Revolutionizing Bantu Lexicography — A Zulu Case Study

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Belgium; Xhosa Department, University of the Western Cape, Bellville, South Africa; and TshwaneDJe HLT, Stellenbosch, South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: Zulu uses a conjunctive writing system, that is, a system whereby relatively short linguistic words are joined together to form long orthographic words with complex morphological structures. This has led to the so-called 'stem tradition' in dictionary making — for Zulu, as well as for most other Bantu languages. Given this lemmatization approach has been found to be inadequate for young learners (who fail to isolate stems), the development of a new approach was imperative for them, but until recently deemed impossible to implement. In this paper it is argued that it is now perfectly possible to reverse the unproductive trend, and to opt for the lemmatization of full words for all but one of the word classes in Bantu. This revolution is made possible thanks to the recent availability of relatively large corpora, with which the really frequent citation options may be pinpointed. Rather than a mission statement, this paper offers the result for all word classes. To do so, an actual guide to the use of a Zulu dictionary is re-represented and annotated.

Keywords: ZULU, BANTU, DICTIONARY, USAGE GUIDE, MINI-GRAMMAR, WORD CLASSES, STEM VS. WORD LEMMATIZATION, CORPUS, USER-FRIENDLY

Samenvatting: Bantoe lexicografie radicaal omgooien — een gevalsanalyse voor Zoeloe. Zoeloe maakt gebruik van een conjunctief schrijfsysteem, d.w.z. een systeem waarbij relatief korte linguïstische woorden vast aan elkaar geschreven worden met lange orthografische woorden tot gevolg, die ook nog complexe morfologische structuren vertonen. Dit heeft geleid tot wat men de 'stam traditie' in de lexicografie is gaan noemen — voor Zoeloe, alsook voor de meeste andere Bantoetalen. Aangezien deze lemmatisatieaanpak ongeschikt is gebleken voor jonge gebruikers (die woordstammen maar niet kunnen isoleren), moest voor hen een nieuwe aanpak ontwikkeld worden. Tot voor kort werd zo'n aanpak echter als niet-implementeerbaar beschouwd. In dit artikel wordt geargumenteerd dat het vandaag de dag perfect mogelijk is om de onproductieve trend om te keren, en om te kiezen voor het lemmatiseren van volledige woorden voor alle woordklassen op één na in Bantoe. Deze radicale ommezwaai werd mogelijk gemaakt dankzij het recent beschikbaar komen van relatief grote corpora, waarmee de echt frequente opties qua trefwoordkeuzes bepaald kunnen worden. In de plaats van louter de beschrijving van een doelstelling, biedt dit artikel oplossingen voor alle woordklassen. Daartoe wordt de effectieve gebruikersgids van een Zoeloe woordenboek voorgesteld en van commentaar voorzien.

Sleutelwoorden: ZOELOE, BANTOE, WOORDENBOEK, GEBRUIKERSGIDS, MINI-GRAMMATICA, WOORDKLASSEN, STAM- VS. WOORD-LEMMATISATIE, CORPUS, GEBRUIKSVRIENDELIJK

Lexikos 20 (AFRILEX-reeks/series 20: 2010): 161-201

1. The one-size-fits-all problem

Although dictionaries for the Bantu languages have been compiled for several centuries now, and although Bantu metalexicography is at least 150 years old (cf. Benson 1964), the field has been plagued throughout by what one could call the one-size-fits-all approach. By and large that 'size' has been to attempt to lemmatize all words from all word classes under their *stems*, no matter the target user envisaged. Arguably, from a strict morphological point of view, this is a perfectly valid and linguistically sound approach. Students at institutes of higher learning are confronted with it from day one of their studies, and end up mastering the system given their otherwise general grounding in linguistics and their year-long exposure to a dictionary culture in a variety of languages. Such students are typically non-Africans studying at universities in the West or East. When that same approach is used to compile dictionaries for elementary learners, who moreover have not had the chance to be exposed to any other dictionaries in their lives, the dictionaries have shown to be too challenging to use. Here the intended user is typically a mother-tongue speaker of a Bantu language, in need of a dictionary with local relevance.

The problem of the one-size-fits-all approach for Bantu lexicography has been noted before in the scientific literature, and it is often presented as a need to choose between stem lemmatization (more often than not seen as 'the right size') vs. word lemmatization (typically 'the wrong size', or at least looked down upon, especially by linguists). Half a century ago Benson, after surveying the field, concluded in favour of stem lemmatization as follows:

> It is now right and proper to [...] make certain suggestions which could help future compilers of dictionaries of African languages, whoever they may be, to avoid some of the more obvious pitfalls. [...] there are no rules laid down for lexicographers, and whatever has been learnt by toil and sweat, by trial and error, is worth passing on. [...] One cardinal principle which emerges from our study is that everything which needs to be said about a stem or root should be channelled into one single full article, complete with citations if needed.
>
> — Benson (1964: 78, 80, 82)

Merely a year later, however, in a discussion of a Luganda–English dictionary, Snoxall argued in favour of word lemmatization, as follows:

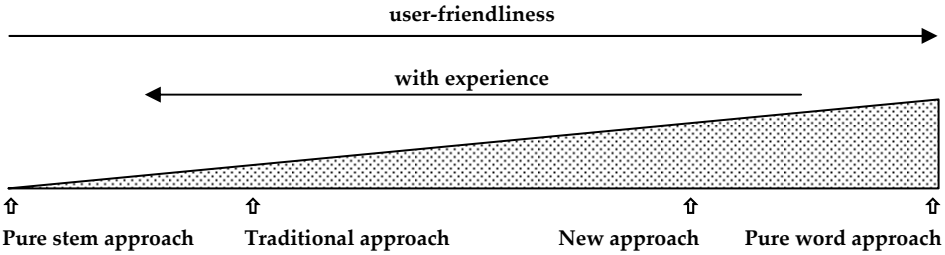
> [E]ven many Baganda would have little idea under what root form they should look up many of the commonest words which they use. [...] The general principle of entering words in a dictionary under roots [...] could never be of great assistance [...] It would seem therefore that, although disappointing perhaps to etymologists, a decision to enter headwords in the form in which they are used in actual speech, as words possessing meaning, [...] will be welcomed by the great majority of the users of the dictionary.
>
> — Snoxall (1965: 27-28)

The debate has raged on ever since, with Bennett writing two decades later:

> 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)

A more recent and excellent overview of the various pros and cons of the two opposing approaches can be found in Van Wyk (1995). The opposition, however, is a false opposition, as no Bantu dictionaries exist that are purely stem-based, neither do Bantu dictionaries exist that are purely word-based. I have pointed this out in an earlier contribution, one in which I had proposed a 'new approach':

> These two extremes are but two poles on a continuum, of course. In reality, a 'traditional' stem-based approach to lemmatization [...] also has word features, and thus moves up on the continuum, while the approach advocated in this research article moves in the other direction of the continuum, away from the sole orthographic word. [The figure below] summarizes this situation, where the shaded triangle illustrates the increase in user-friendliness for *junior* users as one moves from stem-like to word-like lemmatization. With experience, however, one tends to crave for more condensed and more abstract information, and thus the wish to move in the other direction.
>
> — De Schryver (2008a: 86-87)



It is my contention that the great majority of the existing dictionaries for the Bantu languages are actually variations of the 'traditional approach' depicted on the continuum, thus at heart dictionaries in which the lexicon is grouped around stems, with some dictionaries showing word-like features for selected word classes only, at which point they move up the continuum a little. Moving towards the word approach is easier to achieve for some Bantu languages than it is for others, depending on the degree of conjunctivism / disjunctivism of the

Bantu language in question (cf. Prinsloo and De Schryver 2002). Actually, no more than a handful of Bantu lexicographers have consciously tried to approach the continuum from the other (word) pole for *all* word classes. Moreover, and as we will see below, achieving this goal for the highly conjunctive Bantu languages has only become possible in recent years, thanks to the availability of (relatively) large corpora, from which real facts can be derived, rather than confining lexicographers to theoretical conjectures.

New types of dictionaries for the Bantu languages have indeed started to be compiled — that is, dictionaries for mother-tongue speakers, who do not have any previous exposure to a dictionary culture. Interestingly, the pressing need for such dictionaries was not only felt near-simultaneously in both monolingual and bilingual environments, the compilation of new, word-based dictionaries also started around the same time. For example, in Uganda, M. Nabirye undertook to compile the first monolingual dictionary for Lusoga, and in doing so she placed the (non-existent!) reference skills of the two million Basoga first. Her extensive research (Nabirye 2008) quite naturally led her to compile a dictionary in which *words* are listed, the only 'natural' language units according to the mother-tongue speakers she tested her dictionary material on (cf. Nabirye 2009a, 2009b). The result — the *Eiwanika ly'Olusoga* (Nabirye 2009) — is a most-powerful statement by a native speaker of a Bantu language.

Meanwhile in South Africa, several teams led by myself — a non-mother-tongue speaker of the Bantu languages — began work on a series of bilingual dictionaries with in each case one of the official South African Bantu languages and English as treated language pairs. The overarching theoretical framework was described in my PhD thesis (De Schryver 2004), and to date two dictionaries — the *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007), and the *Oxford Bilingual School Dictionary: Zulu and English* (De Schryver 2010) — have been completed and published. An accompanying workbook was also prepared for the Northern Sotho dictionary (Taljard et al. 2008),[1] and both dictionaries have received favourable academic reviews; see for instance Prinsloo (2009) for the Northern Sotho dictionary, and Prinsloo (2010) for the Zulu dictionary.

Compared to any other earlier dictionaries for the Bantu languages at large, the *Oxford Bilingual School Dictionary: Zulu and English* (henceforth OZSD) is the first dictionary for a Bantu language to radically move away from the one-size-fits-all approach to lemmatization, in that the reference skills of the envisaged target user group forced me to come up with a new, tailored, word-based approach. In the present paper this approach is briefly introduced.

2. A daunting finding

Although the new approach is a direct implementation of the proposals which I formulated in my PhD thesis (completed in 2004), it would not have been possible to subsequently release two dictionaries at three-year intervals (for North-

ern Sotho in 2007, and for Zulu in 2010) without the support of large dictionary teams.[2] At this point I would like to salute and thank them, as *without* them my proposals would have remained academic theorizing; *with* them my dreams became a reality. For anyone wishing to make analogous dictionaries, it is instructive to have an idea about some of the practical aspects of these dictionary projects. I will use the OZSD as a case in point.

The OZSD is a bidirectional, bilingual school dictionary with Zulu and English as treated language pair, for use in South African schools in Grades 4 to 9 (i.e. the intermediate and senior phase of GET),[3] covering 5 000 lemmas on each side. The OZSD has been designed to help students read and write Zulu better, if they are learners of Zulu, or read and write English better, if they are learners of English. The fact that it is a school dictionary had far-reaching implications for its design. Chief among those is that, in addition to the two A-to-Z sections, which amount to a total of 582 pages, the OZSD contains extensive extra-matter texts, totalling 58 pages (or ten percent of the A-to-Z sections), the main purpose of which is to teach a full-experience dictionary culture.[4] Where relevant, those extra-matter pages are presented in both Zulu and English. As such, the front matter has a section on 'Dictionary features', in which the structure of the dictionary articles is visualised, and a gentle 'Introduction' on why dictionaries are important, on how the OZSD (a corpus-driven dictionary) differs from other dictionaries, and on what can be found in it. Completing the 12 pages of the front matter are also the title and imprint page, and a table of contents.

The middle matter or 'Study section', placed between the Zulu to English and English to Zulu sides of the dictionary, opens with 'Dictionary activities' in Zulu and English, taking the learners all the way from the sequence of the alphabet to both decoding and encoding exercises with the dictionary. Writing skills are covered in samples of e-mails, letters and electronic (SMS) messages. For Zulu, a mini-grammar and information on pronunciation are included, and for English, irregular verbs, punctuation and spelling are covered. All these sections, together with a detailed table of contents, amount to 30 pages.

The back matter has been conceived as a 'Reference section', and opens with six full-page plates with illustrations (domestic, wild and sea animals; small creatures; fruit and vegetables; etc.), and further contains information on the South African provinces, languages and phases of education, as well as numerous tables that bring together closed-class items (months, days, seasons; the solar system; public holidays; symbols; etc.), the numbering system in Zulu, and weights and measurements. Together with a detailed table of contents, as well as a page with the answers to the dictionary exercises from the middle matter, this back matter contains 16 pages.

Coordinating all these different components of the OZSD is no small matter, and considerable pressure was put on the human as well as financial resources — there never seemed to be enough of either of them. Although a tiny pilot was run in 2005 already (cf. De Schryver 2006), and although a preliminary draft of the Zulu to English side had been compiled by 2006, it is only

with a third team that the finish line was reached. In addition to myself as the editor-in-chief, the final dictionary development team consisted of a chief compiler (Nomusa Sibiya) and a linguist (Arnett Wilkes), six more compilers (Sibusiso Dlamini, Thandeka Cebekhulu, Wo Mthembu, Mduduzi Ndlovu, Moses Biyela, Kholiswa Sitole), two proofreaders for Zulu (Msawakhe Hlengwa, Thokozani Buthelezi), two proofreaders for English (John Linnegar, Celia Slater), a consultant on curriculum entries (Daphne Paizee), and two computational engineers (David Joffe, Malcolm MacLeod). At the publishing house, a publishing manager (Megan Hall), two project managers (Fred Pheiffer, Phillip Louw), an editorial assistant (Lorna Hiles), three designers (Peter Burgess: A-to-Z, Oswald Kurten: extra matter, Sharna Sammy: cover), and two illustrators (Julien Marais, Leigh-Anne Wolfaardt) supported and interacted with the dictionary development team. Finally, two typesetters (Tommy Bell: A-to-Z, Ingrid Richards: extra matter) took care of the final layout. That's a total of 27 people to prepare the manuscript!

I have chosen to list all these team members here, rather than hidden in an endnote, as I want to make sure that it is clear from the start that everything possible was done to ensure that all the best available skills were brought together to create the OZSD. While the team at the publishing house was largely responsible for the development of an English dictionary-template for the English to Zulu side, to be translated into Zulu and coordinated with the material on the Zulu to English side by the dictionary development team, the main task of the dictionary development team itself was the creation of the Zulu to English side from scratch. To do so, that team had a relatively large Zulu corpus at their disposal, with which the Zulu lemma list could first be drawn up, and then queried for each and every lemma. The various meanings for each lemma were mapped directly onto the uses as seen in the corpus, and each main meaning was illustrated with material taken straight from the corpus.

All of this, of course, is easier said (or 'recounted' here) than done, and to do justice to all the details of the processes followed in compiling the OZSD, one would need far more pages than those that are available in a single scientific paper. There is firstly a big difference between the ways in which the two A-to-Z sides of the dictionary were compiled, an aspect with implications for the (non-)reversibility of the dictionary. That would have to be explained in a paper. Detailing how the Zulu lemma list was created would need at least one other paper-length treatment. A further paper could then deal with the description of the overall macrostructural as well as microstructural decisions. More detailed studies for the English side would have to focus on the presentation of the Zulu lexicon in the microstructure, as one can assume that general lemmatization decisions need not be covered for English. Grouping certain word classes or parts of speech (POSs), this could be done in about five papers. Detailing the lemmatization decisions as well as the dictionary structure on the Zulu side, however, can only be done in earnest if each and every word class is considered in isolation first, with generalizations in a second phase. There are

21 main Zulu word classes in the OZSD, and 42 if one includes the sub-word classes. Attention should furthermore also go to all the extra matter texts, conceived as front, middle and back matter in the OZSD. Together, I reckon about ten papers would be needed for a proper coverage of the various extra matter features. In short, then, giving a full account of the many aspects revolving around the creation of a dictionary such as the OZSD would require anything between 40 to 60 scientific papers. At about 20 pages each, this amounts to 800 to 1 200 pages in total, the equivalent of three full-length monographs!

This is a daunting finding, and one that has worried me for some years now. It is all good and well to write overview papers, in which one briefly sketches the general approach (cf. e.g. De Schryver 2008), but if one also wishes to stimulate a healthy academic debate, then more detailed studies are required, studies in which each step of one's reasoning is carefully argued. Specifically for the OZSD, I have presented detailed accounts for four Zulu word classes to date: possessive pronouns (De Schryver and Wilkes 2008), adjectives (De Schryver 2008a), quantitative pronouns (De Schryver 2008b), and ideophones (De Schryver 2009). Rather than to continue with the series of Zulu word classes, which at the current rate will take at least another decade, and rather than to give yet another overview (this time for Zulu, rather than for Cilubà, Swahili and Northern Sotho, as in De Schryver 2008), I have opted for a compromise in this paper. The OZSD itself actually contains a text that is particularly fit for this purpose, and this text is presented next.

3. How to use your dictionary (a Zulu mini-grammar)

On pages S13 to S26 of the Study section, thus right in the middle of the dictionary, the OZSD contains a chapter titled 'How to use your dictionary (a Zulu mini-grammar)', which is both a mini-grammar of Zulu in disguise and a true guide to the proper use of the Zulu dictionary. In the fourteen sub-sections that follow, I will present the text of that chapter in full, and I will intersperse it with additional comments. In doing so I hope to achieve at least four goals. Firstly, by presenting the full text of one of the extended extra-matter texts of a published dictionary, I illustrate the features of an actual *text in action*, rather than a proposal which may or may not materialize. Secondly, by using this particular text, I will automatically cover a *full system*, and run through all parts of speech of a language, thus covering the breadth one expects to find in overview papers. Thirdly, in the process the reader will have been offered a *synoptic view* of the grammar of Zulu relevant for lexicographic purposes, an aspect especially welcome to those not familiar with this Bantu language. And fourthly, with the added comments this paper functions as a stand-alone text, in which just enough depth is presented to serve as a *launching pad* for further academic discussions.

This approach remains an experiment, however, as two writing styles will now alternate throughout Section 3. On the one hand a dictionary user is

addressed, who is a Grade 4 to 9 learner, and for whom everything that is said is assumed to be new. I address that user directly, and avoid, wherever possible, all unnecessary 'difficult words'. On the other hand I am addressing the reader of the present paper, in an academic register, for whom using the correct terminology is crucially important, and for whom enough context and references must frame the work. The presentation starts with the text as found in the OZSD, and the added material is flagged with numbered comment fields in a smaller font size.[5]

3.1 Introduction

In this dictionary English words (in the English to Zulu side) have been listed as in any other English dictionary. If you are uncertain how to use a dictionary and would like some practice, please work through the 'Dictionary activities' first (see pages S4 to S6).

Comments 1: From the start a clear division is made between looking up lemmas in the English side vs. looking up lemmas in the Zulu side of the OZSD. The 'Dictionary activities' the user is referred to have also been prepared in Zulu (to be found on pages S2 to S4). Note the use of the terms 'word' and 'listed' rather than 'lemma' and 'lemmatized'. Following these two opening sentences, the remainder of the mini-grammar deals exclusively with the lemmatization and treatment of the Zulu lexicon, as seen below.

Zulu words (in the Zulu to English side) have been listed in a radically new way, unlike the approach used in any other dictionary for Zulu. In this dictionary you can look up many Zulu words *directly* as they are written and used. This is different to conventional Zulu dictionaries, where you need to break down most words until you reach the roots or stems of the words — and it is the roots or stems which have been listed in those dictionaries. In contrast, in this dictionary all primary prefixes are still attached to the word roots and stems, except in the case of verbs. For verbs you will still need to learn to cut off all verbal prefixes, as well as some other formatives. Therefore, studying and knowing Zulu grammar remains very important. You can only use this Zulu dictionary with success if you use it *together* with your Zulu textbook. What we have done in this dictionary is to make it much easier for you to find the words you need.

Comments 2: This section introduces the main revolutionary aspect of the OZSD, viz. the fact that for all Zulu word classes except one, an orthographic word was chosen as lemma rather than a root or stem.[6] No information is given in the Zulu mini-grammar as to *how* words were chosen, but the use of a corpus is mentioned in the Introduction which is found in the front matter of the dictionary. The OZSD is not the place to explain this, nor is the present paper (as a paper each is needed per word class, cf. the penultimate paragraph of Section 2 above). What the Zulu mini-grammar does do is to explain the *result* for each word class. I can add here that for each word class as a whole, overall corpus frequencies were used in deciding where to cut off the various morphemes for that word class, so as to arrive at the most frequent type of orthographic form to be used as dictionary citation form for the entire paradigm. Making informed decisions to achieve this would

not have been possible without the availability of a relatively large Zulu corpus, consisting of a 7.5-million-word general-language component and a 1-million-word customised component of school textbooks.[7] Throughout the scientific literature (cf. e.g. the references to earlier work in Van Wyk 1995), lemmatizing a conjunctively-written language such as Zulu as words rather than stems, has even been considered a theoretical impossibility due to the multiplication effect. With enough evidence of real use, however, and especially with frequency distributions for all words in all word classes — with which one is able to separate the wheat from the chaff, or thus the truly common (and frequent) from the rare (and infrequent) — it becomes possible to select exactly those (frequent) Zulu *words* that ought to feature in a dictionary aimed at junior learners, rather than having to resort to lemmatizing *roots and stems* only, for which only generic meanings rather than customised meanings and customised examples can be provided. The prime word class for which the word vs. stem debate has raged is that of the nouns, but it is as relevant for nearly all other word classes as well (adjectives, relatives, (derived) adverbs, pronouns, etc.). Not having the space for further elaborations here, I would like to refer the reader to De Schryver (2008a), where, as a case study for one of the word classes traditionally lemmatized as stems only, the treatment of adjectives as words is presented. For the overall procedure to arrive at the corpus-driven selection of all 5 000 Zulu lemmas, see especially page 69 therein.

Not surprisingly, 'the most frequent type of orthographic form' for a particular word class often corresponds with what is intuitively 'the most logical', with the outcome that each resulting lemma is also felt to be 'the most natural word'. As straightforward as this may seem, there were still numerous additional decisions that had to be taken, and many ad-hoc solutions had to be designed due to the fact that language is not as regular as linguists would like it (or force it) to be. As a result, the Zulu lemma list is not only unique in that it presents words rather than stems, but it is also unique to this very dictionary, as it is unlikely that another team of compilers for a Zulu dictionary would arrive at the very same corpus-driven lemmatized frequency list. The dictionary development team felt so strongly about this particular selection of lemmas that a copyright was taken on it, as seen on the imprint page: "© Zulu text, including Zulu headword list, TshwaneDJe HLT 2010".[8] Claiming a copyright on the lexicon of a language is of course ludicrous, but claiming a copyright on a tailored lemma-sign list is not.

Also note the insistence here (and repeated further down) on the fact that the OZSD and the mini-grammar itself need to be used in conjunction with the learners' Zulu textbooks. This should not be seen as a cop-out for the aspects not covered in the mini-grammar, but rather as admitting that a mini-grammar is just that: a *brief* overview in which one simply cannot treat everything.[9]

Here is an easy example of how to look up words in this dictionary. All the words in the following sentence *Phuza amanzi ngazo zonke izikhathi, ikakhulu uma kushisa* 'Drink water at all times, especially when it is hot' are shown below. For each word you need to look up, we have shown the word class and a translation:

word in sentence:	*phuza*	*amanzi*	*ngaz**o***	***zo**nke*	***izi**khathi*
entry in dictionary:	-phuza	amanzi	ngazo	zonke	izikhathi > isikhathi
word class + class no.:	verb	noun 6	temp. adv. 8	incl. q. pron. 8	pl. noun 7/**8**
translation:	drink	water	at	all	times

word in sentence:	*ikakhulu*	*uma*	*ku-*	*-shisa*
entry in dictionary:	ikakhulu	uma	[*not in dictionary*]	-shisa
word class + extension:	conjunction	conjunction	indefinite concord	verb + causative
translation:	especially	when	it	be hot

As you can see, most of the words in this sentence can be found in the dictionary exactly as they appear in the sentence. The main exception is verbs, which need to be looked up under the first letter of their root or stem. To find the root or stem you need to cut off all verbal prefixes, if these are present. The first verb (*-phuza* 'drink') has no prefixes, so you can look it up directly. With the last verb (*kushisa* 'it is hot') you first need to cut off the indefinite concord (*ku-* 'it'), and then look up the verb stem (*-shisa* 'be hot'). Refer to your Zulu textbook for the correct use and meaning of all verbal prefixes, as well as for the sound changes that take place when attaching prefixes to one another and to roots and stems. In this dictionary, the most important verbal prefixes are listed in Tables 4 to 6, but remember that those tables do not replace the need for you to study and know the grammar of Zulu!

Comments 3: Although the example used to illustrate the dictionary system is a real example (and as a matter of fact, is one of the examples under the lemma **ngazo²**), it was chosen for its 'easiness' and the fact that the syntax in Zulu runs parallel to the English syntax. Implicitly, the difference with a pure stem approach is well illustrated, as there each of the orthographic words would need to be looked up under respectively *-phuza*, *-nzi*, *-zo(na)*, *-nke*, *-khathi*, etc. In some of the existing Zulu dictionaries, such as Doke and Vilakazi's *Zulu–English Dictionary* (1953²) or Dent and Nyembezi's *Scholar's Zulu Dictionary* (1995³), some of these indeed need to be looked up under their roots and stems, others have nonetheless been lemmatized under their full words, and for still others both options have been lemmatized.

Although a dictionary is not a grammar, all the important verbal prefixes have nonetheless been tabulated in the mini-grammar, again with the caveat that the learners should still consider their textbooks as well.

From this example, you can see that nouns have been listed as full words in your dictionary, so you will look up *amanzi* 'water' under the letter A in the alphabet (and *not* under N, the first letter of the noun stem, *-nzi*). Even (frequently used) plural nouns have been listed in your dictionary, so you can look up *izikhathi* 'times' directly. This is a unique feature of this dictionary, which makes it very easy to use.

Comments 4: After having briefly introduced the procedure to look up verbs, the next word class given a first brief attention is nouns. While verbs could be said to have been lemmatized like in traditional dictionaries (with this difference that extensive tables of verbal prefixes are included in the mini-grammar), the inclusion of full noun prefixes attached to noun stems is a radical departure from tradition (and also the main aspect singled out for review in Prinsloo 2010).

Verbs and nouns as two different word classes are used to introduce the concept of a 'word class' next.

So far we have said that you need to look up verbs under the first letter of their roots or stems, and that you need to look up nouns under their full forms. Verbs and nouns are two different **word classes**, so you can see that it is important that you know which word class a certain word belongs to. There are many different word classes in Zulu, and you will need to know how to look up words in each type of word class. We will explain this in detail below.

In addition to a word class, you will also often see a number or even a pair of numbers following the word class in your dictionary. This is because nouns

in Zulu are traditionally grouped in different pairs of **noun classes** — which have unique pairs of **noun class prefixes** — and many other words have to be 'in harmony' with those noun class prefixes. In the example sentence, 'at all times' is *nga**zo** **zo**nke **izi**khathi* in Zulu. Because the class 8 plural noun ***izi**khathi* is used, the temporal adverb *nga**zo*** and the inclusive quantitative pronoun ***zo**nke* also have class 8 forms. If the class 7 singular noun ***isi**khathi* 'time' had been used, then the phrase would become *nga**so** **so**nke **isi**khathi* 'all the time', with all words now in class 7 forms. It is very important that you understand this **need for harmony** (the so-called 'concordial agreement system') in Zulu, because only then will you appreciate the reason for assigning both a word class and class numbers to words in your Zulu dictionary. The different concords that are prefixed to verbs also need to be in harmony with the classes of the nouns they refer to, which is why all tables with concords consist of many lines: one line for each class, and a different concord for each class. The numbering system itself has been agreed upon internationally, so it is good you learn and know it.

Comments 5: This paragraph summarizes the core of the Bantu concordial agreement system, which is linked to the classification of Bantu nouns in noun classes, and a Bantu-wide numbering system. Note how the concept is gently introduced, by referring (twice) to a need for harmony, before the proper linguistic description is used. From a lexicographic point of view, the need to indicate the word class for each lemma, *and* (where relevant) the need to also indicate the class number, is also explained. Needless to say, the indication of class numbers across the word classes is missing from all existing dictionaries for Zulu.

3.2 Word classes

Most of the words listed in your dictionary are nouns and verbs. Actually, as many as 45.3% of all the dictionary entries are **nouns**, and 15.5% are **verbs**. Unique to this dictionary is that certain **locative forms** derived from nouns have also been listed as headwords; they make up 12.4% of your dictionary.

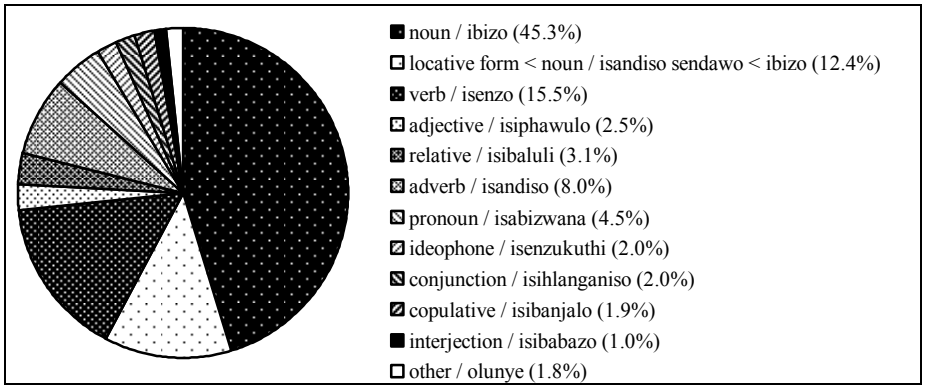


Figure 1: Distribution of the main Zulu word classes in your dictionary

As the pie chart shown in Figure 1 indicates, these three groups of words (nouns, verbs, and locative forms derived from nouns) make up nearly three-quarters of your dictionary. This does not mean that all the other word classes are less important — on the contrary. Many words in the other word classes are used much more frequently than some nouns and verbs. Words that 'accompany' nouns are typically **adjectives** and **relatives**; words that 'accompany' verbs are typically **adverbs**. To make well-formed sentences in Zulu, you also need to make use of **pronouns**, **ideophones**, **conjunctions**, **copulatives**, **interjections**, etc. We will now explain how you can look up words in each of these word classes.

Comments 6: With the concept of a word class introduced, it was now possible to extend the list of word classes to all the major ones in terms of number of members. In typical Zipfian style, some of the smaller word classes (in terms of members) actually contain the words used most often in Zulu (think for example of the class of the conjunctions). In the dictionary, actual word frequencies are indicated with a star rating following the lemma signs, an aspect explained in the Introduction to the OZSD. From the perspective of the dictionary user, the frequency breakdown shown in Figure 1 is most relevant, as it immediately tells that user how large each of the word classes is compared to the other word classes. The order shown in Figure 1 is also the order in which each of the word classes is given attention in the mini-grammar. That order combines both frequency (from most to least populous word class) and linguistic logic (e.g. bringing nouns together, or noun modifiers before verb modifiers, etc.). The coverage of each word class is henceforth fully driven by the relative frequencies seen in Figure 1. A word class with more members gets more attention than one with fewer ones, so on the whole this means that the sections become shorter and shorter as one proceeds through the different word classes in the mini-grammar. What is said about each word class is also fully driven by corpus facts, meaning that only what is most frequent in the corpus ends up being discussed. This is a radical departure from conventional grammatical descriptions, where one is interested in presenting full paradigms, irrespective of whether all items actually occur or not.

3.3 Nouns

Grammatically, a noun consists of a **noun class prefix** and a **noun stem**. The noun class prefix itself consists of a **pre-prefix** and a **basic prefix**. For example, the singular noun *isikole* 'school' consists of the pre-prefix *i-*, the basic prefix *-si-*, and the noun stem *-kole*. To change a singular noun into a plural noun in Zulu, you need to change the form of the noun class prefix, here to *izikole* 'schools'. Nouns that have their singular in the *isi-* class and their plural in the *izi-* class belong to the class pair 7/8.

In your dictionary, you need to look up nouns under the pre-prefix; thus under the letter I for *isikole*. Under *isikole* you will find a full treatment of this word, with frequency of use and grammatical information, a translation, and an example sentence. The full treatment is normally only found under the **singular** form (*isikole*), and not under the plural form (*izikole*). When a plural form is frequent, it has also been listed in your dictionary, but the only information you will find there is a **cross-reference** to the singular form. So it is important

that you learn to recognize the full forms of nouns, with their pre-prefixes, and that you know how to change a plural noun into a singular noun. If you do not learn the system, you will spend more time thumbing through your dictionary, going from one entry to the next. The noun class system is shown in Table 1, which shows the main singular/plural noun pairs in your Zulu dictionary.

Table 1: Distribution of the singular and plural Zulu nouns in your dictionary

Class (pair)	**%**	**Class prefixes ①**	**Class prefixes ②**	**Notes**
1/2	4.6	um-/aba-	umu-/aba-	① Polysyllabic stem ǀ ② Monosyllabic stem
1/**2**	3.7	..	..	..
1a/2a	4.6	u-/o-		
1a/**2a**	2.1	..		
1a/-	1.8	u-		
3/4	6.0	um-/imi-	umu-/imi-	① Polysyllabic stem ǀ ② Monosyllabic stem
3/**4**	3.7	..	..	..
3/-	1.1	um-	umu-	..
5/6	11.4	i-/ama-	ili-/ama-	② Variant for some monosyllabic stems
5/**6**	7.8	..	..	..
5/-	1.1	i-		
6	1.1	ama-		
7/8	8.7	isi-/izi-	is-/iz-	① Consonant-initial stem ǀ ② Vowel-initial stem
7/**8**	7.0	..	..	..
7/-	1.3	isi-	is-	..
9/10	9.2	iN-/iziN-	i-/izi-	① With N = m before b, p, f and v; else N = n
9/**10**	4.7	..	..	② When nasal-initial stem, or with some loanwords
9/-	2.9	in-	i-	..
11/10	2.7	u-/iziN-	u-/izi-	① With N = m before b, p, f and v; else N = n
11/**10**	1.4	..	..	② When stem starts with l or n
11	1.6	u-	ulw-	① Consonant-initial stem ǀ ② Vowel-initial stem
14	2.6	ubu-		+ sound changes (=> *uboya, utshani, utshwala*)
15	4.4	uku-	ukw-	① Consonant-initial stem ǀ ② Vowel-initial stem
Other	4.7			
	100.0			

Table 1 tells you that most nouns in your dictionary — as many as 11.4% — belong to the singular class 5, with their corresponding plural in class 6. The notation **5**/6 is used for these nouns: the number in bold (**5**) refers to the class number of the noun you are looking at, the other number (6) to the corresponding (here a plural) class number. Examples are ***i****khanda* 'head', ***i****langa* 'day; sun', ***i****zwe* 'country'. From Table 1 you can see that when you wish to look up the plural forms ***ama****khanda* 'heads', ***ama****langa* 'days; suns', ***ama****zwe* 'countries', you can go directly to the letter I of your dictionary, to look up the singular forms.

For some class pairs, there are two sets of class prefixes. For example, for **1**/2 one has both ***um-***/*aba-* and ***umu-***/*aba-*. The pair ***um-***/*aba-* is used with noun stems that have more than one syllable (which we call 'polysyllabic stems'). For instance, the noun stem of ***um****felokazi* 'widow' / ***aba****felokazi* 'widows' has four syllables: *-fe-lo-ka-zi*. The pair ***umu-***/*aba-* is used with noun stems that have only one syllable (which we call 'monosyllabic stems'). For instance, the noun stem of ***umu****khwe* 'father-in-law' / ***aba****khwe* 'fathers-in-law' has only one syllable *-khwe*. Therefore, if you want to look up a noun that starts with *aba-*, Table 1 not only tells you that this is a plural noun, but also that the singular needs to be looked up under *um-* for a polysyllabic noun stem, and under *umu-* for a monosyllabic noun stem. For the class pair 3/4, one finds a similar situation: *um-/imi-* for polysyllabic stems, but *umu-/imi-* for monosyllabic stems.

Because nouns have been listed with their prefixes in your dictionary, you will not have problems looking up the alternative forms shown for the singular/plural pairs 5/6, 7/8, 9/10, and 11/10, nor for the alternative forms of class 11. As long as you remember to look up singular forms of nouns in full, you cannot go wrong. Similarly, 'difficult' class 14 nouns like *uboya*, *utshani*, or *utshwala*, can be looked up directly.

Comments 7: Table 1 is a quantified representation of the noun class system in Zulu, a first for this language. Noun classes are not considered in isolation, but are treated as genders, typically linking singular and plural members. In Column 1, dashes replace class numbers for one-class genders. Each number in bold, a notation first introduced in De Schryver (2001: 3-4), indicates the class of the member currently in focus. A noun class, as a member of a gender, is literally weighted (Column 2), and for each the various class prefixes are shown (Columns 3 and 4, with explanations in Column 5). The order of those class prefixes (with Column 3 always more frequent than Column 4) is based on dictionary occurrences, unlike the presentation in traditional grammars, where the so-called full forms are always presented first (e.g. *umu-/aba-* before *um-/aba-*). Only what is actually found in the dictionary is mentioned in Table 1 (e.g. for gender **9**/- one finds the prefix *in-* rather than *iN-*, simply because no nouns in gender **9**/- have been lemmatized that have the prefix *im-*, so there is no need to over-generalize). For linguists used to traditional Zulu grammars, a presentation like the one seen in Table 1 is undoubtedly a radical departure from more familiar presentations; for the user of the OZSD, however, this is simply in direct agreement with the word lemmatization used for nouns.[10]

Note, however, that when the pre-prefix of a noun is missing, as in *lo muntu* 'this person', you will first need to add the pre-prefix before looking up this noun under *umuntu* 'person'. Furthermore, various **morphemes** may be prefixed to a noun, and those need to be cut off. Thus, words like *lomuntu* 'of a person', *ngumuntu* 'she/he/it is a person', *nomuntu* 'with a person', *njengomuntu* 'like a person', *ngomuntu* 'about a person', *okomuntu* 'that of a person', *ngingumuntu* 'I am a person', *wayengumuntu* 'he was a person', etc. all need to be looked up under *umuntu*. Remember, therefore, that all nouns are listed under their pre-prefixes, which are either *a-*, *i-*, *o-*, or *u-*. Reformulated, all nouns have been listed under just four letters of the alphabet in your dictionary: A, I, O, or U.

Comments 8: Although lemmatizing nouns with their full prefixes, thus as complete words carrying meaning, is also the most intuitive lemmatization approach, users should not be led to think that all nouns appear in this canonical form in written Zulu. This paragraph dispels this, and shows typical environments of words and morphemes preceding nouns, either written disjunctively or conjunctively. Cutting off the conjunctively written parts may be challenging for the beginner, restoring the pre-prefix, which is always a reflection of the vowel of the basic prefix, should be more manageable.

In your dictionary you will also find over a hundred so-called **infinitive nouns**. These are all nouns derived from verbs, and always take the noun class prefix of class 15 (*uku-* or *ukw-*). These nouns have been chosen for a combination of two reasons: (i) they have new, independent meanings, and (ii) they are often used in Zulu. Examples are *ukuhlolwa* 'examination', *ukulimala* 'injury', or *ukwenza* 'action'. When the verb from which such an infinitive noun is derived has also been listed in your dictionary, you will see a cross-reference linking the noun to the verb, for instance *ukwenza* 'action' **<** *-enza* 'do, make, act; cause'. In general, all **deverbatives** (meaning all nouns which are derived from verbs) for which the verb is also listed in your dictionary, have been linked with that verb by means of a **cross-reference**. For example *impilo* 'life; health' **<** *-phila* 'live; be in good health'. Again, in this dictionary, you do not need to go to the alphabetic section PH (the first letters of the stem) in order to find the noun *impilo*; simply go directly to the first letter of the full noun, thus I.

Comments 9: The mini-grammar does not cover word formation processes, but cross-references in the dictionary do link deverbatives to their verb roots and stems whenever the latter have also been lemmatized. At the expense of fewer lemmas, cross-references in the other direction, thus from verb roots and stems to all the lemmas derived from those, could also have been considered. This was for example done in my *Cilubà-Dutch Lexicon* (De Schryver and Kabuta 1997) by means of what I termed '(frequency-based) tail slots' in my MA dissertation (De Schryver 1999: 53-54; cf. also De Schryver and Prinsloo 2001). Expecting Grade 4 to 9 learners to follow up on all cross-references away from a node, in addition to those from single spokes (deverbatives) to single nodes (verb roots or stems), was however considered too advanced for the OZSD.

When you do not find a certain noun under its singular form, this may be because the noun is infrequent and has thus not been listed in your dictionary. In some cases, however, the plural form was frequent enough to be listed, while the singular was not. In such cases, you will find a full treatment under the plural form. Examples include: *amaphesenti* 'percent; percentage', *iziphumuzi* 'punctuation marks', or *izinhlobonhlobo* 'different kinds'.

Lastly, note that not all possible singular/plural pairs have been listed in Table 1 (see the 4.7% 'Other'). Another combination is for example 9/6: ***i***fil*imu* 'film, movie' / ***ama***fil*imu* 'films, movies', or ***i***nk*osi* 'chief, king' / ***ama***k*hosi* 'chiefs, kings'. Here too, your dictionary will show you the correct classes (9 and 6), as well as the exact forms (for example *inkosi* and *amakhosi*), so they should not be problematic to look up.

Comments 10: In line with what was noted under Comments 6, what is mentioned in the mini-grammar is restricted to what is both frequent and relevant for using the dictionary. Compared to Table 1, a staggering 19 further genders and combinations of genders are for example attested in the dictionary (nine of them hapaxes however), yet not knowing about those does not hamper successful dictionary use. Keeping them for a scientific article in which the noun class system of Zulu is revisited makes all the more sense. Likewise, there are also more (but infrequent) variant prefixes for some genders than the two columns of variants offered. For example, for class 11 the OZSD has one (and only one) lemma with the noun class prefix *ulu-*, namely *uluthi* 'stick; twig'.

3.4 Locative forms derived from nouns

Normally, a locative meaning can be 'added' to a noun from the class pairs 1/2 or 1a/2a by replacing the pre-prefix with the class 17 locative prefix *ku-* (or the variant *ko-*). As such, *umuntu* 'person' becomes ***ku****muntu* 'to a person', or *omalume* 'uncles' becomes ***ko****malume* 'to the uncles'. Especially with loanwords in the class pair 9/10, one may also find the variant *kwi-*. As such, *inombolo* 'number' becomes ***kwi****nombolo* 'at the number'. Word forms like these are not hard to decode, and have therefore not been listed in your dictionary.

A second way to add a locative meaning to a noun is to replace the pre-prefix with *e-* and to add the suffix *-ini*. As such, *abantu* 'people' becomes ***e****bantw****ini*** 'to/from/among/... the people', or *umthombo* 'fountain, spring' becomes ***e****mthonj****eni*** 'in/at/to/from/... the fountain/spring'. This approach may be used for nouns in all classes, except for those in class 11 or 14, where the pre-prefix is replaced with *o-*, again with the suffix *-ini*. As such *uhlangothi* 'side' becomes ***o****hlangoth****ini*** 'on/at/to/from/... the side', or *ukhetho* 'election' becomes ***o****khethw****eni*** 'in the election'. Numerous sound changes are found when the suffix *-ini* is attached to nouns, as seen in the underlined parts in the examples here. It is mainly for this reason that all frequent locative forms that are derived from nouns using the so-called ***e-/o-...-ini* 'locativisation strategy'** have been listed directly into your dictionary. Grammatically, these locativised nouns have actually become 'locative adverbs'. To save space in the dictionary, only translations into English of the meanings are given for these locative adverbs, without any examples.

A second reason for listing these locative adverbs in your dictionary is that there are locativised nouns where the suffix *-ini* does not appear. The appearance or not of the suffix *-ini* is not predictable, which makes it useful to list the correct frequent forms in the dictionary. As such, all of the following forms for example appear without the suffix *-ini*: ***e****busika* 'in/during/... winter' (< *ubusika* 'winter'), ***e****khaya* 'at/from/... home' (< *ikhaya* 'home'), or ***o****lwandle* 'in/on/to/from/... the ocean/sea' (< *ulwandle* 'ocean, sea').

A third reason for listing these locative adverbs in your dictionary is that there are locativised nouns for which the frequency of the noun itself is extremely low (or may not even appear at all in our 8.5-million-word Zulu corpus). Examples include: ***e****maphandl****eni*** 'in/to/from/... the rural areas', ***e****msamo* 'in/at/to/from/... the back of the hut', or ***e****sidlangalal****eni*** 'in public; openly'. For

all these examples the corresponding noun is not seen in the corpus, so it would be wrong to list such nouns in a dictionary that is corpus-driven and that focuses on frequently used words.

Lastly, as was the case for the nouns, locativised nouns may be preceded by one or more **morphemes**, which need to be cut off before they can be looked up in your dictionary.

Comments 11: This sub-section is entirely self-contained and provides an excellent argued example of the type of choices that had to be made with regard to the lemmatization or not of certain Zulu 'words'. Looking up nouns locativised by means of *ku-* and its variants was considered manageable for the target user group, but locativisation by means of the *e-/o-...-ini* strategy was not. For a linguistic account of the issues involved, see De Schryver and Gauton (2002).

3.5 Verbs

In this dictionary verbs need to be looked up under the first letter of their roots or stems. The same approach is followed in all dictionaries for Zulu. This is because each verb root or stem can combine with very many combinations of prefixes, and the ending of a verb also varies (it can be *-a*, *-e*, *-i*, *-ile*, or *-anga*). Listing all these possibilities in a dictionary would not be practical, as one would need to list many dozens, sometimes hundreds of forms for each verb. This therefore means that you will need to learn to cut off all **verbal prefixes** and **formatives** from verbs (those listed in Tables 4 to 6 as well as all the others listed in your Zulu textbook) before you can look up the basic meaning of a verb.

Comments 12: Van Wyk (1995: 86-87) performed a quick, back-of-the-envelope calculation for Zulu, and claimed that "[t]he number of combinations possible for a suitable transitive verb stem is [...] 18 x 19 x 6 x 2," or thus 4 104. While the actual theoretical figure is many times higher (for one, Van Wyk did not take the variation in verb endings into account), corpus counts indicate that the number of those that are actually used is many times smaller. Disregarding the very rare uses, which would not have to feature in a word-based school dictionary anyway, the number of those that are used is still very high (many dozens to several hundred orthographic forms per verb root or stem, as stated in the mini-grammar), so a word-based approach to the lemmatization of verbs in a paper dictionary for Zulu is indeed not feasible.[11] This thus means that we have come back to square one as far as the lemmatization of verbs in a paper dictionary for Zulu is concerned, in that the traditional approach is stuck too. The exercise was not futile, however, as the decision was arrived at following a study of large amounts of actual language use, rather than being based on linguistic extrapolations. The decision, in other words, can truly be defended.

A verb without any verbal prefixes and without any verbal extensions is known as a **verb root**. In your dictionary the final vowel *-a* is always added to verb roots. Examples are *-anga* 'kiss', *-linga* 'try, attempt', and *-thuma* 'send'. 54.5% of all the verbs listed in your dictionary are verb roots. The other 45.5% of the verbs take one or more **verbal extensions**. When a verb takes a verbal extension, it is known as a **verb stem**. Verb stems are also shown together with the final vowel *-a* in your dictionary.

Comments 13: Some linguists may object to the terminology used here, and may prefer to talk about (formal or base) radicals and extended radicals rather than roots and stems respectively, and may prefer to view a stem as merely a root plus the verbal ending, with that root either a radical or extended radical (cf. e.g. Schadeberg 1992[3]: 8).

In Table 2 the distribution of the verbal extensions in your dictionary, as well as the main combinations of verbal extensions, is shown.

Table 2: Distribution of Zulu verbs with verbal extensions in your dictionary

Verbal extension	Form	Meaning	%	Examples
passive	-(i)w-	adds 'be', 'being'	32.0	*-akhiwa, -khishwa, -shaywa*
applicative	-el-	adds 'for', 'on behalf of'	16.6	*-buyela, -khulumela, -xoxela*
causative	-is-	adds 'cause to', 'help'	13.6	*-balisa, -bhalisa, -qondisa*
neuter	-ek- OR -akal-	adds '-able/-ible'	12.1	*-bhekeka, -bonakala, -tholakala*
reciprocal	-an-	adds 'each other'	4.5	*-bambana, -bhekana, -xhumana*
applicative + passive	-elw-	adds 'be ... for/by'	4.5	*-balelwa, -banjelwa, -phelelwa*
causative + passive	-isw-	adds 'be ...'	3.9	*-aziswa, -fundiswa, -setshenziswa*
causative + reciprocal	-isan-	adds '... each other'	1.8	*-bonisana, -hambisana, -phikisana*
reciprocal + causative	-anis-	adds '... each other'	1.2	*-ehlukanisa, -qondanisa, -xhumanisa*
perfective	-elel-	completes an action	1.2	*-azelela, -bambelela, -phumelela*
intensive	-isis-	expresses intensity	1.2	*-bhekisisa, -qondisisa, -zwisisa*
Other			7.3	
			100.0	

Nearly one third (32.0%) of all the verbal extensions simply add a passive meaning to a verb root, 16.6% add an applicative meaning, 13.6% a causative meaning, etc. Because all these verb stems (thus 'verbs with verbal extensions') have been listed directly into your dictionary, you do not need to memorize the various sound changes that apply when suffixing certain verbal extensions to verb roots. For example: *-khipha* + passive > *-khishwa*. Also, the exact meanings of the resulting verb stems are mentioned in the dictionary articles of those verb stems, with examples to support those meanings. It is a good idea now to look up each of the examples listed in Table 2, and to compare the dictionary information for each verb stem with that found under each verb root from which it is derived. (Note that a **cross-reference** always links a verb stem with its verb root, on the condition that that verb root is frequent enough to be listed in your dictionary.)

Comments 14: Although I said that verbs in the OZSD are lemmatized like in any other dictionary for Zulu, this is not entirely true in that this is only correct for verb *roots* (or for formal or base radicals if one prefers). Existing dictionaries for Zulu will not systematically lemmatize verb *stems* (or extended radicals). Doke and Vilakazi (1953[2]), for example, do not include any verbs with passive extensions, and will furthermore normally only include those verbs with verbal extensions that have undergone some level of lexicalisation. In contrast, the OZSD systematically includes all frequent verb stems.

As an example of the exercises the dictionary user is asked to do, (1) and (2) below show the effect of adding a perfective extension:

(1) **-phumelela** ** *verb + perfective*
C -PHUMA ■ succeed ◆ Kufanele usebenze kanzima ukuze **uphumelele**. *You must work hard in order to succeed.*

(2) **-phuma** *** *verb* ■ go out; come out; exit
◆ UGumede **uphume** emotweni ephethe izimbali ezinhle zikankosikazi wakhe. *Gumede got out of the car carrying beautiful flowers for his wife.*

This exercise will give you a good idea of the meanings added by the verbal extensions, and it will expose you to typical dictionary structures. Three of those structures are commented on below.

Firstly, note that frequent reflexive verbs are mentioned within the dictionary articles of the verbs from which they are derived, and thus need to be looked up there. To form a reflexive verb, the **reflexive concord** ***-zi-*** (or ***-z-*** before vowels) is simply prefixed to a verb root or stem. For example, for the applicative form *-xoxela* 'tell (someone)', which is derived from the verb root *-xoxa* 'talk', the reflexive form is ***-zi****xoxela* 'just talk, simply talk, merely talk'. This reflexive form is explicitly mentioned under the verb stem *-xoxela* as a derivation because most uses of that verb stem take the reflexive concord. Throughout your dictionary, all **derivations** are preceded by a black arrow (►). For examples of other derivations, see the verb *-bona* (> *sawubona*), the noun *imali* (> *malini*), or the adjective *enkulu* (> *enkulukazi*).

Comments 15: Two of the examples mentioned are shown in (3) and (4) below:

(3) **-xoxela** *verb + applicative* C -XOXA ■ tell (someone) ◆ Ungalokothi **umxoxele** uMaNdlovu ngaleli phupho lakho. *You dare not tell MaNdlovu about this dream of yours.*
► **-zixoxela** ■ just talk; simply talk; merely talk ◆ Baqhubeka-ke **bezixoxela** abangane bethu laba bedlalisa nomntwana owayelokhu ehleka. *They continued merely talking, these friends of ours, while playing with the child who was laughing all the time.*

(4) **imali** *** *noun* ***9****/10* (*pl.* **izimali**) ■ money
◆ Ukhokhe **imali** eningi yesikole. *She paid a lot of money in school fees.*
► **malini** ■ how much money? ◆ Kukhokhwa **malini** emathekisini ahambela eMtuba esuka eMpangeni? *How much money is paid for taxis going from eMpangeni to Mtubatuba?*

Secondly, verbs with the reciprocal verbal extension (*-an-*) are mostly, sometimes exclusively, followed by the **adverbial formative** ***na-*** 'with, together with'. For example, look up the verbs *-bhekana*, *-hambisana*, and *-phikisana*, and

you will see that translations and examples are only given for the combinations *-bhekana na-*, *-hambisana na-*, and *-phikisana na-*. Throughout your dictionary, all **combinations** are preceded by an empty diamond (◊). Also, in your dictionary, when verbs combine with nouns to form new meanings, you will need to look up the noun and not the verb for those meanings. See for example the combinations with verbs listed under the nouns *icala*[1], *ithambo*, or *intwala*.

Comments 16: Two of the examples mentioned are shown in (5) and (6) below:

(5) **-phikisana** *verb + causative + reciprocal*
C -PHIKA
◇ **-phikisana na-** ■ argue (with); disagree (with); dispute; oppose ◆ Musa **ukuphikisana** nami! *Don't argue with me!* ◆ Akekho **owaphikisana** nokuqanjwa kabusha kwalo mgwaqo. *There is no one who disputed the renaming of this road.*

(6) **icala**[1] *** *noun* **5**/6 (*pl.* **amacala**) **1** ■ case (*in law*) ◆ Kuyothi uma **icala** liya emajajini wena ube ngufakazi. *When the case goes to the judges, you will be the witness.*
2 ■ offence ◆ Kuthiwa badedelwa kodwa baphinde benza amanye **amacala** anzima kakhulu kunalawo ababeboshelwe wona. *It is said that they were released, but they committed other, more serious offences than those they had been arrested for.*
◇ **-bekwa icala** ■ be accused (of); be charged (with) ◆ Waboshwa abekwe **icala** lokweqa umyalelo wenkantolo. *He was arrested and accused of breaking an order of the court.*
◇ **-vuma icala** ■ admit; plead guilty ◆ Uma elivuma **icala**, uzomiswa imidlalo emithathu. *If he pleads guilty, he will be suspended for three games.*

Thirdly, **grammatical abbreviations** may also be used in derivations and combinations. These are used to summarize a full list (known as a 'paradigm') of possibilities. For example, if you look up the verb stem *-azelela*, you will find the derivation *[SC+]ngazelele*. Here SC stands for subject concord, and this notation means that any of the subject concords seen in Table 4 must precede *-ngazelele* in order to obtain the meaning.

Comments 17: The article for the verb stem *-azelela* is shown in (7) below:

(7) **-azelela** *verb + perfective* C -AZI[1]
▶ [SC+]**ngazelele** ■ unexpectedly; surprisingly ◆ Wathi **engazelele** wabona ikhanda likaShaka livela ngaphezu kothango lwesigodlo. *He unexpectedly saw Shaka's head appearing above the fence of the royal enclosure.*

The various grammatical abbreviations used in your dictionary are listed in Table 3.

Table 3: Grammatical abbreviations used in your dictionary

Code	Meaning	For examples, see under
[DEM+]	preceded by any demonstrative	*lakho, lwethu, wethu*
[NEG+]	preceded by any negative	*elangeni, isinqe, quthu*
[PC+]	any possessive concord prefixed to word	*igolide, isihlanu, ngasese*
[RC+]	any relative concord prefixed to word	*-banda, -mela, -phakama*
[SC+]	any subject concord prefixed to word	*-azelela, ngokuphindwa, uzwelonke*
[+LOC]	followed by any locative	*-balekela, -sebenzela, -shonela*

Comments 18: As before, the dictionary user is expected to look up one or more of the examples given, in order to get acquainted with the dictionary system. As an illustration, one of the examples for the grammatical abbreviation [DEM+] is shown in (8) below:

(8) **wethu** ** *possessive pronoun 1+1p pl* ■ our
◆ Ngisho nomqeqeshi **wethu** unethemba elikhulu ngalo mdlalo. *Even our trainer has a lot of faith in this game. 3+1p pl* ■ our
◆ Kungumgomo **wethu** ukuphatha iziguli ngesizotha nangesineke. *It's our policy to treat patients with dignity and care.*
◇ [**DEM+**] **... wethu** ■ of ours ◆ Kumele senzeni nje nempela kulo mjondolo **wethu**? *What must we really do in this shack of ours?*

Verbs are and will always be the most difficult part of speech to look up in a printed Zulu dictionary. This is so because a verb root or stem is normally 'hidden' in the middle of a much longer word. In order to help you recognize verb roots and stems, Table 5 lists the main verbal tenses and moods in Zulu. Table 5 has to be used together with Table 4, as well as Table 6. In a way, Tables 4, 5 and 6 give you a short overview of the verbal grammar. Please read through this information slowly, and reread it often, until you memorize it.

Table 4: Verbal prefixes (SC and OC) to be cut off before looking up Zulu verbs, versus adjective prefix (AP) and relative concord (RC) to keep when looking up adjectives and relatives

Class no.	Class prefix	+SC +C	+SC +a/e	+SC +o	-SC	~SC	§SC	OC	AP	RC
1; 1a	**um(u)-; u-**	u-	w-	w-	ka-	e-	a-	-m(u)-	om(u)-	o-
2; 2a	**aba-; o-**	ba-	b-	b-	= +SC	be-	= +SC	-ba-	aba-	aba-
3	**um(u)-**	u-	w-	w-	wu-	= +SC	= +SC	-wu-	om(u)-	o-
4	**imi-**	i-	y-	y-	yi-	= +SC	= +SC	-yi-	emi-	e-
5	**i(li)-**	li-	l-	l-	= +SC	= +SC	= +SC	-li-	eli-	eli-
6	**ama-**	a-	ø	ø	wa-	e-	= +SC	-wa-	ama-	a-
7	**is(i)-**	si-	s-	s-	= +SC	= +SC	= +SC	-si-	esi-	esi-
8	**iz(i)-**	zi-	z-	z-	= +SC	= +SC	= +SC	-zi-	eziN-	ezi-
9	**i(N)-**	i-	y-	y-	yi-	= +SC	= +SC	-yi-	eN-	e-
10	**izi(N)-**	zi-	z-	z-	= +SC	= +SC	= +SC	-zi-	eziN-	ezi-
11	**u(lu)-**	lu-	lw-	l-	= +SC	= +SC	= +SC	-lu-	olu-	olu-
14	**ubu-**	bu-	b-	b-	= +SC	= +SC	= +SC	-bu-	obu-	obu-

15	**uku-**	ku-	kw-	k-	= +SC	= +SC	= +SC	-ku-	oku-	oku-
17	**ku-**	ku-	kw-	k-	= +SC	= +SC	= +SC	-ku-	oku-	oku-
1p sg	**—**	ngi-	ng-	ng-	= +SC	= +SC	= +SC	-ngi-	engim(u)-	engi-
1p pl	**—**	si-	s-	s-	= +SC	= +SC	= +SC	-si-	esiba-	esi-
2p sg	**—**	u-	w-	w-	= +SC	= +SC	= +SC	-ku-	om(u)-	o-
2p pl	**—**	ni-	n-	n-	= +SC	= +SC	= +SC	-ni-	eniba-	eni-

Note that +SC depends on whether the initial letter of what follows is a consonant (+C), an 'a' or 'e' (+a/e), or an 'o' (+o).

Meanings of the abbreviations in Tables 4 and 5

Abbreviation	Meaning	Abbreviation	Meaning
p	person	SC	subject concord
sg	singular	OC	object concord
pl	plural	AP	adjective prefix
+	positive	RC	relative concord
-	negative	C	consonant
ø	no concord	V	vowel
~	situative; continuous	N	nasal
§	subjunctive	VERB	verb root or stem

Table 5: Verb formulas for the main Zulu tenses and moods
(Note that an OC, when present, is always found immediately before the verb root or stem.)

Tenses and Moods	+ -		Formula	Examples
Imperative	+	sg	VERB_a	*bhala!* 'write!', *hamba!* 'go!'
		pl	VERB_a_ni	*bhalani!* 'write, all of you!', *hambani!* 'go, all of you!'
VERB = VC...		sg	y_VERB_a	*yakha!* 'construct!', *yenza!* 'do!'
		pl	y_VERB_a_ni	*yakhani!* 'construct, all of you!', *yenzani!* 'do, all of you!'
VERB = C(C)		sg	yi_VERB_a	*yima!* 'stand!', *yidla!* 'eat!'
		pl	yi_VERB_a_ni	*yimani!* 'stand, all of you!', *yidlani!* 'eat, all of you!'
	-	sg	musa uku_VERB_a	*musa ukudlala!* 'don't play!', *musa ukwesaba!* 'don't be afraid!'
		pl	musani uku_VERB_a	*musani ukuhleka!* 'don't laugh, all of you!', *musani ukwethuswa!* 'don't be surprised, all of you!'
Present	+	long	+SC_ya_VERB_a	***u**Zodwa **u**yakhala* 'Zodwa is crying', or 'Zodwa cries'
			+SC_ya_OC_VERB_a	***zi**ya**ba**bulala* 'they are destroying them'
		short	+SC_VERB_a [+object]	***i**gazi **li**gijima kalula* 'the blood is running easily'
			+SC_OC_VERB_a	*uma **i**ntombi **i**ngithanda* 'if a girl loves me'
	-		a_-SC_VERB_i	***in**tsha **ya**le mihla a**yi**thandi ...* 'the youth of today do not like ...'
			a_-SC_OC_VERB_i	***um**fundisi a**ka**baphi amaswidi* 'the pastor doesn't give them sweets'

Perfect / Near past	+	long	+SC_VERB_ile	***lo**wo **m**fana **u**balekile* 'that boy ran away'
			+SC_OC_VERB_ile	***ngi**ku**u**tshelile* 'I've told you'
		short	+SC_VERB_e [+object]	***si**zwe izindaba* 'we heard the news'
			+SC_OC_VERB_e [+object]	***u**thisha **u**ba**u**vumele amaswidi* 'the teacher allowed them sweets'
	-		a_-SC_VERB_anga	***angi**lalanga* 'I could not sleep', or 'I didn't sleep'
			a_-SC_OC_VERB_anga	***we**na a**wu**u**yi**zwanga le ndaba*? 'haven't you heard this story?', ***we**na a**wu**u**yi**zwanga*? 'you have not heard it?'
Stative	+		+SC_VERB_ile	***ku**balulekile* 'it is important'
	-		a_-SC_VERB_ile	***angi**lambile* 'I am not hungry'
Remote past	+		+SC_a_VERB_a	*cishe **ng**afa!* 'I nearly died!'
			+SC_a_OC_VERB_a	*wa**ng**itshela* 'you told me'
	-		a_-SC_a_VERB_a	*a**ka**hlala* 'he didn't sit'
Near future	+		+SC_zo_VERB_a	***ba**zoqhubeka* 'they will continue'
			+SC_zo_OC_VERB_a	***ngi**zo**m**tshela ubaba* 'I will tell father', ***ngi**zo**m**tshela* 'I will tell him'
	-		a_-SC_zu_VERB_a	***aba**zali a**ba**zukhokha* 'the parents will not pay'
			a_-SC_zu_OC_VERB_a	*a**ngi**zu**m**buza* 'I will not ask her'
Remote future	+		+SC_yo_VERB_a	***si**yohlala* 'we shall stay'
			+SC_yo_OC_VERB_a	*umkakho **u**yo**m**thanda* 'you will love your wife', ***u**yo**m**thanda* 'you will love her'
	-		a_-SC_yu_VERB_a	*a**ku**yubuya muntu* 'no one will return'
Present Situative / Participial	+		~SC_VERB_a	*nga**ba**bona **aba**bili **be**guqa* 'I saw the two kneeling down'
	-		~SC_nga_VERB_i	*uma **be**ngaphumeleli empilweni* 'if they are not succeeding in life'
Subjunctive	+		§SC_VERB_e	***a**thathe amanzi **a**phuze* 'then she takes water and drinks'
	-		§SC_nga_VERB_i	***in**tsha **i**ngagcini* 'the youth must not stop'
Past continuous	+		be_~SC_VERB_a	*be**ngi**hamba* 'I was walking'
Future continuous	+		+SC_zo_be ~SC_VERB_a	***u**zobe **e**gijima* 'he will be running', ***ama**kwaya **a**zobe **e**dlana imilala* 'the choirs will be competing'

Table 6: A selection of other verbal formatives, to be cut off before looking up Zulu verbs

Formative	Meaning	Translation	Place in formula
-sa-	+ progressive	'still'	after SC
..	- progressive	'no longer'	..
-nga-	+ potential	'can', 'may', 'could'	after SC
-nge-	- potential	..	..
-bo-	request	'must'	after SC
-ka-	exclusive negative	'not yet'	after SC
ka-, ma-, a-	hortatives	'please', 'let'	before SC
-yo	relative	'who', 'which', 'that'	after VERB
-ni	question	'what?'	after VERB

Comments 19: The verb being the hardest word class to parse in Zulu, Table 5, together with the supporting Tables 4 and 6, is truly the most important summary in the mini-grammar *as a grammar*, which should enable the dictionary users to decode (as well as encode) the language. These tables are undoubtedly overwhelming, a direct result of the highly complex nature of Zulu. No existing grammars of Zulu were consulted in building these tables. Rather, the Zulu corpus was queried and all frequent (and only the frequent) structures seen were described. The use of underscores to indicate morpheme boundaries in the verb formulas, as well as the decision to use a number of different labels for the subject concords (+SC +C, +SC +a/e, +SC +o, -SC, ~SC, §SC) all developed naturally during the effort to describe the corpus patterns seen. Rather than underscores, existing grammatical descriptions in Bantu at large also use dashes, hyphens and dots. Interestingly, for Zulu, the use of a well-defined set of subject concords was utilized to great effect in a series of studies on the Zulu verb published in the 1960s (Beuchat 1963, 1964, 1964a, 1966). In hindsight, the linguists who worked on Zulu half a century ago could very well have come up with overview tables such as Table 5. What they didn't have was evidence on frequency of occurrence. Also observe that all examples used in Table 5 have been selected from the OZSD, so are *real* examples. Parts in bold in the formulas and examples always refer to subjects, underlined parts to objects. While considerable efforts were put into the creation of this section of the mini-grammar, it should be remembered that this is *not* an exhaustive treatment of all aspects of the verb in Zulu, and as such far from complete. In places it is also decidedly too approximate. As a tool to support successful dictionary use, however, thus viewing the mini-grammar *as a guide to use the dictionary*, the level of detail is more or less the furthest one can go for a school dictionary.

The actual use of the material presented in the tables is illustrated next in the mini-grammar, followed by brief sections on auxiliary and copulative verbs.

For example, in the sentence *Izidakamizwa ziyababulala abadlali abancane* 'Drugs are destroying the young players', you should analyze the verb according to the formula +SC_ya_OC_VERB_a.

word in sentence:	***izi****dakamizwa*	***zi****ya*<u>*aba*</u>*bulala*	<u>*aba*</u>*dlali*	<u>*aba*</u>*ncane*
entry in dictionary:	izidakamizwa > isidakamizwa	-bulala	abadlali > umdlali	abancane
word class + class no.:	pl. noun 7/**8**	+SC8_ya_OC2_VERB_a	pl. noun 1/**2**	adjective 2
translation:	drugs	they are destroying them	players	young

The *-be* in the formulas for the past and future continuous, seen in the last two blocks of Table 5, is actually the auxiliary verb stem *-be*. Please consult your Zulu textbook for the correct use of this auxiliary verb stem, as well as for the other frequently used auxiliary verb stem *-se*. Several usage examples may also be found in your dictionary, under the entries -*be* and -*se*. In addition to -*be* and -*se*, which are complex to use, there are also 50 other auxiliary verbs in your dictionary. Auxiliary verbs are typically used together with other verbs. For examples, see under -*azi*[2], -*buye* or -*ngahle* in your dictionary.

Comments 20: In (9) and (10) below, the complete articles for the auxiliary verbs *-be* and *-azi*[2] as found in the OZSD are shown:

(9) **-be** *** *auxiliary verb* **1** ■ was/were (in a state); used to; had; might/may/could/would have ◆ **Wabe** enezizathu zakhe ezamenza wesaba ukubuya emini. *He had his reasons to fear returning in the afternoon. (Literally: He was in a state of being with reasons …)* ◆ **Babe** bephuma naye ngomnyango ongemuva. *They came out with him through the back door. (Literally: They were in a state of coming out …)* **2** ■ will/shall (be in a state); will/shall have ◆ Umqhudelwano **uzokube** use-Stellenbosch. *The competition will be in Stellenbosch.* ◆ **Kuzobe** sekungokwesibili evela kule nkantolo ngala macala. *It will be the second time that he appears in this court in connection with these charges.*

This auxiliary verb actually indicates a continuous action or process, either in the remote past (sense 1) or in the future (sense 2). The meaning of this verb is not always reflected in English translations. Note that ''**-be**'' can also be the past tense of the copulative verb ''-ba''.

(10) **-azi**[2] *** *auxiliary verb* ⊂ -AZI[1] ■ be able to; know how; can ◆ Abafundi **bangakwazi** ukuzicabangela okuqhubekayo. *The students should be able to think for themselves progressively.*

The copulative verb stem *-ba* (with its past tense *-be*) is also complex to use. For examples, see your dictionary under *-ba*, but only your Zulu textbook will teach you the correct use.

Comments 21: There is just one copulative verb stem in Zulu, namely *-ba*, which is why it is this verb that ends the section on verbs. In (11) below, the OZSD article for *-ba* is shown:

(11) **-ba** *** *copulative verb* ■ be; become ◆ Wahleka kwaze **kwaba** buhlungu izimbambo. *He laughed until his ribs became sore.* ◆ Mina bengithi **kungaba** ngcono lokho sikwenze kusasa. *I thought that it would be better if we did that tomorrow.*

3.6 Adjectives

In Zulu, there are only a few adjective stems (such as *-nye*; *-khulu*; *-bili*; etc.). In your dictionary, 17 are covered, and for each of those adjective stems all the frequent full forms have been listed. 'Full form' here means the **adjective stem** together with the **adjective prefix** (AP in Table 4). As such, adjectives like *omunye*, *ezinye*; *omkhulu*, *ezinkulu*; *ababili*, *ezimbili*; etc. can all be looked up directly, thus under the first letter of the adjective prefix.

Comments 22: For an exhaustive treatment of the lemmatization of adjectives in the OZSD, see De Schryver (2008a). In (12) and (13) below, two examples are given:

(12) **omkhulu**[1] *** *adjective cl. 1, cl. 3* ■ big; large; great ◆ Ngithe ngilele ngaphupha ngibona umuntu **omkhulu** evele phakathi kwamafu. *Every time I dreamt, I saw a large person appearing between the clouds.*
◇ **uthisha omkhulu** ■ principal; headmaster ◆ UThisha **omkhulu** wafunda eNcwadini eNgcwele. *The principal read from the Holy Book.*
▸ **mkhulu** ■ master; head (*used to address someone respectfully*) ◆ Ngiyabonga, **mkhulu**. *Thank you, master.*
▸ **omkhulukazi** ■ very big; very large; very great; huge ◆ Inkosi yakwaButhelezi ibize khona umhlangano **omkhulukazi**, lapho wonke umuntu ongumZulu ebelindeleke khona. *The chief of the Buthelezi clan called a great meeting at which every Zulu person was expected.*

(For more on the suffix *-kazi*, as in *omkhulukazi*, see Gauton et al. 2004.)

(13) **ezinkulu** ** *adjective cl. 8, cl. 10* ■ big; large; great ◆ Bayothenga ezitolo **ezinkulu** nasemakethe eku-Warwick Avenue eThekwini. *They are going to shop in the large stores and the market in Warwick Avenue in Durban.* ◆ "Awu! Kanti yinina zinsizwa **ezinkulu**!" *"My word! Is it you, big guys?"*

3.7 Relatives

Compared to adjective stems, there are more relative stems in Zulu (e.g. *-mhlophe*; *-mnandi*; *-thile*). In your dictionary, 26 are covered, and for each of those relative stems all the frequent full forms have been listed. 'Full form' here means the **relative stem** together with the **relative concord** (RC in Table 4). As such, relatives like *emhlophe*, *obumhlophe*; *emnandi*, *obumnandi*; *ethile*, *obuthile* can all be looked up directly (under the first letter of the relative concord).

Comments 23: By and large, the lemmatization of relatives in the OZSD runs parallel to the way in which adjectives have been lemmatized. This is also implicit in the heading of Table 4, where both are contrasted to the way in which verbs have been lemmatized. In (14) and (15) below, two examples are given:

(14) **emnandi** *relative cl. 4, cl. 9* ■ nice; pleasant ◆ Maningi amaphethini emisindo **emnandi** angadaleka ngokuphindaphinda. *There are many pleasant sound patterns that can be created by repetition.* ◆ Wonke umuntu ufuna impilo **emnandi**. *Every person wants a nice life.*

(15) **obumnandi** *relative cl. 14* ■ nice; pleasant ◆ Ngizolala ubuthongo **obumnandi** namhlanje. *I will have a nice sleep today.*

3.8 Adverbs

Zulu **adverbs** (for example *impela*, *kangaka*, or *nakanjani*) can be looked up directly.

Comments 24: Although the lemmatization of what Taljaard and Bosch (1993[2]: 41) call 'true adverbs' should be plainly straightforward in any Zulu dictionary, some lexicographers manage to omit the adverb category altogether from their Zulu dictionaries. For example, the adverb *impela* 'really; actually; truly; indeed' — the 124th most frequent lemma in the OZSD — has not been lemmatized in Mbatha's (2006) monolingual dictionary for Zulu. Instead, the infrequent noun ***ímpéla*** 'the real one' has been lemmatized, as Mbatha, following Nkabinde (1975), does not 'recognize' any word classes besides nouns, verbs, ideophones and interjections! Although not accepted by anyone else working on the Bantu languages, it is unfortunate that the Zulu National Lexicography Unit compiled their dictionary within this framework.[12] The OZSD's dictionary entry for *impela* is shown in (16) below:

(16) **impela** *** *adverb* ■ really; actually; truly; indeed ◆ U-Maurice wenze isimemezelo ukuthi uyamthanda **impela** u-Minah. *Maurice announced that he truly loves Minah.*

Some adverbs have been subdivided into **locative adverbs** (for example *kuwe*, *ngaphambi*, *phansi*), **adverbs of manner** (for example *kabi*, *ngokusemthethweni*, *njengenhlalayenza*), **temporal adverbs** (for example *kudala*, *manje*, *ngomhla*), **instrumental adverbs** (for example *ngabo*[1], *ngathi*, *ngazo*[1]), **associative adverbs** (for example *nabo*[1], *nathi*, *nazo*), **comparative adverbs** (for example *njengabo*, *njengathi*, *njengazo*), and **interrogative adverbs** (for example *kangakanani*, *kanjani*, *nini*).

Comments 25: In addition to 'true adverbs', adverbs may also be derived from words in other word classes. Whenever the words from which such adverbs are derived have also been lemmatized as words in the OZSD, a cross-reference is included. One example each of the adverb subtypes included in the OZSD is shown in (17) to (23) below:

(17) **kuwe** *** *locative adverb 2p sg* ⊂ WENA ■ at/to/from/in/on/... you ◆ Luchazani usuku lwakusasa **kuwe**? *What does tomorrow mean to you?*

(18) **ngokusemthethweni** *adverb of manner* ⊂ EMTHETHWENI ■ officially ◆ Bonke bafungiswe **ngokusemthethweni** yiJaji, uMnu. Vuka Tshabalala. *All were sworn in officially by the Judge, Mr Vuka Tshabalala.*

(19) **kudala** ** *temporal adverb* ■ long ago; in the past; in the old days ◆ Ngakutshela **kudala** ukuthi angifuni ukushada. *I told you a long time ago that I don't want to get married.*

(20) **ngathi** *instrumental adverb 1p pl* ⊂ THINA
■ about us ◆ Ngesaba ukubatshela **ngathi** sebeyoze bazibonele. *I am afraid to tell them about us; they will eventually find out for themselves.* ■ with us ◆ Okwamanje umqeqeshi usebenza kakhulu **ngathi** ukuze sikwazi ukushaya amagoli amaningi. *At the moment the trainer is working very hard with us, so that we can score many goals.*

(21) **nabo**[1] *** *associative adverb* ⊂ BONA
1 ■ (cl. 2) with (them); to (them) ◆ Hamba **nabo**! *Go with them!* ■ also (they); (they) too; even (they) ◆ Bavuma **nabo** ngaleyo ndledlana. *They also agreed in that small way.* **2** ■ (cl. 14) also (it); (it) too; even (it) ◆ Njengazo zonke izinhlobo zesichasiso ubumnini **nabo** bunezakhi ezimbili. *Just like all the other kinds of qualificative, the possessive also has two formatives.*

(22) **njengazo** *comparative adverb cl. 8, cl. 10*
⊂ ZONA ■ (just) like (them) ◆ Phela ngiyazazi mina ukuthi sengimdala **njengazo** nje izalukazi Baba. *Of course, I know that I am old now, just like the old ladies, Father.* ◆ Angilitholanga ithuba lokufunda **njengazo** zonke ezinye izingane. *I didn't get the opportunity to study like all the other children.*

(23) **nini** *** *interrogative adverb* ■ when? ◆ Ufike **nini** lapha? *When did you arrive here?*
◇ **-noma nini** ■ whenever; anytime; any day ◆ Yinto ongangena esitolo uyithenge noma **nini**. *It is something that you can go into a store for any time and buy.*

As a result of the word-based approach, numerous adverbs actually found their way into a dictionary for the very first time. From the above, all of the following have for example not been lemmatized in Doke and Vilakazi (1953[2]): *kuwe* 'at/to/from/in/on/... you', *ngokusemthethweni* 'officially', *njengenhlalayenza* 'like every other day; as usual', *ngomhla* 'on (the day)', *ngabo* 'about (them, whom), etc.', *ngathi* 'about us, etc.', *ngazo* 'about (them, which), etc.', ..., *nabo* 'with (them), etc.', *nathi* 'with us', *nazo* 'with (them), etc.', ...

3.9 Pronouns

All Zulu pronouns can be looked up directly in your dictionary. They have been grouped together in seven different categories. **Possessive pronouns** are formed by combining a **possessive concord** (which refers to what is possessed) with a **possessive stem** (which refers to the possessor). In Table 7, you would therefore combine a PC with a Pstem to obtain forms like *akhe*, *kwaso*, or *lwami*. All these forms can be looked up directly, but you need to understand the numbers in the notation. For example: 6+1 at *akhe* means that someone in class 1 'possesses' something in class 6; 15+7 and 16+7 at *kwaso* means that a noun in class 7 'possesses' a noun in class 15 or class 16; 11+1p sg at *lwami* means that I

(the first person singular) 'possess' a noun from class 11. Please look at these examples in your dictionary so you become familiar with this system.

Comments 26: The lemmatization of possessive pronouns in the OZSD has been treated in De Schryver and Wilkes (2008). In (24) below, one example is given:

(24) **kwaso** ⋆ *possessive pronoun 15+7* ■ its; her/his ◆ Lesi saga sisazi ngokusetshenziswa **kwaso**. *We know this proverb by its usage. 16+7* ■ of it; of her/him ◆ Wamisa phezu **kwaso** inqwaba enkulu yamatshe ekhona nanamuhla. *On top of it he erected a great heap of stones that are there to this day.*

(I.e., the *ukusetshenziswa* 'usage' (cl. 15) of the *isaga* 'proverb' (cl. 7), and *phezu* 'on top' (cl. 16) of it, with 'it' something in cl. 7, to be inferred from the wider context.)

Table 7: Pronouns in Zulu (formatives and full words)

Class no.	Class prefix	PC	Pstem	DEM I	DEM II	DEM III	Abs Pron	Incl Q Pron	Excl Q Pron	Incl Num Pron "2"	Incl Num Pron "3"
1 **1a**	**um(u)-** **u-**	wa- ka-	-khe	lo	lowo	lowaya	ye(na)	wonke	yedwa	—	—
2 **2a**	**aba-** **o-**	ba-	-bo	laba	labo	labaya	bo(na)	bonke	bodwa	bobabili	bobathathu
3	**um(u)-**	wa-	-wo	lo	lowo	lowaya	wo(na)	wonke	wodwa	—	—
4	**imi-**	ya-	-yo	le	leyo	leya	yo(na)	yonke	yodwa	yomibili	*yomithathu*
5	**i(li)-**	la-	-lo	leli	lelo	leliya	lo(na)	lonke	lodwa	—	—
6	**ama-**	a-	-wo	la	lawo	lawaya	wo(na)	(w)onke	(w)odwa	(w)omabili	*(w)omathathu*
7	**is(i)-**	sa-	-so	lesi	leso	lesiya	so(na)	sonke	sodwa	—	—
8	**iz(i)-**	za-	-zo	lezi	lezo	leziya	zo(na)	zonke	zodwa	zo(zi)mbili	zo(zi)ntathu
9	**i(N)-**	ya-	-yo	le	leyo	leya	yo(na)	yonke	yodwa	—	—
10	**izi(N)-**	za-	-zo	lezi	lezo	leziya	zo(na)	zonke	zodwa	zo(zi)mbili	zo(zi)ntathu
11	**u(lu)-**	lwa-	-lo	lolu	lolo	*loluya*	lo(na)	lonke	lodwa	—	—
14	**ubu-**	ba-	*[-bo]*	lobu	lobo	*lobuya*	bo(na)	bonke	bodwa	*bobubili*	*[bobuthathu]*
15	**uku-**	kwa-	*[-kho]*	lokhu	lokho	lokhuya	kho(na)	konke	kodwa	kokubili	*kokuthathu*
16	**pha-**	kwa-	*[-kho]*	lapha	lapho	laphaya	—	—	—	—	—
17	**ku-**	—	—	lokhu	lokho	lokhuya	kho(na)	konke	kodwa	kokubili	*kokuthathu*
1p sg	**—**	(wa-)	-mi	—	—	—	mi(na)	wonke	ngedwa	—	—
1p pl	**—**	(ba-)	-ithu	—	—	—	thi(na)	sonke	sodwa	sobabili	sobathathu
2p sg	**—**	(wa-)	-kho	—	—	—	we(na)	wonke	wedwa	—	—
2p pl	**—**	(ba-)	-inu	—	—	—	ni(na)	nonke	nodwa	nobabili	*nobathathu*

Meanings of the abbreviations and fonts in Table 7

Abbreviation	Meaning	Abbreviation	Meaning
PC	possessive concord	Abs	absolute
Pstem	possessive stem	Pron	pronoun
DEM	demonstrative	Q	quantitative
I, II, III	position I, II, or III	Incl	inclusive
Italics	not in dictionary; frequency too low	Excl	exclusive
[Italics]	not in dictionary; not attested in corpus	Num	numeral

From the possessive pronouns one can also 'derive' **relativised possessive pronouns**. All the frequent ones have been listed directly into your dictionary; see for example *awakhe* (< *akhe*), *eyabo* (< *yabo*), or *okwethu* (< *kwethu*). Also in this category are words like *ezemfundo*, *ezemidlalo*, or *ezempilo*.

Comments 27: In (25) and (26) below, two examples of relativised possessive pronouns are given:

(25) **awakhe** *relativised possessive pronoun*
6+1 ⊂ AKHE ■ her/his (own); hers/his
◆ Kanti **awakhe** amabutho alwa kanjani, Nxumalo? *In fact, how did his own soldiers fight, Nxumalo?*

(26) **ezemfundo** * *relativised possessive pronoun*
⊂ IZINDABA + IMFUNDO ■ education; educational affairs ◆ UNkk. Thembisile Dlamini ovela eMnyangweni **Wezemfundo**, uthe bayaluseka kakhulu lolu hlelo. *Mrs Thembisile Dlamini, who is from the Department of Education, said they support this programme wholeheartedly.*

There are three types of **demonstrative pronouns**, positions I, II and III, depending on the 'distance' relative to the speaker ('here', 'there', and 'over there'). See Table 7 for all the main forms, and your dictionary for corresponding examples. When a demonstrative pronoun precedes a noun, the pre-prefix of the noun is dropped, and the demonstrative pronoun and the following noun are written as two words in the current orthography. The **adverbial formatives** *nga-*, *na-*, and *njenga-*, on the other hand, may be prefixed to the demonstrative pronouns, and thus written as one word. These formatives need to be cut off before looking up a demonstrative pronoun.

Comments 28: The OZSD being a corpus-driven dictionary, paradigms are not necessarily completed for the sake of completeness. In Table 7, Roman type is used to mark all forms frequent enough to be covered in the A-to-Z section of the OZSD. On the other hand, italics is used to mark those forms for which the frequency is too low to warrant inclusion in the A-to-Z section, while italics plus square brackets is used for forms not attested in the corpus at all, and thus most definitely absent from the A-to-Z section. Frequent variants for position I, on the other hand, have been lemmatized.

In (27) to (30) below, one set is given for positions I, II and III for the class 4 and 9 demonstratives, as well as the corresponding variant for position I:

(27) **le**[1] *** *demonstrative pronoun pos. I* **1** ■ (cl. 4) these (ones) ◆ **Le** mithetho izoletha ezayo izinkinga. *These laws will bring their own problems.* **2** ■ (cl. 9) this (one) ◆ Iyiphi imoto edale **le** ngozi? *Which car caused this accident?*

(28) **leyo** *** *demonstrative pronoun pos. II* ⊂ LE[1] **1** ■ (cl. 4) those (ones) ◆ **Leyo** misebenzi yethu isemqoka njengayo yona **leyo** eminye. *Those jobs of ours are as important as those other ones.* **2** ■ (cl. 9) that (one) ◆ Kungabe yini **leyo**? *What is that one?*

(29) **leya** *demonstrative pronoun pos. III* ⇄ LE[1]
1 ■(cl. 4) those (ones) ◆ Umuntu omunye ubengathi sithandaza sibusisa **leya** mikhuba ebigilwa yilaba bafana uyazi? *You know, another person might think we are praying while blessing the tricks played by those boys.* **2** ■(cl. 9) that (one) ◆ **Leya** ndawo iyaziwa ngomlando wayo. *That place is known for its history.*

The person(s) or object(s) referred to is/are far away in time or distance.

(30) **lena** *** *demonstrative pronoun pos. Ia* ⇄ LE[1]
1 ■(cl. 4) these (ones) ◆ Imikhonto **lena** ibazwa ngendlela ethize ebalulekile kulesi sizwe. *These assegais are shaped in a very special way in this nation.* **2** ■(cl. 9) this (one) ◆ Indatshana **lena** ingahlukaniswa ibe yizingxenye eziyisithupha. *This short story can be divided into six parts.*

This pronoun always follows the noun.

In Zulu, the **absolute pronouns** are used for emphasis or contrast. The full forms shown in Table 7 have been listed in your dictionary, together with important usage notes for each of them. Note that a large number of adverbs may be derived from absolute pronouns (for instance: *bona* > *kubo*, *kubona*, *nabo*, *ngabo*, *njengabo*).

Comments 29: In (31) below, one example of an absolute pronoun is given:

(31) **sona** *** *absolute pronoun cl. 7* ■ it (in particular); she/he (in particular) ◆ **Sona** isizwe sakwaZulu siyobuswa ngubani? *By whom will the Zulu nation be governed?* ◆ **Yiso** isipho sakho leso, ngiyakupha. *That is your gift; I am giving it to you.*

The meaning of this pronoun is not always directly visible in an English translation. When "**sona**" is preceded by formatives such as "yi-", "si-", "na-" or "ku-", the final emphasizing syllable "-na" is usually deleted. See the second example.

Finally, there are also three types of quantitative pronouns in Zulu, with which quantities are expressed. All the forms that have been listed in your dictionary are shown in the last four columns of Table 7. The **inclusive quantitative pronouns** (stem *-nke*) mean 'the whole' in the singular and 'all' in the plural. The **exclusive quantitative pronouns** (stem *-dwa*) mean 'alone' or 'only'. The **inclusive numeral pronouns** (only frequent stems *-bili* and *-thathu*) are used to refer to groups of items (here 'both' and 'all three' respectively).

Comments 30: For an exhaustive treatment of the lemmatization of the three types of quantitative pronouns in the OZSD, see De Schryver (2008b). In (32) to (34) below, an example each is given:

(32) **bonke** *** *inclusive quantitative pronoun*
1 ■ (cl. 2) all ◆ Bawethamela **bonke** abantu umngcwabo. *All the people attended the funeral.* **2** ■ (cl. 14) all; the whole ◆ Zilinde **bonke** ubusuku. *They waited the whole night.*

(33) **odwa** *exclusive quantitative pronoun cl. 6*
Compare WODWA ■ alone; on their/its own ◆ Wase ebona amaphoyisa asehlangana **odwa** endleleni. *And then he saw the police assembling on their own in the street.* ■ only ◆ Ukhuluma amanga **odwa** wena. *You are only telling lies.*

(34) **sobabili** *inclusive numeral pronoun 1p pl*
■ both of us ◆ Uthishanhloko ubefuna ukusibona **sobabili** emva kwesikole. *The principal wanted to see both of us after school.*
◇ [**NEG**+] **sobabili** ■ neither of us

3.10 Ideophones

Ideophones as a word class do not exist in English. In a language such as Zulu they are highly characteristic and their use signals good and beautiful Zulu. There are 100 ideophones in your dictionary. Examples are *nya*, *phecelezi*, and *qho*. Ideophones have been described as "marked words that vividly evoke sensations and perceptions". They are not problematic to look up, as no other morphemes are attached to them. However, they are **very hard to translate**. For that reason, only general paraphrases are given, between brackets and always preceded by '*of ...*'. Often, several examples are shown, to illustrate the range of possibilities. Only when combined with verbs and nouns can rather fixed meanings be given. See for example the combinations under *ngci*, *phaqa*, or *qakala*.

Comments 31: For an exhaustive treatment of the lemmatization of ideophones in the OZSD, see De Schryver (2009). In (35) and (36) below, two examples are given:

(35) **nya** ** *ideophone* ■ (*of complete absence*)
◆ Kwathula kwathi **nya**. *It was completely silent.* ◆ Wayefuna ukukhohlwa **nya** yile ndaba kaNomusa. *He wanted to forget this Nomusa affair completely.* ◆ Inyoka uyishaye wayibulala **nya**. *He hit the snake and killed it altogether.*

(36) **ngci** *ideophone* **1** ■ (*of tightness, of density*)
◆ Ithi **ngci** intuthu. *The smoke is very dense.* ◆ Abantu bacinana kuthi **ngci**. Akube kusaphefumuleka nakahle. *There is a great crowd; it is no longer possible to breathe properly.* **2** ■ (*of extreme action*) ◆ Avuke umsindo uloku uthe **ngci**. *Then he woke up while the noise continued to be overwhelming.*

◇ **-gcina ngci** ■ be the very last ◆ U-Ayanda ingane yokugcina **ngci** ekhaya. *Ayanda is the last-born at home.* ◆ Uzogcina **ngci ngci**! *It will be your very last time!*
◇ **-vala ngci** ■ close tightly ◆ Wavala ifasitela **ngci**. *She closed the window tightly.*

The description of ideophones as "marked words that vividly evoke sensations and perceptions" in the mini-grammar is taken from Dingemanse (2009). According to Blench:

> Ideophones are abundant in natural and heightened speech, notably in Africa, but absent from typical example sentences, hence their failure to be treated adequately in typical grammars and dictionaries. They can be difficult to elicit since their existence is unpredictable and speakers have no natural 'hook' to recall them. Their elusive nature, in grammatical terms, has made them poor relations to other word classes and they have been little treated by the schools of grammar dominated by syntax [...] Our understanding of the role they play in natural language (as opposed to elicited examples) is still very preliminary.
>
> — Blench (2009: 1)

As I pointed out in my own study of ideophones, compiling entries for ideophones in the OZSD "took an average three times longer than the compilation of entries in any other word class" (De Schryver 2009: 38). While the lemmatization proper of ideophones does not pose any problems (stable orthographic forms can simply be used as lemmas), I can reconfirm, with Blench, that the analysis and synthesis of large amounts of *natural language data* was simply paramount in order to make any sense at all of ideophones. Without the corpus used, in other words, it would simply not have been possible to treat ideophones adequately.

3.11 Conjunctions

Conjunctions in Zulu, as in English, introduce or link sentences. Examples are *bese*, *ukuze*, and *nxa*. Most conjunctions can be looked up directly, without the need to cut off additional prefixes. Conjunctions are very frequent. The most frequent word in Zulu is a conjunction (*ukuthi*), as well as the third-most frequent (*uma*), and the sixth-most frequent (*ngoba*).

Comments 32: In (37) below, the article for the conjunction *ukuthi* is shown:

(37) **ukuthi** *** *conjunction* ☾ -THI[1] ■ that
◆ Kwezwakala **ukuthi** useshonile. *We heard that he passed away.* ◆ Ngawatshela **nokuthi** u-Matthew wayekhona naye. *I also told them that Matthew was there as well.*
◆ Bebejatshuliswa **wukuthi** yimi owenze lo msebenzi omkhulu. *They were happy that it was I who did this great job.*

Ukuthi is one of the conjunctions to which morphemes may be prefixed. As is generally the approach in the OZSD, the top orthographic forms are illustrated in the examples. To that end, the orthographic forms cum linked corpus frequencies shown in (38) were available to the compilers. Although the frequency of *ukuthi* itself (131 950) is many times higher than the frequency of all the other forms together (5 527 + 4 552 + …), in order to illustrate the conjunctive potential for this lemma, the next few forms (here two) were also selected for illustrative purposes in (37).

(38) ukuthi <131950>, nokuthi <5527>, wukuthi <4552>, yokuthi <1517>, ngukuthi <774>, wokuthi <738>, ukuth <465>, elokuthi <315>, kungukuthi <315>, yikuthi <210>, kuwukuthi <141>, kunokuthi <131>, kwazisukuthi <90>, nawukuthi <62>, kwakungukuthi <44>

3.12 Copulatives

Copulatives in Zulu are typically **derived from words in other word classes**. To derive them, the prefixes are changed, with the goal to add the meaning '**is/are ...**'. Derived from the relative stem *-mhlophe*, for example, the frequent copulatives listed in your dictionary are *imhlophe*, *umhlophe*, and *zimhlophe*. From the relative stem *-mnandi*, the frequent copulatives *amnandi*, *imnandi*, *kumnandi*, and *zimnandi* are derived and listed. Derived from the enumerative stem *-phi*, your dictionary lists the following frequent copulatives: *(y)imuphi*, *yibaphi*, *yimiphi*, *(y)iliphi*, *yimaphi*, *(y)isiphi*, *yiziphi*, *(y)iyiphi*, *yiluphi*, and *(y)ikuphi*. And derived from the interrogative stem *-phi*, your dictionary lists the following frequent copulatives: *uphi*, *baphi*, *i(ku)phi*, *liphi*, *aphi*, *siphi*, *ziphi*, *luphi*, and *(ku)kuphi*. In each case, you can look up all these forms directly, as full words in the dictionary, where you will find precise grammatical information, including a mention of the class numbers, as well as specific examples.

Comments 33: No other dictionaries for Zulu lemmatize copulatives. The decision to consistently lemmatize words wherever possible in the OZSD, however, naturally led to their inclusion. In (39) below, an example is shown:

(39) **umhlophe** *copulative* **1** ■(cl. 1) (she/he) is white ◆ Mfowethu, uma ungibukisisa ungasho ukuthi ngimnyama, yena umlungu **umhlophe** ngempela? *My brother, if you look carefully at me, would you really say that I am black and that the white man is white?* **2** ■(cl. 3) (it) is white ◆ Umyeko wezinwele **umhlophe** wu ngubuhlalu. *The hair of the diviner is sparkling white with beads.* **3** ■(2p sg) you are white ◆ Umuhle kunezihlabathi zolwandle / **Umhlophe** kunobisi lwezimazi zakwethu [imigqa emi-2 yenkondlo] *You are prettier than the sands of the sea / You are whiter than the milk of our cows [two lines of poetry]*

3.13 Interjections

Interjections in Zulu, as in English, are words 'thrown in' to mark surprise, pleasure, amazement, disbelief, disappointment, (dis)approval, (dis)agreement, etc. Examples are *awu*, *hhayi*, and *wo*. All interjections can be looked up directly.

Comments 34: In (40) below, an example of an interjection is shown:

(40) **awu** *** *interjection* ⊂ HAWU ■ my word!; good heavens!; look!; hey!; wow!; oh!; eish! (*marks a surprise or disbelief*) ◆ Waphinde washaya izandla wathi: "**Awu**! **Awu**! **Awu**! Usuyasazi impela isiZulu." *She clapped her hands again and said: "Wow! Wow! Wow! You know Zulu very well."* ◆ **Awu** bayahlupha impela! *Eish, they are really bothersome!*

3.14 Other word classes

Smaller word classes include **enumeratives** (for example all frequent forms of *-ni*: *muni*, *mini*, *lini*, *mani*, *sini*, *zini*, *yini*, *luni*, *buni*, and *kuni*), **locative demonstrative copulatives** (for example for class 1, positions I, II and III: *nangu*, *nango*, and *nanguya*), **possessives** (see for example *alo*[2], *balo*[1], *kwalo*[2], etc.), **interrogatives** (see for example *na*, *ngabe*[2], or *yini*[2]), so-called **suffixes** (for example *jikelele*, *ndini*, or *yansondo*), **enclitics** (see for example *bo*, *-ke*[1], or *nje*[1]), **hortative markers** (*ake* and *make*), and **abbreviations** (*a.m.* and *p.m.*).

Comments 35: In (41) to (48) below, an example for each of these word classes is shown:

(41) **muni** *enumerative cl. 1, cl. 3* ■ what kind of? ◆ Muntu **muni** ongathandi ukusizwa? *What kind of person doesn't like to be helped?* ◆ U-Vincent wenza msebenzi **muni**? *What kind of work does Vincent do?*

(42) **nangu** ** *locative demonstrative copulative pos. I cl. 1* ■ here (she/he) is ◆ **Nangu** umuntu eqeda imali yabantu. *Here is the person who wasted the people's money.*

This word can also be used as a demonstrative, with the meaning "this", e.g. "**nangu** umfowethu" = "this brother of mine/ours".

(43) **alo**[2] *possessive* ⊂ LO **1** *6+1* ■ of/for this ◆ Elinye lamathemba **alo** Marion Jones owawina izindondo ezinhlanu e-Sydney, e-Australia ... *One of the hopes of this Marion Jones who won five medals in Sydney, Australia ...* **2** *6+3* ■ of/for this ◆ Amalungiselelo e-Nigeria **alo** mdlalo aphazanyiswe wukulimala kukakaputeni yayo. *Nigeria's preparations for this game were interrupted by the injury of their captain.*

(For the next three entries, no translation equivalents are available. Note how the *function* is described rather, starting with the formulaic "marks ...".)

(44) **na** *** *interrogative* ■ (*marks interrogative sentences*) ◆ Usaphila **na**? *Are you fine?* ◆ Kwenzenjani **na** khona? *What happened there?* ◆ Ungizwa kahle **na**? *Do you understand me?*

(45) **ndini** * *suffix* ■ (*marks sarcasm or contempt*)
◆ Ake uthule Ndondo **ndini**, akusiwe ophethe lo mhlangano. *Please keep quiet, smart Ndondo, you are not in charge of this meeting.* ◆ Ziphukuphuku**ndini**, niyedelela kanti. *You stupid fools, you are rude after all.*

Although "**ndini**" is a noun suffix, it is generally written separately.

(46) **bo** *** *enclitic* ■ (*marks insistence or anger*)
◆ Yicala lani-ke leli osulenzile? Khuluma **bo**! Ubulale umuntu yini? *What type of crime have you committed? Speak up! Did you kill the person or not?* ◆ Hhayi **bo**! *No, dammit!*

(47) **ake** *** *hortative marker* ■ will you (please) ... ; let ... (please) (*marks a polite request*)
◆ **Ake** uchaze isifundo esitholakala kule ndaba. *Will you please explain the lesson found in this story?* ◆ Manje **ake** sibheke le misho sibone ukuthi izenzo zakhiwa kanjani. *Now let's look at these sentences and see how the verbs are structured.*

(48) **a.m.** *abbreviation* ■ a.m.; in the morning (*any time between midnight and midday*)
◆ Inkonzo izoqala ngesikhathi sika-10:00 **a.m.** *The church service will begin at 10:00 a.m.*

Sehora le-10 ekuseni is the traditional way to express *at ten in the morning*.

4. Discussion

Despite the fact that this paper is now already twice as long as the average scientific paper, the analysis of the OZSD is not all-embracing. This could not have been otherwise, as about a thousand pages are needed to do so (cf. Section 2). It is hoped, however, that all the claims made in the OZSD's Introduction have now been sufficiently substantiated:

> The Zulu mini-grammar teaches you where to find particular words in the Zulu to English side of this dictionary, and thus teaches you *Zulu-specific dictionary skills*. It is very important that you study this section, because the method used to list Zulu words in this dictionary is new and therefore unfamiliar, but certainly more user-friendly. The result is a new type of Zulu dictionary, for the following reasons. **Complete meaningful words** have been entered in this dictionary, rather than parts of words. The selection of Zulu headwords is thus unique to this dictionary. The **modern class numbers** are used for all headwords: for nouns, of course, but also for all other word classes (parts of speech) that need to be in harmony with the nouns they refer to. **Informative cross-references** not only link verb stems with verb roots, but also derived nouns with the verbs they are derived from. In addition to headwords selected for their high frequency, this dictionary also treats all **frequent combinations** and **frequent derivations**. Headwords, combinations and derivations are illustrated with **authen-**

> **tic Zulu examples**, taken from a **large corpus** of sentences that have actually been written or spoken before. A corpus consists of hundreds and hundreds of texts, containing millions of words, that have been taken from both the general language and from school textbooks. All **core and current meanings** have been listed, based on such corpus evidence.
>
> — De Schryver (2010: xi), emphasis in bold as in original

The actual reception of the OZSD will ultimately be the litmus test for the claim that this dictionary is more user-friendly as a school dictionary than any other existing dictionary for Zulu. I had noticed (see e.g. De Schryver 2008a: 64) that the stem approach to the lemmatization of Zulu failed this particular target user group, so I set out to develop a daring word approach. In the process I also introduced the various other novelties marked in bold in the quote from the Introduction above — all of them used for the first time in Zulu lexicography. It is important to recall, however, that this new approach was specifically designed for young learners. Although I am indeed convinced that it is ideal for them, I by no means want to claim that I have solved all look-up problems for a highly conjunctive Bantu language such as Zulu. By lemmatizing all word classes except one as words, I basically bring the entire look-up problem back to recognizing and dealing with the one remaining word class, viz. verbs. Reformulated, in decoding Zulu, the one remaining orthographic word in a sentence after the easier ones will have been looked up, will be the verb. One thus also knows it must be a verb, at which point Tables 4 to 6 can be unleashed. The verb is also the only word class in the OZSD that is lemmatized with a preceding dash, indicating that something was cut off before reaching the lemma in the dictionary.[13]

Compared to stem lemmatization, word lemmatization is undoubtedly more repetitive, even though that repetition is tailored to each sense of each lemma anew. The information that is packed in a stem dictionary has been unpacked in a word dictionary. Some level of generalization is therefore missed, though one could argue that providing that is the task of a grammar, not a dictionary. Up to a point, the mini-grammar restores this, and repacks. But the mini-grammar is not complete: a section on the morphophonological (sound) changes, for example, could have been added had there been space for it. What word lemmatization does do is to put the lexicon centre-stage, with the intricacies of each word dealt with in detail. The lexicon is seen as the pivot in mastering a language, not the grammar. Grammar can be built around the lexicon. This is quite a reversal of Bloomfield's (1933: 274) view of "[t]he lexicon [as] an appendix of the grammar, a list of basic irregularities". Being able to approach Zulu words rather than Zulu roots and stems is a direct result of the corpus revolution. So are most other microstructural innovations, chief among those the authentic examples to illustrate the synthesised analysis.

Abandoning generalizations, in combination with the extra information categories offered in the OZSD, also means that the focus shifts from quantity to quality. Rather than offering 30 000 lemmas on about 920 pages as in Doke and Vilakazi's *Zulu–English Dictionary* (1953[2]), or 13 600 lemmas on about 220

pages as in Dent and Nyembezi's *Scholar's Zulu Dictionary* (1995[3]), the OZSD offers 5 000 lemmas on about 270 pages.[14] The OZSD does fill a gap in the market, however, both in terms of its user-friendly access structure, and in terms of its coverage. Microstructurally, grammar was in effect brought into each dictionary article, systematic exemplification is a first for Zulu, and of course the lexicon was brought in line with current usage: meaning shifts resulted in new meanings which are now recorded, and hundreds of 'new' words were added to the macrostructure.

This paper comes after the lemmatization of some Zulu word classes have already been treated in the literature, and it is hoped that it can serve as a launching pad for all the remaining ones. For those treated before (possessive pronouns, adjectives, quantitative pronouns, and ideophones) comparing the tiny summaries presented in the mini-grammar (respectively Sections 3.9, 3.6, 3.9, and 3.10 above) gives an idea of how a scientific description may be presented to the dictionary user. Conversely, the summaries for the other word classes give an idea of what the scientific descriptions will conclude. Concluding is one level, going through the details another, and performing the actual corpus analysis — in order to first synthesize the facts to then compile the dictionary articles themselves — is yet another, far more complex, level. In order for a thriving Bantu metalexicography to develop, more dictionaries will have to be compiled for these languages, so that more dictionary compilers will be able to share their experiences. As such, the present contribution is but one such attempt, hopefully one that will stimulate debate.

Endnotes

1. A workbook to accompany the Zulu dictionary is being planned as well.
2. While compiling the Northern Sotho dictionary took slightly over two and a half years, for the Zulu one nearly four years were required. From a lexicographic point of view, a conjunctively-written language such as Zulu is indeed much more complex to handle than a disjunctively-written one such as Northern Sotho.
3. With GET the general education and training band. Grades 10 to 12 (the last three years of secondary education) are known as FET, the further education and training band.
4. Together, the two A-to-Z sections and all extra-matter texts add up to 640 pages, or thus in the book-binding jargon, twenty 32-page signatures (20 x 32 = 640). Although tweaking the different sections, as well as the contents of those sections, until such a round multiple is reached is an important aspect of finalising an actual dictionary for the trade, this aspect will not be covered here.
5. Sincere thanks are due to A. Wilkes and D. Gowlett for their critical evaluation of the mini-grammar, and to M. Hall for making sure the English of the mini-grammar is on the level of the intended target user group.
6. Verb lemmas, even as roots or stems, are actually also orthographic words, but only in their imperative forms. As such, one could say that the OZSD managed to lemmatize *all* Zulu words as words. However, given verbs rarely occur in their imperative forms, it is more cor-

rect to stick to the traditional terminology, and to admit that verbs are lemmatized as roots and stems.

7. For more information on the preparation of a tailored lemma-sign list for junior dictionary users, balancing a general and a customised sub-corpus, see De Schryver and Prinsloo (2003).
8. The contract to produce the OZSD was signed between Oxford University Press Southern Africa and TshwaneDJe HLT, makers of the dictionary production system TLex.
9. Ideally, one would also have known which Zulu textbooks the learners would be using, so as to avoid the possibility of a diverging terminology, but in the absence of that knowledge, the best that could be done was to define each term and to use it accordingly in the OZSD.
10. Considerable thought also went into the way to present the information shown in Table 1. One adjudicator in particular, suggested using dedicated columns for mono- vs. polysyllabic stems, additional columns for vowel- vs. consonant-initial stems, and then another column for the remainder of the notes now in the last column. I feel that a single notes column, with numbered notes referring to the earlier columns, is sufficiently clear.
11. In an electronic environment, where space is not an issue, lemmatizing all frequent verb forms becomes a possibility, either with the help of unlimited human resources or computational techniques, but at that point other problems surface, such as the abandonment of important generalizations (cf. De Schryver 2008: 269-270).
12. For a brief analysis of Mbatha's (2006) dictionary, and its implications, see De Schryver and Wilkes (2008: 829-830).
13. There are only seven exceptions in the OZSD: the conjunctions *-thi* and *-the* (unique in that these two only can take verbal prefixes), the relative stems *-mbumbulu*, *-thize*, and *-thizeni* (as these three are always used in compounds), the relative / copulative *-emqoka* (the only such combined word class), and the enclitic *-ke* (always attached to another word, with dash).
14. Extrapolated lemma-sign counts for Doke and Vilakazi as well as Dent and Nyembezi are taken from De Schryver (2009: 52).

References

Bennett, P.R. 1986. Grammar in the Lexicon, Two Bantu Cases. *Journal of African Languages and Linguistics* 8(1): 1-30.

Benson, T.G. 1964. A Century of Bantu Lexicography. *African Language Studies* 5: 64-91.

Beuchat, P.-D. 1963. A Restatement of the Zulu Verb Conjugation. Part 1. *African Studies* 22(4): 137-169.

Beuchat, P.-D. 1964. A Restatement of the Zulu Verb Conjugation. Part 2. *African Studies* 23(1): 35-49.

Beuchat, P.-D. 1964a. A Restatement of the Zulu Verb Conjugation. Part 3. *African Studies* 23(2): 67-87.

Beuchat, P.-D. 1966. The Perfect Tenses in Zulu. *African Studies* 25(2): 61-71.

Blench, R. 2009. *The Sensory World; Ideophones in Africa and Elsewhere*. Draft circulated for comment; 5 October 2009 [online]. http://www.rogerblench.info/Language%20data/Africa/General/Blench%20African%20Ideophones.pdf

Bloomfield, L. 1933. *Language*. New York: Henry Holt & Co.

Dent, G.R. and C.L.S. Nyembezi. 1995[3] [1988[2], 1969]. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.

De Schryver, G.-M. 1999. *Bantu Lexicography and the Concept of* Simultaneous Feedback*, Some Preliminary Observations on the Introduction of a New Methodology for the Compilation of Dictionaries with Special Reference to a Bilingual Learner's Dictionary* Cilubà-Dutch. Unpublished MA Dissertation. Ghent: Ghent University.

De Schryver, G.-M. 2001. *Pukuntšutlhaloši ya Sesotho sa Leboa 1.0* (PyaSsaL's First Parallel Dictionary). Pretoria: (SF)² Press.

De Schryver, G.-M. 2004. *Concepts and Tools for Lexicography in the Electronic Age — A Case Study of Dictionary Compilation in South Africa*. Unpublished PhD Thesis. Ghent: Ghent University.

De Schryver, G.-M. 2006. Compiling Modern Bilingual Dictionaries for Bantu Languages: Case Studies for Northern Sotho and Zulu. Corino, E., C. Marello and C. Onesti (Eds.). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6-9 settembre 2006 / Proceedings XII Euralex International Congress, Torino, Italia, September 6th-9th, 2006*: 515-525. Alessandria: Edizioni dell'Orso.

De Schryver, G.-M. 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.

De Schryver, G.-M. 2008. Why Does Africa Need Sinclair? *International Journal of Lexicography* 21(3): 267-291.

De Schryver, G.-M. 2008a. A New Way to Lemmatize Adjectives in a User-friendly Zulu–English Dictionary. *Lexikos* 18: 63-91.

De Schryver, G.-M. 2008b. The Lexicographic Treatment of Quantitative Pronouns in Zulu. *Lexikos* 18: 92-105.

De Schryver, G.-M. 2009. The Lexicographic Treatment of Ideophones in Zulu. *Lexikos* 19: 34-54.

De Schryver, G.-M. 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi, Esishicilelwe abakwa-Oxford*. Cape Town: Oxford University Press Southern Africa.

De Schryver, G.-M. and R. Gauton. 2002. The Zulu Locative Prefix ku- Revisited: A Corpus-based Approach. *Southern African Linguistics and Applied Language Studies* 20(4): 201-220.

De Schryver, G.-M. and N.S. Kabuta. 1997. *Lexicon Cilubà–Nederlands, Een circa 2500-lemma's-tellend strikt alfabetisch geordend vertalend aanleerderslexicon met decodeer-functie ten behoeve van studenten Afrikaanse Talen & Culturen aan de Universiteit Gent*. Recall Linguistics Series 1. Ghent: Recall.

De Schryver, G.-M. and D.J. Prinsloo. 2001. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots — with Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001. *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.

De Schryver, G.-M. and D.J. Prinsloo. 2003. Compiling a Lemma-sign List for a Specific Target User Group: The Junior Dictionary as a Case in Point. *Dictionaries: Journal of The Dictionary Society of North America* 24: 28-58.

De Schryver, G.-M. and A. Wilkes. 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 827-836. Sèrie Activitats 20. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Dingemanse, M. 2009. *How to Do Things with Ideophones: Observations on the Use of Vivid Sensory Language in Siwu*. Presentation at the SOAS Research Seminar, June 3. London: The School of Oriental and African Studies.

Doke, C.M. and B.W. Vilakazi. 1953[2] [1948]. *Zulu–English Dictionary*. Johannesburg: Witwatersrand University Press.

Gauton, R., G.-M. de Schryver and L. Mohlala. 2004. A Corpus-based Investigation of the Zulu Nominal Suffix -kazi: A Preliminary Study. Akinlabi, A. and O. Adesola (Eds.). 2004. *Proceedings of the 4th World Congress of African Linguistics, New Brunswick 2003*: 373-380. Cologne: Rüdiger Köppe Verlag.

Mbatha, M.O. 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers.

Nabirye, M. 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished MA Dissertation. Kampala: Makerere University.

Nabirye, M. 2009. *Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwega Olusoga* [A Dictionary of Lusoga. For Speakers of Lusoga, and for Those Who Would Like to Learn Lusoga]. Kampala: Menha Publishers.

Nabirye, M. 2009a. Compiling the First Monolingual Lusoga Dictionary. *Lexikos* 19: 177-196.

Nabirye, M. 2009b. Dictionary Compilation for Mother-tongue Speakers of Bantu Languages. Zhu, R. (Ed.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary & Lexicology*: 597-607. Beijing: Beijing Normal University.

Nkabinde, A.C. 1975. *A Revision of the Word Categories in Zulu*. Unpublished PhD Thesis. Pretoria: UNISA.

Prinsloo, D.J. 2009. Current Lexicography Practice in Bantu with Special Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.

Prinsloo, D.J. 2010. Review: *The Oxford Bilingual School Dictionary: Zulu and English*. *Lexikos* 20: 760-766.

Prinsloo, D.J. and G.-M. de Schryver. 2002. Towards an 11 x 11 Array for the Degree of Conjunctivism / Disjunctivism of the South African Languages. *Nordic Journal of African Studies* 11(2): 249-265.

Schadeberg, T.C. 1992[3] [1984[2]]. *A Sketch of Swahili Morphology*. Grammatical Analyses of African Languages 2. Cologne: Rüdiger Köppe Verlag.

Snoxall, R.A. 1965. Some Problems and Principles of Lexicography in Luganda. *African Language Studies* 6: 27-31.

Taljaard, P.C. and S.E. Bosch. 1993[2] [1988]. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.

Taljard, E., G.-M. de Schryver, M. Mogodi and D. Paizee. 2008. *Oxford Bilingual School Dictionary Workbook: Northern Sotho and English / Puku ya Mešomo: Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.

TLex (formerly TshwaneLex). 2010. TshwaneLex Dictionary Production System [online]. http://tshwanedje.com/tshwanelex/

Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.