A New Way to Lemmatize Adjectives in a User-friendly Zulu–English Dictionary

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

Abstract: Traditionally, Zulu adjectives have been lemmatized under their stems only. In this research article, an in-depth analysis is undertaken to make a case for the lemmatization of all frequent adjectival forms *with their adjective concords* rather. It is shown that the supposed explosion in size of the dictionary may be contained within a corpus-driven Sinclairian framework. The advantages of such a word-like treatment far outnumber the generalizations that have hitherto characterized the lexicographic treatment of adjectives in Zulu. The study is supported by ample dictionary extracts from a Zulu–English dictionary project aimed at junior users. Comparisons with existing dictionaries and textbook data are also made.

Keywords: LEXICOGRAPHY, LINGUISTICS, GRAMMAR, DICTIONARY, BILINGUAL, CORPUS, LEMMATIZATION, FREQUENCY, ZULU (ISIZULU), ENGLISH, ADJECTIVE, ADJECTIVE STEM, QUALIFICATIVE ADJECTIVE, COPULATIVE ADJECTIVE, USER-FRIENDLY, REAL EXAMPLE, COLLOCATION, COMBINATION, DERIVATION, IDIOMATIC USE, SEMANTIC PROSODY

Samenvatting: Een nieuwe manier om adjectieven te lemmatiseren in een gebruiksvriendelijk Zoeloe–Engels woordenboek. Traditioneel worden adjectieven in Zoeloe enkel onder hun stam gelemmatiseerd. In dit onderzoeksartikel wordt een grondige analyse uitgevoerd met het oog op de invoering van een nieuwe methode waarbij alle frequente adjectieven *met hun adjectiefschakel* in het woordenboek worden geplaatst. Er wordt aangetoond dat de vooronderstelde explosie in grootte van het woordenboek beperkt kan worden binnen een corpusgedreven Sinclairiaans kader. De voordelen van zo een woordachtige behandeling overstijgen ruimschoots de veralgemeningen die totnogtoe de lexicografische behandeling van adjectieven in Zoeloe hebben gekarakteriseerd. De studie wordt ondersteund door een groot aantal passages uit een Zoeloe–Engels woordenboekproject gericht op jonge gebruikers. Vergelijkingen met bestaande woordenboeken, alsook handboeken worden ook gemaakt.

Sleutelwoorden: LEXICOGRAFIE, LINGUÏSTIEK, GRAMMATICA, WOORDENBOEK, TWEETALIG, CORPUS, LEMMATISATIE, FREQUENTIE, ZOELOE, ENGELS, ADJECTIEF, ADJECTIEF STAM, KWALIFICEREND ADJECTIEF, COPULATIEF ADJECTIEF, GEBRUIKSVRIENDELIJK, ECHT VOORBEELD, COLLOCATIE, COMBINATIE, AFLEIDING, IDIOMATISCH GEBRUIK, SEMANTISCHE PROSODIE

Lexikos 18 (AFRILEX-reeks/series 18: 2008): 63-91

1. From Bloomfield to Sinclair via Doke

Half a century ago, two excellent dictionaries for Zulu appeared, viz. Doke and Vilakazi's (1953) *Zulu–English Dictionary*, and Doke, Malcolm and Sikakana's (1958) *English–Zulu Dictionary*. The coverage, detail and meticulousness of these two dictionaries are of such a high standard that they had the ironic effect of stalling all future lexicographic efforts for Zulu. Indeed, to this date not a single dictionary for Zulu — whether bilingual or monolingual — has been compiled that comes even close to the quality of Doke's pair of dictionaries. Doke's pair remains the standard against which all current Zulu dictionaries are compared, and will likely remain the standard for many years to come.

In Doke and Vilakazi's Zulu to English dictionary, the so-called 'stem approach' to lemmatization is used, meaning that (a section of) the Zulu lexicon is grouped around word stems. The multitude of (often stacked) prefixes, suffixes and circumfixes which characterize a conjunctively written language such as Zulu have thus been cut off, with (supposed) meanings assigned to the resulting (extracted) stems. For linguists such an approach is arguably a magnificent and efficient lemmatization approach; for the average user it is problematic.

For about a decade now, we have informally observed the use of this Zulu dictionary at university level as well as within different language services of various government departments. We have noticed that, on average, as many as two look-up procedures are required before a user also finds what he/she is looking for. The main reason for this is not so much the result of inconsistencies in the lemmatization proper, but simply because a large amount of grammatical knowledge is presupposed before one can successfully consult this dictionary. This is valid for both decoding (receptive) and encoding (active) use, and for learners as well as mother-tongue speakers. Two random, straightforward examples follow to illustrate these points.

Zulu nouns in the gender 9/10 have the noun class prefixes *iN-* for the singular (class 9), and *iziN-* for the corresponding plural (class 10) — with *N* a nasal, i.e. *n* or *m*. A user of a stem-based dictionary may conclude that 9/10 nouns are lemmatized under the nasal *N*. So when wishing to look up, say, *indlovu/izindlovu* 'elephant/elephants' this user will go to the alphabetic stretch **N**. In this case, however, these words cannot be found there, as Doke realized that the stem here is not *-ndlovu*, but rather *-dlovu*, calling in the Ur-Bantu form of this noun stem (*-γoγû*) to substantiate this. Neither learners nor mother-tongue speakers, however, can be expected to be versed in comparative or historical Bantu linguistics, so the finer points of Doke's lemmatization approach are entirely lost on all but a few of the most ardent users.

As an example to illustrate the encoding use of a Zulu dictionary, consider the ordinal 'fourth'. When used neutrally (as in 'she came fourth'), the form is *isine*; while a possessive concord needs to be prefixed to this form for definite uses (as in 'the fourth quarter'), resulting in forms such as *yesine*, *wesine*, *lesine*,

sesine, etc. In Doke, one needs to look up all these forms under *-ne* (the reasoning being that these forms are derived from the adjective stem *-ne* 'four'), but under *-ne* the differing ordinal uses (neutral vs. definite) are not stated explicitly. Linguists, of course, will see nothing wrong with this, as they will refer the dictionary user to the grammar for the actual *use*.

One solution is indeed to dissociate the grammar from the lexicon, recalling Bloomfield (1933: 274): 'The lexicon is really an appendix of the grammar.' At this point one could focus on, say, just nouns and verbs in a dictionary, and relegate all other word classes to the grammar. If this sounds too far-fetched, consider the latest monolingual dictionary for Zulu, *Isichazamazwi sesiZulu* (Mbatha 2006). In this dictionary's front matter, one reads that (a) only content words belong in a dictionary, and that (b) this means only four word classes are recognized in Zulu: noun, verb, exclamation or interjection, and ideophone. Probably realizing that this proposition is untenable, the compilers somehow 'forced' meanings onto extremely low-frequent to non-existing verb and noun stems. As such, one for instance finds the noun *í(li)***nîngi** 'the majority' but not the adjective stem *-ningi* 'much/many'. Likewise, the extremely-low-frequent noun *í***mpéla** 'the real one' — which is mostly used in possessive constructions, at which point it is a possessive — is found instead of the highly-frequent adverb *impela* 'really'.[1]

Even though there are days on which the prospect surfaces to 'get rid of' all lemmatization and presentation problems in Bantu lexicography by this means, it is exactly the lexicographer's task not to give in here. Indeed, no sooner has one finished contemplating Bloomfield than Sinclair (1966: 422-423) must be considered:

> We speak casually about 'fully grammatical items' or 'function words' as if there were items which were entirely irrelevant in the study of lexis. … Every morpheme in a text must be described both grammatically and lexically … Each successive form in a text is a lexical item or part of one, and there are no gaps where only grammar is to be found.

2. A user-friendly Zulu dictionary: mission statement

Against the background sketched in Section 1, a new type of (bilingual) Zulu dictionary has been envisaged, one which would also and for the first time be pitched at the level of junior users. The mission statement for this project has been described by De Schryver and Wilkes (2008: 831) as follows:

> An approach which cuts down to the smallest morpheme level (as in Doke & Vilakazi) is user-unfriendly for the target user group envisaged, while an approach which throws out most word categories, and forces so-called core Zulu meanings onto the remaining section (as in Mbatha) is even more user-unfriendly. While the former is linguistically sound, the latter moreover is not.
>
> The user-friendly approach/solution advocated here revolves around two notions: (a) except for verbs and a few exceptions (such as the conjunction *-thi*

(when), which behaves like a verb), all items from all word classes can be lemmatised *with their primary prefix(es)* included, as well as *with their suffixes* included; (b) overall *corpus frequencies* may be used in order to make a decision on the number of prefixes as well as which prefixes to include for each word class as a whole, and thus on how to organise/lemmatise the lexicon.

Implicit in this mission statement is that one has access to a large Zulu corpus, that one has a procedure to lemmatize this corpus (while keeping track of all individual as well as summed and overall corpus frequencies), and that one has a clear approach to the lexicographic treatment of each and every Zulu word class. Critically analyzing each of these aspects is a massive undertaking, one that cannot be achieved within the ambit of just one research article. The current contribution, therefore, is one in a series.

At face value one would have thought that the logical starting point would have been to discuss macrostructural aspects, and thus to defend the creation of an entire user-friendly lemma-sign list which is word-like rather than stem-like. However, to truly appreciate this effort, it was found that it is more advantageous to analyze the lexicographic treatment of selected Zulu word classes first, and only then to turn to the full macrostructure. As such, De Schryver and Wilkes (2008) concentrated on the treatment of the *possessive pronouns* in a user-friendly Zulu–English dictionary, in this article the focus is on the treatment of *adjectives* in such a dictionary, and in De Schryver (2008a) the focus will be on *quantitative pronouns*.

In order to pick up the thread started in Section 1, and before analyzing the adjectives themselves, the extracts below compare the entries for 'elephant/elephants' in Doke (1)(a) with those in a projected user-friendly Zulu–English dictionary (1)(b).

(1)(a) **-dlovu** (**indlovu,** 2.9.9, **izindlovu**) n. [< dlóvu; Ur-B. ***-γoγû***. > umdlovu; indlovudalana; indlovudawana; indlovukazi; indlovunda; indlovundwane.

1. Elephant. *Indlovu iwile, ziphelele zonke izizwe ziye kuxephula kuyo* (The elephant has fallen, and every single one from the tribes has gone to pull off a bit from it; i.e. where the carcass is there will the vultures be gathered together). *Indlovu idla aßasondezeli* (The elephant eats up those who go too near; i.e. don't play with fire). *indlovu enesihlonti* (the elephant with a burning torch — used in *izißongo zikaMßuyaze*).

2. term used of a very stout person.

(1)(b) **indlovu** ** *noun* ***9/10*** [*pl.* **izindlovu**] ► **elephant** ♦ Nansi-ke inganekwane elandwa nguNanana. Kwasukela kwathi indlovu ilambile yahamba ifuna ukudla. • *Here is the folk tale told by Nanana. Once upon a time, there was a hungry elephant who went looking for food.*

izindlovu *plural noun 9/**10*** See singular **indlovu**

As may be seen from (1)(b), and in contrast to (1)(a), nouns are lemmatized with (and may be found under) their full noun class prefixes, with cross-refer-

ences from the plural to the singular forms.

Extracts (2)(a) and (2)(b) show 'fourth' in the same two dictionaries.

(2)(a) **-ne** (**isine,** 3.2.9, **izine**) n. [< adj. **-ne**.]
The fourth place, the fourth. *usuku lwesine* (the fourth day); *ngokwesine* (fourthly).

(2)(b) **isine** * *adverb* ► **fourth** *(used neutrally)* ♦ UVeli uphume isine ku-5000 m. • *Veli came fourth in the 5000 metres.*
◊ **[PC +]isine** ► **fourth** *(used definitely)* ♦ Uginqwe phansi ngomzuliswano wesine. • *He was knocked down in the fourth round.*

The information given under (2)(b) is more explicit — 'spelled out' even — compared to (2)(a). Grammatical guidance is not shunned, and is offered there where the dictionary user will most likely need it (compare this with Sinclair's observation). Here '[PC +]' stands for any prefixed possessive concord. The number of such codes the dictionary user should master has been kept to an absolute minimum.[2]

A lot more can be said about the lemmatization of the word classes (nouns and adverbs) used as illustrations here, but this will be done in forthcoming studies. Important to note, however, is that all the data shown in (1)(b) and (2)(b) is corpus-driven. The selection of the lemma signs, for instance, is based on overall corpus frequencies, with the top 500 lemmas marked with three stars (***), the next 500 with two stars (**), and the third 500 with one star (*). Meanings have been 'mapped onto use' as seen in the corpus (Hanks 2002). These meanings were then ordered according to individual frequencies and translated into English. Needless to say, the Zulu examples are 'real' (Fox 1987) because they are extracts from the Zulu corpus. For a detailed discussion of the use of this Sinclairian apparatus to dictionary making for the Bantu languages, the reader is referred to De Schryver (2008).

3. True adjective stems in Zulu

Bantu languages have about twenty to thirty so-called 'true adjective stems', and in most existing Bantu dictionaries these are (a) simply (and only) lemmatized as stems, (b) given a basic (or generic) meaning, and, for the larger dictionaries, (c) exemplified with one or more (often invented) phrases. Given Zulu's conjunctive writing system, the required agreement morphemes — known as adjective concords (ACs) — are physically attached to the front of these stems. In such dictionaries, it is thus left to the dictionary user to consult a grammar in addition, where information must be sought on the form and use of the adjective concords, as well as on the morphophonological rules (i.e. sound changes) applicable when attaching an adjective concord to an adjective stem. It is further also assumed that the dictionary user will be able to adapt the meaning depending on class membership of the noun that is being described.

In line with the mission statement presented in Section 2, our claim is that the lemmatization of adjective stems *with their adjective concords* will result in a more user-friendly dictionary. At face value, this may look like a waste of space and resources, as instead of, say, just 25 dictionary articles for adjectives, one will end up with 20 x 25 or thus 500 articles (assuming 16 classes, plus first and second persons). We will come back to this explosion of orthographic forms in Section 4.

At this point, it is instructive to look at the lemmatization of adjectives in a desktop dictionary for Zulu, and to compare the coverage found there with the list of adjectives in a standard Zulu textbook.

Table 1: Adjectives in Zulu: Textbook vs. corpus vs. dictionary data (w*ith T = textbook (Taljaard and Bosch 1993: 99); Z-E = user-friendly Zulu–English dictionary; Freq. = lemmatized corpus frequency (in 8.5 million words); Lemma sign, POS and Translation equivalent(s) as in Dent and Nyembezi's (1995) dictionary*)

T	Z-E	Freq.	Lemma sign	POS	Translation equivalent(s)
✓	✓	5 192	**-bi**	(adj.)	ugly; bad; evil.
✓	✓	11 667	**-bili**	(adj)	two.
✓	✓	4 937	**-dala**	(adj)	old; aged.
-	-	9	**-daladala**	(adj)	ancient; very old.
✓	✓	4 801	**-de**	(adj)	long; tall; high; deep.
-	-	4	**-fisha**	(adj)	short.
-	*x-ref*	439	**-fishane**	(adj)	short.
✓	-	18	**-fuphi**	(adj)	short.
-	-	1	**-fusha**	(adj)	short.
✓	✓	534	**-fushane**	(adj)	short.
✓	✓	1 921	**-hlanu**	(adj)	five.
✓	✓	10 534	**-hle**	(adj)	good; beautiful; pretty.
✓	✓	18 216	**-khulu**	(adj)	large; great.
✓	✓	6 875	**-ncane**	(adj)	small; few; young. *kwaba kuncane indawo* — keen competition; outcome difficult to predict.
✓	-	69	**-nci**	(adj)	very small; minute.
✓	-	21	**-ncinyane**	(adj)	very small.
-	-	3	**-ncu**	(adj)	minute; very small.
✓	✓	2 818	**-ne**	(adj)	four.
✓	✓	690	**-ngaki**	(adj)	how many?
-	*der*	*cf. next*	**-ningana**	(adj)	quite a fair number; not too few.
✓	✓	16 515	**-ningi**	(adj)	many; much.
-	-	29	**-ninginingi**	(adj)	numerous.
✓	✓	56 971	**-nye**	(adj)	some; other.
✓	✓	6 132	**-sha**	(adj)	new; young.
✓	✓	5 338	**-thathu**	(adj)	three.
		153 734			

The last three columns of Table 1 list all the adjective stems, 25 in all, as well as their lexicographic treatment, found in the Zulu to English side of Dent and Nyembezi's (1995) *Scholar's Zulu Dictionary*. Of these 25 adjective stems, 7 have not been mentioned in Taljaard and Bosch's (1993) *Handbook of isiZulu*, namely the two reduplicated stems *-daladala* (< *-dala*) and *-ninginingi* (< *-ningi*), the

derived stem *-ningana* (< *-ningi*), and the variants *-fusha*, *-fisha*, *-fishane* (~ *-fushane*) and *-ncu* (~ *-nci*). Looking at summed 'lemmatized corpus frequencies', this is defendable, except for *-fishane* and *-ningana*, which are frequent and should have been mentioned. Conversely, these same frequencies also indicate that the forms *-fuphi*, *-nci* and *-ncinyane* are infrequent, so these adjective stems could have been left out as well. This pattern, whereby some frequent forms of a closed set of items are missing while infrequent ones are mentioned instead, is often encountered in textbooks not based on corpus data.[3] The first three columns in Table 1 summarize these statistics.

4. Using a corpus to map adjectives onto a user-friendly Zulu dictionary

Given the importance of a corpus within a Sinclairian approach to dictionary making, a few words about the corpus used for this study are necessary. A Zulu corpus totalling 8.5 million running words (tokens) was built, much along the lines described in De Schryver and Gauton (2002: 202-203). A corpus of this size contains a massive 800 000 unique orthographic words (types), of which the top 20 000 were lemmatized. This section represents roughly 70% of the tokens in the Zulu corpus.[4] Lemmatized corpus frequencies in this article therefore represent the summed frequencies of all items brought together during lemmatization. To complete some of the tables in this article, also lower corpus frequencies are shown (and counted).

In Table 1, one sees that all the lemmatized corpus frequencies together represent about 150 000 running words. Expressed as a percentage of the tokens used for this study, this corresponds to roughly 2.5% of these tokens. Reformulated, this article — which deals with the adjectives in Zulu — is a lexicographic study of about 2.5% of the Zulu lexicon. Conversely, this also means that an average of 2.5 adjectives for each 100 words is used in any spoken or written Zulu.

For the envisaged user-friendly Zulu–English dictionary, the idea is to describe the most frequent 5 000 lemmas only (5 000 in Zulu, and 5 000 in English). The minimum frequency of each Zulu orthographic form before lemmatization was 42, after lemmatization this figure climbed to 50. In other words, the lemmatized corpus frequency must be at least 50 for any Zulu lemma to be considered for inclusion. Applied to the adjectives, one obtains the data shown in Table 2. In this table, the top row lists the various Zulu class numbers as well as the first and second persons singular and plural, 20 in all, while the first column lists the same 25 adjective stems from Table 1.

The ticks (✓) in Table 2 indicate that of the 500 candidate adjectives to be lemmatized, only 160 are left. (Note that the line at *-ningana* was left blank, cf. Section 5.1 below.) Furthermore, given the adjective concords for classes 1 and 3 (and the 2nd person singular) are equal — namely *om(u)-*, as well as those for classes 8 and 10 — namely *eziN-*, and for classes 15 and 17 — namely *oku-*, these 160 collapse to just 126 articles from the point of view of the number of diction-

ary articles. Within a corpus-driven framework, therefore, the explosion of truly important adjectives is not necessarily so dramatic.

Table 2: Adjectival forms in a user-friendly Zulu dictionary with 5 000 lemmas (*with Adj. = adjective stem; Cl. = noun class number and 1st and 2nd persons*)

Adj. ⇩ Cl. ⇨	1	2	3	4	5	6	7	8	9	10	11	14	15	16	17	18	1sg	1pl	2sg	2pl
-bi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-bili	-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	✓	-	-	-	-	-
-dala	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	-	-
-daladala	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-de	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-
-fisha	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-fishane	✓	-	✓	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-
-fuphi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-fusha	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-fushane	✓	-	✓	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-
-hlanu	-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-
-hle	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-khulu	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-ncane	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	✓	-	-	-
-nci	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
-ncinyane	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-ncu	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-ne	-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-
-ngaki	-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-
-ningana																				
-ningi	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	-	-	-	-
-ninginingi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-nye	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-sha	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-thathu	-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-

5. Advantages of lemmatizing word-like adjectives rather than stems

Observe that 126 entries for adjectives out of a total of 5 000 dictionary articles, corresponds to 2.5% of the total. A word-based approach to the lemmatization of adjectives (in contrast to the traditional stem-based approach) thus also gives a far better reflection of the distribution of the lexicon: Zulu speech and text contains 2.5% adjectives; the number of articles for adjectives in a user-friendly Zulu dictionary is also 2.5%. This finding, of course, is a kind of self-fulfilling prophesy.

There are a number of additional advantages to lemmatizing adjectives with their adjective concords; the main ones are discussed in the next four sections. Each of these sections is accompanied by detailed corpus statistics and star ratings, aimed at shedding further light on the soundness of lemmatizing word-like adjectives. In order not to overload the tables that follow, the ticks (✓) from Tables 1 and 2, which indicated the presence of certain forms, are replaced with the background shading of the corresponding cells ( ).

5.1 On varying semantics and diminutives

A first semantic aspect that is lost when one lists adjective stems only, with one overarching meaning (as for instance seen in Table 1), is the different meanings some singular vs. plural forms take.[5] This is the case for all the adjectives shown in Table 3, and is illustrated for *-nye* in (3).

Table 3: Adjectives expressing 'another/other'; 'much/many'; and 'small/few'

Cl.	AC	-nye	-ningi	-ninginingi	-ncane	-nci	-ncinyane	-ncu
1;3; 2sg	om(u)-	omunye 9 103 ***	omningi 168	omninginingi 0;1;0	omncane 1 460 **	omunci 0	omncinyane 3;1	omuncu 0;1;0
2	aba-	abanye 12 908 ***	abaningi 5 267 ***	abaninginingi 1	abancane 735 *	abanci 0	abancinyane 7	abancu 0
4	emi-	eminye 2 211 ***	eminingi 1 477 **	emininginingi 6	emincane 79	eminci 0	emincinyane 1	emincu 0
5	eli-	elinye 4 375 ***	eliningi 46	elininginingi 0	elincane 465 *	elinci 0	elincinyane 1	elincu 0
6	ama-	amanye 4 358 ***	amaningi 1 819 **	amaninginingi 5	amancane 582 *	amanci 0	amancinyane 0	amancu 0
7	esi-	esinye 3 004 ***	esiningi 426	esininginingi 0	esincane 513 *	esinci 0	esincinyane 1	esincu 0
8; 10	eziN-	ezinye 8 418 ***	eziningi 4 513 ***	ezininginingi 10	ezincane 769 *	ezinci 0	ezincinyane 0	ezincu 0
9	eN-	enye 5 569 ***	eningi 627 *	eninginingi 1	encane 1 527 **	enci 0	encinyane 4	encu 2
11	olu-	olunye 1 322 **	oluningi 82	oluninginingi 0	oluncane 216	olunci 69	oluncinyane 1	oluncu 0
14	obu-	obunye 118	obuningi 47	obuninginingi 0	obuncane 15	obunci 0	obuncinyane 0	obuncu 0
15; 17	oku-	okunye 5 585 ***	okuningi 2 039 **	okuninginingi 1;4	okuncane 427	okunci 0	okuncinyane 1;0	okuncu 0
1sg	engim(u)-	engimunye 0	engimningi 0	engimninginingi 0	engimncane 84	engimunci 0	engimncinyane 1	engimuncu 0
1pl	esiba-	esibanye 0	esibaningi 4	esibaninginingi 0	esibancane 1	esibanci 0	esibancinyane 0	esibancu 0
2pl	eniba-	enibanye 0	enibaningi 0	enibaninginingi 0	enibancane 2	enibanci 0	enibancinyane 0	enibancu 0
	Freq.	**56 971**	**16 515**	**29**	**6 875**	**69**	**21**	**3**

(3) **esinye** *** *adjective cl. 7* ► **another (one)** ♦ Isigcino umfana lowo waba yinkosi yesizwe esinye. • *Eventually that boy became the king of another nation.*

ezinye *** *adjective cl. 8, cl. 10* ► **other(s)** ♦ Nazi ezinye izibonelo. • *Here are other examples.* ♦ Zathi ezinye izingane ushaywe nguMdingi. • *Other children said he was hit by Mdingi.*

For classes 15 (the infinitive class) and 17 (the locative class, with 16, 17 and 18 all collapsed into 17) the meaning often deviates even further, as may be seen when comparing (4) with (3).

(4) **okunye** *** *adjective* **1** *cl. 15* ► **some other**; **certain** ♦ Izangoma okunye ukudla azikudli, ziyakuzila. • *Diviners do not eat certain foods, they abstain from them.* **2** *cl. 17* ► **something else** ♦ Okunye okubalulekile yindlela umlobi abhala ngayo. • *Something else that is important is the way the writer writes.* ♦ OkukaMagwababa kwaba okunye ngoba yena akadonswanga muntu. • *In Magwababa's case it was something else, because* he *was not dragged by anyone.*

For *-ncane* one core meaning is present for all classes, but for the plural classes corpus evidence points to an additional meaning. Compare (5) with (6).[6]

(5) **omncane** ** *adjective cl. 1, cl. 3* ► **small**; **young**; **little** ♦ Uma unezinyawo ezivuvukele uthele usawoti omncane emanzini uma ugeza. • *If you have swollen feet, pour a little salt in the water when you take a bath.* ♦ Indoda yayinomkhaba omncane. • *The man had a small protruding stomach.*

(6) **abancane** * *adjective cl. 2* **1** ► **small**; **young**; **little** ♦ Izidakamizwa ziyababulala abadlali abancane. • *Drugs are destroying the young players.* ♦ Uma ungumqeqeshi wabadlali abasebancane kufanele ube nesineke. • *If you are a coach of players who are still young, you should be patient.* **2** ► **a few**; **a small number** ♦ Siyazi ukuthi unabalandeli abancane. • *We know that she has a few followers.*

Whereas 'another' alternates with 'other' for *-nye*, 'much' alternatives with 'many' for *-ningi*. Recall that Dent and Nyembezi had also listed *-ningana* as an adjective. Actually, this is the diminutive of *-ningi*, and is only frequent enough for classes 8 and 10. Given it is a derivative, it may handily be treated under the form from which it is derived, as shown in (7).

(7) **eziningi** *** *adjective cl. 8, cl. 10* ► **many**; **a lot of** ♦ Wayevamise ukudla yedwa ezikhathini eziningi. • *He used to eat alone many times.* ♦ Izindlu eziningi azakhekanga kahle emalokishini. • *A lot of houses are not built properly in the townships.*
◊ **eziningana** ► **a small number**; **quite a few** ♦ Sekuyizikhathi eziningana wena uhlala lapha ekhishini ulinde uZokwenzani. • *It is now quite a few times that you have been sitting here in the kitchen waiting for Zokwenzani.* ♦ Ngenkathi befika egalaji kwakukhona izimoto eziningana. • *When they arrived at the garage there were a small number of cars.*

The frequencies — whether summed or individually — for the adjective stems *-ninginingi*, *-nci*, *-ncinyane* (the diminutive of *-nci*) and *-ncu* clearly indicate that these adjectives should not be entered in a user-friendly dictionary, where one attempts to cover what users are most likely to need. There is one exception, however. Although the frequency of *olunci* is just 3, there are 66 occurrences of this adjective with the associative formative *na-* 'with' prefixed to it. (8), therefore, is a possible treatment.

(8) **olunci** *adjective cl. 11*
▪ **(lutho) nolunci ► small thing** *(always used in negative sentences)* ♦ Akukho lutho nolunci olukhona phakathi kwethu. • *There is not even the smallest thing between us.* ♦ Nya! kungasali nolunci phansi. • *Nothing! Not even the smallest thing must remain on the floor.* ♦ Akukho nolunci olwaluyosindisa uCetshwayo. • *There is absolutely nothing that would have saved Cetshwayo.*

The dictionary article shown in (8) is interesting in various ways. Firstly, note that *olunci* has not been given a meaning — this is in line with its extremely low frequency, combined with the fact that the combination that follows *is* given a meaning. Secondly, every single example in the corpus indicates that the form *nolunci* collocates with *lutho* (< *utho* 'something; anything'), which is either physically present in the sentence or, more often, implied — hence the brackets around *lutho*. Thirdly, *(lutho) nolunci* 'something small' is *only* used in environments with a negative 'semantic prosody' — see Sinclair (1998: 16-22) for the full meaning of this term, and De Schryver (2008: 284-285) for a Bantu-language example. Fourthly, this negative semantic prosody is actually carried over from the noun *utho*, as there is nothing inherently negative about the adjective stem *-nci*. Extra contextual guidance is thus required — achieved by means of the text '*always used in negative sentences*' that follows the translation equivalent. Ample example sentences further illustrate the various ways in which the negativity is brought about — here amongst others by means of a negative copulative (*akukho* 'there is/are no(t)'), a negative verb (*kungasali* 'must not remain') and even a negative ideophone (*nya* 'of nothingness, disappearance, ending, silence').

Clearly, in a stem-based dictionary, where just *-nci* is lemmatized, it is sheer impossible one could have reached this level of customized accuracy. In comparison, (9) reproduces the full entry for *-nci* in the all-encompassing Doke and Vilakazi (1953).

(9) **-ncí,** adj. [cf. Ur-B. ***-nî***. > kancí; uɓuncí; ncipha; -ncincincí; -ncinyane.]
Tiny, minute, very small. [cf. *-ncane*, *-ncú*.]

Lemmatizing word-like adjectives, then, allows for far more precise meanings to be conveyed, adapted to the particular class of the adjective. Also, derived adjectives such as diminutives can be described exactly there where they occur.

5.2 On morphophonological rules and augmentatives

Table 4: Adjectives expressing 'big'; 'good' vs. 'bad'; and 'new' vs. 'old'

Cl.	AC	-khulu	-hle	-bi	-sha	-dala	-daladala
1;3; 2sg	om(u)-	omkhulu 3 818 ***	omuhle 1 751 **	omubi 604 *	omusha 1 380 **	omdala 1 423 **	omdaladala 0;4;0
2	aba-	abakhulu 444 *	abahle 264	ababi 144	abasha 817 **	abadala 1 397 **	abadaladala 0
4	emi-	emikhulu 316	emihle 410	emibi 250	emisha 165	emidala 48	emidaladala 1
5	eli-	elikhulu 2 198 ***	elihle 1 382 **	elibi 403	elisha 576 *	elidala 249	elidaladala 3
6	ama-	amakhulu 875 **	amahle 480 *	amabi 166	amasha 571 *	amadala 243	amadaladala 0
7	esi-	esikhulu 1 658 **	esihle 1 206 **	esibi 447 *	esisha 281	esidala 149	esidaladala 1
8; 10	eziN-	ezinkulu 1 162 **	ezinhle 1 096 **	ezimbi 469 *	ezintsha 563 *	ezindala 257	ezindaladala 0
9	eN-	enkulu 4 213 ***	enhle 2 081 **	embi 740 *	entsha 1 086 **	endala 702 *	endaladala 0
11	olu-	olukhulu 1 225 **	oluhle 184	olubi 372	olusha 350	oludala 180	oludaladala 0
14	obu-	obukhulu 277	obuhle 240	obubi 106	obusha 93	obudala 66	obudaladala 0
15; 17	oku-	okukhulu 2 026 **	okuhle 1 438 **	okubi 1 487 **	okusha 250	okudala 63	okudaladala 0
1sg	engim(u)-	engimkhulu 1	engimuhle 1	engimubi 1	engimusha 0	engimdala 149	engimdaladala 0
1pl	esiba-	esibakhulu 2	esibahle 1	esibabi 2	esibasha 0	esibadala 8	esibadaladala 0
2pl	eniba-	enibakhulu 1	enibahle 0	enibabi 1	enibasha 0	enibadala 3	enibadaladala 0
	Freq.	**18 216**	**10 534**	**5 192**	**6 132**	**4 937**	**9**

The data in Table 4 was presented first to see whether or not readers would notice that the form of the stem *-khulu* 'big; large; great' for classes 8 to 10 has changed. Indeed, one of the morphophonological rules in Zulu forbids the succession of **n** + **kh**, with the result that the **h** is dropped. Likewise, **n** + **sh** is not allowed, so a **t** is inserted between the *N* of the adjective concord and the initial consonant of the adjective stem. This affects *-sha* 'new; young' for classes 8 to 10. Rather than expecting that dictionary users remember such rules, lemmatizing word-like adjectives immediately gives them the correct forms, as seen in (10) and (11).

(10) **enkulu** *** *adjective cl. 9* ► **big**; **large**; **great** ♦ Ubaba wenze ibhena enkulu yesikole. • *Father made a big banner for the school.*
◊ **enkulukazi** ► **very big**; **very large**; **very great**; **huge** ♦ Babulale inyoka enkulukazi, ngiyabona yinhlwathi. • *They killed a very large snake; I think it is a python.*

(11) **ezintsha** * *adjective cl. 8, cl. 10* ► **new**; **young** ♦ Kuzokwakhiwa izibhedlela ezimbili ezintsha eSoweto. • *Two new hospitals will be built in Soweto.* ♦ Batsheleke izimali emabhange ukuze bathenge lezi zimoto ezintsha. • *They borrowed money from the banks in order to buy these new cars.*

Other 'orthographic rules' which were implicit so far concern *N* for classes 8 to 10 — *m* before **b** or **f**, *n* elsewhere; and the form of the adjective concord for classes 1 and 3 (and the 2nd person singular) — *omu-* vs. *om-*, as well as the form for the 1st person singular — *engimu-* vs. *engim-*. The first prefix in each series is used for monosyllabic stems, the second for polysyllabic stems. See for instance (12) and (13), respectively (14) and (15), applied to the adjective stems *-hle* 'good; beautiful; nice' and *-bi* 'bad; ugly; evil' vs. *-dala* 'old'.

(12) **omuhle** ** *adjective cl. 1, cl. 3* ► **good**; **beautiful**; **nice** ♦ Ngitshele konke mama wami omuhle. • *Tell me everything my good mother.* ♦ UKhanyi unomzimba omuhle. • *Khanyi has a beautiful body.*

(13) **omubi** * *adjective cl. 1, cl. 3* ► **bad**; **ugly**; **evil** ♦ Nguye umuntu omubi kubalandeli baleli qembu. • *He is the bad person among the followers of this group.* ♦ Uwuthandelani umdlalo omubi kangaka? • *Why do you like such an ugly game?*

(14) **omdala** ** *adjective cl. 1, cl. 3* ► **old** ♦ Umuntu omdala akanawo amandla okuphikisana nomuntu osemusha. • *An old person doesn't have the strength to compete with a person who is still young.* ♦ Ubona lapho umgwaqo omdala. • *She sees an old road there.*

(15) **engimdala** *adjective 1p sg* ► **I who am old** ♦ Wacabanga ukuthi ngimdala. • *She thought that I am old.* ♦ Wayesethi: 'Bheka, sengimdala, angilwazi usuku lokufa kwami.' • *And then he said: 'Look, although I am already old, I do not know the day of my death.'*

Note that adjectives for the first and second persons singular and plural are *very rare* overall. There are just 9 in all for the 1st person singular, 18 for the 1st person plural, and 8 for the 2nd person plural.[7] Finding 2nd person singular adjectives is very difficult, given that the orthographic form of these is the same as for class 1 and 3 adjectives. They probably have the same order of magnitude as the other first and second person adjectives.

As was the case for the reduplicated stem *-ninginingi* 'numerous' (< *-ningi* 'much/many'), also the frequency of the reduplicated stem *-daladala* 'ancient' (< *-dala* 'old') is too low for it to be included in a dictionary covering the most frequent words only.

Further note that (10) above also listed *enkulukazi* 'very big; very large; very great; huge' as a derivation. Indeed, with adjectives the suffix *-kazi* is used for augmentative purposes. Augmentative adjectives being rather rare (cf. Gau-

ton, De Schryver and Mohlala 2004: 374), they can again best be included directly under those adjectives with which they actually occur.[8]

5.3 On class restrictions

The next group of adjectives is peculiar because they only occur with certain classes, namely the plural classes 2, 4, 6, 8 and 10, as seen in Table 5.[9]

Table 5: Adjectives expressing 'how many?', 'two', 'three', 'four', and 'five'

Cl.	AC	-ngaki	-bili	-thathu	-ne	-hlanu
1;3;2sg	om(u)-	—	—	—	—	—
2	aba-	abangaki 199	ababili 1 973 **	abathathu 727 *	abane 382	abahlanu 256
4	emi-	emingaki 124	emibili 1 705 **	emithathu 993 **	emine 542 *	emihlanu 677 *
5	eli-	—	—	—	—	—
6	ama-	amangaki 112	amabili 4 276 ***	amathathu 1 856 **	amane 1 034 **	amahlanu 575 *
7	esi-	—	—	—	—	—
8;10	eziN-	ezingaki 252	ezimbili/mbili 3 380 ***/120	ezintathu 1 725 **	ezine 851 **	ezinhlanu 407
9	eN-	—	—	—	—	—
11	olu-	olungaki 0	olubili 0	oluthathu 1	olune 0	oluhlanu 0
14	obu-	obungaki 1	obubili 7	obuthathu 8	obune 0	obuhlanu 0
15;17	oku-	okungaki 1;1	okubili 0;86	okuthathu 1;27	okune 2;7	okuhlanu 0;6
1sg	engim(u)-	—	—	—	—	—
1pl	esiba-	—	—	—	—	—
2pl	eniba-	—	—	—	—	—
	Freq.	**690**	**11 667**	**5 338**	**2 818**	**1 921**

Clearly, one cannot 'count' singular things, so the distribution seen in Table 5 is not so surprising. This said, when assigning a meaning to adjective stems in isolation, without truly considering all and only those *possible* forms that belong to the paradigm, it is rather easy to err in this regard. Taljaard and Bosch (1993: 99), for instance, assign the meaning 'how much/many?' to *-ngaki*. This is incorrect, as 'how *much?' would only be used for singular adjectives, of which there are none for this adjective stem! Compare with the adjective stem *-ningi* 'much/many' in Section 5.1 which, conversely, does have both singular and plural forms. (16) shows a possible treatment for one of the forms of *-ngaki* 'how many?'

(16) **emingaki** *adjective cl. 4* ► **how many?** ♦ Linemibala emingaki ifulegi laseNingizimu Afrika? • *How many colours does the South African flag have?*
- **iminyaka emingaki** ► **how old?** ♦ Waqala uneminyaka emingaki ukucula? • *How old were you when you started singing?*

In (16) one can also see how frequent combinations may be included in a user-friendly dictionary — again directly under the relevant lemma (here: 'how old?' < 'how many years?', with 'years' a plural noun in class 4).

The other forms in Table 5 are used for counting: *-bili* 'two', *-thathu* 'three', *-ne* 'four' and *-hlanu* 'five'. An extra morphophonological rule applies here: in the combination **n** + **th**, the **h** needs to dropped. This affects *-thathu* in classes 8 and 10. Interestingly, going from 2 to 5, the overall frequency decreases. People seem to talk more often about a few things rather than about many things. (17) is a straightforward example.

(17) **ezimbili** *** *adjective cl. 8, cl. 10* ► **two** ♦ Kwakuxabana ziphi lezi zizwe ezimbili kule mpi? • *Which of these two countries were at loggerheads in this war?* ♦ Izimbuzi ezimbili ngezami. • *The two goats are mine.*
mbili *adjective cl. 8, cl. 10* Short form of **ezimbili**

Corpus evidence indicates that *mbili* (frequency = 120), a short form of *ezimbili*, is frequent enough to be lemmatized. A straightforward cross-reference to the full form suffices here, see (17). Needless to say, a form such as *mbili* is not lemmatized nor covered in traditional Zulu dictionaries.

Only one adjectival form 'breaks' the symmetrical pattern seen in Table 5, namely *okubili*, for class 17, see (18).

(18) **okubili** *adjective cl. 17* ► **two things** ♦ Benzani laba bafana? Shono okubili. • *What are these boys doing? Name two things.* ♦ Nokho kubili athanda ukukugqa-misa lapha. • *Nevertheless, there are two things that he wants to highlight here.*

Two reasons may be offered for the relatively high frequency of *okubili*, the first being that people tend to count up to two rather than higher, the second being that this effect is doubled as a result of the copulative use (cf. Section 6 below).[10]

5.4 On cross-references

The adjectives *-fushane*, *-fishane*, *-fusha*, *-fisha* and *-fuphi* may all be used to refer to 'short' people or things. The last three, however, are clearly not frequent enough to be included in even the larger Zulu dictionaries. The first two are synonyms of one another, and overall summed frequencies indicate that *-fishane* should be considered a variant of *-fushane*.[11] All this information, then, leads to a treatment like (19).

(19) **esifishane** *adjective cl. 7* = **esifushane**
esifushane *adjective cl. 7* ► **short**; **brief** ♦ Isitimela sithatha isikhathi esifushane uma sisuka eDanawozi bese sifika eGlencoe. • *The train takes a short time going from Dannhauser to Glencoe.*

So far, the following 'opposite adjective pairs' were discussed: *-khulu* 'big' and *-ningi* 'much/many' vs. *-ncane* 'small/few'; *-hle* 'good' vs. *-bi* 'bad'; and *-sha* 'new' vs. *-dala* 'old'. As the last in this series, *-de* 'long' may be contrasted with *-fushane* 'short'. See Table 6 for the full picture, and (20) for one example.

Table 6: Adjectives expressing 'long' vs. 'short'

Cl.	AC	-de	-fushane	-fishane	-fusha	-fisha	-fuphi
1;3;2sg	om(u)-	omude 304	omfushane 60	omfishane 65	omfusha 0	omfisha 1;1;0	omfuphi 1;2;0
2	aba-	abade 39	abafushane 6	abafishane 3	abafusha 0	abafisha 0	abafuphi 1
4	emi-	emide 111	emifushane 30	emifishane 37	emifusha 0	emifisha 0	emifuphi 0
5	eli-	elide 362	elifushane 20	elifishane 12	elifusha 0	elifisha 0	elifuphi 5
6	ama-	amade 210	amafushane 19	amafishane 29	amafusha 0	amafisha 0	amafuphi 3
7	esi-	eside 2 571 ***	esifushane 98	esifishane 56	esifusha 1	esifisha 0	esifuphi 0
8;10	eziN-	ezinde 237	ezimfushane 102	ezimfishane 63	ezimfusha 0	ezimfisha 0	ezimfuphi 0;1
9	eN-	ende 637 *	emfushane 189	emfishane 160	emfusha 0	emfisha 1	emfuphi 3
11	olu-	olude 241	olufushane 7	olufishane 9	olufusha 0	olufisha 0	olufuphi 1
14	obu-	obude 54	obufushane 1	obufishane 1	obufusha 0	obufisha 0	obufuphi 0
15;17	oku-	okude 34	okufushane 1;0	okufishane 1;3	okufusha 0	okufisha 1;0	okufuphi 1;0
1sg	engim(u)-	engimude 0	engimfushane 1	engimfishane 0	engimfusha 0	engimfisha 0	engimfuphi 0
1pl	esiba-	esibade 0	esibafushane 0	esibafishane 0	esibafusha 0	esibafisha 0	esibafuphi 0
2pl	eniba-	enibade 1	enibafushane 0	enibafishane 0	enibafusha 0	enibafisha 0	enibafuphi 0
	Freq.	**4 801**	**534**	**439**	**1**	**4**	**18**

(20) **amade** *adjective cl. 6* ► **long**; **tall**; **high**; **deep** ♦ Abantu abaningi basebenza amahora amade kodwa bahola amakinati. • *Many people work long hours but earn peanuts.*

6. Qualificative adjectives versus copulative adjectives

In the picture sketched so far, although dealing with complex issues already, a few extra parameters have purposely been avoided. Firstly, in all but three of the examples from (3) to (20), the orthographic form illustrated in the example sentences is exactly the lemma sign. As a result, it may now appear as if the lemma signs are also the only members of each paradigm. Of course, this is not the case.

During dictionary compilation, the lexicographers have at their disposal the full list of all the forms which were brought together during lemmatization, as well as the frequencies for each of these forms. For instance, for *abancane*, see (6) above, these forms are:

(21) abancane <483>, abasebancane <135>, nabancane <66>, besebancane <51>

As one can see, here the most frequent form of the lemma (*abancane*, with a frequency of 483) equals the lemma sign (*abancane*, with a summed lemmatized frequency of 735). This pattern is seen for 113 of the 126 adjectives. In other words, for about 90% of the adjectives, the lemma sign is also the most frequent form of the adjective. This, then, is another good and user-friendly consequence of lemmatizing adjective stems with their full adjective concords.

Rather than choosing random forms to illustrate the lemma signs, the lexicographers try to pick frequent forms from lists such as (21). If one now returns to the article shown in (6), then one notices that the second example exemplifies the second-most frequent form of the lemma, namely *abasebancane*. This form can be analyzed as follows: *aba-* (relative concord class 2, RC2) + *se-* (progressive formative) + *ba-* (adjective prefix class 2, AP2) + *-ncane* (adjective stem) 'who are still small/young/little'. Hence the example: *Uma ungumqeqeshi wabadlali abasebancane kufanele ube nesineke*. 'If you are a coach of players who are still young, you should be patient.'

The last form in (21), *besebancane*, is actually a *copulative adjective*. This is the second aspect that has been kept out of the discussion so far. Under 'adjectives', then, both the qualificative (i.e. the form with the adjective concord) and the copulative uses are brought together. In some rare cases, a copulative adjective is even more frequent than its corresponding qualificative adjective. In (22), for instance, the frequencies are: bahle <153>, abahle <111>; which explains the order of the examples.

(22) **abahle** *adjective cl. 2* ► **good**; **beautiful**; **nice** ♦ Abantu besifazane bahle ngezindlela ezingafani. • *Women are beautiful in different ways.* ♦ Khetha abangani abahle abangeke bakudukise. • *Choose good friends who will never lead you astray.*

In (18) above, the lemma was formed from: okubili <49>, kubili <37>; which should again make the treatment clear.

Also above, (15) is an extreme case. The forms brought together during lemmatization are: ngimdala <73>, sengimdala <72>, engimdala <4>. In other words, the qualificative entry was 'created' to cater for the two copulative uses. Both examples, of course, only illustrate the copulative uses (with 'se' in *sengimdala* the auxiliary verb *-se*).

To all intents and purposes both qualificative and copulative adjectives may be covered by the same translation equivalents (even though the copulative use includes the meaning 'to be' in addition). To turn a qualificative adjec-

tive into a copulative adjective it suffices to drop the initial vowel for all classes, except for class 9 where the initial **e** becomes an **i** (cf. Section 7). This is a feature that *can* and *must* be explained in the integrated 'corpus-based dictionary mini-grammar' (compare with De Schryver and Taljard 2007). It must be explained, because a user who encounters a copulative use of an adjective will need to be able to add the initial vowel in order to look up the lemmatized qualificative use.

7. The tension between linguistics and lexicography

It is now time to depart from the gentle linguistic introduction which has characterized the discussion so far, and to look at some hardcore linguistic facts. What is really the case with the adjective in Zulu? One first needs to know that the **adjective concord** (AC) is actually composed of two formatives, the **relative concord** (RC) plus the **adjective prefix** (AP):

(23) AC = RC + AP

The RC is the *abbreviated* RC. The RC itself is formed by prefixing the relative formative **a-** to the **subject concord** (SC). As such, one for instance obtains *aba-* for class 2 (< *a-* + *ba-*, abbreviated form: *a-*), or *e-* for class 9 (< *a-* + *i-*). The AP for class 2 is *ba-*, so the AC for this class becomes *aba-* (< *a-* + *ba-*); the AP for class 9 is *iN-*, so the AC for this class becomes *eN-* (< *e-* + *iN-*).

With 'AStem' the **adjective stem**, the basic structure of a **qualificative adjective**, respectively **copulative adjective** is:

(24) Basic qualificative adjective = AC + AStem
Basic copulative adjective = AP + AStem

In other words, to turn a qualificative adjective into a copulative adjective, one basically drops the RC. For instance, in (22), *abahle* 'good' becomes *bahle* '(they) are good'. Likewise, the form from (10), *enkulu* 'big', becomes *inkulu* '(it) is big'.

This brief sketch summarizes most adjectival forms seen so far. These forms can however also be preceded by various other prefixes. In order to streamline the presentation, we can divide these into three groups. Firstly, qualificative adjectives may be preceded by a **possessive concord** (PC):

(25) PC + AC + AStem

Secondly, the qualificative adjectives can also be preceded by any of the following formatives: **locative** (*kwa-*/*ku-*), **associative** (*na-*), **instrumental** (*nga-*), **comparatives** (*kuna-* (< *ku-* + *na-*), *njenga-*), and combinations thereof (attested for the top adjectives are: *ngakwa-*/*ngaku-* (< *nga-* + *ku-*), *nakwa-* (< *na-* + *ku-*), *nanga-* (< *na-* + *nga-*)).

For instance, all the forms seen at the bottom of (26) are also the forms seen by the lexicographers during dictionary compilation in *TshwaneLex* (for more on this software, cf. Joffe et al. 2008). An analysis is shown in (27).

(26) **abanye** **** adjective cl. 2* ► **other(s)** ♦ Akufanele uhleke abanye abantu uma bengaphumeleli empilweni. • *You should not laugh at other people who are not succeeding in life.*

12908, abanye <8480>, nabanye <2243>, kwabanye <977>, ngabanye <548>, yabanye <158>, zabanye <149>, kunabanye <136>, njengabanye <111>, nakwabanye <106>

(27)

abanye	= AC2 + AStem	= 'other(s)'
nabanye	= ass + AC2 + AStem	= 'and/with other(s)'
kwabanye	= loc + AC2 + AStem	= 'to other(s)'
ngabanye	= instr + AC2 + AStem	= 'concerning/with other(s); …'
yabanye	= PC4or9 + AC2 + AStem	= 'of other(s)'
zabanye	= PC8or10 + AC2 + AStem	= 'of other(s)'
kunabanye	= comp + AC2 + AStem	= 'than other(s)'
njengabanye	= comp + AC2 + AStem	= 'just like other(s)'
nakwabanye	= ass + loc + AC2 + AStem	= 'and to other(s)'

Thirdly, corpus evidence — as summarized in the bottom slots such as the one seen in (26) — further indicates that all the structures shown in (28) are possible (this is a selection of ten only).

(28) RC + progressive formative *se-* + AP + AStem
RC + negative formative *nge-* + AP + AStem
RC/SC + copulative formative *ng-* + AC + AStem
SC in situative mood + progressive formative *se-* + AP + AStem
(SC in situative mood + auxiliary verb *-se* +) SC in situative mood + AP + AStem
SC in remote past tense (+ auxiliary verb *-se*, optionally dropped) + SC in situative mood + AP + AStem
SC + potential formative *nga-* + AP + AStem
negative morpheme in indicative mood *a-* + SC in indicative mood + AP + AStem
auxiliary verb *-se* (+ SC, obligatory in situative mood) + AP + AStem
[for class 9] copulative formative (terminal depressor) *y-* + AP9 *iN-* + AStem

As a random example, some of the forms from (29) are analyzed in (30).

(29) **omunye** **** adjective cl. 1, cl. 3* ► **another (one)** ♦ Kubulawe omunye osomatekisi KwaZulu-Natali. • *Another taxi man was murdered in KwaZulu-Natal.* ♦ Umehluko omunye ukuthi wayengasakhulumeli futhi. • *Another difference is that she was no longer talkative.*

9103, omunye <5608>, komunye <906>, ngomunye <729>, nomunye <519>, ungomunye <515>, ongomunye <215>, ngingomunye <89>, lomunye <85>, yomunye <79>, njengomunye <73>, engomunye <70>, wayengomunye <66>, ubengomunye <56>, womunye <47>, kunomunye <46>

(30) ungomunye = SC1or3 *u-* + copulative formative *ng-* + AC1or3 *omu-* + AStem *-nye* = 'he/she/it is one of the others'
ongomunye = RC1or3 *o-* + copulative formative *ng-* + AC1or3 *omu-* + AStem *-nye* = 'he/she/it is another one'
wayengomunye = SC1 in remote past tense *wa-* + time auxiliary *-be* (dropped here) + *-y-* (bridging sound) + SC1 in situative mood *e-* + copulative formative *ng-* + AC1 *omu-* + AStem *-nye* = 'he/she was another one' (in the remote past tense)
ubengomunye = SC1 in present tense *u-* + time auxiliary *-be* + SC1 in situative mood *e-* + copulative formative *ng-* + AC1 *omu-* + AStem *-nye* = 'he/she was another one' (in the near past tense)

When considering the examples listed under (27) and (30) — which, it must not be forgotten, are but a tiny selection of the full spectrum —, it is easy to understand why traditional lexicographers for the Bantu languages in general, and for Zulu in particular, decided to collapse all of these forms into just 'a single dictionary article', here the adjective stem *-nye*. Lemmatizing *all* forms, even only all *frequent* forms, remains an impossibility. Yet, meeting the dictionary user halfway is a realistic proposition, as has been shown in Section 5.

Indeed, as is obvious from the full statistics listed under (26) and (29), lemmatizing the basic qualificative adjectives only, truly covers the most important uses. The other forms may be relegated to the integrated mini-grammar. There too, corpus statistics with regard to the frequency of the various *structures* may be used in the endeavour to present the core issues.

The tension, then, between a detailed, all-encompassing linguistic coverage on the one hand, and a user-friendly, tailored lexicographic treatment on the other, has been eased by a study of overall corpus statistics. What is of prime importance ends up in the dictionary A-to-Z section; what is secondary ends up in the attached grammar.

8. Getting the adjective frequencies right

Frequencies such as those shown in the two previous sections are not always as straightforward as they may seem. At face value, several adjectival forms may also be other parts of speech. When one actually sets out to compile a dictionary article, it is not exceptional to browse through literally hundreds of concordance lines in order to extract the meaning(s) and to select appropriate example sentences for the lemma one is working on. However, when one needs to get an idea of the relative frequencies of different forms — be these on homonym level, sense level, or both simultaneously —, sampling techniques are used for all frequent items in order to limit the number of concordance lines to be studied.[12] Typically, the lexicographers aim at studying about fifty KWIC lines at this point. In Figure 1, *okudala* is being analyzed, an item which can be both an adjective and a verb (marked with 'a' and 'v' respectively during the analysis).

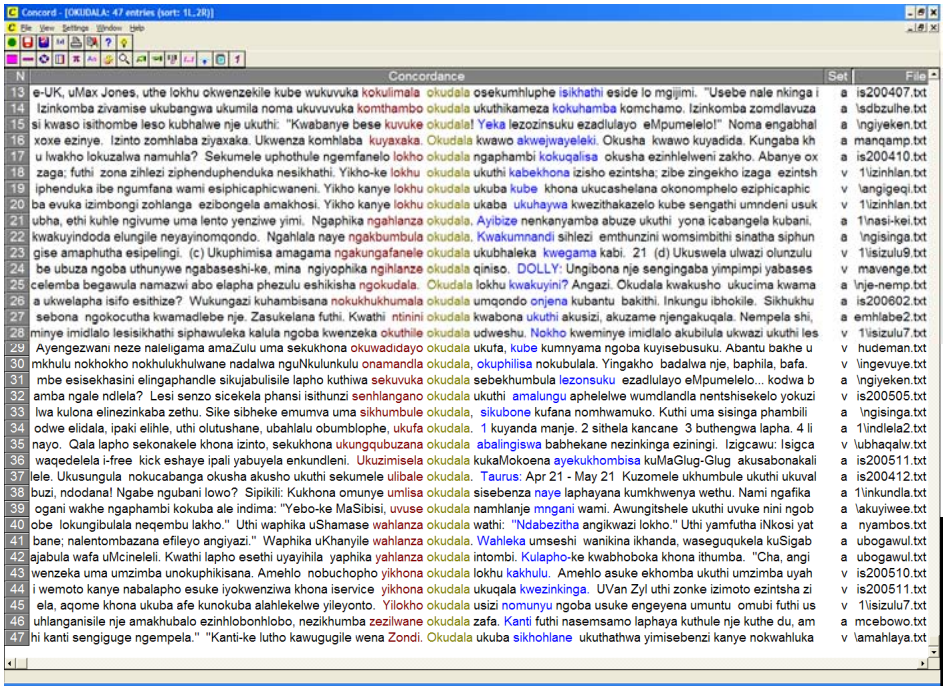

Figure 1: Sampling *okudala*, which is both an adjective (a) and a verb (v)

In Figure 1, the corpus software, *WordSmith Tools* (Scott 2008), was requested to randomly select one out of every three occurrences, and the allocation seen in the sample was then used to distribute the total frequency across the verb *-dala* 'create', and the adjective *okudala*, shown in (31).

(31) **okudala** *adjective* **1** *cl. 15* ► **old** ♦ Ukuzimisela okudala kukaMokoena ayekukhombisa kuMaGlug-Glug akusabonakali njengoba emaningi amaphutha awenzayo. • *The old determination of Mokwena which he had shown with the Team of the Crocks is no longer visible because of the many mistakes that he made.* **2** *cl. 17* ► **something old**; **long ago** ♦ UNondela wayesekhumbule okudala ngempela kusabusa inkosi uNdaba. • *Nondela had remembered the really old things during the reign of chief Ndaba.* ♦ Kukhona omunye umlisa okudala sisebenza naye laphayana. • *There is another male person with whom we worked together long ago.*

Focusing on the adjective: The meanings for the different senses were 'derived' from the corpus, and at the same time one of course keeps an eye on all other items within the same paradigm too — compare for instance (14) and (15). Further observe that two of the three examples in (31) were also selected from the sample seen in Figure 1 (viz. lines 36 and 38). As another example, the frequency of *kubili*, see (18), was split over the adjectival and nominal use.

9. Pinpointing idiomatic uses with adjectives

In Table 1, one could see that Dent and Nyembezi (1995) covered one instance of idiomatic use with an adjective, reprinted in (32).

(32) **-ncane** (adj) small; few; young.
kwaba kuncane indawo — keen competition; outcome difficult to predict.

Coverage of idiomatic use is of course commendable, but in a user-friendly dictionary, this usage should at least be truly frequent too. A corpus-wide search through 8.5 million words of Zulu returns just six instances of *-ba/-be kuncane indawo*. The meaning 'keen competition' cannot be derived from these lines, however, rather something like 'it is not comprehensible what the outcome will be'. The latter is also the meaning listed in Nyembezi's (1992: 317) monolingual dictionary *Isichazimazwi sanamuhla nangomuso*, as well as in Nyembezi and Nxumalo's (1966: 223) miscellany of Zulu culture *Inqolobane yesizwe*. In any case, there are certainly better candidates; (33) is an example.

(33) **oludala** *adjective cl. 11* ► **old** ♦ Indibilishi nosheleni uhlobo oludala lwemali. • *A penny and a shilling are an old type of money.*
▪ **kusadliwa ngoludala ► old customs are still followed** *(Literally: there (things are) still being eaten with an old one (referring to a spoon))* ♦ Kusadliwa ngoludala eMsinga. • *Old customs are still followed at Msinga.*

While the frequency of *oludala* is 60, that of *ngoludala* is twice as high, 120. Of these 120 all but one of the occurrences refer directly to the idiomatic use. The adjective *oludala*, then, has clear open and idiomatic uses, roughly one-third being open, two-thirds being idiomatic (compare with Sinclair 1987: 319-320).

10. Overruling strict principles for the sake of user-friendliness

A lexicographer's job is one of repetitious systematicity. Every now and then, however, flexibility is called for in the user's interest. (34) is a case in point.

(34) **okuhle** ** *adjective* Compare **kuhle[1] 1** *cl. 15* ► **good**; **beautiful**; **nice** ♦ Bamfisela ukuhlolwa okuhle. • *They wished him a good examination.* **2** *cl. 17* ► **something good / beautiful / nice** ♦ Siyifisela okuhle le ngane. • *We wish this child good luck.*
▪ **okuhle kodwa ► only the best** ♦ Umfisele okuhle kodwa nempilo ende. • *He wished her only the best and a long life.*
◊ **kungakuhle ► it would be good / beautiful / nice** ♦ Kungakuhle uma uthisha engabanika amaphuzu aphezulu. • *It would be nice if the teacher could give them high marks.*
Note: *For the copulative use of 'okuhle', see 'kuhle' (it's good/beautiful/nice).*

1438, okuhle <964>, kungakuhle <253>, kukuhle <175>, ngokuhle <46>

The article shown in (34) has a bit of everything: two senses (one for class 15, one for 17), a frequent collocation (*okuhle kodwa*), and a frequent derivation (*kungakuhle*). What this article does not cover is the copulative use. In this one exceptional case, the copulative adjective has been lemmatized in its own right, and this for two main reasons: (a) its high frequency, and (b) as a copulative adjective, it is homonymous with two other words — see (35).

(35) **kuhle[1]** ** *copulative cl. 15, cl. 17* < **okuhle ► it's good / beautiful / nice** ♦ Ukulobola kuhle ngezizathu eziningi. • *Lobola is a good thing for various reasons.* ♦ Kuhle ukushada umuntu omthandayo. • *It's nice to marry someone you love.*

kuhle[2] ** *conjunction* ► **must**; **ought to** ♦ Zaqala ukweluleka izingane zazo zithi kuhle zilingise uGabha. • *They began to advise their children, saying they ought to imitate Gabha.*

kuhle[3] *

▪ **kuhle kwa- / okwa-** *adverb* ► **(just) like**; **as** ♦ Bajamelana kuhle kwamaqhude amabili. • *They stared at each other just like two cocks do.*

The various forms (*kuhle*, *akukuhle*, *kwakuhle*, *kusekuhle*, and *kwakukuhle*) were sampled, and the frequencies redistributed as 1,944, 1,169 and 741 respectively. The use as a copulative adjective thus turns out to be the most frequent of the three. In comparison, a dictionary user who consults Dent and Nyembezi's dictionary, will only find '**kuhle** (adv) like' and '**kuhle** (conj) ought', in this order, while Doke and Vilakazi only treat the adverbial use. Both these existing dictionaries also fail to provide a crucial (encoding) feature, namely that as an adverb, *kuhle* is always followed by the PC17 *kwa-*, or the pronominalized indefinite PC15 *okwa-*. In our user-friendly dictionary, these are all provided for. A user who looks up the copulative use under *okuhle* (which is the 'normal' thing to do given the dictionaries' lemmatization policy), will be referred to **kuhle[1]**: see the cross-reference before the first sense in (34), as well as the usage note at the bottom there.

11. Other words formed from adjective stems

Sections 3 to 10 introduced a new way to lemmatize adjectives in a user-friendly Zulu–English dictionary. Before we conclude, one last important point must be made. As has no doubt become clear from the discussion so far, words that belong together, no matter the size of the set, are best *treated together* — 'in one go', so to say. In this way one makes sure that one has truly considered everything that is common to each member, while highlighting what makes certain forms different from what is common — a variant of the well-known lexicographic tool *per genus proximum et differentia(e) specifica(e)*. Once one has completed this job, one must however also consider the wider picture, and treat all related forms. In the case of adjectives, a large number of words can be derived from the adjective stems, words that end up in other word classes. The Addendum shows all the 'derivations' belonging to the top 5 000 lemmas.

A total of 82 lemmas may be said to be linked to and derived from the adjective stems, five of which are not covered in any of the existing dictionaries for Zulu (these are marked in bold in the Addendum). The overall frequency for these 82 forms is about 100 000 (97 430 to be exact), so two-thirds of the overall frequency of the adjectives themselves. It is interesting to see that one only finds derivations with the frequent adjective stems (those with a tick (✓) in the Z-E column of Table 1), except for *ngamafuphi* 'in brief' (286), a 'new' word which may be analyzed as follows: instrumental formative *nga-* + adjective concord *ama-* (referring to *amagama* 'words') + adjective stem *-fuphi* 'short', or thus 'with short words'. (Note that all 'derivations' with *-nye* are derived from the enumerative stem *-nye*, rather than from the adjective stem *-nye*.)

12. Pros and cons of the user-friendly lemmatization of adjectives in Zulu

Bringing the various strands together, and polarizing the extremes first, one may imagine at one end of the spectrum a *purely stem-based* lemmatization approach to the Bantu languages, whereby only the smallest meaningful morphemes are lemmatized and used as entry points for all members of the lemma as well as for all 'derived' items. Applied to the adjectives that would mean lemmatizing *core adjective stems* only, and under each of these twenty-odd stems, one would not only provide detailed guidance on the various qualificative and copulative uses (as discussed in Section 7), but also list all adjectives with extensions (such as diminutives and augmentatives), as well as all (main) derivations (such as all the items with other parts of speech listed in the Addendum). An approach like this would result in massive articles, each several pages long, the contents of which would need to be hierarchically and logically structured, but for the linguist and all language enthusiasts, this presentation would likely be the most rewarding one.

At the other end of the spectrum, one may imagine a *purely word-based* lemmatization approach, whereby each and every orthographic word is entered 'as is' into the dictionary. This effort, too, would be massive, and for all conjunctively written languages simply impracticable. Although extremely user-friendly for any beginner or even anyone with no knowledge whatsoever of the language concerned, such an approach would of course not only be endlessly repetitive, but would also miss out on important generalizations.

These two extremes are but two poles on a continuum, of course. In reality, a 'traditional' stem-based approach to lemmatization such as Doke's 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. Figure 2 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.

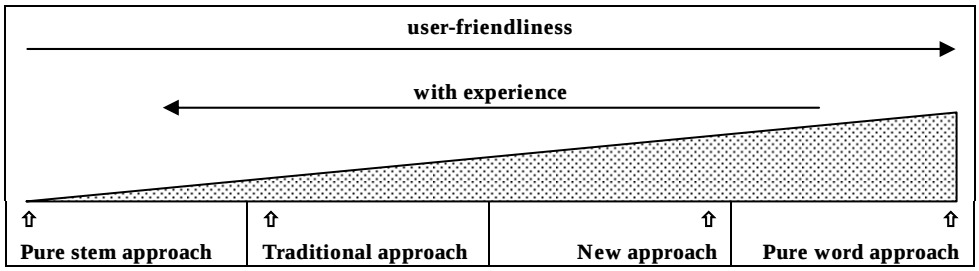

Figure 2: Stem versus word lemmatization for the Bantu languages

In the initial list of 20 000 items to be lemmatized (cf. Section 4), there were 332 adjectival forms. These were collapsed into 126 adjective articles — a move away from the 'pure word' pole, but still a *long* way from the 'pure stem' pole. Indeed, we settled for an approach that includes the adjective concord, as overall frequencies indicated that this form is also the most frequently used one. Note that of the 126 adjectives, about half (68) also have a star rating (cf. Tables 3 though 6: 17 x ***, 30 x **, 21 x *).

Given one is moving over a continuum, no matter which approach one settles for, there will always be pros and cons. The main '**cons**' of our new approach may be summarized as follows (with, between square brackets, a cross-reference to the relevant section where it was discussed above):

— Given the focus on top-frequent members only, none of the paradigms is ever complete. [*4*]
— For copulative adjectives, one needs to 'guess' the (abbreviated) relative concord. [*6*]
— For all adjectives with further prefixes, one needs to know or consult a (or 'the attached') grammar anyway. [*7*]
— Some of the (implicit) connections between words derived from the same adjective stem are lost. [*11 and Addendum*]
— One misses out on generalizations. [*12*]

In our view the '**pros**', which we list by way of conclusion below, far outweigh these few 'cons':

— Excellent reflection of the true distribution of the lexicon. [*5*]
— Precise translation equivalents are provided, rather than general ones. [*5.1*]
— The exact spelling for each form is given, without the need to apply the various morphophonological rules (useful for both receptive and active use). [*5.2*]
— The correct (and modern) class numbers are indicated, while class restrictions are implied. [*5.3*]

- Only the frequent core adjectives are treated, with (frequent) variant forms being cross-referenced to their more frequent forms. [*5.4*]
- Diminutive and augmentative adjectives are listed directly there where they are also used. [*5.1 and 5.2*]
- Typical combinations and collocations are entered where they are relevant. [*5.3 (16) and 10 (34)*]
- Frequent derivations are listed there where they are relevant. [*10 (34)*]
- Idiomatic use is pinpointed and covered there where it is relevant. [*9*]
- Real examples illustrate each and every lemma, collocation, combination, derivation and idiomatic use. [*5 through 10*]
- The most salient form of each lemma is illustrated, which in 90% of the cases is the lemma itself. [*6*]
- The star rating (a logical by-product of the approach advocated here) gives a visual clue as to each adjective's relative importance. [*2, 8 and 12*]
- The detailed analysis of corpus evidence also allows for a move towards the inclusion of supra-semantic features, such as the attention to semantic prosody. [*5.1 (8)*]
- There is less dependency on a grammar for successful dictionary consultation. [*5 through 10*]
- The ultimate user-friendliness is flexibility: for instance, when it comes to the need to differentiate between various homographs in a text, word-like adjectives may now be juxtaposed with words in other word classes through the inclusion of homonymous forms. [*10 and -khulu in the Addendum*]

It is our contention, then, that this new approach to the lemmatization of adjectives in Zulu will result in a more successful dictionary look-up experience.

Acknowledgements

Heartfelt thanks to my Zulu mentor, Prof. A. Wilkes. Thanks too, to the Publisher who was willing to embark on this innovative project, as well as to Ghent University for its continued support of my field trips to South Africa.

Endnotes

1. For an analysis of this dictionary, and its implications, cf. De Schryver and Wilkes (2008).
2. Comparing (2)(a) with (2)(b) one also notices that Doke assigned the part of speech (POS) 'noun' to *isine/*izine*. This is incorrect, *isine* (at least when used neutrally) is an adverb. There is also no *izine* (corpus frequency = 0). Furthermore, in our user-friendly dictionary the form for 'fourthly' is lemmatized in its own right, under *okwesine*, which is the basic form from which Doke's *ngokwesine* is derived. Further note that the numbers between the (supposed) singular and plural forms in (2)(a), as well as in (1)(a), are tone markings, one per syllable.

3. Poulos and Msimang (1998: 142) list three more adjectives, namely: *-ngakanani* 'how big?', *-ngaka* 'so big, as big as this' and *-ngako* 'as big as that'. In Doke and Vilakazi (1953), these three stems are given both relative and adjective status. Given the adjective concords are different from the relative concords for classes 1 + 3, 4 and 6, corpus frequencies for adjectival forms in these specific classes can pinpoint whether or not these stems are indeed (frequently-used) adjectives. The results are: *omngakanani* (0), *emingakanani* (0), *amangakanani* (2); *omngaka* (1), *emingaka* (3), *amangaka* (2); *omngako* (0), *emingako* (0), *amangako* (0). Extrapolating from this, one can safely say that these three stems are *not* adjective stems. Conversely, Poulos and Msimang fail to mention the third most frequent adjective, *-ningi* (and its derivation *-ningana*), as well as *-ngaki* and *-fushane*. Although not a textbook, but a full-blown linguistic analysis, Poulos and Msimang erred in the same way as Taljaard and Bosch.
4. The 'proof' that this is a valid approach for Zulu lexicography will be given in a forthcoming article, but see De Schryver (2003) for a similar approach, applied to Ndebele.
5. For those not familiar with the numbering system for the Bantu noun classes: Classes 1, 3, 5, 7 and 9 contain singular nouns, with the corresponding plurals in classes 2, 4, 6, 8 and 10. Classes 11 and 14 contain mostly singular nouns, with some of them having plural features. In Zulu, classes 12 and 13 do not exist. Class 15 is the so-called infinitive class, containing (all the) verbs and (a few) lexicalized nouns. Classes 16 to 18 are the so-called locative classes which, for various Zulu parts of speech, can often be collapsed into just one class, class 17. This phenomenon is known as noun class reduction.
6. The various senses are always viewed from the Zulu point of view, which explains a series like 'small; young; little' under a single sense number, as one is dealing with a single concept in Zulu here.
7. These are, for the 1st person singular: *engimdala* (frequency = 4), *engimfushane* (1), *engimkhulu* (1), *engimncinyane* (1), *engimubi* (1) and *engimuhle* (1); for the 1st person plural: *esibabi* (2), *esibadala* (8), *esibahle* (1), *esibakhulu* (2), *esibancane* (1) and *esibaningi* (4); and for the 2nd person plural: *enibabi* (1), *enibadala* (3), *enibade* (1), *enibakhulu* (1) and *enibancane* (2).
8. The only other augmentative adjective that is frequent enough to be included is *omkhulukazi* (frequency = 53), with the same meaning as *enkulukazi* (56). *Eziningana* (101), mentioned in Section 5.1, is the only frequent diminutive adjective.
9. Cf. Endnote 5 for more on the Zulu classes.
10. Observe that this very paragraph is another example of this. Also, the hapaxes and other low frequencies, in Table 5 and elsewhere, are always 'suspect' — all class 15 forms in Table 5, for instance, come from textbooks only. The Bible is another source of many low-frequency words.
11. The frequencies of the adjectives *omfushane*, *omfishane* and *ezimfishane* is actually lower than 2 x 50, but these adjectives are nonetheless lemmatized for both classes in each case. Idem for *okudala* in Table 4.
12. If POS-tagged corpora were available for Zulu — as is for instance the case for Northern Sotho (De Schryver and De Pauw 2007) — the distribution for the different parts of speech would immediately be known. Sampling techniques would still be needed, however, and this (a) to pinpoint the relative distribution of the various senses for polysemous items, and (b) to make sure one has a truly mixed set of KWIC lines, randomly sampled from the various sources, for monosemous items. In general, some homonyms also have the same POSs, and for these sampling is also often a must.

References

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. 2003. Drawing up the Macrostructure of a Nguni Dictionary, with Special Reference to isiNdebele. *South African Journal of African Languages* 23(1): 11-25.

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

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

De Schryver, G.-M. and G. De Pauw. 2007. Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*. *Lexikos* 17: 226-246.

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 E. Taljard. 2007. Compiling a Corpus-based Dictionary Grammar: An Example for Northern Sotho. *Lexikos* 17: 37-55.

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: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Doke, C.M., D.M. Malcolm and J.M.A. Sikakana. 1958. *English–Zulu Dictionary*. Johannesburg: Witwatersrand University Press.

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

Fox, G. 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London: Collins ELT.

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.

Hanks, P. 2002. Mapping Meaning onto Use. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 156-198. EURALEX.

Joffe, D. et al. 2008. *TshwaneLex Suite* [online]. http://tshwanedje.com/tshwanelex/.

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

Nyembezi, C.L.S. 1992. *Isichazimazwi sanamuhla nangomuso*. Pietermaritzburg: Reach Out Publishers.

Nyembezi, C.L.S. and O.E.H. Nxumalo. 1966. *Inqolobane yesizwe*. Pietermaritzburg: Shuter & Shooter.

Poulos, G. and C.T. Msimang. 1998. *A Linguistic Analysis of Zulu*. Cape Town: Via Afrika.

Scott, M. 2008. *WordSmith Tools* [online]. http://www.lexically.net/wordsmith/.

Sinclair, J.M. 1966. Beginning the Study of Lexis. Bazell, C.E., J.C. Catford, M.A.K. Halliday and R.H. Robins (Eds.). 1966. *In Memory of J.R. Firth*: 410-430. London: Longman.

Sinclair, J.M. 1987. Collocation: A Progress Report. Steele, R. and T. Threadgold (Eds.). 1987. *Language Topics: Essays in Honour of Michael Halliday*: 319-331. Amsterdam: John Benjamins.

Sinclair, J.M. 1998. The Lexical Item. Weigand, E. (Ed.). 1998. *Contrastive Lexical Semantics*: 1-24. Current Issues in Linguistic Theory 171. Amsterdam: John Benjamins.

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

Addendum: All words, with a lemmatized corpus frequency ≥ 50, 'derived' from adjective stems

-bi 'bad' > adverb: *kabi* 'badly; very (much)' (5432); noun: *ububi* 'evil' (578); locative adverb: *ebubini* 'from the evil' (83)

-bili 'two' > adverbs: *isibili* 'second' (3016), *kabili* 'two times' (1024), ***ngambili*** 'both' (63), *okwesibili* 'secondly; for the second time' (841); conjunction: *nambili* 'and two' (460); noun: *uLwesibili* 'Tuesday' (457); inclusive numeral pronouns (cf. De Schryver 2008a): *bobabili* 'both (of them)' (1445), *kokubili* 'both' (166), *nobabili* 'both of you' (164), *omabili* 'both (of them)' (214), *sobabili* 'both of us' (351), *womabili* 'both (of them)' (375), *yomibili* 'both (of them)' (168), *zombili* 'both (of them)' (943), *zozimbili* 'both (of them)' (74)

-dala 'old' > adverb: *kudala* 'long ago' (993); noun: *ubudala* 'old; age' (184)

-de 'long' > adverbs: *kade* 'long ago' (4338), ***kakade*** 'long ago' (193), *kude* 'far' (2331), *ngesikade* 'at long last' (193), *phakade* 'forever' (382); noun: *ubude* 'length; height; depth' (442)

-fuphi 'short' > adverb: ***ngamafuphi*** 'in brief' (286)

-fushane 'short' > adverb: *kafushane* 'shortly' (173)

-hlanu 'five' > adverbs: *isihlanu* 'five; fifth' (909), *kahlanu* 'five times' (61), conjunction: *nanhlanu* 'and five' (178), noun: *uLwesihlanu* 'Friday' (1099)

-hle 'good' > adverbs: *kahle* 'well; carefully' (20382), ***kahlehle*** 'very well; very much; precisely' (100); nouns: *isihle* 'kindness' (79), *ubuhle* 'goodness' (1231); locative adverb: *ebuhleni* 'near/in/… beauty' (62)

-khulu 'big' > adverb: *kakhulu* 'very much' (19249); conjunction: *namakhulu* 'and hundreds' (117); nouns: *ikhulu/amakhulu* 'hundred/~s' (256/898) [in the plural a (more frequent) homonym of the class 6 adjective *amakhulu* 'big'], *indlunkulu* 'main hut; royal house' (133), *isikhulu/izikhulu* 'important person/~s' (1228/2050), *onkulunkulu* 'gods' (136), *ubabamkhulu/obabamkhulu* '(my/our) grandfather/~s' (77/80), *ubukhulu* 'greatness; size' (374), *umdlunkulu* 'chief's wife/wives' (67), *umkhulu/omkhulu* 'grandfather/~s' (330/83) [in the plural a (lesser frequent) homonym of the class 1or3 adjective *omkhulu* 'big'], *undlunkulu* 'member of the royal family' (65), *undunankulu* 'premier' (606), *uNkulunkulu* 'God' (4473), *uthishomkhulu* 'principal' (51), *uyisemkhulu/oyisemkhulu* '(her/his, their) grandfather/~s' (95/50); locative adverbs: ***ekomkhulu*** 'in/at/to/from/… the head office' (50), *endlunkulu* 'in/at/to/from… the main hut; in/at/to/from/… the royal house' (237), *ezikhulwini* 'to/from/among/… important persons' (122), *komkhulu* 'in/at/to/from/… the chief/king's place' (1000)

-ncane 'small / few' > adverb: *kancane* 'a little; slowly' (5572)

-ne 'four' > adverbs: *isine* 'fourth' (493), *kane* 'four times' (73), *okwesine* 'fourthly; for the fourth time' (55); conjunction: *nane* 'and four' (182); noun: *uLwesine* 'Thursday' (352)

-ngaki 'how many?' > adverb: *kangaki* 'how often?' (62)

-ningi 'much/many' > adverb: *kaningi* 'many times' (656); nouns: *iningi* 'the majority' (2549), *ubuningi* 'abundance; plural' (734)

-sha 'new' > adverb: *kabusha* 'anew' (805); noun: *intsha* 'youth' (1328); locative adverb: *entsheni* 'to/among/… the youth' (126)

-thathu 'three' > adverbs: *isithathu* 'third' (1380), *kathathu* 'three times' (459), *okwesithathu* 'thirdly; for the third time' (251); conjunction: *nantathu* 'and three' (126); noun: *uLwesithathu* 'Wednesday' (608); inclusive numeral pronouns (cf. De Schryver 2008a): *bobathathu* 'all three (of them)' (209), *sobathathu* 'all three of us' (52), *zontathu* 'all three (of them)' (61)