The Lexicographic Treatment of the Demonstrative Copulative in Sesotho sa Leboa — An Exercise in Multiple Cross-referencing[*]

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Ghent, Belgium and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (gillesmaurice.deschryver@UGent.be),*
Elsabé Taljard, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (etaljard@postino.up.ac.za),*
M.P. Mogodi, *Sesotho sa Leboa National Lexicography Unit, Pretoria Branch and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (pmogodi@postino.up.ac.za)*, and
Salmina Maepa, *Department of Arts and Culture and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (salmina.nong@dac.gov.za)*

Abstract: In this research article an in-depth investigation is presented of the lexicographic treatment of the demonstrative copulative (DC) in Sesotho sa Leboa. This one case study serves as an example to illustrate the so-called 'paradigmatic lemmatisation' of closed-class words in the African languages. The need for such an approach follows a discussion, in Sections 1 and 2 respectively, of the present and missing directions in African-language metalexicography. A theoretical conspectus of the DC in Sesotho sa Leboa is then offered in Section 3, while Section 4 examines the treatment of the DC in the four existing desktop dictionaries for this language. The outcomes from the two latter sections are then used in Section 5, which analyses the problems of and options for a sound lexicographic treatment of the DC in bilingual and monolingual dictionaries. The next two sections proceed with a review of the practical implementation of the DC lemmatisation suggestions in PyaSsaL, i.e. the *Pukuntšutlhaloši ya Sesotho sa Leboa* 'Explanatory Sesotho sa Leboa Dictionary' — with Section 6 focussing on the hardcopy and Section 7 on the online version. In the process, the very first fully monolingual African-language dictionary on the Internet is introduced. Section 8, finally, concludes briefly.

Keywords: LEXICOGRAPHY, PARADIGMATIC LEMMATISATION, AFRICAN LANGUAGES, SESOTHO SA LEBOA (NORTHERN SOTHO, SEPEDI), DEMONSTRATIVE COPULATIVE, CROSS-REFERENCING, CORPUS, MONOLINGUAL DICTIONARY, BILINGUAL DIC-

[*] An earlier version of this article was presented at the Eighth International Conference of the African Association for Lexicography, organised by the Department of Germanic and Romance Languages, University of Namibia, Windhoek, Namibia, 7–9 July 2003.

TIONARY, HARDCOPY, ONLINE, INTERNET, EXPLANATORY SESOTHO SA LEBOA DICTIONARY (PYASSAL), SIMULTANEOUS FEEDBACK (SF)

Senaganwa: Tokelotlhalošo ya lešalašupi-leba ka mo pukuntšung ya Sesotho sa Leboa — Tirišo ka go šupana go gontši. Ka go sengwalwana se sa nyakišišo, nyakišišo yeo e tseneletšego e laetšwa ka ga go lokelwa le go hlalošwa ga lešalašupi-leba ka mo pukuntšung ya Sesotho sa Leboa. Thutwana ya mohuta wo ya nyakišišo e šoma bjalo ka mohlala go laetša seo se bitšwago 'tokelo ya mantšu ka lenaneo' (paradigmatic lemmatisation) ya mantšu a legoro leo le tswaletšwego ka go maleme a Afrika. Tlhokego ya nyakišišo ya mohuta wo e tla ka morago ga therišano ya ditaetšo tša gonabjale le tšeo di sego gona ka go tlhamopukuntšu ya teori ya maleme a Afrika. Ditaba tše di hlalošwa ka go dikarolo 1 le 2. Tlhalošo ya teori ya lešalašupi-leba ka go Sesotho sa Leboa e fiwa ka go karolo 3, mola karolo 4 e hlahloba tokelo le tlhalošo ya lešalašupi-leba ka go dipukuntšu tše nne tšeo di lego gona mo polelong ye. Dipoelo tša dikarolo 3 le 4 di šomišwa karolong ya 5, yeo e sekasekago mathata le dikgonego tša tokelotlhalošo ya lešalašupi-leba ka go dipukuntšu tša polelopedi le tša polelotee. Dikarolo tše pedi tšeo di latelago di tšwela pele ka tekolo tirišong ya dikakanyetšo tša tšhomišo ya lešalašupi-leba ka go PyaSsaL, e lego *Pukuntšutlhaloši ya Sesotho sa Leboa*. Karolo 6 e lebane le taodišwana ya pampiri mola karolo 7 e lebane le taodišwana ya Inthanete. Ka go dira bjalo, pukuntšu ya mathomothomo ya polelotee ya maleme a Afrika e tsebagatšwa mo Inthaneteng. Mafelelong karolo 8 e fa kakaretšo ka bokopana.

Mantšu a bohlokwa: TLHAMOPUKUNTŠU, TOKELO YA MANTŠU KA LENANEO, MALEME A AFRIKA, SESOTHO SA LEBOA, LEŠALAŠUPI-LEBA, TŠHUPANO, KHOPHASE, PUKUNTŠU YA POLELOTEE, PUKUNTŠU YA POLELOPEDI, PUKUNTŠU YA PAMPIRI, KA GO INTHANETE, INTHANETE, PUKUNTŠUTLHALOŠI YA SESOTHO SA LEBOA (PYASSAL), SIMULTANEOUS FEEDBACK (SF)

1. Present directions in African-language metalexicography[1]

For over a decade now, African-language metalexicography has become increasingly popular in South Africa. No doubt, the new lexicographic dispensation in the now officially eleven-lingual South Africa has been instrumental in boosting interest in this field. Three main directions may be observed. Firstly it is noticed that a substantial number of corpus-based lexicographical studies for the African languages are being produced, starting with Prinsloo's (1991) 'computer-assisted word frequency studies', and culminating a decade later in a string of suggestions for corpus-building as well as considerations for dictionary-making on the macro- and microstructural levels (e.g. De Schryver and Prinsloo 2000b, 2000c, 2000d, 2001, 2003; Prinsloo and De Schryver 2001; De Schryver 2002). A second direction of research has been the development of concepts and tools for lexicography in the modern age. These include, inter alia, the concepts of *Simultaneous Feedback* or 'SF' (De Schryver 1999; De Schryver and Prinsloo 2000, 2000a) and *Fuzzy SF* (De Schryver and Prinsloo 2001a), as well as tools such as *Multidimensional Lexicographic Rulers* and *Block Systems* (Prinsloo and De Schryver 2002, 2003, 2004, 2004a; De Schryver 2003b) and the dictionary compilation software *TshwaneLex* (Joffe et al. 2003, 2003a). On a third level lemmatisation studies proper can be grouped. Research articles in this field are typically entitled 'Lemmatisation of ...', and this 'formula' has been particularly successful for Sesotho sa Leboa. As with corpora for the South African languages, the formula was first set out by D.J. Prinsloo, with colleagues following suit. By way of example, Table 1 lists the most influential attempts for Sesotho sa Leboa in this regard.

Table 1: The 'lemmatisation of ...'-formula in the case of Sesotho sa Leboa

Topic: Lemmatisation of ...	**Author(s)**	**Year**	**Journal/Proc.**
Reflexives	Prinsloo	1992	*Lexikos* 2
Verbs	Prinsloo	1994	*SAJAL* 14(2)
Verbs (..ga/sa/se..~ convention)	Prinsloo and Gouws	1996	*SAJAL* 16(3)
Adjectives	Gouws and Prinsloo	1997	*Lexikos* 7
Nouns	Prinsloo and De Schryver	1999	*SAJAL* 19(4)
Days	De Schryver and Lepota	2001	*Lexikos* 11
Verbs (freq.-based tail slots)	De Schryver and Prinsloo	2001	*Kiswahili* 2000
Abbreviated nouns	Bosch and Prinsloo	2002	*SAJAL* 22(1)
Loan words	Nong, De Schryver and Prinsloo	2002	*Lexikos* 12
Copulatives	Prinsloo	2002	*Lexikos* 12
Adverbs	Prinsloo	2003	*Lexikos* 13

As can be deduced from Table 1, with this formula African source-language lexical items belonging to distinct word classes are analysed, and suggestions for lemmatisation are offered for each of them, viz. for different types of verbs (Prinsloo 1992, 1994; Prinsloo and Gouws 1996; De Schryver and Prinsloo 2001b), for different types of nouns (Prinsloo and De Schryver 1999; Bosch and Prinsloo 2002), for adjectives (Gouws and Prinsloo 1997), for adverbs (Prinsloo 2003), for copulatives (Prinsloo 2002), and even for specific lexical sets (De Schryver and Lepota 2001) and loan words (Nong et al. 2002).

The 'lemmatisation of ...'-formula is typically concerned with methods to 'enter' or 'list' African-language items in the macrostructure of a dictionary, or thus in ways to 'lemmatise' various related forms under a single 'dictionary citation form'. This process can also be seen as a search for the most suitable 'canonical form' of each lexical item. Given that lexical items can successfully be grouped into types of words, known as 'grammatical classes' or 'parts of speech' (POSs), it does make sense to suggest lemmatisation approaches for each main type of African-language POS. Implicitly, the 'lemmatisation of ...'-formula is most relevant to dictionaries with an African language as source language, typically a bilingual dictionary treating an African language in the macrostructure with translation equivalents in a language of wider diffusion, or else an explanatory African-language dictionary. Note that even though this type of lemmatisation deals mainly with macrostructural aspects, certain suggestions also have representation repercussions on the microstructural level (such as, cf. Table 1, the introduction of the ..ga/sa/se..~ convention or of frequency-based tail slots).

2. Missing directions in African-language metalexicography

There seem to be two crucial aspects that have received very little if any metalexicographical attention so far, viz. (a) the treatment of an African language in the reverse side of a bilingual dictionary, where it is thus used for translating and/or paraphrasing another language, and (b) the 'paradigmatic lemmatisation' of closed-class words in the African languages. A metalexicographical discussion of a combination of (a) and (b) is of course even harder to come by.

These aspects can best be illustrated with an example from Rycroft's (1981) *Concise SiSwati Dictionary*. In the 100-million-word *British National Corpus* (BNC), 'its' — the third person singular possessive determiner (det-poss) — has a frequency of 163 081, which makes it the sixty-second most frequent word of the English language (Kilgarriff 1996). No dictionary with English as one of its (!) treated language pairs may thus omit this word, not even a junior dictionary (cf. De Schryver and Prinsloo 2003). However, the det-poss 'its' has not been entered into the *English–siSwati* side of Rycroft's dictionary. The (very) diligent dictionary user may however realise that 'its' can be derived from that dictionary's table in the front matter which summarises the concordial agreement system in siSwati. Table 2 shows the relevant (and simplified) section.

Table 2: Deriving the translation equivalents for 'its' in siSwati (patterned on Rycroft 1981: xxiv, with highlighting added)

3rd p:	**Possessive**		
Class	**Stem**	**Concord**	
1	-âkhe	w-(e/a)-	
2	-âbo	b-(e/a)-	
3	-âwo	w-(e/a)-	
4	-âyo	y-(e/a)-	POSSESSIVE STEMS: These, denoting the 'possessor', occur only with a possessive concord agreeing with the class of the item possessed. The initial vowels shown here were relegated to the concord by previous analysts. But as their tone (high or falling) is determined by the particular stem, it is better to include them with the stems.
5	-âlo	l-(e/a)-	
6	-âwo	ø-(e/a)-	
7	-âso	s-(e/a)-	
8	-âto	t-(e/a)-	
9	-âyo	y-(e/a)-	
10	-âto	t-(e/a)-	
11	-âlo	lw-(e/a)-	
14	-âbo	b-(e/a)-	
15	-âko	kw-(e/a)-	
16-18	-âko	kw-(e/a)-	

From Table 2 one may derive that the det-poss 'its' for a possessor from class 5 or 11 possessing something in classes 8 or 10 translates as *talo*, while it for instance becomes *bayo* for a possessor in class 9 possessing people (in class 2) or say abstract nouns (in class 14).[2]

With this brief example the very heart of the African-language system, in this case siSwati, is being touched. Indeed, it is not enough to lemmatise the items from the second column of Table 2 (*-âkhe* to *-âko*) in the siSwati–English side (as has been done in Rycroft's dictionary); when reversing the dictionary one must also provide the necessary clues under the relevant reversed entries. In this case, each highlighted possessive stem (*-âwo*, *-âlo*, *-âso*, *-âyo*, *-âlo*, *-âbo*, *-âko* and *-âko*) can combine with all the possessive concords (*w-(e/a)-* to *kw-(e/a)-*), which thus means that there are 8 times 14 or 112 possible ways to say 'its' in siSwati. In other words, in the reverse side of an African-language dictionary, the lemmatisation aspects/problems become microstructural design aspects/ problems. The dictionary compiler will have to decide whether or not to give all 112 forms in full, or only the frequent ones, and whether or not to provide examples for those treated. This decision will of course have to take the intended target user into account, and in today's dictionary landscape also whether the output is to paper (with severe space restrictions) or to an electronic format (with virtually no space restrictions).

In the previous discussion, the notion of 'paradigmatic lemmatisation' was implicit. Given that all nouns belong to classes in the African languages, and given that this membership drives the entire concordial agreement system, it is actually surprising that so little attention has been paid so far to this aspect in African-language metalexicography. The basic question is: If certain lexical items belong to a closed set or 'paradigm' driven by the class system, is it enough to randomly sample a few of its members for purposes of (a) lemmatisation, (b) microstructural representation in the source language, and, for a bilingual dictionary, (c) reversing the African-language source side? A random approach, even if based on intuition, does not seem appropriate in today's corpus-based/-driven lexicographical framework.

Paradigms of closed-class items, which by definition contain a *limited* number of members, are numerous in the African languages: object concords, subject concords, possessive concords, various types of demonstratives, various types of pronouns, etc. When there are moreover two dimensions, as in the case of possessives (cf. 'its' above), the paradigms may also become relatively large. Even though such paradigms are core blocks of the grammatical systems of the African languages — and thus important to both mother-tongue speakers learning another language (who e.g. have to be able to *map* 112 environments of 'its' on a single one) and learners of an African language (who have to do the reverse) — dictionaries notoriously fail to take the paradigms seriously.

Although paradigms are important when reversing a dictionary with an African language as the source language, whereby paradigms enter the microstructure on the reverse side, these will not be the focus of this article. Paradigms can further also play a role in the African-language source side, whether in a bilingual or a monolingual dictionary, as illustrated for the various concord and pronoun paradigms introduced in Prinsloo and De Schryver (2002a: 81, 2002b: 173-174, 177). These will not be the focus of this article either. In its most narrow sense 'paradigmatic lemmatisation' refers to the 'lemmatisation of ...'-formula, but now applied to paradigms. For the African languages these for instance include various types of adjectives, but also numerous more traditional closed-class words. One instance of the latter, namely the demonstrative copulative in Sesotho sa Leboa, will now be presented. It should be clear from the outset, however, that even though just one single case study is presented for one African language, the implications are generic.

3. The demonstrative copulative (DC) in Sesotho sa Leboa: A theoretical conspectus

Demonstrative copulatives are primarily nominal determiners, appearing in either the pre-nominal or post-nominal position, the post-nominal position being the dominant one. Like all nominal determiners in Sesotho sa Leboa, they can also function as pronominal forms in cases where the nominal antecedent is deleted. Although the term 'demonstrative copulative' (DC) is a somewhat cumbersome one, it does give an apt description of the semantic nature of these forms. Its demonstrative meaning is vested in the fact that it specifies the locality of a person or object relative to the position of the speaker and addressee in terms of different positions. It also has a copulative meaning and is thus loosely translated as 'here is/are', 'there is/are', etc., representing a predicative form of the demonstrative. Compare the examples in (1).

(1)

Bašemane *šeba*	'*Here are* the boys'	[*šeba* = DC position I, class 2]
Kgoši *šeo*	'*There is* the chief'	[*šeo* = DC position II, class 9]
Mohlare wo mogolo *šola*	'*Over there is* the big tree'	[*šola* = DC position III, class 3]

Sesotho sa Leboa grammarians do not agree as to the number of positions that are to be distinguished, nor is there any consistency in the numbering/labelling of the positions which they do distinguish. It would seem that all grammarians recognise at least three basic positions, i.e. (a) a basic form consisting of a root *š(e)-* (*se-* for class 7), followed by a concordial morpheme, (b) a second form consisting of the basic form to which a raised *-ô* has been suffixed, and (c) a third form, consisting of the basic form plus the suffix *-la* or *-lê*.[3] Compare Ziervogel et al. (1969: 85), Ziervogel and Mokgokong (1975: 104-105, *Introduction*), Louwrens (1994: 49) and Poulos and Louwrens (1994: 87) in this regard. For the purposes of the current discussion, it has been decided to split the third position into two, since it would seem that both variants could (theoretically) occur in every class. Compare the examples for classes 1/2 and 5/6 in Table 3.

Table 3: Basic DC positions I, II, III and IIIa for classes 1/2 and 5/6

Class	I	II	III	IIIa
1	šo	šoô	šola	šolê
2	šeba	šebaô	šebala	šebalê
5	šele	šeleô	šelela	šelelê
6	šea	šeaô	šeala	šealê

The different positions are characterised by highly specific semantic distinctions, with each suffix carrying a particular semantic nuance, which distinguishes it from the other suffixes. As has been mentioned, these demonstratives are used to pinpoint the position of some person(s) or object(s) in relation to the position of both speaker and addressee. Thus, the DC of position I would be used to refer to some referent(s) close to both speaker and addressee, who are in turn in close proximity to one another. The translational equivalent of these forms would therefore be 'here (s)he/it is, close to us', or in the case of the plural 'here they are, close to us'. The DC of position II would be used in a situation where the speaker and the addressee are relatively far apart, while the person(s) or object(s) referred to is/are nearer to the addressee, but not right next to him/her. It would thus be translated as 'there (s)he/it is, close to you' or 'there they are, close to you'. Should the addressee and the speaker be in very close proximity to one another and the person(s) or object(s) being referred to is/are far away from the interlocutors, the DC of position III or IIIa would be used, carrying the meaning of 'there (s)he/it is, over yonder' or 'there they are, over yonder'.

Four additional forms are also mentioned in the literature, but these are generally regarded as dialectal forms. The first two of these four additional forms, the occurrence of which specifically excludes the eastern dialects (Kotzé 1985: 85), consist of the basic form (position I) to which the suffix *-nô* is added. The suffix *-khwi* occurs as a further dialectal variant of *-nô*. These two forms are labelled Ia and Ib respectively in Table 4.

Table 4: Basic DC positions I, II, III and IIIa, plus dialectal variants Ia and Ib, for classes 1/2 and 5/6

Class	I	**Ia**	**Ib**	II	III	IIIa
1	šo	**šonô**	**šokhwi**	šoô	šola	šolê
2	šeba	**šebanô**	**šebakhwi**	šebaô	šebala	šebalê
5	šele	**šelenô**	**šelekhwi**	šeleô	šelela	šelelê
6	šea	**šeanô**	**šeakhwi**	šeaô	šeala	šealê

According to Kotzé (1985: 85) and Louwrens (1991: 106), the DCs with suffixes *-nô* and *-khwi* would be used in a situation where the interlocutors are at a distance from one another and where the person(s) or object(s) being referred to is/are right next to the speaker. The meaning expressed by these forms is thus 'here (s)he/it is, right next to me' or 'here they are, right next to me'.

The other two additional forms are also regarded as dialectal, with the suffix *-uwê* found in the dialects spoken in the vicinity of Polokwane, and its variant *-wê* found only in the dialects of Sekhukhuniland, particularly the Sepedi dialect (Ziervogel and Mokgokong 1975: 104, *Introduction*; Kotzé 1985: 86). These are labelled IIa and IIb respectively in Table 5.

Table 5: Basic DC positions I, II, III and IIIa, dialectal variants Ia and Ib, plus dialectal variants IIa and IIb, for classes 1/2 and 5/6

Class	I	Ia	Ib	II	**IIa**	**IIb**	III	IIIa
1	šo	šonô	šokhwi	šoô	**šouwê**	**šowê**	šola	šolê
2	šeba	šebanô	šebakhwi	šebaô	**šebauwê**	**šebawê**	šebala	šebalê
5	šele	šelenô	šelekhwi	šeleô	**šeleuwê**	**šelewê**	šelela	šelelê
6	šea	šeanô	šeakhwi	šeaô	**šeauwê**	**šeawê**	šeala	šealê

In a situation where the speaker and addressee are quite far apart from one another, these demonstratives would be used to refer to an object that is very close or directly next to the addressee. It can therefore be translated as 'there (s)he/it is, right next to you' or 'there they are, right next to you'.

Not all grammar books list the full paradigm for all the classes of Sesotho sa Leboa, and scrutiny of those that do, reveals that a number of differences exist with regard to especially the forms to be distinguished for classes 15 to 18. Ziervogel et al. (1969: 86), Lombard et al. (1985: 166), Nokaneng and Louwrens (1988: 221) and Poulos and Louwrens (1994: 88) indicate that the basic DC for class 15 is *šefa*, whereas Ziervogel and Mokgokong (1975: 104, *Introduction*) list *šego*. With regard to the basic DC for classes 16 to 18, Ziervogel and Mokgokong (1975: 104, *Introduction*) again differ from other scholars in that they are the only ones who distinguish the form *šego* for class 17. All other grammars indicate that the demonstrative for these classes is *šefa*. A search through a 6.1-million-word Sesotho sa Leboa corpus, henceforth 'the corpus', brought no results for a DC *šego*, thus the information provided by Ziervogel and Mokgokong with regard to both class 15 and class 17 seems to be unverified. Interestingly though, a basic demonstrative form not listed in any of the standard Sesotho sa Leboa grammars was thrown up by a corpus search. Indeed, seven instances were found of a demonstrative ***šemo***, in all probability belonging to class 18, with a meaning similar to that of *šefa*, i.e. 'here (s)he/it is, here they are, here I am, etc.'

Another DC identified during a corpus search is ***šetše***, containing the raised vowel [ẹ], i.e. [ʃẹtʃi]. This seems to be a variant of *šidi,* which is the so-called standard DC for classes 8 and 10. The two examples of *šetše* that were found in the corpus both have nouns in class 10 as antecedents, thus making it difficult to ascertain whether *šetše* could also have a noun from class 8 as antecedent. According to L.J. Louwrens (*personal communication*, 26 May 2004), this particular variant is found as *sitši* in the Tlokwa dialect as a DC for both classes 8 and 10. Mother-tongue speakers also agree that this form may be used together with an antecedent from class 8. They furthermore indicate that the use of this form is widespread in the spoken language but, due to its non-standard status, is replaced in the written language with the standardised form.

In the last instance, it was noticeable that the variant ***šedi*** of the DC *šidi* for classes 8 and 10 is more often used than the so-called standard form. All grammar books list only *šidi* and both Ziervogel et al. (1969: 86) and Lombard et al. (1985: 166) clearly state that in the case of classes 8 and 10, assimilation between the vowel [e] of the root and the [i] of the concordial morpheme takes place, resulting in the assimilated form *šidi*. Despite the standard status of this form, only 36 concordance lines were found in the corpus in which *šidi* appears. Its distribution is furthermore relatively limited in that it is found in only 7 different sources. The 'non-standard' form *šedi*, however, appears in 209 concordance lines, spread across 56 different sources.

The full paradigm for the DC in Sesotho sa Leboa is therefore as shown in Table 6.

Table 6: Full paradigm for the DC in Sesotho sa Leboa (using the standard orthography) [**bold** = frequency of at least three in a 6.1-million-word corpus; *italics* = frequency of two and one in the corpus; (...) = only occurs in grammars]

	I	**Ia**	**Ib**	**II**	**IIa**	**IIb**	**III**	**IIIa**
1	**šo**	**šono**	**šokhwi**	**šoo**	(šouwe)	**šowe**	**šola**	**šole**
2	**šeba**	*šebano*	(šebakhwi)	**šebao**	(šebauwe)	*šebawe*	**šebala**	**šebale**
3	**šo**	(šono)	*šokhwi*	**šoo**	(šouwe)	*šowe*	**šola**	**šole**
4	**še**	*šeno*	(šekhwi)	**šeo**	(šeuwe)	(šewe)	(šela)	(šele)
5	**šele**	(šeleno)	(šelekhwi)	**šeleo**	(šeleuwe)	(šelewe)	(šelela)	*šelele*
6	**šea**	(šeano)	(šeakhwi)	**šeao**	(šeauwe)	(šeawe)	(šeala)	**šeale**
7	**sese**	(seseno)	(sesekhwi)	**seseo**	(seseuwe)	(sesewe)	*sesela*	(sesele)
8	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	*šedile*
8'	**šidi**	(šidino)	(šidikhwi)	(šidio)	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
8''	(šetše)	(šetšeno)	(šetšekhwi)	(šetšeo)	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
9	**še**	(šeno)	**šekhwi**	**šeo**	(šeuwe)	**šewe**	**šela**	**šele**
10	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	**šedile**
10'	**šidi**	(šidino)	(šidikhwi)	**šidio**	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
10''	*šetše*	(šetšeno)	(šetšekhwi)	*šetšeo*	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
14	**šebo**	(šebono)	(šebokhwi)	(šeboo)	(šebouwe)	*šebowe*	(šebola)	(šebole)
15	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
16	**šefa**	(šefano)	*šefakhwi*	**šefao**	(šefauwe)	(šefawe)	(šefala)	(šefale)
17	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
18	**šemo**	(šemono)	(šemokhwi)	(šemoo)	(šemouwe)	(šemowe)	(šemola)	(šemole)

Even though all forms enumerated in Table 6 are theoretically possible in Sesotho sa Leboa it is important to recall that, given that some positions are only found in particular dialects, one will not normally find all of them used by any single speaker. Actually, when comparing each DC against a 6.1-million-word corpus of Sesotho sa Leboa, only those forms marked in bold and italics do occur, with the bold items having a frequency of three or more, and the italicised ones a frequency of two or one only. Bracketed items have a zero frequency. Note further that each DC may be prefixed by the non-standard *a-* in the spoken language, effectively doubling the (theoretical) size of Table 6. In the corpus, however, not a single example was found of these *a-* forms. Given that dictionaries are mainly based on the standardised language, and that this is also the orthography represented in the corpus (since the corpus contains mainly written material), it is defendable not to include any of these *a-* forms in Sesotho sa Leboa dictionaries. If only a selection of members of the DC paradigm is to be treated in a dictionary, it further seems logical to focus on the truly frequent ones (i.e. the 42 bold items in Table 6).

4. Treatment of the DC in the four existing desktop dictionaries for Sesotho sa Leboa

An investigation into the treatment of the DC in the four existing desktop dictionaries for Sesotho sa Leboa brings to light that these forms are dealt with in an inconsistent and sometimes even idiosyncratic manner. The four dictionaries, three of them bilingual and one trilingual, are (a) the third edition of the *Pukuntšu woordeboek* (Kriel 1983³), (b) the revision by Van Wyk of the *Pukuntšu woordeboek* (Kriel et al. 1989⁴) — being the latest edition of this dictionary, (c) the last edition of *The New English–Northern Sotho Dictionary* (Kriel 1976⁴), and (d) the trilingual *Comprehensive Northern Sotho Dictionary* (Ziervogel and Mokgokong 1975) — which saw only one edition.

4.1 Treatment of the DC in Kriel's (1983³) *Pukuntšu woordeboek, Noord-Sotho–Afrikaans*

One of the first problems one encounters when investigating the *Pukuntšu woordeboek* is the fact that indication of class membership is inconsistent and erratic, and the user is often left to his/her own devices to ascertain to which class a particular DC belongs, the only clue being provided by the translation equivalent (TE). The correct interpretation of the information contained in the TE furthermore often presupposes a thorough grammatical background. A case in point is the treatment of *šo*, which is the DC position I for both classes 1 and 3. The TE provided is *hier is hy/sy* 'here he/she is' which in all likelihood will be interpreted as referring to class 1, since class 1 is the one class containing nothing but nouns referring to humans. This implies that the DC position I of class 3 is not treated. In the case of *še*, which can be either class 4 or class 9, no indication of class membership is given. The TE *hier is dit* 'here it is', however, suggests class 9, since it refers to the singular, class 4 being a plural class. Again, the implication is that the DC position I of class 4 is not treated in the dictionary. With regard to classes 8 and 10, the form *šidi* is listed, but with a label *ou spelling* 'old spelling' and with a cross-reference to *šedi*. This is contrary to the information uniformly provided by the Sesotho sa Leboa grammars that *šidi* is the standard form. Furthermore, class membership is incorrectly indicated as class 7 instead of class 8, while the DC of class 10 remains untreated. For the locative classes (16 to 18) only the form *šefa* is listed, with no class membership indicated and the TE given as *hier is dit* 'here it is'. For position Ia only one DC is treated, i.e. *šono*, which can belong to either class 1 or 3, the TE *hier is hy/sy* 'here he/she is' again suggesting class 1. None of the position Ib forms is found in the dictionary.

For position II the DCs of classes 5 (*šeleo*), 6 (*šeao*), 7 (*seseo*), 9 (*šeo*) and 16 (*šefao*) are treated. Under the entry *šeo* (class 9), mention is made of *šewe*, the DC position IIb of class 9. This is the only (implicit) cross-reference found to any of the DCs of position IIb in this particular dictionary; *šewe* itself, however, just as all other DCs from IIb, has not been entered. With regard to *šoo* the TE again suggests class 1 only, leaving the DC of class 3 untreated. It is further noticeable that although no DC position I *šego* is listed, the position II form *šegoo* is treated, albeit without any indication of class membership. No entries are found for any of the position IIa DCs. This is to a certain extent understandable, since the only grammar in which these variants are enumerated, is that of Nokaneng and Louwrens (1988: 221). These forms might have been regarded as (non-standard) dialectal forms and might therefore not have been included in the dictionary.

The DC position III class 1 listed is *šolaa*, a spelling not attested by any of the other sources.[4] Again, class membership is vested in the TE *daar is hy/sy* 'there he/she is'. The only other form of position III that is treated is *šegola*, but yet again without class indication, and this even though the basic form *šego* was not treated. For position IIIa, finally, only the DCs for classes 1 and 2 are listed, as well as *šefale* for the locative classes. A summary of these findings is presented in Table 7.

Table 7: Treatment of the DC in Kriel's (1983^3) *Pukuntšu woordeboek, Noord-Sotho–Afrikaans* [dark shade = lemmatised; light shade = wrongly spelt]

	I	**Ia**	**Ib**	**II**	**IIa**	**IIb**	**III**	**IIIa**
1	**šo**	**šono**	**šokhwi**	**šoo**	(šouwe)	**šowe**	**šola**	**šole**
2	**šeba**	*šebano*	(šebakhwi)	**šebao**	(šebauwe)	*šebawe*	**šebala**	**šebale**
3	**šo**	(šono)	*šokhwi*	**šoo**	(šouwe)	*šowe*	**šola**	**šole**
4	**še**	*šeno*	(šekhwi)	**šeo**	(šeuwe)	(šewe)	(šela)	(šele)
5	**šele**	(šeleno)	(šelekhwi)	**šeleo**	(šeleuwe)	(šelewe)	(šelela)	*šelele*
6	**šea**	(šeano)	(šeakhwi)	**šeao**	(šeauwe)	(šeawe)	(šeala)	**šeale**
7	**sese**	(seseno)	(sesekhwi)	**seseo**	(seseuwe)	(sesewe)	*sesela*	(sesele)
8	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	*šedile*

8'	**šidi**	(šidino)	(šidikhwi)	(šidio)	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
8''	(šetše)	(šetšeno)	(šetšekhwi)	(šetšeo)	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
9	**še**	(šeno)	**šekhwi**	**šeo**	(šeuwe)	**šewe**	**šela**	**šele**
10	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	**šedile**
10'	**šidi**	(šidino)	(šidikhwi)	**šidio**	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
10''	*šetše*	(šetšeno)	(šetšekhwi)	*šetšeo*	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
14	**šebo**	(šebono)	(šebokhwi)	(šeboo)	(šebouwe)	*šebowe*	(šebola)	(šebole)
15	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
16	**šefa**	(šefano)	*šefakhwi*	**šefao**	(šefauwe)	(šefawe)	(šefala)	(šefale)
17	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
18	**šemo**	(šemono)	(šemokhwi)	(šemoo)	(šemouwe)	(šemowe)	(šemola)	(šemole)

From Table 7 it is clear that the DCs were lemmatised in a rather haphazard way in this dictionary. If corpus statistics are used as a guideline, then 42 forms (the bold ones) should be entered. In this dictionary, only 25 were lemmatised, including the wrongly spelt form. The overlap, however, is only 20 out of 42, or thus 48%. Conversely, 20 out of 25 lemmatised forms, or 80%, means that Kriel did a rather remarkable job on intuition alone.

4.2 Treatment of the DC in Kriel, Van Wyk and Makopo's (1989^4) *Pukuntšu woordeboek, Noord-Sotho–Afrikaans*

This revision of the *Pukuntšu woordeboek* does represent some improvement in the lexicographical treatment of DCs. Class membership, for instance, is explicitly stated, thus easing the burden placed on the target user. Unfortunately, in a number of cases indication of class membership is incorrect. For position I, the DCs of classes 1 to 14 are treated consistently; however, the form *še* is indicated as belonging to classes 3 and 9, which is incorrect as it should be 4 and 9. Also, they claim that the DCs *šea* and *šeao* belong to class 8, whereas these forms are in fact the position I and II DCs for class 6. For classes 8 and 10, both *šedi* and *šidi* are given, but not *šetše*. No DC for class 15 is listed, and for the locative classes only *šefa* is treated. As far as the treatment of DCs of position Ia is concerned, only *šono* belonging to classes 1 and 3 is included. No DCs of position Ib have been lemmatised. With regard to DCs of position II, those of classes 2, 8, 10 and 14 have not been entered and, as was pointed out previously, *šeao* is incorrectly labelled with regard to class membership. No DC for class 15 is recorded and for the locative classes *šefao* and *šegoo* are treated. As was the case for the third edition of this dictionary, it is thus noticeable that the DC position II of class 17 is treated, but not the corresponding form for position I, i.e. *šego*. No DCs of positions IIa and IIb were found in the central lemma-sign list of this dictionary. Of all the position III DCs, only the form for class 17 is treated, i.e. *šegola*. With regard to position IIIa DCs, only those forms belonging to classes 1, 2, 3 and 16 have been entered into the dictionary. Compare Table 8 for a summary.

Table 8: Treatment of the DC in Kriel, Van Wyk and Makopo's (1989^{4}) *Pukuntšu woordeboek* [dark shade = lemmatised; light shade = class wrongly assigned]

	I	Ia	Ib	II	IIa	IIb	III	IIIa
1	**šo**	**šono**	**šokhwi**	**šoo**	(šouwe)	**šowe**	**šola**	**šole**
2	**šeba**	*šebano*	(šebakhwi)	**šebao**	(šebauwe)	*šebawe*	**šebala**	**šebale**
3	**šo**	(šono)	*šokhwi*	**šoo**	(šouwe)	*šowe*	**šola**	**šole**
4	**še**	*šeno*	(šekhwi)	**šeo**	(šeuwe)	(šewe)	(šela)	(šele)
5	**šele**	(šeleno)	(šelekhwi)	**šeleo**	(šeleuwe)	(šelewe)	(šelela)	*šelele*
6	**šea**	(šeano)	(šeakhwi)	**šeao**	(šeauwe)	(šeawe)	(šeala)	**šeale**
7	**sese**	(seseno)	(sesekhwi)	**seseo**	(seseuwe)	(sesewe)	*sesela*	(sesele)
8	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	*šedile*
8'	**šidi**	(šidino)	(šidikhwi)	(šidio)	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
8"	(šetše)	(šetšeno)	(šetšekhwi)	(šetšeo)	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
9	**še**	(šeno)	**šekhwi**	**šeo**	(šeuwe)	**šewe**	**šela**	**šele**
10	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	**šedile**
10'	**šidi**	(šidino)	(šidikhwi)	**šidio**	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
10"	*šetše*	(šetšeno)	(šetšekhwi)	*šetšeo*	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
14	**šebo**	(šebono)	(šebokhwi)	(šeboo)	(šebouwe)	*šebowe*	(šebola)	(šebole)
15	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
16	**šefa**	(šefano)	*šefakhwi*	**šefao**	(šefauwe)	(šefawe)	(šefala)	(šefale)
17	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
18	**šemo**	(šemono)	(šemokhwi)	(šemoo)	(šemouwe)	(šemowe)	(šemola)	(šemole)

From Table 8 it is clear that as far as the lemmatisation proper is concerned, this fourth edition does not really improve much on the third edition. Including the three wrongly assigned forms, only 30 forms were lemmatised. The overlap with the frequent forms is 26 out of 42, or thus 54%. With 26 out of the 30 lemmatised forms also being frequent, the intuition score is however as high as 87%.

4.3 Treatment of the DC in Kriel's (1976^{4}) *The New English–Northern Sotho Dictionary, Northern Sotho–English*

Whereas the two *Pukuntšu* editions above treated Afrikaans as second language pair, *The New English–Northern Sotho Dictionary* is currently the only bidirectional desktop dictionary with English and Sesotho sa Leboa as treated language pairs. Unfortunately, users of this dictionary are presented with an even more erratic treatment of the DCs. Table 9 summarises the lemmatisation status.

Table 9: Treatment of the DC in Kriel's (1976^4) *The New English–Northern Sotho Dictionary, Northern Sotho–English* [dark shade = lemmatised; light shade = wrongly spelt]

	I	Ia	Ib	II	IIa	IIb	III	IIIa
1	**šo**	**šono**	**šokhwi**	**šoo**	(šouwe)	**šowe**	**šola**	**šole**
2	**šeba**	*šebano*	(šebakhwi)	**šebao**	(šebauwe)	*šebawe*	**šebala**	**šebale**
3	**šo**	(šono)	*šokhwi*	**šoo**	(šouwe)	*šowe*	**šola**	**šole**
4	**še**	*šeno*	(šekhwi)	**šeo**	(šeuwe)	(šewe)	(šela)	(šele)
5	**šele**	(šeleno)	(šelekhwi)	**šeleo**	(šeleuwe)	(šelewe)	(šelela)	*šelele*
6	**šea**	(šeano)	(šeakhwi)	**šeao**	(šeauwe)	(šeawe)	(šeala)	**šeale**
7	**sese**	(seseno)	(sesekhwi)	**seseo**	(seseuwe)	(sesewe)	*sesela*	(sesele)
8	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	*šedile*
8'	**šidi**	(šidino)	(šidikhwi)	(šidio)	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
8"	(šetše)	(šetšeno)	(šetšekhwi)	(šetšeo)	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
9	**še**	(šeno)	**šekhwi**	**šeo**	(šeuwe)	**šewe**	**šela**	**šele**
10	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	**šedile**
10'	**šidi**	(šidino)	(šidikhwi)	**šidio**	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
10"	*šetše*	(šetšeno)	(šetšekhwi)	*šetšeo*	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
14	**šebo**	(šebono)	(šebokhwi)	(šeboo)	(šebouwe)	*šebowe*	(šebola)	(šebole)
15	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
16	**šefa**	(šefano)	*šefakhwi*	**šefao**	(šefauwe)	(šefawe)	(šefala)	(šefale)
17	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
18	**šemo**	(šemono)	(šemokhwi)	(šemoo)	(šemouwe)	(šemowe)	(šemola)	(šemole)

Including two wrongly spelt forms, 19 DCs were entered in this dictionary, which, with an overlap of 15 items with the frequent forms, gives a coverage of 15 out of 42, or only 36%. The intuition score is still rather high, however, at 79% (15 out of 19). As was the case for the third edition of the *Pukuntšu*, there is virtually no class information, but more importantly, the labelling and data provided are anything but consistent. Apart from the fact that the POS of all forms is given as 'dem.', which means that these demonstrative *copulatives* are not distinguished from the real demonstratives, some are labelled 'pron.' and in addition even 'adj.' Other forms, although variants, are nonetheless treated very differently, e.g. '**še′diˈle,** dem., pl., there they are, yonder.' versus '**šidiˈle-e,** dem., there they are.' Or as a last example: '′**šo,** dem., here he is, here she is.' versus '′**šono,** dem., here he (she) is.'

4.4 Treatment of the DC in Ziervogel and Mokgokong's (1975) *Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*

This trilingual dictionary is the only one of the desktop dictionaries that treats the prefixal element of the DCs. Both variants, i.e. *še-* and *ši-* are found in the lemma-sign list: *še***-** is fully treated and defined as a 'prefixal element in formation of cop. dem.', whereas *ši***-** is defined as the 'assimilated form of **še-**', thus implicitly cross-referring the user to the canonical form of the prefix, i.e. *še*-. In this dictionary, the DCs of position I for classes 1 to 14 are consistently treated, with indication of class adherence. Only the dialectal variant *šetše* for classes 8 and 10 is not listed. Although both other forms of the DCs of classes 8 and 10 are found in the lemma-sign list, the 'non-standard' *šedi* is not treated, but cross-referred to the 'standard' form *šidi*, which is then fully treated. No DC is distinguished for class 15. For the locative classes, only *šefa* is listed. The DCs of positions Ia and Ib do not appear in this dictionary, and neither do those of positions IIa and IIb. The same DCs which are treated under position I, are also treated for position II, with the addition of *šegoo*, which is distinguished for the locative classes. It is labelled as a dialectal form, but it is (again) not clear why the position I form (*šego*) is not treated. The form *šedio* (classes 8 and 10) is listed, but cross-referred to *šidio*, which is then treated. With regard to position III, all DCs of classes 1 to 14 are treated, with the exception of those of classes 2 and 6. For the locative classes, only *šegola* is listed, again labelled as a dialectal form. For position IIIa, only classes 1 and 3 (*šole*), class 2 (*šebale*), class 6 (*šeale*) and class 16 (*šefale*) have been lemmatised. Compare Table 10 for an overview.

Table 10: Treatment of the DC in Ziervogel and Mokgokong's (1975) *Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English* [dark shade = lemmatised; light shade = wrongly spelt]

	I	Ia	Ib	II	IIa	IIb	III	IIIa
1	**šo**	**šono**	**šokhwi**	**šoo**	(šouwe)	**šowe**	**šola**	**šole**
2	**šeba**	*šebano*	(šebakhwi)	**šebao**	(šebauwe)	*šebawe*	**šebala**	**šebale**
3	**šo**	(šono)	*šokhwi*	**šoo**	(šouwe)	*šowe*	**šola**	**šole**
4	**še**	*šeno*	(šekhwi)	**šeo**	(šeuwe)	(šewe)	(šela)	(šele)
5	**šele**	(šeleno)	(šelekhwi)	**šeleo**	(šeleuwe)	(šelewe)	(šelela)	*šelele*
6	**šea**	(šeano)	(šeakhwi)	**šeao**	(šeauwe)	(šeawe)	(šeala)	**šeale**
7	**sese**	(seseno)	(sesekhwi)	**seseo**	(seseuwe)	(sesewe)	*sesela*	(sesele)
8	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	*šedile*
8'	**šidi**	(šidino)	(šidikhwi)	(šidio)	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
8''	(šetše)	(šetšeno)	(šetšekhwi)	(šetšeo)	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
9	**še**	(šeno)	**šekhwi**	**šeo**	(šeuwe)	**šewe**	**šela**	**šele**
10	**šedi**	(šedino)	(šedikhwi)	**šedio**	(šediuwe)	(šediwe)	(šedila)	**šedile**
10'	**šidi**	(šidino)	(šidikhwi)	**šidio**	(šidiuwe)	(šidiwe)	(šidila)	*šidile*
10''	*šetše*	(šetšeno)	(šetšekhwi)	*šetšeo*	(šetšeuwe)	(šetšewe)	(šetšela)	(šetšele)
14	**šebo**	(šebono)	(šebokhwi)	(šeboo)	(šebouwe)	*šebowe*	(šebola)	(šebole)
15	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
16	**šefa**	(šefano)	*šefakhwi*	**šefao**	(šefauwe)	(šefawe)	(šefala)	(šefale)
17	(šego)	(šegono)	(šegokhwi)	(šegoo)	(šegouwe)	(šegowe)	(šegola)	(šegole)
18	**šemo**	(šemono)	(šemokhwi)	(šemoo)	(šemouwe)	(šemowe)	(šemola)	(šemole)

As can be seen from Table 10, a total of 46 DCs (which includes two misspelled ones) were lemmatised in this dictionary, 33 of which belong to the frequent DCs. With 33 out of 42 frequent items, the coverage is rather satisfactory at 79%; the intuition ratio stands at slightly less with 72% (33 out of 46).

Of all the desktop dictionaries, this reference work is the only one that has a relatively extended front matter, including a mini-grammar of Sesotho sa Leboa. In this section, the DC is discussed, but the full paradigm is only provided for position I. For some other positions, information is provided on their morphological composition only. Even though the DCs of positions Ia (-*nô*) and Ib (-*khwi*) do not appear in the central lemma-sign list, reference is made in the front matter to these forms, but not to the DCs of positions IIa (-*uwê*) and IIb (-*wê*). Unfortunately, there is no system of cross-referencing from the central lemma-sign list to the information provided in the front matter. This is particularly relevant for the DCs of positions Ia and Ib, which are not treated in the central text, but which are discussed in the front matter. Even if these forms are not treated in the dictionary proper, they could simply have been entered with a cross-reference to the information provided in the front matter.

5. Towards a sound lexicographic treatment of the DC: Problem analysis and options

From the above review of the treatment of the DC in the current desktop dictionaries for Sesotho sa Leboa, one actually realises that there are three main issues at stake. Firstly there is the problem of consistency, secondly there is the problem of which data categories to include and how, and thirdly there is the problem of which DCs to treat and where.

5.1 The problem of consistency

As far as consistency is concerned, a quick glance at an overview of the actual treatment of the DC in the four desktop dictionaries, reproduced *verbatim* in Addendum 1, makes the current erratic approach very evident. Even the treatment that clearly received the most detailed attention, viz. the one found in Ziervogel and Mokgokong's (1975) *Comprehensive Northern Sotho Dictionary*, is still full of inconsistencies. The very label for the POS, to begin with, is found 22 times as 'dem. kop.' versus 8 times as 'kop. dem.' (plus once as 'dem. kop ') in Afrikaans, and 22 times as 'dem. cop.' versus 8 times as 'cop. dem.' (plus once as 'dem. cop ') in English. On the semantic level one finds 'there are they over there' for *šebale* where it should be 'there they are over there', or 'doer is dit' at *sesela* where it should be 'dáár is dit'. On the cross-reference level, the abbreviation 'v.' is for instance used at *še-*, but 'cf.' at *ši-*. Lastly, as illustrations of general layout infelicities, in articles such as *sese* or *seseo* the Afrikaans and English sections are not correctly separated from one another, while in *šeleo* the brackets do not come in pairs. The inconsistencies in the other three dictionaries are more abundant, as can be seen from Addendum 1.

With modern dictionary writing systems such as *TshwaneLex*, most of these problems are taken care of by the software. POSs, for example, are chosen from a list of options and ought not to be typed in anywhere by the lexicographers, cross-references are also chosen from a finite list of options and can only be inserted when the reference address physically exists, the data distribution structure is mainly an output aspect about which lexicographers should not have to worry during compilation, while basic typographical issues such as the pairing of brackets belong to standard error checks built into the software.

With specific reference to African-language paradigms, experience has shown that the only truly sensible way to treat them is to work through all forms consistently, going down the list of all the classes in a principled way. It is impossible to produce a coherent text if one member of a paradigm is treated today, and another a week later — as they happen to cross the compiler's way. In other words, there must not only be a strict lemmatisation approach to paradigms, paradigms must also be treated 'in block'.

5.2 The problem of which data categories to include and how

Apart from the inconsistent and erratic treatment of DCs in the dictionaries under discussion, an additional shortcoming is the insufficient attention that is paid to the semantic implication carried by each of the DCs. In all of these dictionaries, the meaning of the DC is defined in terms of the relative distance between two reference points only, i.e. the speaker and the person(s) or object(s) being referred to. As was indicated above, the crucial aspect in defining the exact semantic implication expressed by the DCs, is the position of the object in relation to the positions of both speaker and addressee. TEs and definitions — for bilingual and monolingual dictionaries respectively — of these forms should therefore be formulated in such a manner that the semantic nuances that distinguish the different DCs from one another are clearly indicated. In practical terms this actually means that there are *two* levels of data that need to be given due attention in the comment on semantics (CS) of the DCs; these are the *semantic content* and the *spatial relation*, as summarised in Table 11.

When dealing with a bilingual dictionary, the 'semantic content' column in Table 11 actually provides a rather precise set of TEs, but the full picture is only obtained when the information presented in the 'spatial relation' column is also available to the person consulting the reference work. For a monolingual dictionary, the TE-like column is close to irrelevant, precisely as a result of the structure of African languages, while it is the information from the 'spatial relation' column that should receive prime attention. While both a translation and an explanatory dictionary should thus try to incorporate as much as possible from the two main columns of Table 11, translation dictionaries will rather focus on the left column, while explanatory dictionaries will rather focus on the right column.

Table 11: Semantic content and spatial relation for each of the different positions of the DCs [with sg = singular, pl = plural, S = speaker, A = addressee, * = person(s) and/or object(s) being referred, ↔ and ↕ = relative distances]

DC		**Semantic content**		**Spatial relation**
I	sg	here (s)he/it is, close to us	SA *	used in a situation where speaker and addressee are in close proximity to one another, with the person(s) and/or object(s) being referred to also close to the interlocutors
	pl	here they are, close to us		
Ia	sg	here (s)he/it is, right next to me	S↔A *	used in a situation where speaker and addressee are at a distance from one another, while the person(s) and/or object(s) being referred to is/are right next to the speaker
	pl	here they are, right next to me		
Ib	sg	here (s)he/it is, right next to me		
	pl	here they are, right next to me		
II	sg	there (s)he/it is, close to you	S⟷A *	used in a situation where speaker and addressee are relatively far apart, while the person(s) and/or object(s) being referred to is/are nearer to the addressee but not right next to him/ her
	pl	there they are, close to you		
IIa	sg	there (s)he/it is, right next to you	S⟷A *	used in a situation where speaker and addressee are quite far apart from one another, while the person(s) and/or object(s) being referred to is/are very close or directly next to the addressee
	pl	there they are, right next to you		
IIb	sg	there (s)he/it is, right next to you		
	pl	there they are, right next to you		
III	sg	there (s)he/it is, over yonder	SA ↕ *	used in a situation where speaker and addressee are in very close proximity to one another, while the person(s) and/or object(s) being referred to is/are far away from the interlocutors
	pl	there they are, over yonder		
IIIa	sg	there (s)he/it is, over yonder		
	pl	there they are, over yonder		

Besides the comment on semantics (CS), one will also have to decide what to include under the comment on form (CF), selecting from aspects such as POS label, class affiliation, pronunciation, tone marking, etc.

5.3 The problem of which DCs to treat and where

The most important issue of all, however, remains how to decide which members of a paradigm like the one of the DCs to treat, and where to treat those members. In the case of the lemmatisation of a one-dimensional paradigm such as, say, the possessive concords, it seems rather trivial to just make sure all members are included in a dictionary's macrostructure. With a two-dimensional paradigm such as the one of the DCs, the choice is less obvious. One could argue that with two linked tables such as for example an adaptation of Tables 6 and 11, any dictionary user should have enough information regarding the DCs in Sesotho sa Leboa. Such tables could be presented in the dictionary's front or back matter, have the advantage that they display *all* forms, and there would be no need to lemmatise any DC. That this approach is actually not successful is shown by the above discussion of the absence of 'its' in Rycroft's dictionary: one either has to know the meaning and/or the orthographic form prior to looking up such members of a paradigm.

When dealing with complex paradigms, it is thus clear that one will have to lemmatise (a selection of) the members in the macrostructure. Apart from using intuition only, which is not advised, one could follow a principled approach in selecting certain forms. Principled approaches could for instance include: all members with certain characteristics only, all irregular forms, all forms known to be problematic or confusing, etc. Another approach could be to use frequency data in the selection process, and to only lemmatise those that are more frequent than a certain threshold. A further refinement could be to do the latter, and to also 'mention' all attested forms in the macrostructure, but without full treatment of these. If space is of no concern, one may even include and fully treat all forms in the central text. For each of these lemmatisation approaches one may of course in addition also incorporate a set of tables in the front or back matter, preferably with cross-references from the central text to these tables. A last feasible option could be to simply include all forms in the macrostructure, but to only provide a cross-reference to the tables instead of a full treatment of each. All these options are summarised in Table 12.

Table 12: Modern lemmatisation options for the members of a (complex) paradigm

Option		**Lemmas**	**Tables**
• only in table-form (*all* members of the paradigm by default)			✓
• only as lemma signs (*which ones?*:)		✓	
○ principled selection of certain members	(+ tables)	✓	(✓)
○ frequent members only	(+ tables)	✓	(✓)
○ frequent members only + mention of attested ones	(+ tables)	✓	(✓)
○ all (theoretically possible) members	(+ tables)	✓	(✓)
• in table-form + all members with cross-references to the tables		✓	✓

6. A practical implementation: The treatment of the DC in the *Pukuntšutlhaloši ya Sesotho sa Leboa* (PyaSsaL)

In order to best see how the different options translate into a real dictionary project, the treatment of the DC in the PanSALB-sponsored PyaSsaL, i.e. the monolingual *Pukuntšutlhaloši ya Sesotho sa Leboa* 'Explanatory Sesotho sa Leboa Dictionary', will now be analysed.

From its inception the compilation of PyaSsaL has been fully corpus-based (cf. De Schryver and Lepota 2001: 3). More recently the results from fieldwork for especially cultural lexical items have supplemented the corpus data. The current policy is that each lexical item that occurs at least three times in a 6.1-million-word corpus be considered for inclusion.[5] Each defined lemma sign is further illustrated with at least one example *culled from the corpus*. The latter approach, which is considered of paramount importance as it ensures that one is dealing with *authentic* usage of *attested* forms, has actually prescribed the way in which to treat the DC in PyaSsaL. Each DC with a frequency of at least three received full treatment in the central text. In addition, it was decided to 'list' all members of the DC paradigm with lower frequencies as well and to cross-refer these to DC Tables, but *not* to lemmatise any forms not attested in the corpus. This thus means that one is effectively dealing with a three-tier structure in PyaSsaL, i.e. (a) frequent DCs are provided with both a CF and CS, (b) infrequent DCs are provided with a CF (but no CS) and a reference to DC Tables, and (c) non-attested DCs are tabulated outside the central lemma-sign list in DC Tables only. The last group includes those forms that are only mentioned in grammar books, with no examples in the other corpus sources.

Given that automatic POS taggers have as yet not been developed for Sesotho sa Leboa, both the meaning and the class affiliation of each DC 'form' had to be deduced from a meticulous scrutiny of concordance lines. That this is not a trivial process is shown by the detailed corpus data presented in Addendum 2. Firstly, note that some DC forms are *homonymous* with non-DC forms, such as for example the DC *šeba* which is homonymous with the verb *šeba* 'relish, eat as a titbit'. In this case the DC use is slightly more frequent, so this form carries the homonym number 1, while the verb is assigned homonym number 2. Even though some DCs are only attested in grammar books, their homonymous form might be relatively frequent, as in the case of *sesele* 'badger', which means that only this non-DC form is lemmatised. Secondly, note that corpus statistics also indicate in which order DCs that are morphologically similar (which is the case for classes 1 and 3, 4 and 9, and 8 and 10) ought to be ordered *within* an article. For classes 4 and 9, for example, the class 9 form is more frequent than the class 4 form for all positions except for position Ia. Thirdly, a combination of these first two observations also occurs, as for *šele*. This item occurs 817 times in all in the corpus, 719 times as an enumerative stem, 80 times as position I class 5 DC, 18 times as position IIIa class 9 DC, but not once as position IIIa class 4 DC. In a Sesotho sa Leboa–English dictionary, the minimal treatment (i.e. without example sentences) could thus take the form of the articles shown in (2), where 'SEE DC Tables' is a cross-reference to a set of tables analogous to Tables 6 and 11 above, to be found in the front or back matter of the dictionary.[6]

(2)

še **1.** *dem. cop. I cl. 9* here it is, close to us; **2.** *dem. cop. I cl. 4* here they are, close to us
šele[1] *enumerative stem* strange, foreign, different
šele[2] *dem. cop. I cl. 5* here it is, close to us
šele[3] (< **še**) **1.** *dem. cop. IIIa cl. 9* there it is, over yonder; **2.** *dem. cop. IIIa cl. 4* SEE DC Tables

Observe that in the case of the position IIIa class 4 DC, a CF is provided, even though there are no attestations of this form in the corpus. This is an example of a case where practical dictionary making overrides corpus data for the sake of consistency. It would indeed seem awkward to leave out the second sense of **šele**[3], unlike leaving out fully unattested forms such as say *šelekhwi*, *šeawe* or *šemole* from the central dictionary text.

Frequency data thus enable one to decide in which order to present homonymous items, and also in which order to present the data within single articles. Such statistics further also enable one to cross-refer lesser frequent variants to more frequent ones, as can be seen from the cross-reference from the reference position in **šetše**[4] to the reference address **šedi**[2] in (3).

(3)

šedi[1] *n. cl. 9* care, attention
šedi[2] **1.** *dem. cop. I cl. 10* here they are, close to us; **2.** *dem. cop. I cl. 8* here they are, close to us
šetše[1] *aux.* already
šetše[2] *v.* **1.** remain (behind); **2.** follow (behind)
šetše[3] *v.* must pay attention; **..ga/sa/se..~** not pay attention
šetše[4] **= šedi**[2] **1.** *dem. cop. I cl. 10* SEE DC Tables; **2.** *dem. cop. I cl. 8* SEE DC Tables

The treatment of the DC and its homonymous forms as shown in (2) and (3) is also how, mutatis mutandis, the DC is treated in PyaSsaL. If one carefully scrutinises these examples, one realises that one is actually dealing with a complex set of *multiple cross-references* that is partially driven by corpus data. One firstly sees from (2) that all higher-order positions (Ia, Ib, II, IIa, IIb, III and IIIa) may be linked to their base form (i.e. position I) through the use of the non-typographical structural marker '<'. Viewed from the DC Table angle, column-forms are thus cross-referenced. Secondly, row-forms may also be cross-referenced, such as in the case of the variant forms for which the non-typographical structural marker '=' is used, as can be seen in (3). The third type of cross-reference used to link some of the forms of the full DC paradigm are the links between the central text and the DC Tables in the front or back matter. Properly treating a complex African-language paradigm, in casu that of the DC in Sesotho sa Leboa, thus leads to an exercise in multiple cross-referencing.

7. *PyaSsaL ka Inthanete*: A pioneer in the untrodden forest

In the previous section, it was pointed out that a corpus has been consulted since compilation of PyaSsaL began. Also present since its inception has been the theoretical framework of *Simultaneous Feedback* or 'SF', which can be understood as entailing a method in terms of which the release of several small-scale parallel dictionaries triggers off feedback that is instantly channelled back into the compilation process of a main dictionary. To date, three parallel dictionaries have been released in hardcopy format.

The electronic adaptation of SF is known as *Fuzzy SF*, and PyaSsaL indeed went electronic on 10 June 2004. This new online dictionary is freely available from http://africanlanguages.com/psl/ and is known as *PyaSsaL ka Inthanete* 'Online PyaSsaL' (Mojela et al. 2004). The online data show work in progress — a selection of around 7 500 articles out of the 10 000 currently in preparation — and the main purpose is to retrieve various types of feedback, both of the implicit type through a study of the dictionary-use log files, and of the explicit type through receiving comments. An online feedback form can be filled in to that effect. The Online PyaSsaL is a pioneer in that it is the very first monolingual African-language dictionary on the Internet for which also the interface and the metalanguage of all macro- and microstructural elements are presented in the African language (cf. De Schryver 2003a: 9). A screenshot of the start page of this online dictionary can be seen in Addendum 3.

The multifarious advantages of an electronic medium for dictionaries are well known (cf. De Schryver 2003), and this is no different when it comes to the treatment of the DC. It is important to realise that even though the Online PyaSsaL is presented on the Internet, the work remains primarily compiled with a printed dictionary in mind. It would thus be naive for example to decide to treat *all* DCs online, as this would make the hardcopy version unnecessarily bulky, apart from the fact that one cannot find a single corpus example for six out of each ten theoretically possible DCs. Nonetheless, even though one is effectively dealing with the same data, the electronic environment *does* provide an array of useful extra lexicographical devices. The focus will be on one of them: cross-referencing. In line with the concepts of SF and Fuzzy SF, some of the options discussed below are already being implemented, while others are only experimented with — the idea exactly being to attempt to tailor the approach to the users of the dictionary.

Whereas the various reference relations are, to save space, typically established by a set of symbols in printed dictionaries, such a symbol set can easily be replaced with a set of user-friendlier text segments in an electronic environment. In the dictionary compilation software, one only needs to create a 'parallel set of cross-references', without touching any of the dictionary data. As such, the reference marker '<' can for example be replaced throughout with *Go tšwa go* 'Derived from'. This thus means that any *textual* information of any length can be placed within the reference marker itself, greatly enhancing the readability. On a second level one can point out that whereas paging *to* a reference address in a paper dictionary is cumbersome and time-consuming, that process is sped up with what is known as *hyperlinking* in a computational environment. A further user-friendly enhancement in an electronic medium is to display all related cross-references of an item. Here cross-references *from*, but also cross-references *to* an article can simply be culled out of the database automatically and be presented together with the looked-up item. If one looks up *šidi* in the Online PyaSsaL, for example, the article *šedi* (which is referred to from *šidi*) as well as the articles *šidio* and *šidile* (which contain cross-references to *šidi*) are shown on the same output page, enabling the dictionary user to quickly obtain a good overview of the preferred variant and derived forms. A simpler case can be seen in Addendum 4, where a search for *šokhwi* also returns the article for the base form *šo*.

At least as powerful is the possibility to call up tables at any point, so clicking on BONA Lenaneo la mašalašupi-leba 'SEE Table of demonstrative copulatives' in the article of *šokhwi* (cf. Addendum 4), for example, will currently display a table similar to Table 6 above. At present this DC Table is generated as a static table within the same dictionary page, but plans include experimenting with a dynamically generated page, opening in a pop-up window and with the particular DC being looked up immediately highlighted. Such a cross-reference (hyperlink) will instantly allow the dictionary user to frame the DC at hand within the full paradigm of all DCs — an exploit that cannot be achieved in the paper dimension. A further option could be to make each item from the DC Table that is also lemmatised clickable, so that dictionary users could even more easily browse through the dictionary.

8. In conclusion

In 2004, the year South Africa celebrates and looks back on 10 years of democracy, it seemed appropriate to also assess some of the achievements in African-language metalexicography. It was shown that great strides have been made during the past decade, in particular when it comes to Sesotho sa Leboa. New directions of research were then suggested for the future, and one of them, namely the 'paradigmatic lemmatisation' of closed-class words, was singled out for this article. As a case study within this field, the lexicographic treatment of the demonstrative copulative (DC) in Sesotho sa Leboa was discussed in depth. To that end, the DC as lemmatised in the current desktop dictionaries for Sesotho sa Leboa was first analysed, and following an identification of the problem areas, options for a sound treatment in both bilingual and monolingual dictionaries were then suggested. It was indicated that a sound approach to 'paradigmatic lemmatisation' should preferably (a) make use of professional dictionary software such as *TshwaneLex* to ensure consistency, (b) take both the semantic content and the spatial relation of each member of the paradigm into account, and (c) put corpus frequency data to good use when deciding on which members to lemmatise and how to order homonyms and senses. It was also pointed out how the use of multiple cross-referencing (hyperlinking) can successfully link the alphabetically dispersed members of a paradigm. In addition, the benefit was emphasised of also making such 'links' with overview tables that list all theoretically possible forms.

An actual implementation of a modern treatment of the DC was then presented for PyaSsaL, i.e. the *Pukuntšutlhaloši ya Sesotho sa Leboa* 'Explanatory Sesotho sa Leboa Dictionary'. The compilation of PyaSsaL is both corpus-based and undertaken within the theoretical framework of *Simultaneous Feedback* (SF). In order to speed up the process of retrieving feedback, a selection of the PyaSsaL data is currently presented online as work in progress at http://africanlanguages.com/psl/, making *PyaSsaL ka Inthanete* 'Online PyaSsaL' the only truly monolingual dictionary on the Internet for any African language at present. Various presentation options for the DC are currently experimented with online, including multiple ways of cross-referencing and hyperlinking to and from static as well as dynamically generated tables, with and without highlighting, etc. Given that the developed approaches are generic, and given that there are many more paradigms in Sesotho sa Leboa — as well as in all other African languages — the considerable multiplication factor of this study can hardly be underestimated.

Endnotes

1. Since this article is being submitted for publication in a South African journal, necessary sensitivity with regard to the term 'Bantu' languages is exercised in our choice rather to use the term *African* languages. Bear in mind, however, that the latter includes more than just the 'Bantu Language Family'.
2. Observe that tone is not normally marked in the siSwati orthography, which is why it is left out in the running text of this article.
3. In the standard Sesotho sa Leboa orthography the 'raised ê' and 'e' are collapsed to 'e', and likewise the 'raised ô' and 'o' are collapsed to 'o'. This practice is also followed in the running text of this article.
4. A long final vowel is however suggested for positions III and IIIa by Ziervogel and Mokgokong (1975: 104-105, *Introduction*), but only in their grammatical outline, not in their dictionary proper.
5. This 6.1-million-word corpus for Sesotho sa Leboa is an organic corpus that is continuously being expanded and revised. It is built by staff members in the Department of African Languages at the University of Pretoria, in cooperation with lexicographers from the Sesotho sa Leboa National Lexicography Unit (NLU) and the authors of this article.
6. Depending on the intended target user group, the type and amount of information provided in such tables, as well as the way in which this information is presented, will differ. The different members of the DC paradigm that are also treated in the central lemma-sign list could for example be presented in a different colour, typeface or even be highlighted, to differentiate these forms from those that are not attested in the corpus. It could even be considered in-

cluding the actual corpus frequencies at each paradigm member, as has been done in Addendum 2. The minimal treatment, conversely, is as a simple enumeration of all forms in a two-dimensional table.

References

Bosch, Sonja E. and D.J. Prinsloo. 2002. 'Abbreviated Nouns' in African Languages: A Morphological, Semantic and Lexicographic Perspective. *South African Journal of African Languages* 22(1): 92-104.

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

De Schryver, Gilles-Maurice. 2002. Web for/as Corpus: A Perspective for the African Languages. *Nordic Journal of African Studies* 11(2): 266-282.

De Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

De Schryver, Gilles-Maurice. 2003a. Online Dictionaries on the Internet: An Overview for the African Languages. *Lexikos* 13: 1-20.

De Schryver, Gilles-Maurice. 2003b. Drawing up the Macrostructure of a Nguni Dictionary, with Special Reference to isiNdebele. *South African Journal of African Languages* 23.

De Schryver, Gilles-Maurice and B. Lepota. 2001. The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails. *Lexikos* 11: 1-37.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000. Dictionary-Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 197-209. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Stuttgart University.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000a. The Concept of 'Simultaneous Feedback': Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000b. The Compilation of Electronic Corpora, with Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000c. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 1: The *Macrostructure*. *South African Journal of African Languages* 20(4): 291-309.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000d. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 2: The *Microstructure*. *South African Journal of African Languages* 20(4): 310-330.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2001. Corpus-based Activities versus Intuition-based Compilations by Lexicographers, the Sepedi Lemma-Sign List as a Case in Point. *Nordic Journal of African Studies* 10(3): 374-398.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2001a. Fuzzy SF: Towards the Ultimate Customised Dictionary. *Studies in Lexicography* 11(1): 97-111.

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

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

Gouws, Rufus H. and D.J. Prinsloo. 1997. Lemmatisation of Adjectives in Sepedi. *Lexikos* 7: 45-57.

Joffe, David, Gilles-Maurice de Schryver and D.J. Prinsloo. 2003. Introducing TshwaneLex — A New Computer Program for the Compilation of Dictionaries. De Schryver, G-M (Ed.). 2003. *TAMA 2003 South Africa: CONFERENCE PROCEEDINGS*: 97-104. Pretoria: (SF)^2 Press.

Joffe, David, Gilles-Maurice de Schryver and D.J. Prinsloo. 2003a. Computational Features of the Dictionary Application "TshwaneLex". *Southern African Linguistics and Applied Language Studies* 21(4) [Special issue on 'Human Language Technology in South Africa: Resources and Applications']: 239-250.

Kilgarriff, Adam. 1996. *BNC Database and Word Frequency Lists* [online]. Available: <http://www.itri.bton.ac.uk/~Adam.Kilgarriff/bnc-readme.html>.

Kotzé, Albert E. 1985. Herinterpretasie van die betekenis van demonstratiewe in Noord-Sotho. *South African Journal of African Languages* 5(3): 82-87.

Kriel, Theunis J. 1976^4. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English.* Johannesburg: Educum Publishers.

Kriel, Theunis J. 1983^3. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.

Kriel, Theunis J., Egidius B. van Wyk and Staupitz A. Makopo. 1989^4. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.

Lombard, Daniël P., Egidius B. van Wyk and Pothinus C. Mokgokong. 1985. *Inleiding tot die grammatika van Noord-Sotho*. Pretoria: J.L. van Schaik.

Louwrens, Louis J. 1991. *Aspects of Northern Sotho Grammar.* Pretoria: Via Afrika Ltd.

Louwrens, Louis J. 1994. *Dictionary of Northern Sotho Grammatical Terms*. Pretoria: Via Afrika Ltd.

Mojela, M.V. (Editor-in-Chief), M.P. Mogodi, M.C. Mphahlele and M.R. Selokela (Compilers). 2004. *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete (Explanatory Sesotho sa Leboa Dictionary on the Internet)* [online]. Available: <http://africanlanguages.com/psl/>.

Nokaneng, Mogobo B. and Louis J. Louwrens. 1988. *Segagešo Mphato* 9. Pretoria: Via Afrika Ltd.

Nong, Salmina, Gilles-Maurice de Schryver and D.J. Prinsloo. 2002. Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective. *Lexikos* 12: 1-20.

Poulos, George and Louis J. Louwrens. 1994. *A Linguistic Analysis of Northern Sotho.* Pretoria: Via Afrika Ltd.

Prinsloo, D.J. 1991. Towards Computer-assisted Word Frequency Studies in Northern Sotho. *South African Journal of African Languages* 11(2): 54-60.

Prinsloo, D.J. 1992. Lemmatization of Reflexives in Northern Sotho. *Lexikos* 2: 178-191.

Prinsloo, D.J. 1994. Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 14(2): 93-102.

Prinsloo, D.J. 2002. The Lemmatization of Copulatives in Northern Sotho. *Lexikos* 12: 21-43.

Prinsloo, D.J. 2003. The Lemmatisation of Adverbs in Northern Sotho. *Lexikos* 13: 21-37.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2001. Monitoring the Stability of a Growing Organic Corpus, with Special Reference to Sepedi and Xitsonga. *Dictionaries: Journal of The Dictionary Society of North America* 22: 85-129.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2002. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with Special Reference to Afrikaans and English. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002*: 483-494. Copenhagen: Center for Sprogteknologi, University of Copenhagen.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2002a. The Use of Slashes as a Lexicographic Device, with Special Reference to the African Languages. *South African Journal of African Languages* 22(1): 70-91.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2002b. Reversing an African-language Lexicon: The *Northern Sotho Terminology and Orthography No. 4* as a Case in Point. *South African Journal of African Languages* 22(2): 161-185.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2003. Effektiewe vordering met die *Woordeboek van die Afrikaanse Taal* soos gemeet in terme van 'n multidimensionele Liniaal. Botha, W. (Ed.). 2003. *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 106-126. Stellenbosch: Bureau of the WAT.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2004. Crafting a Multidimensional Ruler for the Compilation of Sesotho sa Leboa Dictionaries. Mojalefa, J. (Ed.). 2004. *Rabadia Ratšhatšha: In-depth Literature, Linguistics, Translation and Lexicography Studies in African Languages. Festschrift in Honour of P.S. Groenewald*. Pretoria: J.L. van Schaik.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 2004a. Managing Eleven Parallel Corpora and the Extraction of Data in all Official South African Languages. Daelemans, W. and T. du Plessis (Eds.). 2004. *Multilingualism and Electronic Language Management*. Pretoria: J.L. van Schaik.

Prinsloo, D.J. and Rufus H. Gouws. 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.

Rycroft, David K. 1981. *Concise SiSwati Dictionary. SiSwati–English/English–SiSwati*. Pretoria: J.L. van Schaik.

Ziervogel, Dirk, Daniël P. Lombard and Pothinus C. Mokgokong. 1969. *Handboek van Noord-Sotho*. Pretoria: J.L. van Schaik.

Ziervogel, Dirk and Pothinus C. Mokgokong. 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane / Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels / Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English.* Pretoria: J.L. van Schaik.

Addendum 1: Treatment of the DC in the four existing desktop dictionaries for Sesotho sa Leboa [reproduced verbatim, including all inconsistencies and errors]

Pukuntšu woordeboek, Noord-Sotho–Afrikaans
— Kriel (1983^3)

še, hier is dit.
šea, *še.a,* hier is hulle.
šeao, *še a.o,* daar is hulle(ma-klas).
šeba, *še'ba,* hier is hulle.
šebale, *še ba.le,* daar is hulle.
šebo, *še.'bo,* hier is dit (bo-klas).
šedi, *še.'di,* hier is hulle (*se* – klas).
šefa, *še.'fa,* hier is dit.
šefale, *še fa.lê,* daar is dit.
šefao, *še fa.o,* daar is dit.
šegola, *še go.la,* daar is dit.
šegoo, *še go.o,* daar is dit (go-klas).
šele, *še.'le,* hier is dit (le-klas); *leeba* **-,** hier is die duif.
šeleo, *še'le.o,* (le-klas), daar is dit.
šeo, *še.o, (šewe),* daar is dit/hy, (die *n-*, *di-*klas).
sese, *se'se,* hier is dit.
seseo, *se se.o,* daar is dit.
šidi, *šidi,* ou spelling, kyk: *šedi,* hier is hulle.
šo, hier is hy/sy.
šolaa, *šo'la.a,* daar is hy/sy.
šole, *šo.'le,* daar is hy/sy.
šono, *'šo.no,* hier is hy/sy.
šoo, *šo.o,* daar is hy/sy.

Pukuntšu woordeboek, Noord-Sotho–Afrikaans
— Kriel, Van Wyk and Makopo (1989^4)

še, kop. dem. 1, kl 3/9, H: hier is (dit).
šea, kop, dem. 1, kl 8, HH: hier is (hulle).
šeao, kop. dem 2, kl 8, HHL: daar is (hulle).
šeba, kop. dem. 1, kl 2, H: hier is (hulle).
šebalê, kop. dem. 3, kl 2, HHL: dáár is (hulle).
šebo, kop. dem. 1, kl 14, HH: hier is (dit).
šedi, kop. dem. 1, kl 8/10, HH: hier is (hulle).
šefa, kop. dem. 1, kl 16, HH: hier is (dit).
šefalê, kop. dem. 3, kl 16, HHL: dáár is (dit).
šefao, kop. dem. 3, kl 16, HHL: daar is (dit).
šegola, kop. dem. 3, kl 17, HHL: dáár is (dit).
šegoo, kop. dem. 2, kl 17, HHL: daar is (dit).
šele, kop. dem. 1, kl 5, HH: hier is (dit).
šeleo, kop. dem. 2, kl 5, HHL: daar is (dit).
šeo, kop. dem. 2, kl 4/9, HL: daar is (dit/hulle).
sese, kop. dem. 1, kl 7, HH: hier is (dit).
seseo, kop. dem. 2, kl 7, HHL: daar is (dit).
šidi, kop. dem. 1, kl 8/10, HH: hier is (hulle).
šo, kop. dem. 1, kl 1/3, H: hier is (hy/sy/dit.
šolê, kop. dem. 3, kl 1/3, HL: dáár is (hy/sy/dit).
šono, kop. dem. 1a, kl 1/3, HL: hier(so) is (hy/sy/dit).
šoo, kop. dem. 2, kl 1/3, HL: daar is (hy/sy/dit).

The New English–Northern Sotho Dictionary, Northern Sotho–English
— Kriel (1976^4)

'še, dem., pron., here it is.
še'a, dem., here they are.
še'ba, dem., here they are.
'še'di, dem., **di–**class, here they are.
še'di'le, dem., pl., there they are, yonder.
še'di'o, dem., there they are.
še'fa, dem., here it is, here you are.
še'le, dem., adj., here it is.
še'leo, dem., adj., there it is.
šidi'le-e, dem., there they are.
ši'dio, – se'dio, dem., there they are (animal or things).
'šo, dem., here he is, here she is.
'šole, dem., there he (she) is.
'šono, dem., here he (she) is.

Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English
— Ziervogel and Mokgokong (1975)

ŠÉ- [prefigale element by vorming van kop. dem.] // [prefixal element in formation of cop. dem.] v. **ŠÉBÁ, ŠÉO,** etc.

ŠÉ [dem. kop. I kl. **n-, me-**] hier is hy/sy/dit/hulle // [dem. cop. I cl. **n-, me-**] here he/she/it is, there they are

ŠÉÁ [dem. kop. I kl. **ma-**] hier is hulle // [dem. cop. I cl. **ma-**] here they are

ŠÉÁLE (**šealê**) [dem. kop. III kl. **ma-**] dáár is hulle // [dem. cop. III cl. **ma-**] there they are over there

ŠÉÁO (**šeaô**) [dem. kop. II kl. **ma-**] daar is hulle // [dem. cop. II cl. **ma-**] there they are

ŠÉBÁ [dem. kop. I kl. **ba-**] hier is hulle // [dem. cop. I cl. **ba-**] here they are

ŠÉBÁLE (**šebalê**) [kop. dem. III kl. **ba-**] dáár is hulle // [cop. dem. III cl. **ba-**] there are they over there

ŠÉBÁO (**sebaô**) [kop. dem. II kl. **ba-**] daar is hulle // [cop. dem. II cl. **ba-**] there they are

ŠÉBÓ [kop. dem. I kl. **bo-**] hier is dit // [cop. dem. I cl. **bo-**] here it is

ŠÉBÓLA [dem. kop. III kl. **bo-**] dáár is dit // [dem. cop. III cl. **bo-**] there it is over there

ŠÉBÓO (**šeboô**) [dem. kop. II kl. **bo-**] daar is dit // [dem. cop. II cl. **bo-**] there it is

ŠÉDÍ v. **ŠÍDÍ**

SÉDÍLA v. **ŠÍDÍLA**

ŠÉDÍO v. **ŠÍDÍO**

ŠÉFÁ [dem. kop. I vir lo. klasse] hier is dit // [dem. cop. I for lo. classes] here it is

ŠÉFÁLE (**šefalê**) [dem. kop. III vir lo. klasse] dáár is dit // [dem. cop. III for lo. classes] there it is over there

ŠÉFÁO (**šefaô**) [dem. kop. II vir lo. klasse] daar is dit // [dem. cop. II for lo. classes] there it is

ŠÉGÓLA [dem. kop. III vir lo. klasse (dia.)] dáár is dit // [dem. cop. III for lo. classes (dia.)] there it is over there

ŠÉGÓO (**šegoô**) [dem. kop. II vir lo. klasse (dia.)] daar is dit // [dem. cop. II for lo. classes (dia.)] there it is

ŠÉLA [dem. kop. III kl. **n-, me-**] dáár is hy/sy/dit/hulle // [dem. cop. III cl. **n-, me-**] there he/she/it is over there, there they are over there

ŠÉLE [dem. kop. I kl. **le-**] hier is hy/sy/dit // [dem. cop. I cl. **le-**] here he/she/it is

ŠÉLÉLA [dem. kop. III kl. **le-**] dáár is hy/sy/dit // [dem. cop. III cl. **le-**] there he/she/it is over there

ŠÉLÉO (**šeleô**) [dem. kop. II kl. **le-**] daar is hy/sy/dit // (dem. cop. II cl. **le-**] there he/she/it is

ŠÉO (**seô**) [kop. dem. II kl. **n-, me-**] daar is hy/sy/dit/hulle // [cop. dem. II cl. **n-, me-**] there he/she/it is, there they are

SÉSE [kop. dem. I, kl. **se-**] // [cop. dem. I, cl. **se-**] hier is dit // here it is

SÉSÉLA [kop. dem. III, kl. **se-**] // [cop. dem. III, cl. **se-**] doer is dit // there it is over there

SÉSÉO (**seseô**) [kop. dem. II, kl. **se-**] // [cop. dem. II, cl. **se-**] daar is dit // there it is

ŠÍ- [geassimileerde vorm van **še-**] // assimilated form of **še-**], cf. **ŠIDI**

ŠÍDÍ (< **šedi**) [dem. kop. I kl. **di-**, **din-**] hier is hulle // [dem. cop I cl. **di-, din-**] here they are

ŠÍDÍLA [dem. kop. III kl. **di-, din-**] dáár is hulle // [dem. cop. III cl. **di-, din-**] there they are over there

ŠÍDÍO (**šidiô**) [dem. kop. II kl. **di-, din-**] daar is hulle // [dem. cop. II cl. **di-, din-**] there they are

ŠO [dem. kop. I kl. **mo-**] hier is hy/sy/dit // [dem. cop. I cl. **mo-**] here he/she/it is

ŠÓLA [dem. kop. III kl. **mo-**] dáár is hy/sy/dit // [dem. cop. III cl. **mo-**] there he/she/it is over there

ŠÓLE (**šolê**) [dem. kop. III kl. **mo-**] dáár is hy/sy/dit // [dem. cop. III cl. **mo-**] there he/she/it is over there

ŠOO (**šoô**) [dem. kop II kl. **mo-**] daar is hy/sy/dit // [dem. cop. II cl. **mo-**] there he/she/it is

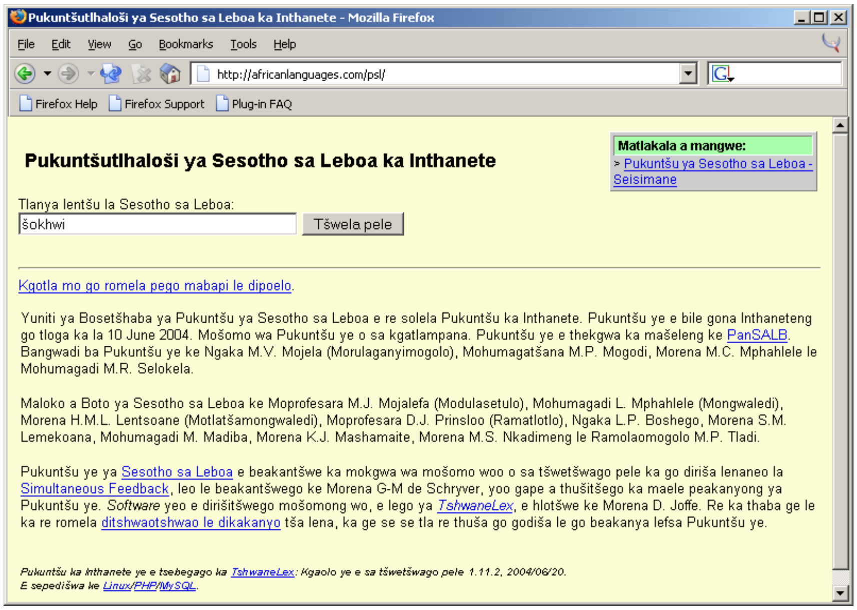
Addendum 2: Frequencies of all DCs and homonymous items in a 6.1-million-word Sesotho sa Leboa corpus [∑ = total frequency; ? = not possible to see class affiliation; G = occurs in grammar book(s) only; *v.* = verb(s); *aux.* = auxiliary verb; *n.* = noun(s); *adj.* = adjective; *e.* = enumerative stem]

	I	**Ia**	**Ib**	**II**	**IIa**	**IIb**	**III**	**IIIa**
1&3	**šo** ∑ 564 I.1 348 I.3 198 ? 18	**šono** ∑ 3 Ia.1 3 Ia.3 0	**šokhwi** ∑ 18 Ib.1 16 Ib.3 2	**šoo** ∑ 83 II.1 76 II.3 7	**šouwe** ∑ 3 IIa.1 2G IIa.3 1G	**šowe** ∑ 19 IIb.1 17 IIb.3 2	**šola** ∑ 20 *v.* 9 III.1 8 III.3 3	**šole** ∑ 102 IIIa.1 97 IIIa.3 5
2	**šeba** ∑ 248 I.2 125 *v.* 123	**šebano** ∑ 2 Ia.2 2	**šebakhwi** ∑ 1 Ib.2 1G	**šebao** ∑ 22 II.2 22	**šebauwe** ∑ 2 IIa.2 2G	**šebawe** ∑ 1 IIb.2 1	**šebala** ∑ 5 III.2 5	**šebale** ∑ 25 IIIa.2 25
4&9	**še** ∑ 412 I.9 244 I.4 153 ? 15	**šeno** ∑ 4 *aux.* 3 Ia.4 1 Ia.9 0	**šekhwi** ∑ 7 Ib.9 7 Ib.4 0	**šeo** ∑ 45 II.9 41 II.4 4	**šeuwe** ∑ 2 IIa.9 1G IIa.4 1G	**šewe** ∑ 4 IIb.9 4 IIb.4 0	**šela** ∑ 12 III.9 7 *n.* 5 III.4 0	**šele** ∑ 817 *e.* 719 I.5 80 IIIa.9 18 IIIa.4 0
5	**šele** ∑ 817 *e.* 719 I.5 80 IIIa.9 18	**šeleno** ∑ 0	**šelekhwi** ∑ 0	**šeleo** ∑ 7 II.5 7	**šeleuwe** ∑ 1 IIa.5 1G	**šelewe** ∑ 0	**šelela** ∑ 0	**šelele** ∑ 2 IIIa.5 2
6	**šea** ∑ 135 I.6 133 *v.* 2	**šeano** ∑ 0	**šeakhwi** ∑ 0	**šeao** ∑ 17 II.6 17	**šeauwe** ∑ 1 IIa.6 1G	**šeawe** ∑ 0	**šeala** ∑ 1 III.6 1G	**šeale** ∑ 5 III.6 5
7	**sese** ∑ 100 I.7 97 *adj.* 3	**seseno** ∑ 0	**sesekhwi** ∑ 0	**seseo** ∑ 15 II.7 15	**seseuwe** ∑ 1 IIa.7 1G	**sesewe** ∑ 0	**sesela** ∑ 34 *n.* 32 III.7 2	**sesele** ∑ 20 *n.* 19 IIIa.7 1G
8&10	**šedi** ∑ 502 *n.* 293 I.10 185 I.8 24	**šedino** ∑ 0	**šedikhwi** ∑ 0	**šedio** ∑ 75 II.10 70 II.8 5	**šediuwe** ∑ 0	**šediwe** ∑ 0	**šedila** ∑ 0	**šedile** ∑ 15 IIIa.10 13 IIIa.8 2
8'&10'	**šidi** ∑ 36 I.10 25 I.8 7 ? 4	**šidino** ∑ 0	**šidikhwi** ∑ 0	**šidio** ∑ 8 II.10 8 II.8 0	**šidiuwe** ∑ 2 IIa.10 1G IIa.8 1G	**šidiwe** ∑ 0	**šidila** ∑ 8 *v.* 6 III.10 1G III.8 1G	**šidile** ∑ 4 IIIa.10 2 IIIa.8 2

8"&10"	**šetše** ∑ 6846 *aux.* 6069 *v.* 775 I.10 2 I.8 0	**šetšeno** ∑ 0	**šetšekhwi** ∑ 0	**šetšeo** ∑ 1 II.10 1 II.8 0	**šetšeuwe** ∑ 0	**šetšewe** ∑ 0	**šetšela** ∑ 0	**šetšele** ∑ 0
14	**šebo** ∑ 24 I.14 24	**šebono** ∑ 0	**šebokhwi** ∑ 0	**šeboo** ∑ 1 II.14 1G	**šebouwe** ∑ 1 IIa.14 1G	**šebowe** ∑ 1 IIb.14 1	**šebola** ∑ 1 III.14 1G	**šebole** ∑ 0
15&17	**šego** ∑ 6 *n.* 6	**šegono** ∑ 0	**šegokhwi** ∑ 0	**šegoo** ∑ 0	**šegouwe** ∑ 0	**šegowe** ∑ 0	**šegola** ∑ 0	**šegole** ∑ 0
16	**šefa** ∑ 106 I.loc 106	**šefano** ∑ 0	**šefakhwi** ∑ 2 Ib.loc 2	**šefao** ∑ 16 II.loc 16	**šefauwe** ∑ 1 IIa.loc 1G	**šefawe** ∑ 0	**šefala** ∑ 1 III.loc 1G	**šefale** ∑ 0
18	**šemo** ∑ 7 I.loc 7	**šemono** ∑ 0	**šemokhwi** ∑ 0	**šemoo** ∑ 0	**šemouwe** ∑ 0	**šemowe** ∑ 0	**šemola** ∑ 0	**šemole** ∑ 0

*__šea__ → **sea** *v.* examine – 2
šeba *v.* relish, eat as a titbit – 123
šedi *n. cl. 9* care, attention – 293
*__šego__ → **lešego** *n. cl. 5* blessing – 3;
→ **bošego** *n. cl. 14* night – 2;
→ **sešego** *n. cl. 7* granary – 1
*__šela__ → **lešela** *n. cl. 5* cloth – 5
šele *enumerative stem* strange, foreign, different – 719
*__šeno__ → **seno** *aux.* as soon as – 3
-sese *adj.* thin, small – 3
Sesela (person's name) – 32
sesele *n. cl. 7* badger – 19
šetše^{1} *aux.* already – 6069
šetše^{2} *v.* **1.** remain (behind) – 520; **2.** follow (behind) – 160
šetše^{3} *v.* must pay attention; **..ga/sa/se..~** not pay attention – 95
šidila *v.* **1.** iron – 5; **2.** massage – 1
šola *v.* bring bad luck – 9

Addendum 3: Screenshot of a search for *šokhwi* in the Online PyaSsaL – input


Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://africanlanguages.com/psl/
Firefox Help
Firefox Support
Plug-in FAQ
Tšwela pele

Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete

Matlakala a mangwe:
> Pukuntšu ya Sesotho sa Leboa - Seisimane

Tlanya lentšu la Sesotho sa Leboa:
šokhwi

Kgotla mo go romela pego mabapi le dipoelo.

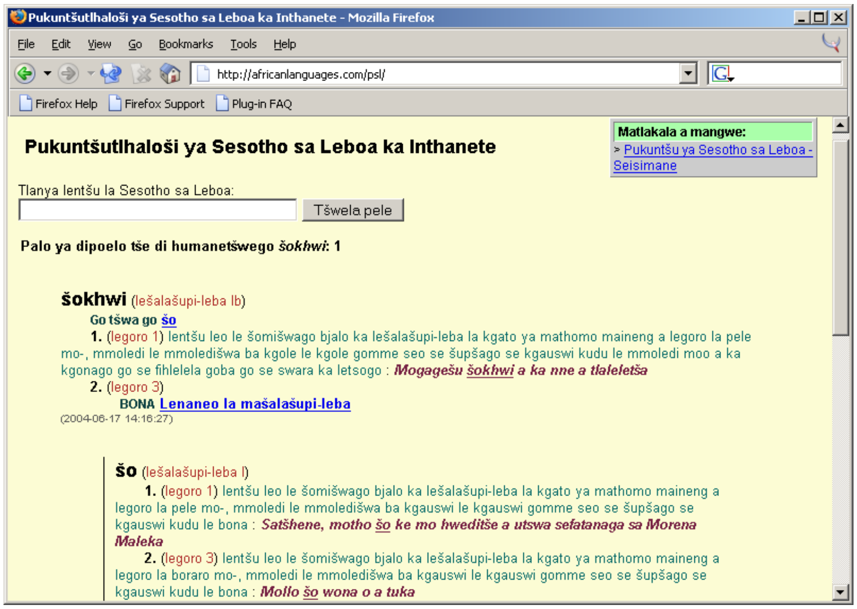
Yuniti ya Bosetšhaba ya Pukuntšu ya Sesotho sa Leboa e re solela Pukuntšu ka Inthanete. Pukuntšu ye e bile gona Inthaneteng go tloga ka la 10 June 2004. Mošomo wa Pukuntšu ye o sa kgatlampana. Pukuntšu ye e thekgwa ka mašeleng ke PanSALB. Bangwadi ba Pukuntšu ye ke Ngaka M.V. Mojela (Morulaganyimogolo), Mohumagatšana M.P. Mogodi, Morena M.C. Mphahlele le Mohumagadi M.R. Selokela.

Maloko a Boto ya Sesotho sa Leboa ke Moprofesara M.J. Mojalefa (Modulasetulo), Mohumagadi L. Mphahlele (Mongwaledi), Morena H.M.L. Lentsoane (Motlatšamongwaledi), Moprofesara D.J. Prinsloo (Ramatlotlo), Ngaka L.P. Boshego, Morena S.M. Lemekoana, Mohumagadi M. Madiba, Morena K.J. Mashamaite, Morena M.S. Nkadimeng le Ramolaomogolo M.P. Tladi.

Pukuntšu ye ya Sesotho sa Leboa e beakantšwe ka mokgwa wa mošomo woo o sa tšwetšwago pele ka go diriša lenaneo la Simultaneous Feedback, leo le beakantšwego ke Morena G-M de Schryver, yoo gape a thušitšego ka maele peakanyong ya Pukuntšu ye. *Software* yeo e dirišitšwego mošomong wo, e lego ya *TshwaneLex*, e hlotšwe ke Morena D. Joffe. Re ka thaba ge le ka re romela ditshwaotshwao le dikakanyo tša lena, ka ge se se tla re thuša go godiša le go beakanya lefsa Pukuntšu ye.

Pukuntšu ka Inthanete ye e tsebegago ka TshwaneLex: Kgaolo ye e sa tšwetšwago pele 1.11.2, 2004/06/20.
E sepedišwa ke Linux/PHP/MySQL.

Addendum 4: Screenshot of a search for *šokhwi* in the Online PyaSsaL – output


Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://africanlanguages.com/psl/
Firefox Help
Firefox Support
Plug-in FAQ
Tšwela pele

Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete

Matlakala a mangwe:
> Pukuntšu ya Sesotho sa Leboa - Seisimane

Tlanya lentšu la Sesotho sa Leboa:

Palo ya dipoelo tše di humanetšwego *šokhwi*: 1

šokhwi (lešalašupi-leba lb)
Go tšwa go **šo**
1. (legoro 1) lentšu leo le šomišwago bjalo ka lešalašupi-leba la kgato ya mathomo maineng a legoro la pele mo-, mmoledi le mmoledišwa ba kgole le kgole gomme seo se šupšago se kgauswi kudu le mmoledi moo a ka kgonago go se fihlelela goba go se swara ka letsogo : ***Mogagešu šokhwi a ka nne a tlaleletša***
2. (legoro 3)
BONA **Lenaneo la mašalašupi-leba**
(2004-06-17 14:16:27)

šo (lešalašupi-leba l)
1. (legoro 1) lentšu leo le šomišwago bjalo ka lešalašupi-leba la kgato ya mathomo maineng a legoro la pele mo-, mmoledi le mmoledišwa ba kgauswi le kgauswi gomme seo se šupšago se kgauswi kudu le bona : ***Satšhene, motho šo ke mo hweditše a utswa sefatanaga sa Morena Maleka***
2. (legoro 3) lentšu leo le šomišwago bjalo ka lešalašupi-leba la kgato ya mathomo maineng a legoro la boraro mo-, mmoledi le mmoledišwa ba kgauswi le kgauswi gomme seo se šupšago se kgauswi kudu le bona : ***Mollo šo wona o a tuka***