
Linguistic Assumptions and Lexicographical Traditions in the African Languages

E.B. van Wyk, *Jeffreys Bay, South Africa*

Abstract: Two lexical traditions exist in the African languages of South Africa. According to the word tradition, lemmas are based on complete written words, and there is a one-to-one correspondence between written words and lemmas. According to the stem tradition, lemmas are based on the stems of written words without their prefixes. It is pointed out that the difference between these traditions lies mainly in the treatment of nouns. It is also shown that the stem tradition, uncritically regarded by many linguists as more scientific, is based on incorrect assumptions with reference to the morphology of nouns, that it is not applied consistently and that it is less user friendly. It is concluded that the word tradition is based on sounder lexicographical principles.

Keywords: LEXICOGRAPHY, DICTIONARY, AFRICAN LANGUAGES, NGUNI LANGUAGES, ZULU, SOTHO LANGUAGES, NORTHERN SOTHO, SOUTHERN TRADITION, SOTHO, VENDA, TSONGA, MORPHOLOGY, MORPHOPHONOLOGY, WORD DIVISION, TRADITION, ASSUMPTION

Opsomming: Linguistiese aannames en leksikografiese tradisies in die Afrikatale. Twee leksikografiese tradisies kom in die Suid-Afrikaanse Afrikatale voor. Volgens die woord-tradisie word lemmas gebaseer op geskrewe woorde en is daar 'n een-een-verhouding tussen geskrewe woorde en lemmas. Volgens die stam-tradisie word lemmas gebaseer op woordstamme wat ontgaan is van prefikse. Daar word aangetoon dat die verskil slegs op naamwoordlemmas betrekking het en dat die stambenadering, wat deur baie linguïste onkrities as meer wetenskaplik beskou word, gebaseer is op foutiewe aannames i.v.m. die morfologie van naamwoorde, dat dit nie konsekvent toegepas kan word nie, en dat dit minder gebruikersvriendelik is. Die gevolgtrekking is dat die woordtradisie op suiwerder leksikografiese beginsels berus.

Sluitelwoorde: LEKSIKOGRAFIE, WOORDEBOEK, AFRIKATALE, NGUNITALE, ZULU, SOTHOTALE, NOORD-SOTHO, SUID-SOTHO, VENDA, TSONGA, MORFOLOGIE, MORFOFONOLOGIE, WOORDVERDELING, TRADISIE, AANNAME

A dictionary is a compilation of lexical items, not a grammar. It is impossible to learn a language by consulting a dictionary only. All dictionaries therefore assume grammatical knowledge on the part of the user. At the very least the user is required to know the rules of the syntax of the language concerned. In the case of most languages a knowledge of important facts concerning the mor-

phology and even some morphophonological processes are also regarded as a necessary prerequisite to the use of a dictionary.

Lexicography requires, amongst others, a balancing act between linguistic assumptions on the one hand and user friendliness on the other. These two requirements are seldom compatible. It could be stated axiomatically, therefore, that the more specialized the linguistic knowledge required of the user, the less user friendly a dictionary will be. And conversely, the more user friendly a dictionary is, the less sophisticated the linguistic knowledge it assumes.

In most languages the linguistic assumptions underlying dictionaries are agreed upon by tacit conventions giving rise to more or less standard lexicographical traditions. In most European languages, for instance, verbs are lemmatized in their infinitive forms. Thus the equivalent of English *speak* will be found under *sprechen* in German, under *spreken* in Dutch, under *parler* in French, under *parlare* in Italian, under *hablar* in Spanish, and under *falar* in Portuguese. In Japanese the informal positive forms of verbs serve as lemmas and in Biblical Hebrew the second person singular masculine of the perfect (Qal forms) and/or where such forms are not found, the stem forms without vowels. The equivalent of English *speak* is lemmatized as *hanasu* in Japanese and the equivalents of English *say* and *speak* as *'amar* and *dbr* respectively in Hebrew. In all these cases users are assumed to be able to derive the correct tenses, moods, persons, negatives and other verbal forms from such lemmas.

In the case of the African languages two lexical traditions exist, based upon different linguistic assumptions. For the lack of better terms I shall, inappropriately as it will appear later, refer to them as the "word" and "stem" traditions respectively. The word tradition originated in the Sotho languages, Venda and Tsonga, and requires that lexical entries be based on complete written words. In the stem tradition, which is characteristic of the Nguni languages, but is also followed in some Sotho dictionaries, the stems of written words form the basis of lexical entries.

It will be noticed that I refer specifically to *written* words, the reason being that the languages concerned have different traditions of word division. The Nguni languages employ a conjunctive system, whereas the Sotho languages, Venda and Tsonga are written disjunctively. The difference between conjunctivism and disjunctivism concerns the status of certain linguistic elements, which are joined to following elements in the Nguni languages, but written separately in the other languages. Thus the equivalent of the English sentence *the woman is speaking to the child* is written as *umfazi ukhuluma nomntwana* in Zulu, but as *mosadi o bolela le ngwana* in Northern Sotho.

Although the phonological structures of the languages concerned are to a certain extent predisposing factors,¹ conjunctivism or disjunctivism is purely a matter of orthographical convention. The Nguni languages could just as well have been written disjunctively, as it was in older Bible translations, and the other languages conjunctively, as Cole (1955) did for Tswana and Doke and Mofokeng (1957) for Southern Sotho. The two sentences above could in other

words also have been rendered as *umfazi u khuluma no mntwana* and *mosadi obolela lengwana* respectively.

Conjunctive writing has had the benefit of having been championed by C.M. Doke, who attempted to justify it on phonological grounds,² whereas no authors tried to provide a theoretical basis for disjunctivism. As a result, conjunctivism has been accepted uncritically by many African linguists as the only linguistically correct method, on the unjustifiable assumption that there is a one-to-one correspondence between conjunctively written, phonologically defined "words" on the one hand and grammatical words on the other. It has been proven conclusively, however, that neither conjunctivism nor disjunctivism is linguistically justified and that, like disjunctivism, conjunctivism is merely an orthographical convention. Cf. Van Wyk 1994.

The conjunctive system of word division has given rise to the stem tradition in the Nguni languages and disjunctivism to the word tradition in the other South African languages. The five words of the Northern Sotho sentence *mosadi o bolela le ngwana* are therefore entered as five separate lemmas in Northern Sotho word dictionaries, whereas the three words of the Zulu sentence *umfazi ukhulumu nomntwana* are lemmatized as *-fazi*, *-khuluma* and *-ntwana* respectively in stem dictionaries of Zulu.

Assuming uncritically that conjunctivism is the only linguistically valid method of word division, many lexicographers have come to the erroneous conclusion that only the stem tradition is linguistically justified. Ziervogel (1965: 45), for example, claims that it is scientifically sound, and Ziervogel and Mokgokong (1975: 87) (henceforth referred to as ZM) state categorically that it is the only scientific method. It is for this reason that it has been applied in some dictionaries of non-Nguni languages, notably ZM (1975) for Northern Sotho, Mabile et al (1974) and Du Plessis et al (nd) for Southern Sotho, and Snyman et al (1990) for Tswana.

In the word tradition, of which Kriel (1976) for Northern Sotho, Mabile and Dieterlen (1937) for Southern Sotho, Matumo (1993) for Tswana, Van Warmelo (1989) for Venda, and Cuenod (1976) for Tsonga are examples, each written word is entered as a separate lexical item. There is therefore, a one-to-one correspondence between written words and lexical entries in these dictionaries. A one-to-one relation is theoretically also possible in conjunctively written languages, but would result in prohibitive redundancy. It would require, for instance, that verbs with the same stem be entered separately for each different combination of a stem plus subject marker, object marker, modal morpheme and negative morpheme, and this could result in thousands of entries per stem.

The solution which is adopted in the stem tradition is to base lexical entries on the positive forms of infinitive verb stems, leaving it to the user to combine such stems with the required prefixal morphemes. The user is therefore assumed to be au fait with the prefixal morphology and morphophonology of the verb in the language concerned. Now, this is precisely what is

assumed by word dictionaries too. Verbal lemmas are based on the positive infinitive "verbs" and the user is expected to be able to combine these with the appropriate prefixal "words". The two types do not differ on this point, therefore, except for the notational device of a hyphen to indicate verb stems in most stem dictionaries. The entry for 'see' will therefore be found as *bona* in word dictionaries and as *-bona* in stem dictionaries.

The main difference between word and stem dictionaries lies in the treatment of nouns. Nouns pose no problem in the case of the word tradition. They are entered in their complete forms, i.e. prefix plus stem, and classified under their prefixes. The Northern Sotho word for a person, consisting of the prefix *mo-* and the stem *-tho* is consequently lemmatized as *motho*, and the related word for humanity, consisting of *bo-* and *-tho* as *botho*. In the case of stem dictionaries it does pose a problem, however. Since the principle is accepted that lexical entries should be shorn of their prefixes, as in the case of verbs, it is assumed to be logical to do the same with regard to nouns. Like verbs, nouns are therefore classified according to their stems and not to their class prefixes. The two Northern Sotho words referred to above will therefore both be entered under the lemma *-tho*. The terms "word tradition" and "stem tradition", it will now be obvious, refer only to the treatment of nouns and not on other parts of speech.

I do not know of any attempt to justify word dictionaries linguistically. Although stem dictionaries are uncritically regarded as being scientifically superior, not many authors have tried to justify this assumption either. I am acquainted with only two attempts. Ziervogel (1965: 45) claims that:

Entries must be arranged under their stems with cross-references where necessary. This method is scientifically sound. A systematized survey of word-formation in the language is given; it shows word and lexical relationships and prevents repetition.

Paroz states the following in Mabille et al (1974: xii-xiii) a propos of his approach:

... this has the great advantage of bringing words together which are similar in origin and related in meaning, and of showing better the relative place of a given word in the language.

These authors offer four arguments in favour of the stem principle, viz. (1) it gives a survey of word-formation, (2) it shows the relative place of a word in the language, (3) it reveals lexical relations, and (4) it avoids repetition. The implication is that word dictionaries are lacking, or at least inferior, in all these respects.

In this paper I will critically analyze the claims, implicit as well as explicit, of the stem tradition, and compare stem dictionaries with word dictionaries with regard to their linguistic justification and user friendliness. I will not be concerned with practical issues such as the selection of lexical entries or the

handling of exceptions, but only with the morphological structure of lemmas. I shall base my discussion mainly on two representative stem dictionaries, viz. Doke and Vilakazi's *Zulu-English Dictionary* (henceforth DV), and ZM's *Comprehensive Northern Sotho Dictionary*. Examples from these and other dictionaries will contain only those details which are pertinent to the context in which they are quoted.

Of the four arguments mentioned above, the first two can be rejected immediately. It is the task of a grammar, not a dictionary, to give an account of word-formation.³ Showing the relative place of a word in a language can only be taken to mean the same as giving an account of word-formation, or to refer to the grammatical categorization of a word, which can be indicated as well in a word dictionary as in a stem dictionary. Only the remaining two arguments therefore merit consideration.

The basic assumption of stem dictionaries is that the morphology of the verb and the noun is identical in that prefixal elements can be attached freely to stems in both cases. Paroz, explaining his approach in Mabile, Dieterlen and Paroz (p. xii-xiii), says:

... we have ... considered the prefix as *a mobile and exchangeable element*, and classified words according to their stems, or as near to them as could be ascertained or was practicable. (Italics mine, EBvW).

This assumption is, however, wrong; *the morphology of the noun differs in crucial ways from that of the verb. The noun prefix is not mobile or freely exchangeable as Paroz claims.*

The morphology of the verb in the South African languages is completely regular and fully productive. The prefixes concerned, words according to the disjunctive approach, are without exception grammatical morphemes, i.e. negative morphemes, modal and aspectual morphemes, subject markers and object markers. These are arranged as follows, with reference to the Zulu stem *-bona* 'see':

NEG	SUBJ	NEG	MOD	OBJ	STEM (= Root ⁴ + Final Vowel)
(k)a	<i>ngi</i>	<i>nga</i>	<i>ya</i>	<i>ngi</i>	bona (bon + a)
	<i>u</i>	<i>nge</i>	<i>zo</i>	<i>ku</i>	
	<i>si</i>		<i>yo</i>	<i>si</i>	
	<i>ni</i>		<i>sa</i>	<i>ni</i>	
	<i>u/a</i>		<i>ka</i>	<i>m(u)</i>	
	<i>ba</i>		<i>nga</i>	<i>ba</i>	
	<i>u</i>			<i>wu</i>	
	.			.	
	.			.	
	.			.	
	(18)		(19, including the reflexive marker)		

Any verb root can be combined with any subject marker, any modal or aspectual morpheme plus a compatible final vowel, and any appropriate negative morpheme. If it is a transitive root, it can moreover be combined with any object marker (or the reflexive morpheme). The number of combinations possible for a suitable transitive verb stem is, therefore, $18 \times 19 \times 6 \times 2$.

The morphophonological processes to which verbal prefixes are subject in Zulu, are fairly simple too. These are again regular, and involve mainly vowel deletion and glide formation. The user is therefore assumed to know only such comparatively simple rules as $*u + a \rightarrow wa$ and $*zi + a \rightarrow za$. There are similar rules in the Sotho languages, Venda and Tsonga, but they occur in fewer contexts than in the Nguni languages.

None of this applies to the noun. Noun stems can be preceded by only one type of prefix, the class prefix. There are anything between 13 and 15 of these in the South African languages, excluding the three locative classes. In the Nguni languages class prefixes are in most contexts preceded by an individualizing morpheme which is euphonicly determined in that it duplicates the vowel of the prefix. In conjunctive word division prepositions, which are strictly speaking words and will therefore be ignored here, are joined to nouns. The structure of the Zulu noun can be schematized as follows:⁵

(Prep.)	IND	PREF	STEM
(na)	u	m(u)	zi
(nga)	a	ba	
(ku)	i	mi	
.	.	.	
.	.	.	
.	.	.	
		(13)	

Class prefixes are partially semantic and partially grammatical classifiers, comparable to the gender markers of European languages. The crucial difference with verbs is that *noun class prefixes are combined largely in an ad hoc manner with stems*. It is not possible to combine any noun stem freely with any class prefix as Paroz claims. In the case of the Zulu stem *-ntu*, for example, only four prefixes are found, i.e. *umu-ntu* 'person', *isi-ntu* 'African culture', *ubu-ntu* 'human nature' and *u(lu)-ntu* 'common people'. It is also difficult, and in some cases indeed impossible, to abstract the meaning of a noun stem from the meanings of the complete words in which it occurs. Thus Zulu *-ntu*, could be defined only vaguely and artificially as 'human' or 'African', whereas the meanings of verb stems can be defined or translated unambiguously.

This results in a *fundamentally different handling of verbs and nouns in stem dictionaries*. For any given verb stem only the positive infinitive form is given, with an unambiguous meaning or meanings. In the case of nouns, on the other

hand, it is necessary to specify all the prefixes with which each stem is found and to give the meaning of each combination of prefix plus stem separately. This means, on the one hand, that separate entries must be made for each combination of a prefix plus a stem, or that sub-entries added under the main entry, and that the meaning of each entry or sub-entry must be given separately. The stem of the word for a person, for instance, is handled as follows in four different stem dictionaries of South African languages:

DV: -**ntu** (**isintu**), [< umuntu.] African characteristics, culture.
 -**ntu** (**ubuntu**), [< umuntu.] Human nature.
 -**ntu** (**u(lu)ntu**), [< umuntu.] Common people.
 -**ntu** (**umuntu, abantu**) [**ubuntu; isintu; u(lu)ntu**] Human being.

ZM: -**THO, bo-** humanity.
 -**THO, mo-/ba-** human being.
 -**THO, n-** pl. **mantho** person, mankind.
 -**THO, se-** African culture.

Mabille et al:

mo.tho (ba.), human being, person.
setho, human manners or customs.
botho, condition of a person.

Snyman et al:

tho, bo-, personality, **mo-**, mankind, **mo- ba-**, person, human being, **se-**, humanely.

This brings no gain in economy compared with word dictionaries. The number of entries is the same for both types, the only difference being the structure and the alphabetic classification of the entries. Kriel, for example, has the following entries corresponding to those of ZM:

botho n., state of being a human being.
motho n., human being, a person.
ntho n., mankind.
setho n., human, human manners or customs.

This does certainly not avoid repetition, as Ziervogel would have it. The only advantage of stem dictionaries would then seem to be that they reveal the lexical relations between nouns with the same stem by grouping them together in one way or another.

It is questionable, however, whether linguistically less sophisticated users of a dictionary would be interested in lexical relations per se. Would the user of an English dictionary, for example, take kindly to the lemmatization of *perception*, *conception*, *reception* and *deception* as *-ception* just because they are lexically related? Or to *pre-empt* and *empty* being lemmatized as *-empt-* (as ZM would do)? Imagine entries like the following:

-ception (conception)	-CEPTION, con-	<i>con.ception</i>
-ception (deception)	-CEPTION, de-	<i>deception</i>
-ception (perception)	-CEPTION, per-	<i>perception</i>
-ception (reception)	-CEPTION, re-	<i>reception</i>

ception, con-, de-, per-, re-

The reader will recognize the formats of the four stem dictionaries referred to above in these examples.⁶

Let us assume that revealing lexical relations is nevertheless an important advantage and that the user is only required to get accustomed to an unfamiliar format. The question then arises whether stem dictionaries are more user friendly than word dictionaries in their handling of nouns. The answer to this seems also to be negative. The prefix morphology of nouns, apart from being irregular, is also subject to fairly complex morphophonological rules.

This is especially true of Classes 9 and 10, where the nasal of the prefix causes various changes in the initial consonants, and in some languages also the initial vowels, of stems. In Zulu the nasal causes affricativization of fricatives, [v], for example, becoming [bv]. This is not reflected in the orthography, however, with the fortunate result that stems with fricatives can be entered under the fricative letter, irrespective of whether it represents a fricative or an affricative. *Imvula* 'rain' and *inhloko* 'head' are therefore lemmatized as *-vula* and *-hloko* respectively in DV. Nasals also cause egressivization of the implosive labial [ɓ]. This change was reflected in the old orthography, but not in the current orthography, so that this does not cause problems either. DV were, however, compelled to enter *imbongi* 'praiser' as *-bongi*, since it is derived from the verb stem *-bonga* 'praise', but *imbuzi* 'goat' as *-buzi*, because it is a basic stem.

It is fortunate for stem dictionaries of Zulu that the orthography ignores the above-mentioned two morphophonological rules. The orthography is not so cooperative in the case of stems with aspirated explosives, however. Nasals require these consonants to be de-aspirated. This accounts for the fact that the plural of *inkosi* 'king' is *amakhosi*. This leaves the lexicographer with the dilemma to decide whether the stem of this word should be lemmatized as *-kosi* or as *-khosi*. DV decided on *-khosi* on the grounds that it represents the "basic" form, but lemmatized *inkabi* 'bull' as *-nkaḽi*, because there is no basic form **-khaḽi* corresponding to it.

The least this approach requires of the user, is that he should know the de-aspiration rule in order to look up nouns of Classes 9 and 10. This in itself does not, of course, decrease user friendliness seriously. The problem confronting the user, however, is that not all stems with voiceless explosives in Classes 9 and 10, are deverbatives like *impilo* 'health' which is derived from *phila* 'live', or have counterparts with the corresponding aspirated consonants in other classes, like *inkosi*. This forces the lexicographer to choose between four options. (1) He may lemmatize all such stems with aspirated consonants, irrespective of whether they occur in this form or not, e.g. *impala* 'impala', *intaba* 'mountain', *impilo* and *inkosi* as **-phala**, **-thaba**, **-thombi** and **-khosi** respectively. (2) He may enter all stems with unaspirated consonants, i.e. **-pala**, **-taba**, **-tombi** and **-kosi**. (3) He may enter those stems which also occur with aspirates under the aspirates, and stems which do not under the unaspirated explosives, e.g. **-thombi**, **-khosi**, **-pala** and **-taba**. (4) Or he may enter stems with corresponding aspirates under the aspirates, and those without under the relevant nasal compounds, e.g. **-thombi**, **-khosi**, **-mpala** and **-ntaba**.

DV, and virtually all other stem lexicographers, opt for the fourth option. This puts the onus on the user to know which nouns have stems to which the de-aspiration rule applies and which not. The result is quite confusing, as the following examples show:

NOUN	LEMMA
<i>impala</i> 'impala'	-mpala (impala)
<i>impilo</i> 'health'	-philo (impilo) [<phila]
<i>intaba</i> 'mountain'	-ntaba (intaba)
<i>intombi</i> 'girl'	-thombi (intombi) [<thomba]
<i>ubuntombi</i> 'maidenhood'	-ntombi [<intombi]
<i>inkosi</i> 'king'	-khosi (inkosi, amakhosi)
<i>inkabi</i> 'ox'	-nkabi (inkabi)

The paradox of this option is that it involves the abandonment of the stem principle and a compromise with the word principle, as the nasal which is included with such stems is in fact the prefix of Class 9, shorn of its individualizer.

The prefixes of Classes 3 and 14 are subject to palatalization in certain environments in Zulu. This masks the prefix to the extent that the uninitiated will not recognize it. **Umu-ezi*, for example, becomes *unyezi* 'moonlight' and **ubu-ala utshwala* 'beer'. Consequently the stem principle has again to be abandoned in favour the word principle by entering such nouns under the prefix without its individualizer. DV lemmatize these words as follows:

<i>unyezi</i>	-nyezi (unyezi)
<i>utshwala</i>	-tshwala (utshwala, ... with palatalization of original prefix <i>ubu-</i>)

The confusion is compounded in the case of nouns with irregular correspondences between their singular and plural forms. *Iso* 'eye' is entered by DV under its stem *-so*, with its plural *amehlo* given between brackets, i.e. *-so (i(li)so, ... amehlo)*. There is no corresponding stem entry for the plural form, however, evidently because of the difficulty of disentangling it orthographically from the underlying **ama+ihlo*.

It is no wonder, then, that Doke, with the admirable honesty typical of this great scholar, had to admit that this is a problem which requires more or less subjective decisions. Cf. DV (1948: viii):

There are cases where the stem-form is in doubt, particularly with certain nouns of Classes 9 and 10 with a nasal compound in the prefix. In such doubtful cases the entry has been made under the nasal. The word *intini* (otter) ... has been entered under *-thini*, the variant form *umthini* demonstrating that *-thini* is the stem; but in the case of *intindili* (useless person), the derivation being unknown, the entry is made under *-ntindili*. Naturally further information upon derivation *may necessitate a later alteration* in the position of certain words. (Italics mine, EBvW).

If the situation is confusing in Zulu, it is even more so in the Sotho languages, where more extensive changes are caused by the nasals of Classes 9 and 10, and where the process is complicated by a nasal deletion rule. The following examples from ZM illustrate the problems of looking up words under "stem" lemmas in Northern Sotho:

NOUN

kgomo 'bovine'
magomo 'cattle'
thaba 'mountain'
thuto 'lesson'
phala 'impala'
phošo 'mistake'

LEMMA

KGOMO, (n-)/di-
 -GOMO, ma- (magomo), cf. kgomo
 THABA, (n-)/di-
 thuto, (n-)di- v. RUTA
 PHALA, (n-)/di-
 -phošo, (n-)/di- v. FOŠA

This certainly does not prevent repetition. On the contrary it causes redundancy by having to resort to unnecessary cross-referencing.⁷

Certain nouns of Classes 1, 3 and 14 are subject to morphological changes which render the prefixes opaque. The following examples show how confusing ZM's handling of such nouns can be:

NOUN	STRUCTURE	LEMMA
<i>ngwana</i> 'child'	* <i>mo+ana</i>	NGWANA [cl. 1]
<i>mмуšo</i> 'government'	* <i>mo+bušo</i>	mмуšo pl. mebušo v. BUŠA
<i>mmutla</i> 'hare'	* <i>mo+butla</i>	MMUTLA pl. mebutla
<i>muši</i> 'smoke'	* <i>mo+uši</i>	muši v. muŠI
<i>moya</i> 'wind, spirit'	* <i>mo+oya</i>	MOYA ... pl. meoya
<i>meoya</i> 'winds, spirits'	* <i>me+oya</i>	-OYA, m-/me-
<i>bjako</i> 'haste'	* <i>bo+ako</i>	BJAKO ... (<*boako) ... cf. nako
<i>bjang</i> 'grass'	* <i>bo+ang</i>	BJANG pl. mabjang
<i>bjala</i> 'beer'	* <i>bo+ala</i>	BJALA [cl. bo-]

In most of these Northern Sotho examples, and there are many others, the lemmas are words, not stems. This proves that the stem principle cannot be applied consistently, in a "scientific" manner, to this language either.

In the case of Venda and Tsonga morphophonological rules applying to nouns are even more complicated than in the Sotho languages. The prefix of Class 3 is changed drastically in certain morphophonological contexts in Tsonga. Compare the following cases:

<i>nambu</i> 'river'	pl. <i>milambu</i>
<i>ndzoho</i> 'vegetables'	pl. <i>miroho</i>
<i>ndzhwalo</i> 'load'	pl. <i>mirhwalo</i>
<i>mongo</i> 'marrow'	(<i>mu + ongo</i>)
<i>munyu</i> 'salt'	(<i>mu + unyu</i>)
<i>n'wala</i> 'nail'	(<i>mu + ala</i>)
<i>mombo</i> 'face'	(<i>mu + ombo</i>)

In Venda the stems of Class 5 nouns are subject to extremely complicated morphophonological changes, e.g.

<i>fumi</i> 'ten'	pl. <i>mahumi</i>
<i>furi</i> 'pumpkin'	pl. <i>mafhuri</i>
<i>pfumo</i> 'spear'	pl. <i>mafumo</i>
<i>shambo</i> 'bone'	pl. <i>marambo</i>
<i>voho</i> 'arm'	pl. <i>mavhoho</i>
<i>dzembe</i> 'hoe'	pl. <i>malembe</i>
<i>gole</i> 'cloud'	pl. <i>makole</i>

In this language there is also a so-called partial denasalization rule applying to Class 20, which is responsible for word-pairs like the following:

<i>mulambo</i> 'river'	: <i>kudambo</i> 'small river'
<i>khali</i> 'pot'	: <i>kukali</i> 'small pot'
<i>lufhafha</i> 'wing'	: <i>kupapa</i> 'small wing'

When partial denasalization co-occurs with derivations in Class 5, the situation gets extremely complicated. Cf e.g.

<i>kholomo</i> 'bovine'	cl. 9
<i>golomo</i> 'large, ungainly bovine'	cl. 5
<i>makolomo</i> 'large, ungainly bovines'	cl. 6

All this can be handled in a stem dictionary, of course, but it will require considerable ingenuity if the stem principle is to be applied without resorting to word entries and extensive cross-referencing. It can only result in unnecessary repetition and increased user unfriendliness. No wonder, then, that no serious lexicographer has as yet attempted a stem dictionary for either of these two languages.

All these problems are avoided in word dictionaries, since nouns are entered as complete words, as they appear in the first columns of the examples above. One has to conclude, then, that the debatable gain of stem dictionaries in indicating lexical relations is more than offset by the complexity of the linguistic knowledge they assume, and the consequent loss of user friendliness.

The application of the stem principle results in still other linguistic inconsistencies. No stem dictionary enters pronouns such as Zulu *bona* 'they', *bonke* 'all' or *bodwa* 'only' under their stems. According to DV (1948: viii) it is impossible to do so. Yet the rules concerning their prefixal elements are the same as those that apply to subject markers of verbs. Morphophonologically there is no difference between, for instance, *bonke* (**ba+onke*) and the verb *bosa* (**ba+osa*) in *bosa inyama* 'they are frying (meat)'. Again, this implies a concession to the word principle.

For some strange reason which I cannot discover, derived adverbs are not listed under their stems either. *Kathathu* 'three times' and *kamnandi* 'pleasantly' are entered by DV as full words and not under the stems *-thathu* 'three' and *-mnandi* 'sweet', as one would expect. The Northern Sotho equivalents, *gararo* and *gamonate* are treated in the same manner by ZM. The word principle is again resorted to in these cases.

The final proof that the stem principle is based on erroneous linguistic assumptions lies in the fact that, unlike verbs, nouns are handled differently in monolingual dictionaries and, for instance, Zulu-English dictionaries, on the one hand, and, English-Zulu dictionaries on the other. In DV *umuntu* 'person', for example, is lemmatized as *-ntu*. In Doke, Malcolm and Sikakana, however, the translation of *person* is given as *umuntu*, not as the "scientifically correct" *-ntu*! This inconsistency is avoided in the word approach, since it appears in the same form in all these types of dictionaries.

It is clear, then, that neither the basic assumptions of, nor the claims made for stem dictionaries are valid. The stem principle is, in fact, less "scientific" than the word principle, since it is based on erroneous assumptions as regards noun morphology. It is not and cannot be applied consistently, as many nouns have of necessity to be lemmatized with their prefixes. Its only possible gain is

that it reveals lexical relations, which is in any case questionable from the point of view of the general user. In addition, it looses in user friendliness in that it assumes considerable linguistic sophistication on the part of the user. ZM (1975: 87) admit this by implication in the introduction to their dictionary, where they specify that Section 1, consisting of 8 pages of fine print, must be studied before using the dictionary.

No wonder, then, that ZM's stem dictionary, the only one of its kind in Northern Sotho, has proved to be unpopular with the general public. For the linguist it is a valuable source of information. If one wants to look up linguistic niceties and has enough time to do so, it is the obvious source to consult. If, however, one needs to look up the meaning of a Northern Sotho word, one is much better served by Kriel's supposedly unscientific word dictionary. And this includes the linguist.⁸

It has been argued, by Benson (1964) amongst others, that word dictionaries cause an imbalance in the alphabet classification through an accumulation of lemmas under specific letters. This is, of course, not a linguistic argument, but could conceivably bear on user friendliness. On closer analysis it proves to be unconvincing. If correct, it would mean that the letters **M**, **L**, **S** and **B** would carry an inordinately high number of entries in the Sotho languages, these being the consonants that occur in singular class prefixes of nouns. What are the facts, however?

The letter with the most entries in Snyman, Shole and Le Roux's stem dictionary of Tswana is **T**, accounting for 21,6% of all entries.⁹ In Matumo's word dictionary **M**, **L**, **S**, and **B** account for 13,7%, 6,6%, 11,9% and **B** for 4,4% of all entries respectively, and **T** for 11,9%. The lower percentage for **T** is largely due to the fact that noun stems starting on this letter are found under their prefix letters in Matumo.¹⁰

In Zulu the imbalance will be much more marked, since all nouns will be found under **U**, **I** and **A**, due to the occurrence of the euphonically determined individualizing morpheme. At a rough guess **U** would contain about 18% of all entries, with **I** accounting for a maximum of 20% and **A** for perhaps 5%, i.e. a total of approximately 43%. I am not sure that the high percentages for the first two vowels is a serious objection either. No language seems to have a balanced distribution of entries over the alphabet in the sense that each letter in a 26 letter alphabet has approximately 4% of all entries. The following percentages of letters with most entries, calculated from dictionaries in my possession, indicate that alphabetic imbalance is unavoidable in most, if not all, languages:

Afrikaans:	S	14%
English:	S	12,6%
Italian:	C	12,3%
Japanese:	H	18%
	S	15,3%
	T	10,1%
Nama:	G	21,8%

In the standard Romanized Japanese dictionary consulted, 43,4% of all entries are concentrated under three letters, which is about the highest percentage which U, I and A together would carry between them in a word dictionary of Zulu. Yet this does not seem to pose serious problems to users of the Japanese dictionary.

Should alphabetic imbalance nevertheless be felt to be an obstacle, there are options that could be considered to assist the user in overcoming the problem. One possibility is to introduce a hyphen after the individualizing morpheme, e.g. **u-muntu** 'person', **u-thando** 'love', **u-buntu** 'human nature', **u-muzi** 'village', **i-hashhi** 'horse', **i-silo** 'wild animal', **i-mvu** 'sheep', **a-manzi** 'water', **a-makhosi** 'kings'. Once under U, I or A, it will be easy to find the next consonant alphabetically.

Another option is to drop the individualizer and to enter nouns under the first letter of the prefix, e.g. **muntu**, **thando**, **buntu**, **muzi**, **hashhi**, **silo**, **mvu**, **manzi**, **makhosi**, etc. This is surely easier to handle than being required to drop the prefix, with all the attendant morphophonological problems. This format will bring the alphabetic distribution of word dictionaries of Zulu in line with the Sotho languages, where it appears to be quite normal.

It has to be concluded, then, that stem dictionaries are in no way superior to word dictionaries, and that claims on their behalf are not only exaggerated, but also unfounded. On the contrary, word dictionaries prove to reflect the differences between the morphology of nouns and verbs more accurately, and to be more consistent and more user friendly, i.e. they are based on sounder linguistic and lexicographical principles.

Notes

1. Vowel deletion and vowel coalescence occurs across word boundaries in Nguni languages, making it difficult to separate the words concerned orthographically. These rules apply rarely outside words in the Sotho languages, Venda and Tsonga.
2. Doke put forward his views in a number of publications. A fairly complete exposition is found in Doke 1932, and his definition of the Zulu word in Doke 1945: 28.
3. Word-formation is, in fact, treated by ZM in an extensive grammatical introduction to their dictionary.
4. Roots may be extended by adding suffixal morphemes such as the passive and perfect morphemes. Since these morphemes do not bear on the problem, they are ignored here.
5. Noun stems may include a diminutive and/or a locative suffix. These are ignored here, because they have no bearing on the problem.
6. I am, of course, aware that morphemes like *per-*, *con-*, *de-* and *re-* are not classifiers like the class prefixes of the African languages. The assumptions regarding noun morphology having been proved to be unfounded, it is the indication of lexical relations in the case of nouns which is at issue here. These examples are therefore to the point.

7. Unlike DV, ZM lemmatize deverbatives under the verb stems involved. This necessitates additional cross-references.
8. It must be mentioned that the lack of user friendliness of ZM is aggravated by a so-called phonemic classification, by extensive cross-referencing and by the inclusion of virtually all conceivable derived stems under verbal lemmas.
9. The high percentage is due to Snyman et al's strict adherence to alphabetical classification. The letter T represents no fewer than 8 phonemes in Tswana, viz. /t/, /th/, /tl/, /tlh/, /ts/, /tsh/, /tj/ and /tʃ/. Another contributing factor is the fact that Tswana has /tlh/ where other Sotho languages have /t/. All this applies to Matumo as well.
10. I realize, of course, that these percentages can be misleading, as they depend to a large extent on the authors' selection of lexical entries. They nevertheless serve to prove that the problem is not as serious as Benson implies.

References

- Benson, T.G. 1964. A Century of Bantu Lexicography. *African Language Studies* V: 64-91.
- Cole, D.T. 1955. *An Introduction to Tswana Grammar*. London: Longmans Green.
- Cuenod, R. 1976. *Tsonga-English Dictionary*. Braamfontein: Sasavona Books.
- Doke, C.M. 1935. *Bantu Linguistic Terminology*. London: Longmans Green.
- Doke, C.M. 1945. *Text-Book of Zulu Grammar*. London: Longmans Green.
- Doke, C.M., D.McK. Malcolm, and J.M.A. Sikakana. 1958. *English and Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- Doke, C.M. and S.M. Mofokeng. 1957. *Textbook of Southern Sotho Grammar*. Cape Town: Longman.
- Doke, C.M. and B.J. Vilakazi. 1948. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Du Plessis, J.A, J.G. Gildenhuys, and J.J. Moiloa. nd. *Bukantswe ya maleme-pedi*. Cape Town: Via Afrika.
- Kriel, T.J. 1976. *The New Northern Sotho Dictionary*. King William's Town: Educum.
- Mabille, A. and H. Dieterlen. 1937. *Sesuto-English Dictionary*. Morija: Sesuto Book Depot.
- Mabille, A., H. Dieterlen, and R.A. Paroz. 1974. *Southern Sotho-English Dictionary*. Morija: Sesuto Book Depot.
- Matumo, Z.I. 1993. *Setswana English Setswana Dictionary*. Gaborone: MacMillan Boleswa.
- Snyman, J.W., J.S. Shole and J.C. le Roux. 1990. *Dikišinare ya Setswana*. Pretoria: Via Afrika.
- Van Warmelo, N.J. 1989. *Venda Dictionary*. Pretoria: Government Printer.
- Van Wyk, E.B. 1994. C.M. Doke. A Critical Review by a Believing Outsider. Herbert, R.K. (Ed.). 1994. *Not with One Mouth*: 21-34. Johannesburg: Witwatersrand University Press.
- Ziervogel, D. 1965. Probleme van leksikografie in die Suid-Afrikaanse Bantoetale. *Taalfasette* 1: 45-53.
- Ziervogel, D. and P.C. Mokgokong. 1975. *Comprehensive Northern Sotho Dictionary*. Pretoria: JL van Schaik.