi]) and 'umsebenzi/imisebenzi' (n, 3/4 (-sebenzi) [umse'be:nzi] [imise'be:nzi]). The right screenshot shows the search results for 'ngomsebenzi', listing 'Potential Decompositions (verbatim)' for 'umsebenzi/abasebenzi' and 'umsebenzi/imisebenzi', and 'ngomsebenzi' with its morphological breakdown: [ngomse'be:nzi] ← nga + umsebenzi, and 'ngomsebenzi' with its morphological breakdown: [ngomse'be:nzi] ← nga + umsebenzi.

The question, however, is how comprehensive the *isiZulu.net* is in terms of coverage of entire paradigms of words, such as the paradigm for *-sebenz-* as given in the appendix? Formulated differently, can all orthographic forms of nouns and verbs in isiZulu be looked up in the *isiZulu.net*? To answer these questions, the *isiZulu.net* was subjected to a number of random tests in terms of the paradigm for *-sebenz-*, as well as to random selections from a number of published isiZulu dictionaries.

For the first test, the top 10 frequencies were taken, i.e., *umsebenzi* (5 583), *emsebenzini* (1 456), *imisebenzi* (1 009), *ukusebenza* (548), *msebenzi* (470), *lomsebenzi* (446), *nomsebenzi* (309), *usebenza* (299), *esebenza* (290), *ngomsebenzi* (253). All of these were found in the *isiZulu.net*. See, for example, the treatment of *umsebenzi* and *ngomsebenzi* in (16).

For the inexperienced user, the automatic guidance from *msebenzi* to *umsebenzi*, *abasebenzi* and *imisebenzi* is excellent because, as mentioned above, no addition of characters to the word is required to look it up.

(17)



In the case of the successful automatic retrieval of *umsebenzi* from *lomsebenzi* (lo+umsebenzi), *nomsebenzi* (na+umsebenzi) and *ngomsebenzi* (nga+umsebenzi), the results are equally satisfying, because the search was successful and morpho-phonological processes are reversed by the dictionary and presented to the user in a clear and user-friendly way.

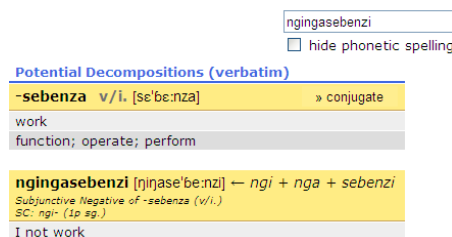
For the second test, ten words were selected from the paradigm for *-sebenz-* that occur five times in the PZC as given in Table 1 and even though these words occur with a low frequency, six were found in the *isiZulu.net*.

Table 1: A random selection of derivations of *-sebenz-* occurring 5 times in the PZC and their presence or absence in *isiZulu.net*

ayisebenzele	YES		singasebenzi	YES
ngingasebenzi	YES		uzosebenzani	NO
ngisebenzile	YES		sisebenzile	YES
nisasebenza	YES		wayesasebenza	NO
njengokwemisebenzi	NO		wayesesebenza	NO

Example (18) reflects the quality of treatment for *ngingasebenzi* in the *isiZulu.net*.

(18)



First, decomposition of *ngingasebenzi* to *-sebenza* gives a useful morphological breakdown into stem with prefixes. The user learns that (s)he is dealing with a derivation of *-sebenza*. Secondly, a translation equivalent of the full word *ngingasebenzi*, 'I not work' (even though the latter is a very direct translation) is given. So, for this example one could say that the problem of stem/word identification has been resolved and the user finds sufficient comment on form as well as comment on semantics of the full word.

For the third test, a study of the first lemma on every 25th page of the WAZ (nine lemmas) was done in terms of its inclusion or omission from the *isiZulu.net* and its presence in the PZC. It was found that five of the nine lemmas occurred in the PZC. Only three were lemmatised and treated in the *isiZulu.net*. A similar selection of the first lemma on every 50th page of the ZED and its inclusion or omission from the *isiZulu.net* and the PZC revealed that from the 19 lemmas in question seven occurred in the PZC and three in the *isiZulu.net*. From these three tests, it is clear that the *isiZulu.net* electronic dictionary performs well on the more frequently used words but substantial enlargement will be required to cover less frequently used words as well.

The least electronic dictionaries could do is to link paradigms, such as those in the appendix to the stem/word, i.e., *-sebenza* in this case.

Conclusion

The weakest option for lemmatising nouns and verbs in isiZulu is to lemmatise verbal stems without suffixes and in the case of nouns, noun stems without their prefixes and without the augmentative and diminutive nominal suffixes. This lemmatisation strategy is not user-friendly; stem identification is a major obstacle; a vast amount of knowledge of morphophonetics is presupposed; and the user is often in doubt whether (s)he has successfully retrieved information. Even if the users do manage to identify the stem and to look it up, all the additional information conveyed by the affixes have to be 'added back on' and the user will not know for sure whether (s)he has come to the right conclusion. Lemmatising verb stems represents a slight improvement. At least the meanings of the suffixes need not be artificially added on as in the case of lemmatising stems without their suffixes.

Lemmatising stems with their prefixes merely added on (left-expanded) is a better option, because the user has the advantage of seeing the full form of infinitive verbs and the full forms of nouns with additional information, such as tonal indication. This strategy is more user-friendly, but stem identification remains problematic and a substantial amount of knowledge of morphophonetics is still presupposed.

Word lemmatisation applicable to nouns is by far the better strategy, because nouns can be looked up under the first letter. For given non-derived nominal forms, the problem of stem identification is solved for all nouns. This strategy is especially beneficial for those nouns where stem identification is

problematic. The strategy is user-friendly and no knowledge of the grammar is presupposed. However, for nominal and verbal derivations, especially those where nominal and verbal stems occur with multiple prefixes, the problem of stem/word identification remains unsolved.

The problem of word/stem identification which is present in all of the lemmatisation strategies employed for isiZulu can only be solved in electronic dictionaries. Most electronic dictionaries are mere translated word lists and are not of much use to the target users especially for their productive needs. A clear exception is the *isiZulu.net* online dictionary, where the problem of stem/word identification has been solved for most of the frequently used words in isiZulu, but more comprehensive electronic isiZulu dictionaries are required to alleviate the need for stem/word identification for less frequently used words as well.

References

A. Dictionaries and corpora

- COBUILD = Sinclair, J. (Ed.). 1995². *Collins COBUILD English Dictionary*. London: HarperCollins.
- CSD = Rycroft, D.K. 1982. *Concise SiSwati Dictionary*. Second impression. Pretoria: J.L. van Schaik.
- CZD = Dent, G.R. and C.L.S. Nyembezi. 1961. *Compact Zulu Dictionary*. Pietermaritzburg: Shuter and Shooter.
- DTS = Elliott, W.A. 1897. *Dictionary of the Tebele & Shuna Languages. With Illustrative Sentences and Some Grammatical Notes*. Frome and London: Butler & Tanner.
- EZD = Doke, C.M., D. McK. Malcolm and J.M.A. Sikakana. 1958. *English-Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- ISZ = Mbatha, M.O. (Ed.). 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers.
- OZSD = De Schryver, G.-M. (Ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. Cape Town: Oxford University Press Southern Africa.
- SSED = Mabile, A. and H. Dieterlen. 1988. *Southern Sotho-English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.
- WAZ = Dekker, A.M. and J.H. Ries. 1958. *Woordeboek Afrikaans-Zoeloe Zoeloe-Afrikaans/Isichazamazwi Isibhunu-IsiZulu IsiZulu-Isibhunu*. Johannesburg: Afrikaanse Pers Boekhandel.
- ZED = Doke, C.M. and B.W. Vilakazi. 1948. *Zulu-English Dictionary*. First Edition. Johannesburg: Witwatersrand University Press.
- Dicts.info*: <http://www.dicts.info/>
- Freelang dictionaries*: <http://www.freelang.net/>
- isiZulu.net*: <http://isizulu.net>
- PSC: Pretoria Sepedi Corpus. University of Pretoria
- PZC: Pretoria Zulu Corpus. University of Pretoria
- Webster's Online Dictionary*: <http://www.websters-online-dictionary.org/>

B. Other literature

- Atkins, B.T. Sue. 1996. *Bilingual Dictionaries: Past, Present and Future*. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Rödger Pap-

- mehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.
- Bennett, P.R.** 1986. Grammar in the Lexicon, Two Bantu Cases. *Journal of African Languages and Linguistics* 8(1): 1-30.
- Bolinger, D.** 1990. Review of the *Oxford Advanced Learner's Dictionary of Current English*. *International Journal of Lexicography* 3(2): 133-45.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M.** 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.
- Dodd, W. Steven.** 1989. Lexicocomputing and the Dictionary of the Future. *Lexicographers and their Works*. James, Gregory (Ed.). 1989. *Lexicographers and Their Works*: 83-93. Exeter Linguistic Studies 14. Exeter: University of Exeter Press.
- Doke, C.M.** 1945. *Text-Book of Zulu Grammar*. London: Longmans, Green and Co. Ltd.
- Gauton, Rachelle.** Forthcoming. Lexicography of the Nguni Languages. *Handbooks of Linguistics and Communication Science, Volume Dictionaries: An International Encyclopedia of Lexicography*. HSK Vol. 5.4. Dictionaries: Supplementary Volume. Berlin: Mouton de Gruyter.
- Geeraerts, D.** 2000. Adding Electronic Value: The Electronic Version of the *Grote Van Dale*. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 75-84. Stuttgart: Stuttgart University.
- Gouws, R.H. and D.J. Prinsloo.** 2005a. *Principles and Practice of South African Lexicography*. Stellenbosch: African Sun Media.
- Gouws, R.H. and D.J. Prinsloo.** 2005b. Left-expanded Article Structures in Bantu with Special Reference to IsiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.
- Harley, Andrew.** 2000. Software Demonstration: Cambridge Dictionaries Online. Heid, Ulrich et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 85-88. Stuttgart: Stuttgart University.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 102-111. Berlin: De Gruyter.
- Masubelele, Rose.** 2007. [Review of] Mbatha, M.O. (Ed.). 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers. *Lexikos* 17. 458-461.
- Meijs, W.J.** 1990. Morphology and Word-Formation in a Machine-Readable Dictionary: Problems and Possibilities. *Folia Linguistica: Acta Societatis Linguisticae Europaeae* 24(1-2): 45-71.
- Mpungose, M.H.** 1998. Analysis of the Word-Initial Segment with Reference to Lemmatizing Zulu Nasal Nouns. *Lexikos* 8(AFRILEX series 8:1998): 65-87.
- Mtuzze, P.T.** 1992. A Critical Survey of Xhosa Lexicography 1772–1989. *Lexikos* 2 (AFRILEX series 2:1992): 165-177.
- Nesi, Hilary.** 1999. A User's Guide to Electronic Dictionaries for Language Learners. *International Journal of Lexicography* 12(1): 55-66.
- Prinsloo, D.J.** 1994. Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 14(2): 93-102.

- Prinsloo, D.J.** 2001. The Compilation of Electronic Dictionaries for the African Languages. *Lexikos* 11: 139-159.
- Prinsloo, D.J.** 2005. Electronic Dictionaries Viewed from South Africa. *Hermes, Journal of Linguistics* 34: 11-35.
- 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.** 2010. [Review of] G.-M. de Schryver (Ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. Cape Town: Oxford University Press Southern Africa. *Lexikos* 20 (AFRILEX series 20:2010). 760-766.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Prinsloo, D.J. and Rufus H. Gouws.** 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.

Appendix A: Words containing *-sebenz-* in the PZC with their frequencies indicated in brackets

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), BASEBENZA (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), USEMSEBENZINI (19), EZISEBENZINI (19), ENGISEBENZA (19), ANGASEBENZI (19), ZISEBENZE (18), UKULUSEBENZISA (18), OSEBENZAYO (18), ESASEBENZA (18), ABASEBENZISA (18), WOKUSEBENZA (17), USASEBENZA (17), UNGASEBENZI (17), SOKUSEBENZA (17), SIYASEBENZA (17), AYISEBENZI (17), WUMSEBENZI (16), USEBENZANI (16), UMSEBENZANA (16), OSEBENZELA (16), NGISEBENZELA (16), ZIYASEBENZA (15), UKUZISEBENZISA (15), UKUSISEBENZISA (15), SINGASEBENZISA (15), NISEBENZE (15), NGISEBENZISE (15), ESASISEBENZA (15), BAYOSEBENZELA (15), ZEZISEBENZI (14), YABASEBENZI (14), WEZISEBENZI (14), UNOMSEBENZI (14), UNGASEBENZISA (14), NISEBENZA (14), NGIYOSEBENZA (14), ISEBENZI (14), EZISEBENZELA (14), EYAYISEBENZA (14), ESEBENZE (14), ANGASEBENZISA (14), ABANGASEBENZI (14), YAYISEBENZA (13), WAYOSEBENZA (13), UWASEBENZISE (13), UKUZOSEBENZA (13), UKUYISEBENZISA (13), NJENGOMSEBENZI (13), NGANGISEBENZA (13), KWAKUSEBENZA (13), KUWUMSEBENZI (13), KUNOMSEBENZI (13), ESEMSEBENZINI (13), ESEBENZELA (13), EMSEBENZI (13), AZOSEBENZA (13), UNGASEBENZISI (12), SIYOSEBENZA (12), SEKUSEBENZA (12), OLUSEBENZAYO (12), NGIZOSEBENZA (12), NGISASEBENZA (12), INSEBENZO (12), INOMSEBENZI (12), EZAZISEBENZA (12), BENGISEBENZA (12), BAYOSEBENZA (12), ZOKUSEBENZISA (11), WUKUSEBENZISA (11), WASEBENZISA (11), WABASEBENZI (11), USEYASEBENZA (11), UBESEBENZA (11), SIZOSEBENZA (11), SASEBENZA (11), ABASEBENZELA (11), YIZISEBENZI (10),

WAWUSEBENZA (10), YEZISEBENZI (10), WEMISEBENZI (10), UZISEBENZISA (10), UYISEBENZISE (10), UKUWASEBENZISA (10), SENGISEBENZA (10), SEBENZANI (10), OKUNGUMSEBENZI (10), NGISEBENZISA (10), NASEMISEBENZINI (10), NANGOMSEBENZI (10), KUMSEBENZI (10), INGUMSEBENZI (10), ELISEBENZA (10), BESEBENZELA (10), AYASEBENZA (10), ZISEBENZI (9), WOKUSEBENZISA (9), USEBENZELE (9), UKUNGASEBENZI (9), SEZISEBENZI (9), SEBESEBENZA (9), SEBENZELA (9), NANGEMISEBENZI (9), KWASEBENZA (9), KUSEBENZISA (9), KULOWOMSEBENZI (9), ISISEBENZA (9), INGASEBENZISA (9), ESIZISEBENZISAYO (9), ESISEBENZISA (9), ESISEBENZAYO (9), ENGASASEBENZI (9), BENGASEBENZI (9), BASEBENZELE (9), BABESEBENZISA (9), AWUSEBENZI (9), ASISEBENZELI (9), ASEBENZELE (9), ASEBENZELA (9), YOKUSEBENZELA (8), WAYESEBENZISA (8), WAYENGASEBENZI (8), USESEBENZE (8), UKUSEBENZELANA (8), UKULISEBENZISA (8), UBUSEBENZA (8), OZOWASEBENZISA (8), OYISEBENZISAYO (8), NIZOSEBENZA (8), NGOKOMSEBENZI (8), NGISEMSEBENZINI (8), NESISEBENZI (8), NAMSEBENZI (8), LUSEBENZA (8), KAYISEBENZI (8), IZOSEBENZA (8), IYOSEBENZA (8), EZISEBENZAYO (8), ESEBENZAYO (8), ELISEBENZISA (8), BEMISEBENZI (8), AYISEBENZISE (8), AWASEBENZISE (8), ALISEBENZISE (8), YISEBENZI (7), YASEMSEBENZINI (7), UYISEBENZISA (7), UYAKUSEBENZA (7), USUSEBENZA (7), USESEBENZA (7), UNGASEBENZA (7), ULISEBENZISE (7), UKUZISEBENZELA (7), UKUSEBENZISANA (7), SISEBENZELA (7), SILISEBENZISE (7), SEBENZE (7), OZOSEBENZA (7), OKOMSEBENZI (7), NGINAMSEBENZI (7), NEZABASEBENZI (7), LEYOMISEBENZI (7), KWEZISEBENZI (7), KOKUSEBENZISA (7), INGASEBENZI (7), ESEBENZISE (7), ENGANGISEBENZA (7), BOKUSEBENZISA (7), BAZOSEBENZA (7), BASEBENZELA (7), AZISEBENZISE (7), ASISEBENZI (7), ANGAZISEBENZISA (7), AKASASEBENZI (7), ZASEBENZA (6), YALOMSEBENZI (6), UZISEBENZELA (6), USUSEBENZE (6), UNGAZISEBENZISA (6), SINGASISEBENZISA (6), SINGASEBENZA (6), SIKUSEBENZISE (6), NISEBENZILE (6), NGIZOSEBENZISA (6), NGINEMISEBENZI (6), NGAEMSEBENZINI (6), MAKASEBENZE (6), LIZOSEBENZA (6), LIYASEBENZA (6), LISEBENZISA (6), LASEBENZA (6), ILUSEBENZISE (6), EYOSEBENZA (6), ESIWASEBENZISA (6), ESESEBENZE (6), ESEBENZEKAYO (6), ENGASENAMSEBENZI (6), BOKUSEBENZA (6), BESEMSEBENZINI (6), BASEBENZISANE (6), AYISASEBENZI (6), ANGENAMSEBENZI (6), ANGASEBENZA (6), ALUSEBENZISE (6), ZIWUMSEBENZI (5), ZILISEBENZISA (5), ZENSEBENZO (5), WAYESESEBENZA (5), WAYESASEBENZA (5), UZOSEBENZANI (5), UWASEBENZISA (5), USISEBENZISE (5), USISEBENZISA (5), SIYISEBENZISA (5), SISISEBENZISE (5), SISEBENZILE (5), SINGASEBENZI (5), SESISEBENZA (5), SEMSEBENZINI (5), SEMISEBENZI (5), SEKUNGUMSEBENZI (5), SASISEBENZA (5), OYISEBENZI (5), OWASEBENZA (5), OKUSEBENZISA (5), NJENGOKWEMISEBENZI (5), NJENGOKUSEBENZISA (5), NJENGEMISEBENZI (5), NISASEBENZA (5), NGOKWEMISEBENZI (5), NGISEBENZILE (5), NGINGASEBENZI (5), NGAKUSEBENZISA (5), NABASEBENZISA (5), NABASEBENZA (5), LISEBENZILE (5), LINGASEBENZA (5), LASEMSEBENZINI (5), LABASEBENZI (5), KUSEBENZISEKE (5), ISEBENZELE (5), EZOSEBENZA (5), EZISEBENZI (5), ESIZISEBENZISA (5), ENGUMSEBENZI (5), ENGASEBENZA (5), BESASEBENZA (5), BASASEBENZA (5), AYISEBENZELE (5), AWUSEBENZISE (5), ASEBENZELANA (5), ASEBENZAYO (5), ABAZOSEBENZA (5)