The Lexicographic Treatment of Quantitative Pronouns in Zulu

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: In Zulu, there are three kinds of quantitatives: inclusive, exclusive and numeral. For the lemmatization of these, even existing traditional dictionaries felt the need to move away from a pure 'stem' approach towards a 'word' approach. In a new Zulu–English dictionary project, this is not only confirmed, but is taken one step further with particular attention to the microstructure.

Keywords: LEXICOGRAPHY, DICTIONARY, BILINGUAL, CORPUS, LEMMATIZATION, FREQUENCY, ZULU (ISIZULU), ENGLISH, QUANTITATIVE PRONOUN, INCLUSIVE QUANTITATIVE PRONOUN, EXCLUSIVE QUANTITATIVE PRONOUN, (INCLUSIVE) NUMERAL QUANTITATIVE PRONOUN, USER-FRIENDLY

Samenvatting: De lexicografische behandeling van kwantitatieve pronomina in Zoeloe. In Zoeloe zijn er drie soorten kwantitatieven: inclusieve, exclusieve en numerieke. Voor de lemmatisatie daarvan voelden zelfs bestaande traditionele woordenboeken de nood om stam-lemmatisatie te verlaten voor woord-lemmatisatie. In een nieuw Zoeloe–Engels woordenboekproject wordt dit niet alleen bevestigd, doch wordt ook een stap verder gegaan, met bijzondere aandacht voor de microstructuur.

Sleutelwoorden: LEXICOGRAFIE, WOORDENBOEK, TWEETALIG, CORPUS, LEMMATISATIE, FREQUENTIE, ZOELOE, ENGELS, KWANTITATIEF PRONOMEN, INCLUSIEF KWANTITATIEF PRONOMEN, EXCLUSIEF KWANTITATIEF PRONOMEN, (INCLUSIEF) NUMERIEK KWANTITATIEF PRONOMEN, GEBRUIKSVRIENDELIJK

1. Distribution of Zulu parts of speech

This short contribution is one in a series of research articles that deal with the lemmatization of the different Zulu word classes, or parts of speech (POSs), in a user-friendly Zulu–English dictionary for junior users. In De Schryver and Wilkes (2008) the *possessive pronouns* were discussed, in De Schryver (2008a) an in-depth analysis of the *adjectives* followed, while the *quantitative pronouns* are the topic of the present endeavour.

The project proceeds within a Sinclairian, corpus-driven lexicographic framework (De Schryver 2008). For the Zulu to English side of the dictionary, a corpus of 8.5 million Zulu words was first lemmatized, and the top 5 000 lem-

mas — together with their lemmatized corpus frequencies — constitute the backbone of the dictionary's Zulu macrostructure. As it turns out, each Zulu lemma with a lemmatized corpus frequency of at least 50 needs to be considered for inclusion in the dictionary.

The three largest categories of Zulu parts of speech are — unsurprisingly — the different types of nouns, verbs and adverbs. Together, and in a dictionary that contains the top 5 000 lemmas only, these three POS categories cater for about 80% of the Zulu lexicon. We refer to these three as 'Group 1'. The other POS categories can be divided into two further groups based on the number of members in each POS category. 'Group 2' consists of those POS categories with around 100 or (slightly) more members — these are the relatives, adjectives, conjunctions, possessive pronouns, and ideophones. 'Group 3' consists of all the rest, thus POS categories with (much) fewer than 100 members each — these include the interjections, enumeratives, demonstrative pronouns, quantitative pronouns, relativized possessive pronouns, locative demonstrative copulatives, absolute pronouns, etc.

A detailed study of the top three POS categories, thus Group 1, will be undertaken in future contributions. Given two POS categories from Group 2 (viz. adjectives, with 126 members, and possessive pronouns, with 99 members) have already been looked into, it is now appropriate to briefly engage with one of the smaller POS categories from Group 3 — *in casu* the quantitative pronouns, with just 33 members. In dictionary terms, the three categories that will have been discussed so far amount to just 5.16% of the planned total of 5 000 dictionary articles, as illustrated in Figure 1.
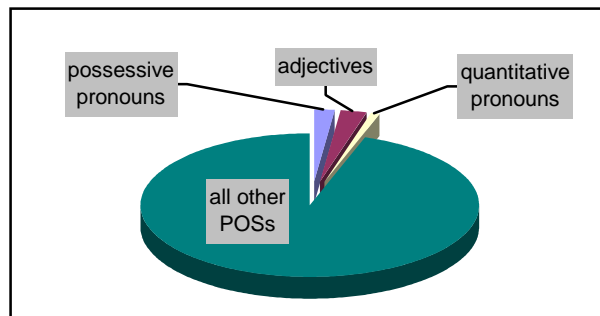

Figure 1: Zulu POS categories studied from a lexicographic point of view

The possessive pronouns amounted to 1.98% of the total, the adjectives to 2.52%, and the quantitative pronouns are good for 0.66%. There is no reason to believe that there is a correlation between the size of a particular POS category and the lexicographic difficulty of that category. Each POS category deserves a discussion in its own right, and once one will have covered all categories, cross-POS discussions will surely be required for it all to make even more sense. Nor is there a reason to believe that there is a strong correlation between the size of

a particular POS category and the summed frequency of its members. The final say on these aspects will only be possible near the end of the project, however. This said, the quantitative pronouns seem not to pose too many lexicographic problems. Yet neither are they trivial.

2.   Zulu quantitative pronouns: A brief linguistic perspective

There are three types of quantitative pronouns in Zulu, all used to express *quantity*, viz. the inclusive quantitative pronouns, the exclusive quantitative pronouns, and the numeral quantitative pronouns. The inclusive quantitative stem is *-nke*, which means 'the whole' when referring to singular nouns and 'all' when referring to plural nouns. The exclusive quantitative stem is *-dwa*, for which the basic meaning is 'alone; only'. For the numeral quantitatives, any of the following adjective stems may be used: *-bili* 'two', *-thathu* 'three', *-ne* 'four', and *-hlanu* 'five'. The quantitative pronouns are formed as shown in Table 1.

Table 1: The formation of the inclusive, exclusive and numeral quantitative pronouns in Zulu
(*with Cl. = noun class number and 1st and 2nd persons; SC = subject concord; PR = pronominal root; QStem = quantitative stem; AP = adjective prefix; AStem = adjective stem (illustrated for* -bili *only);* N = nasal, i.e. n *or* m)

Cl. ⇩	Inclusive quantitative pronouns					Exclusive quantitative pronouns					Numeral quantitative pronouns					
	SC	PR	QStem		⇩	SC	PR	QStem		⇩	SC	PR	AP	AStem		⇩
1	u-	-o-	-nke	>	wonke	u-	-o-	-dwa	>	**yedwa**	—	—	—	—		—
2	ba-	-o-	-nke	>	bonke	ba-	-o-	-dwa	>	bodwa	ba-	-o-	ba-	-bili	>	bobabili
3	u-	-o-	-nke	>	wonke	u-	-o-	-dwa	>	wodwa	—	—	—	—		—
4	i-	-o-	-nke	>	yonke	i-	-o-	-dwa	>	yodwa	i-	-o-	mi-	-bili	>	yomibili
5	li-	-o-	-nke	>	lonke	li-	-o-	-dwa	>	lodwa	—	—	—	—		—
6	a-	-o-	-nke	>	onke	a-	-o-	-dwa	>	odwa	a-	-o-	ma-	-bili	>	omabili
7	si-	-o-	-nke	>	sonke	si-	-o-	-dwa	>	sodwa	—	—	—	—		—
8	zi-	-o-	-nke	>	zonke	zi-	-o-	-dwa	>	zodwa	zi-	-o-	ziN-	-bili	>	zozimbili
9	i-	-o-	-nke	>	yonke	i-	-o-	-dwa	>	yodwa	—	—	—	—		—
10	zi-	-o-	-nke	>	zonke	zi-	-o-	-dwa	>	zodwa	zi-	-o-	ziN-	-bili	>	zozimbili
11	lu-	-o-	-nke	>	lonke	lu-	-o-	-dwa	>	lodwa	—	—	—	—		—
14	bu-	-o-	-nke	>	bonke	bu-	-o-	-dwa	>	bodwa	bu-	-o-	bu-	-bili	>	bobubili
15	ku-	-o-	-nke	>	konke	ku-	-o-	-dwa	>	kodwa	ku-	-o-	ku-	-bili	>	kokubili
17	ku-	-o-	-nke	>	konke	ku-	-o-	-dwa	>	kodwa	ku-	-o-	ku-	-bili	>	kokubili
1sg	ngi-	-o-	-nke	>	**wonke**	ngi-	-o-	-dwa	>	**ngedwa**	—	—	—	—		—
1pl	si-	-o-	-nke	>	sonke	si-	-o-	-dwa	>	sodwa	si-	-o-	ba-	-bili	>	sobabili
2sg	u-	-o-	-nke	>	wonke	u-	-o-	-dwa	>	**wedwa**	—	—	—	—		—
2pl	ni-	-o-	-nke	>	nonke	ni-	-o-	-dwa	>	nodwa	ni-	-o-	ba-	-bili	>	nobabili

In Table 1, one sees that subject concords (SCs) consisting of a vowel only, change to their semivowel (*u-* > *w-*; *i-* > *y-*), while *a-* is dropped. The vowel of the other SCs is elided. The 1st person singular of the inclusive quantitative takes the form of class 1, while class 1 as well as the 1st and 2nd persons sin-

gular of the exclusive quantitative is irregular. These four forms, which do not follow the pattern, have been marked in bold. The formation of only one of the numeral quantitatives is illustrated in Table 1, namely with the adjective stem *-bili* 'two'; the meaning of which becomes 'both'. The formation and meaning of the other three numeral quantitatives is similar. For classes 8 and 10 morpho-phonological rules apply: the nasal *N* is *m* before **b** (applies to *-bili*), *n* else-where; the combination **n** + **th** becomes **nt** (applies to *-thathu*).

The information presented so far is what one typically finds in textbooks and linguistic analyses of Zulu (cf. e.g. Taljaard and Bosch (1993: 83-85), or Poulos and Msimang (1998: 124-129)). Such sources will also list a few more features, some of which will be encountered below. Conversely, corpus evidence reveals other features which none of the existing sources mention. Before these can be discussed, we need to turn from linguistics to lexicography.

3.   Moving from the 'stem pole' to the 'word pole' in lemmatizing quantitative pronouns

As is well known, the lemmatization policy adopted in all existing Zulu dictionaries is to group the lexicon around word stems. So, in a dictionary such as Doke and Vilakazi's (1953) *Zulu–English Dictionary*, a user is able to look up the six stems *-nke, -dwa, -bili, -thathu, -ne* and *-hlanu*. Under both the inclusive and exclusive quantitative stems, this user is even given all the full forms listed in Table 1. Surprisingly, the full forms themselves have *also* been lemmatized in addition. When it comes to the adjective stems *-bili, -thathu, -ne* and *-hlanu*, while not all full quantitative forms are listed within the articles of the stems, a cross-reference to the first (few) of the series is given. Here too, the full forms themselves have been lemmatized *in addition*. Clearly, then, this is a hybrid approach — one in which Doke and Vilakazi is simultaneously acting as a *stem* and *word* dictionary. If the stem and word approaches are viewed as two poles on a continuum (De Schryver 2008a: 86-87), then one could say that Doke and Vilakazi physically move about on this continuum in their dictionary. No doubt, this hybrid approach was followed for reasons of retrievability, or thus user-friendliness.

Interestingly, in Dent and Nyembezi's (1995) *Scholar's Zulu Dictionary*, a dictionary in which an attempt is made to make it easier to find words, the inclusive stem has not been lemmatized, while the exclusive stem, just like the adjective stems, has been. No list of all the forms may be found under the exclusive stem, however, while no further guidance at all with regard to the quantitative pronouns is given under the adjective stems. Yet again, the full forms (or at least, some of them, cf. the Addenda) have been lemmatized. It seems that, in an attempt to lower the threshold, the compilers of this dictionary moved even further away from the 'stem pole'.

Indeed, in a truly user-friendly dictionary there is little point in listing the six stems used in forming quantitative pronouns. This is even true for the

adjective stems when used as adjectives (De Schryver 2008a). In a user-friendly dictionary, one thus moves radically *away* from the 'stem pole', towards the 'word pole'. Given one is dealing with a continuum, the next obvious question is: 'Where to make the cut?' In other words, which formatives and/or prefixes does one keep for lemmatization? Here overall corpus frequencies quickly reveal that the forms as shown in Table 1 are also the ones that are best lemmatized. This will become clear in the discussion below.

4.   Inclusive quantitative pronouns

Of the six quantitative pronouns, the inclusive quantitative pronoun is by far the most frequent, as may be seen from the data in Table 2.

Table 2: Distribution of the quantitative pronouns
(*with Freq. = the summed frequency of all (lemmatized) forms*)

Quantitative pronoun	Freq.	%
inclusive (*-nke*)	47 799	72.33
exclusive (*-dwa*)	13 700	20.73
2 (*-bili*)	3 931	5.95
3 (*-thathu*)	451	0.68
4 (*-ne*)	174	0.26
5 (*-hlanu*)	34	0.05
	66 089	**100.00**

The actual breakdown of the inclusive quantitative pronoun has been tabulated in Addendum 1, where the left side of the table summarizes the corpus statistics, and links these to the user-friendly Zulu–English dictionary under construction; while the right side shows the data as seen in Dent and Nyembezi's dictionary.

Clearly, all forms are frequent enough to be included in any user-friendly dictionary, which was consequently also done in both dictionaries. However, given Dent and Nyembezi do not indicate for which classes certain translation equivalents apply, there is considerable room for confusion. For *sonke*, for example, their equivalents are 'all of us; all of it'. A more user-friendly approach is (1).

(1)   **sonke** *** *inclusive quantitative pronoun* **1** *1p pl* ▶ **all of us** ♦ La manzi asetshenzi-swa yithina <u>sonke</u> lapha ekhaya. • *This water is used by all of us here at home.* **2** *cl. 7* ▶ **the whole**; **each**; **every** ♦ Wabonakala efunda izincwadi zesikole <u>sonke</u> isikhathi. • *He was seen reading school books the whole time.*

In (1), sense numbers are used to present each different class in its own right, while corpus examples illustrate each class. The same applies, *mutatus mutandis*, to all other classes.

Poulos and Msimang (1998: 126) point out that 'it is … not uncommon to hear people say' *wonke* rather than *onke* (and likewise for the other class 6 quan-

titatives). The use of a corpus enables one (a) to see whether or not this is also reflected in the orthography, (b) if it is, to see how (un)common it really is, and (c) to use the results during corpus-driven dictionary compilation.

The orthographic form *wonke* occurs a staggering 6 414 times in 8.5 million words, so it is obviously not feasible to read through all concordance lines. What one can easily do is to sample, and this is what is done in Figure 2 — where the software (*WordSmith Tools*, Scott 2008) has been instructed to randomly select every one-hundredth instance only.
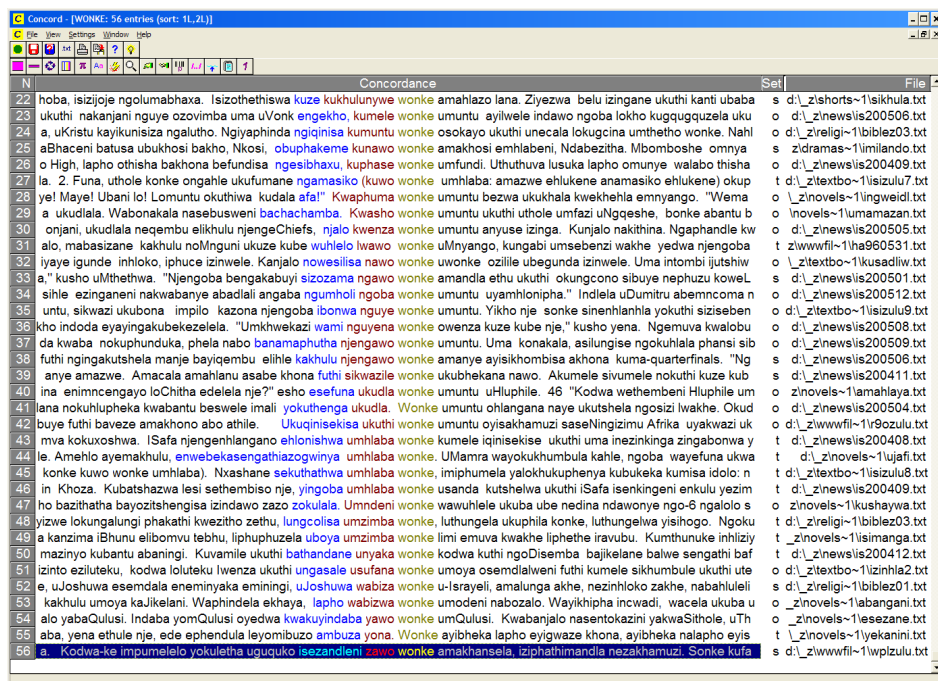

Figure 2: Sampling *wonke*, which occurs in several classes (in the screenshot, under 'Set', 'o' stands for class <u>o</u>ne, 't' for class <u>t</u>hree, and 's' for class <u>s</u>ix)

Rather surprisingly, and extrapolating from the sample, not only does *wonke* indeed occur in class 6, its frequency is as high as 1 489. The distribution across the different subcorpora (cf. the last column in Figure 2) is also even, with instances in short stories, dramas, newspapers, religious texts, etc. Furthermore, although grammars claim that the basic form for class 6 is *onke*, given the frequency of *onke* is 1 659, while that of *wonke* in this class is 1 489, it is clear that both forms are simply used interchangeably. This new information may be embedded into the respective dictionary articles. Compare (2) and (3).

(2)   **onke** ** *inclusive quantitative pronoun cl. 6* Compare **wonke** ► **all** ♦ <u>Onke</u> amehlo
aphenduka abheka le moto. • *All eyes turned around and looked at this car.*

(3)   **wonke** *** *inclusive quantitative pronoun* **1** *cl. 1, cl. 3* ► **the whole**; **each**; **every** ♦
Lokhu kudalela umndeni <u>wonke</u> inkinga. Kuzokwenziwa njani manje? • *This
caused a problem for the whole family. What is going to happen now?* ♦ <u>Wonke</u>
umuntu owayelapho wamangala kabi. • *Each person who was there was very sur-
prised.* **2** *cl. 6* ► **all** ♦ Amandla <u>wonke</u> asemahlombe kaMnuz Bamba
Ndwandwe. • *All the authority rests on the shoulders of Mr. Bamba Ndwandwe.* **3** *1p
sg* ► **the whole of me** ♦ Sengiyibonile mina <u>wonke</u>. • *The whole of me has now
seen it.* **4** *2p sg* ► **the whole of you** ♦ Abakithi bangilethele wena <u>wonke</u>
ngogqoko. • *My friends brought the whole of you to me on a meat tray.*
▪ **wonke** <u>uwonke</u> *cl. 1, cl. 3* ► **everyone**; **everybody** ♦ <u>Wonke</u> <u>uwonke</u> owaye-
khona lapho ngilanda le ndaba, wabamba ongezansi. • *Everybody who was there
when I told the story was very surprised.*
◊ **esewonke** *cl. 6* ► **all together**; **the total** ♦ <u>Esewonke</u> amafulethi alapha angu-
80. • *All together, there are 80 flats here.* ♦ <u>Esewonke</u> amaphepha okuvota angu-12
million. • *The total number of ballot papers is 12 million.*
*Note: For class 6, the pronoun '<u>wonke</u>' also has the variant form 'onke', which is only
slightly more frequent in this class.*

Note how a cross-reference and a usage note have been used in (2) and (3)
respectively to bring all the information together. Further observe that frequent
combinations (*wonke uwonke*) as well as derivations (*esewonke*) may all be
treated under a single lemma such as *wonke*.

For the inclusive quantitatives, the latter is the exception rather than the
norm, as for four of the nine inclusive quantitatives, the lemma sign is the *only*
member of the paradigm. For the other five, (4) shows all the corpus forms that
were brought together — during lemmatization — to obtain the lemma.

(4) Lemma signs with members other than the lemma sign itself
konke <8769>   konke <8717>, konk <52>
sonke <3074>   sonke <3012>, sonkana <62>
wonke <6660>   wonke <6414>, esewonke <108>, wonkana <78>, uwonke <60>
yonke <7823>   yonke <7612>, yonkana <124>, isiyonke <87>
zonke <8815>   zonke <8577>, zonkana <74>, zonk <71>, sezizonke <50>,
kuzozonke <43>

The forms without the final vowel are mostly found in poetry (in written Zulu;
they *are* frequent in everyday speech), while those with the diminutive suffix
-ana are used for extra emphasis. This leaves the instances in (5) to analyze.

(5) Analysis of some of the forms from (4) (*with SC = subject concord*)
esewonke = SC6 in situative mood *e-* + auxiliary verb *-se* + SC6 in situative mood
e- + pronoun *wonke* = 'if they are now all together' (i.e. the sum/total, e.g. in
exam papers)' [cf. derivation under (3)]
uwonke = SC3 *u-* + pronoun *wonke* = 'everyone' [cf. combination under (3)]

isiyonke = SC4or9 *i-* + auxiliary verb *-se* + SC4or9 in situative mood *i-* + pronoun *yonke* = 'it is now all/complete'

sezizonke = auxiliary verb *-se* + SC8or10 in situative mood *zi-* + pronoun *zonke* = 'they are now all/complete'

kuzozonke = locative prefix *ku-* + short form of absolute pronoun 'zona' *zo* + pronoun *zonke* = 'at/to/... all of them'

If one keeps a perspective on the various frequencies as seen in (4), however, then it is clear that the most productive way to lemmatize the inclusive quantitative pronouns is indeed under their basic forms. Low-frequency members of some of the paradigms, then, should only be illustrated when their meanings are lexicalized, as was the case in (3).

5.   Exclusive quantitative pronouns

The corpus statistics and Dent and Nyembezi's treatment for the exclusive quantitative pronouns have been tabulated in Addendum 2. In addition to the data seen there, Dent and Nyembezi also lemmatized two diminutives: *yedwana* (corpus frequency = 109) and *yodwana* (43). These are indeed the two most frequent diminutive exclusive quantitative pronouns, but assigning them lemma-sign status does not seem warranted. Not tying certain translation equivalents to particular classes is again problematic as well. What they missed outright, and what even Doke and Vilakazi overlooked to cover explicitly, is an extra meaning which corpus data clearly reveals for all singular classes (1, 3, 5, 7, 9 and 11). Compare (6).

(6)   **lodwa** *** *exclusive quantitative pronoun cl. 5, cl. 11* ► **alone**; **on its own** ♦ Sibon'ikhwezi seliphezu kwentaba, Seliphume <u>lodwa</u> lathi qekelele [imidwa emi-2 yenkondlo] • *We saw the morning star above the mountain, It came out alone and stood out [2 poetry lines]* ♦ Lolu hlelo belungahambi <u>lodwa</u> njengoba belugcwele amaphutha. • *This programme could not proceed on its own because it was full of flaws.* ► **only** ♦ Wayegqoke ibhulukwe <u>lodwa</u> ngaphandle kweshethi. • *He was wearing trousers only, without a shirt.* ♦ Ungadla uphuthu <u>lodwa</u> uluqede. • *You can eat the thick porridge only and finish it.*

◊ **elilodwa** *cl. 5* ► **(only) one**; **(only) a single** ♦ Empeleni la maqembu asehlukaniswe yiphuzu <u>elilodwa</u>. • *In fact, these teams are now separated by only a single point.*

◊ **olulodwa** *cl. 11* ► **(only) one**; **(only) a single** ♦ Ezinye izichazamazwi zinolimi <u>olulodwa</u> vo, kanti ezinye zinezilimi ezimbili. • *Some dictionaries include only one language, while others include two languages.*

When a singular exclusive quantitative is preceded by a relative concord (RC) then the meaning becomes '(only) one; (only) a single', as may be seen from the article for *lodwa* in (6). *Lodwa* is actually an extreme example, as the frequencies of both *elilodwa* and *olulodwa* are higher than that of the lemma sign itself — see (7).

(7) **lodwa < 2270 >**   elilodwa <1206>, olulodwa <479>, lodwa <355>, nelilodwa <168>, lilodwa <62>

Clearly, then, it is *absolutely crucial* to use this information during the compilation of articles such as *lodwa*. During the project, the lexicographers are in the fortunate position to have the data shown in (7) at their disposal. Indeed, for each and every lemma and linked lemmatized corpus frequency, all the members of each paradigm (together with their individual frequencies) are available right there where they need it in *TshwaneLex* (Joffe et al. 2008), the dictionary writing system used.

For class 6, corpus data indicates that the 'variant form' *wodwa* (175) is actually slightly more frequent than what grammarians consider to be the basic form, *odwa* (163). The dictionary articles for *odwa* and *wodwa* may therefore be treated in a similar way as seen in (2) and (3).

Finally, if one extracts the various structures from each and every exclusive paradigm such as (7), one obtains all the possibilities listed in (8).

(8)   SC in indicative or situative mood + excl. pronoun
       RC + excl. pronoun
       excl. pronoun + diminutive *-ana*
       negative morpheme in indicative mood *a-* + negative SC1 in indicative mood *ka-*
         + excl. pronoun
       locative formative *ku-* + RC + excl. pronoun
       associative formative *na-* + RC + excl. pronoun
       instrumental formative *nga-* + excl. pronoun (+ diminutive *-ana*)

If one now considers the frequencies of each of the structures listed in (8) compared to the frequencies of the basic exclusive quantitative pronouns, then the statistics indicate that the latter is more frequent overall. This, then, is also why lemmatization was undertaken around the basic forms. In addition, for the two low-frequent exclusive quantitative pronouns, viz. *odwa* (163) and *nodwa* (73), the basic forms are also the only ones in the paradigm.

6.   Numeral quantitative pronouns

The corpus and dictionary facts for the four numeral quantitative pronouns have been summarized in Addenda 3 to 6. Note the dramatic decrease in overall frequency going from 'both' to 'all five', viz. 3 931 → 451 → 174 → 34. Clearly, and right away, none of the forms listed in Addendum 6 ('all five') qualifies to be lemmatized within the top 5 000 Zulu lemmas, given the minimum lemmatized frequency is 50. Looking at the breakdown in Addendum 5 ('all four'), one concludes that none of these forms qualifies either.

This leaves us with only 'both' and 'all three'. For these two numeral quantitative pronouns, one immediately notices that Dent and Nyembezi overlooked to lemmatize the *most frequent* form in each case! These are *bobabili* 'both

(of them)' (cl. 2) and *bobathathu* 'all three (of them)' (cl. 2). This once again confirms why one needs a corpus rather than intuition in order to decide on what to include in and what to omit from a dictionary. As another example, *womabili* 'both (of them)' (cl. 6), has not been lemmatized, while the infrequent *yombili* has been. Two more points must be considered. Firstly, the frequency of *womabili* is 375, higher than that of *omabili*, which has a frequency of 214. Compare in this regard the full 'variant' status of the other class 6 quantitatives (*wonke/onke* and *wodwa/odwa*) discussed above. Secondly, *yombili*, with a frequency of just 18, is 'suspect'. In the corpus, it appears once in a textbook, once in the Bible, three times in novels, and 13 times in newspaper articles. The textbook example is exactly that: a textbook example. It was taken from Doke and Vilakazi's dictionary, *imfe yombili* 'both pieces of sweet corn', which thus quantifies a class 9 noun, a singular — this while all numeral quantitative pronouns are only *by definition* supposed to quantify items in the plural classes. Newspaper text is always suspect, and when in a huge text like the Bible only a hapax appears, one again has reason to doubt the status of that particular form. This leaves just three occurrences in novels, too few to make any linguistic claims, and far too few anyway to describe in a dictionary. (For completeness, the forms *yomthathu*, *yomne* and *yomhlanu* (luckily) do not occur in the corpus.)

Three of the four class 14 numeral quantitative pronouns do not occur at all, for 'all three', 'all four', and 'all five', while 'both' in class 14 has the lowest frequency of all forms in Addendum 3. Not treating any of these is thus the proper procedure. Corpus statistics further reveal that the so-called variant forms for classes 8 and 10 are actually more frequent than their basic forms: *zozimbili* (74) vs. *zombili* (943); *zozintathu* (14) vs. *zontathu* (61); *zozine* (30) vs. *zone* (30); *zozinhlanu* (2) vs. *zonhlanu* (7). This has direct implications for dictionary making, as the cross-reference must go from the lesser frequent to the most frequent form. Compare (9) and (10).

(9)   **zombili** ** *inclusive numeral pronoun cl. 8, cl. 10* ▶ **both (of them)** ♦ Akuzange kulimale muntu kuzo <u>zombili</u> lezi zehlakalo. • *Nobody was injured on both these occasions.* ♦ Undebembili: Leli gama lisho ukuthi izindebe <u>zombili</u> ziyasetshenziswa ekuphinyisweni komsindo. • *Bilabial: This word means that both lips are used in the pronunciation of a sound.*

(10)   **zozimbili** *inclusive numeral pronoun cl. 8, cl. 10* = **zombili**

Perhaps a note on the examples is necessary at this stage. Just as was the case for all other dictionary articles, whenever two different classes need to be exemplified in the main section of a dictionary article, an example for each has been selected. This was done, so that the dictionary is also a *didactic tool*, which conveys information both explicitly and implicitly. Further observe that the Zulu word for 'bilabial' (obviously) also includes the adjective stem *-bili* 'two' (*undebembili (Cl. 1a/2a) < izindebe (Cl. 11/10)* 'lips' + *mbili (Cl. 10)* 'two'). The frequency of *undebembili* in the general language is however too low to be lemmatized.

Further note that the POS in (9) and (10) is 'inclusive numeral pronoun', as the numeral quantitatives may indeed be seen as being derived from the inclusive quantitatives (for which the inclusive stem has been dropped) followed by the adjectives.

Lastly, the reason why the numeral quantitatives (or thus the inclusive numeral pronouns) have been lemmatized under their basic forms is simply because each of these forms is also the only member of the paradigm. Nothing else is pre- or suffixed here.

7.   Sinclairian lexicography

Within a Sinclairian, corpus-driven approach to dictionary making, sound lexicographic decisions accompany every step of the compilation; from the use of a corpus for the construction of the macrostructure, which includes decisions on *how to lemmatize* each and every part of speech, all the way to a *detailed analysis of meaning* and the *presentation* thereof in a dictionary. This has been exemplified for the Zulu quantitative pronouns in this contribution.

Acknowledgements

References

Dent, G.R. and C.L.S. Nyembezi. 1995³ [1988², 1969]. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.

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. 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. and B.W. Vilakazi. 1953² [1948]. *Zulu–English Dictionary*. Johannesburg: Witwatersrand University Press.

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

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/.

Taljaard, P.C. and S.E. Bosch. 1993² [1988]. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.

[Abbreviations used in the **Addenda**: ***Z-E*** *= user-friendly Zulu–English diction-ary;* ***Freq.*** *= lemmatized corpus frequency (in 8.5 million words);* ***Band*** *= frequency band in Z-E, with *** the top 500 lemmas, ** the next 500, and * the third 500;* ***Cl.*** *= noun class number and 1st and 2nd persons;* ***CP*** *= noun class prefix;* ***Lemma sign****,* ***POS*** *and* ***Translation equivalent(s)*** *as in Dent and Nyembezi's (1995) dictionary*]

Addendum 1:   Inclusive quantitative pronouns

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
✓	6 660	***	1	um(u)-	wonke	(pron)	the whole of it.
✓	7 859	***	2	aba-	bonke	(pron)	all; all of them.
(✓)	(6 660)	(***)	3	um(u)-	wonke	(pron)	the whole of it.
✓	7 823	***	4	imi-	yonke	(pron)	all of them; the whole of it.
✓	2 642	***	5	i(li)-	lonke	(pron)	all of it.
✓	1 659	**	6	ama-	onke	(pron)	all; the whole of.
(✓)	(6 660)	(***)	..	..	wonke	-	-
✓	3 074	***	7	isi-	sonke	(pron)	all of us; all of it.
✓	8 815	***	8	izi-	zonke	(pron)	all of them.
(✓)	(7 823)	(***)	9	iN-	yonke	(pron)	all of them; the whole of it.
(✓)	(8 815)	(***)	10	iziN-	zonke	(pron)	all of them.
(✓)	(2 642)	(***)	11	u(lu)-	lonke	(pron)	all of it.
(✓)	(7 859)	(***)	14	ubu-	bonke	(pron)	all; all of them.
✓	8 769	***	15	uku-	konke	(pron)	all of it; everything.
(✓)	(8 769)	(***)	17	uku-	konke	(pron)	all of it; everything.
(✓)	(6 660)	(***)	1p sg		wonke	(pron)	the whole of it.
(✓)	(3 074)	(***)	1p pl		sonke	(pron)	all of us; all of it.
(✓)	(6 660)	(***)	2p sg		wonke	(pron)	the whole of it.
✓	498	*	2p pl		nonke	(pron)	all of you.
	47 799						

Addendum 2:   Exclusive quantitative pronouns

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
✓	3 373	***	1	um(u)-	yedwa	(pron)	he/she alone.
✓	1 115	**	2	aba-	bodwa	(pron)	they alone.
✓	1 239	**	3	um(u)-	wodwa	(pron)	it alone.
✓	1 993	**	4	imi-	yodwa	(pron)	they alone; he alone; it alone.
✓	2 270	***	5	i(li)-	lodwa	(pron)	it alone.
✓	163	-	6	ama-	odwa	(pron)	they only; only them.
(✓)	(1 239)	(**)	..	..	wodwa	-	-
✓	972	**	7	isi-	sodwa	(pron)	we alone.
✓	383	-	8	izi-	zodwa	(pron)	they alone.
(✓)	(1 993)	(**)	9	iN-	yodwa	(pron)	they alone; he alone; it alone.
(✓)	(383)	(-)	10	iziN-	zodwa	(pron)	they alone.
(✓)	(2 270)	(***)	11	u(lu)-	lodwa	(pron)	it alone.
(✓)	(1 115)	(**)	14	ubu-	bodwa	(pron)	they alone.
✓	998	**	15	uku-	kodwa	(pron)	alone; only.
(✓)	(998)	(**)	17	uku-	kodwa	(pron)	alone; only.
✓	548	*	1p sg		ngedwa	(pron)	me alone.
(✓)	(972)	(**)	1p pl		sodwa	(pron)	we alone.
✓	573	*	2p sg		wedwa	(pron)	you alone.
✓	73	-	2p pl		nodwa	(pron)	you alone.
	13 700						

Addendum 3:   Numeral quantitative pronoun '2'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
✓	1 445	**	2	aba-	bobabili	-	-
—	—	—	3	um(u)-	—	—	—
✓	168	-	4	imi-	yomibili	(pron)	both.
-	18	-	..	..	yombili	(pron)	both.
—	—	—	5	i(li)-	—	—	—
✓	214	-	6	ama-	omabili	(pron)	(pron)
✓	375	-	..	..	womabili	-	-
—	—	—	7	isi-	—	—	—
✓	74	-	8	izi-	zozimbili	(pron)	both.
✓	943	**	..	..	zombili	(pron)	both.
—	—	—	9	iN-	—	—	—
(✓)	(74)	(-)	10	iziN-	zozimbili	(pron)	both.
(✓)	(943)	(**)	..	..	zombili	(pron)	both.
—	—	—	11	u(lu)-	—	—	—
-	13	-	14	ubu-	bobubili	-	-
✓	166	-	15	uku-	kokubili	(pron)	both.
(✓)	(166)	(-)	17	uku-	kokubili	(pron)	both.
—	—	—	1p sg		—	—	—
✓	351	-	1p pl		sobabili	(pron)	both of us.
—	—	—	2p sg		—	—	—
✓	164	-	2p pl		nobabili	(pron)	both of you.
	3 931						

Addendum 4:   Numeral quantitative pronoun '3'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
✓	209	-	2	aba-	bobathathu	-	-
—	—	—	3	um(u)-	—	—	—
-	15	-	4	imi-	yomithathu	(pron)	three of them.
—	—	—	5	i(li)-	—	—	—
-	27	-	6	ama-	omathathu	(pron)	all three.
-	27	-	..	..	womathathu	-	-
—	—	—	7	isi-	—	—	—
-	14	-	8	izi-	zozintathu	(pron)	all three of them.
✓	61	-	..	..	zontathu	(pron)	all three of them.
—	—	—	9	iN-	—	—	—
(-)	(14)	(-)	10	iziN-	zozintathu	(pron)	all three of them.
(✓)	(61)	(-)	..	..	zontathu	(pron)	all three of them.
—	—	—	11	u(lu)-	—	—	—
-	0	-	14	ubu-	**bobuthathu**	-	-
-	15	-	15	uku-	kokuthathu	(pron)	all three.
(-)	(15)	(-)	17	uku-	kokuthathu	(pron)	all three.
—	—	—	1p sg		—	—	—
✓	52	-	1p pl		sobathathu	(pron)	all three of us.
—	—	—	2p sg		—	—	—
-	31	-	2p pl		nobathathu	(pron)	all three of you.
	451						

Addendum 5:   Numeral quantitative pronoun '4'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
-	29	-	2	aba-	bobane	-	-
—	—	—	3	um(u)-	—	—	—
-	13	-	4	imi-	yomine	(pron)	the four of them.
—	—	—	5	i(li)-	—	—	—
-	35	-	6	ama-	omane	(pron)	all four.
-	27	-	..	..	womane	-	-
—	—	—	7	isi-	—	—	—
-	30	-	8	izi-	zozine	(pron)	all four of them.
-	30	-	..	..	zone	(pron)	all four.
—	—	—	9	iN-	—	—	—
(-)	(30)	(-)	10	iziN-	zozine	(pron)	all four of them.
(-)	(30)	(-)	..	..	zone	(pron)	all four.
—	—	—	11	u(lu)-	—	—	—
-	0	-	14	ubu-	**bobune**	-	-
-	2	-	15	uku-	kokune	(pron)	all four.
(-)	(2)	(-)	17	uku-	kokune	(pron)	all four.
—	—	—	1p sg		—	—	—
-	6	-	1p pl		sobane	(pron)	all four of us.
—	—	—	2p sg		—	—	—
-	2	-	2p pl		nobane	(pron)	all four of you.
	174						

Addendum 6:   Numeral quantitative pronoun '5'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
-	11	-	2	aba-	bobahlanu	-	-
—	—	—	3	um(u)-	—	—	—
-	1	-	4	imi-	yomihlanu	(pron)	the five of them.
—	—	—	5	i(li)-	—	—	—
-	4	-	6	ama-	omahlanu	(pron)	all five.
-	1	-	..	..	womahlanu	-	-
—	—	—	7	isi-	—	—	—
-	2	-	8	izi-	zozinhlanu	(pron)	all five of them.
-	7	-	..	..	zonhlanu	(pron)	all five of them.
—	—	—	9	iN-	—	—	—
(-)	(2)	(-)	10	iziN-	zozinhlanu	(pron)	all five of them.
(-)	(7)	(-)	..	..	zonhlanu	(pron)	all five of them.
—	—	—	11	u(lu)-	—	—	—
-	0	-	14	ubu-	**bobuhlanu**	-	-
-	5	-	15	uku-	kokuhlanu	(pron)	all five.
(-)	(5)	(-)	17	uku-	kokuhlanu	(pron)	all five.
—	—	—	1p sg		—	—	—
-	1	-	1p pl		sobahlanu	(pron)	all five of us.
—	—	—	2p sg		—	—	—
-	2	-	2p pl		nobahlanu	(pron)	all five of you.
	34						