CRAFTING A MULTIDIMENSIONAL RULER FOR THE COMPILATION OF SESOTHO SA LEBOA DICTIONARIES

D.J. PRINSLOO
Department of African Languages,
University of Pretoria, Pretoria.

GILLES-MAURICE DE SCHRYVER
Department of African Languages and Cultures,
Ghent University, Belgium &
Department of African Languages,
University of Pretoria, Pretoria.

Abstract

The aim of this article is (a) to reflect on the contributions made by P.S. Groenewald to the field of lexicography in South Africa, focusing on the importance of determining the relative frequency of individual words in Sesotho sa Leboa, (b) to indicate how this initial basic need for knowing the frequency of use of words gradually grew and motivated the creation of text corpora, first for Sesotho sa Leboa, later for all official South African languages, (c) to illustrate how frequency counts and keywords in context can be used to improve dictionary compilation at the macrostructural and microstructural level, and (d) to utilise frequency counts for a novel cutting-edge dimension, namely to craft a multidimensional Ruler for the compilation of 'major Sesotho sa Leboa dictionaries' – such dictionaries being the Groenewald dream.

Senaganwa

Go hlama Sethaledi sa Tekanyo ya bontši sa go ngwala Dipukuntšu tša Sesotho sa Leboa. *Maikemišetšo a taodišwana ye ke (a) go bonagatša dineo tšeo di dirilwego ke P.S. Groenewald mo thutong ya go ngwalwa ga dipukuntšu mo Afrika Borwa go lebeletšwe bohlokwa bja go laetša tswalano ya go tšwelela ga mantšu kgafetšakgafetša mo go Sesotho sa Leboa, (b) go bontšha ka moo motheo wo wa go nyaka go tseba tšhomišo ya tšwelelo ya mantšu kgafetšakgafetša e godilego ka dikgato gape e hlohleleditšego tlhamo ya sešego sa mantšu sa dingwalwa, go thoma ka ya Sesotho sa Leboa gomme ka morago ya ba ya maleme ka moka a Afrika Borwa, (c) go laetša ka moo tšwelelo ya mantšu kgafetšakgafetša le mantšu a bohlokwa ka gare ga diteng a ka dirišwago go kaonafatša tlhamo ya pukuntšu go maemo a go beakanya le go hlaloša mantšu ka gare ga pukuntšu, gape (d) le go diriša tšwelelo ya mantšu kgafetšakgafetša maemong a mafsa a godimo, se se ra gore go hlama Sethaledi sa*

Tekanyo ya bontši ya go hlama 'dipukuntšu tše dikgolo tša Sesotho sa Leboa' – gomme dipukuntšu tša mohuta woo e tla ba tšona tša toro ya Groenewald.

Towards the first monolingual dictionary of Sesotho sa Leboa – a brief historical overview

In 1988 Prof. P.S. Groenewald, then Head of the Department of African languages at the University of Pretoria, approached the Human Sciences Research Council (HSRC) with a well-founded request to fund a multi-volume monolingual dictionary for Sesotho sa Leboa.[1] This was envisaged as a major dictionary project similar to the *Woordeboek van die Afrikaanse Taal* (WAT), and a budget of approximately R 1 million per year was envisaged. The idea was well received and he was praised for the initiative, but no money was granted. The University of Pretoria was subsequently approached for funding and they undertook to contribute R 8 000 annually for a number of years. In addition, the University of Pretoria also provided comprehensive infrastructure in the form of an office, a telephone, a fax machine and up-to-date computer technology, as well as the free services of computer programmers. The then Departmental Northern Sotho Language Board regarded the project as a top priority and it was prepared to act as the controlling body.

An initial Dictionary Committee was formed, and supplemented from time to time, consisting of P.S. Groenewald, J. Maripane, M.J. Mojalefa, D.J. Prinsloo and B.P. Sathekge. These members worked on the dictionary in their spare time. At the end of the 1990s, a substantial amount of work was also done by G.-M. de Schryver. The annual cash allocation of R 8 000 was hardly enough to hire a typist for 360 hours per year, not to mention employing a single full-time lexicographer, but it was utilised in the most effective way to maintain the project. It was decided to compile a bilingual dictionary Sesotho sa Leboa to English first, later officially referred to as the *Sesotho sa Leboa Dictionary Project* (SeDiPro), as a forerunner to the envisaged main monolingual dictionary. The eventual main project only started in 1999 under the auspices of the Pan South African Language Board (PanSALB) and is currently managed by the Board of the Sesotho sa Leboa National Lexicography Unit (NLU) on the campuses of the University of Limpopo and the University of Pretoria, with V.M. Mojela as the Editor-in-Chief. Pilot dictionaries for both projects have already been published, namely *SeDiPro 1.0* (Prinsloo & De Schryver, 2000) and the *Pukuntšutlhaloši ya Sesotho sa Leboa 1.0* [Explanatory Sesotho sa Leboa Dictionary 1.0] (De Schryver, 2001) respectively.

[1] Note that some debate is going on whether the official name should be *Sesotho sa Leboa* [Northern Sotho] or *Sepedi*. Cf. respectively the Government Gazette, Number 22343, June 2001, page 23, versus the Constitution of the Republic of South Africa 1996 (as adopted on 8 May 1996 and amended on 11 October 1996 by the Constitutional Assembly), Section 6, point 1.

Lexicographic difficulties faced by the pioneers

It has often been mentioned that dictionaries for the African languages lack a solid lexicographic tradition.[2] Gouws (1990:55) says that these dictionaries are unfortunately 'the products of limited efforts not reflecting a high standard of lexicographical achievement'. The compilation of a monolingual dictionary for Sesotho sa Leboa had never been attempted before. Bilingual dictionaries such as Ziervogel and Mokgokong's (1975) *Comprehensive Northern Sotho Dictionary*, as well as a number of other Sesotho sa Leboa to English or Sesotho sa Leboa to Afrikaans dictionaries by Kriel, for example *The New English – Northern Sotho Dictionary* (Kriel, 1976[4]) or the *Pukuntšu woordeboek* (Kriel, 1983[3]) respectively, did exist and could be used as guides. A detailed overview listing all major (mostly bilingual) dictionaries for Sesotho sa Leboa can be found in the Addendum.

Dictionaries such as those of Ziervogel and Mokgokong or those of Kriel, however, had been compiled according to the so-called 'traditional method': although appreciating the many virtues of these pioneering works, they were written solely on intuition, without any lexicographic planning, policies for inclusion versus omission of words, or knowledge of lexicographic problems unique to the African languages in general or to Sesotho sa Leboa in particular. In this regard, Snyman (1990:preface), for instance, honestly admits that 'common and even essential words may easily be omitted during the compiling of a dictionary', simply because they were not encountered by the lexicographer.

This means that no sound strategy for the inclusion or omission of lemma signs was available or employed at the time. Today it is generally accepted that setting up a dictionary's lemma-sign list, or in terms of Tomaszczyk (1983:51) 'to decide what to put in the dictionary and what to exclude', is the first major problem with which any lexicographer is confronted. Gove (1961[3]:4a) builds the selection to be made on the term *usefulness* as 'determined by the degree to which terms most likely to be looked for are included'.

Deciding on what to include and what to exclude also proved to be a major stumbling block for P.S. Groenewald and his team. For each candidate lemma sign, the members of the Dictionary Committee were often at odds when it came to its importance, and thus in deciding on inclusion versus omission on the grounds of relative frequency of occurrence in Sesotho sa Leboa. An initial 3-point relative scale of high (H), medium (M) or low (L) importance / frequency was introduced and each member had to intuitively decide whether a particular lemma sign was

---

[2] Since this article is being submitted for publication in South Africa, necessary sensitivity with regard to the term 'Bantu' languages is exercised in the authors' choice rather to use the term *African* languages. Keep in mind, however, that the latter includes more than just the 'Bantu Language Family'.

H, M or L. In an effort to resolve disputes, in-between categories, namely H-M and M-L, were introduced. These were not very successful and, of course, in the end remained based on the intuition of a few individuals rather than on any scientific grounds or considerations by a large number of speakers and writers of the language.

With this one has arrived at the two core issues to be investigated in this article, namely the power and value of the lexicographer's intuition and the contribution of word-frequency studies to the compilation of Sesotho sa Leboa dictionaries. The latter is addressed first, by describing word-frequency studies for Sesotho sa Leboa from their very beginnings, and by illustrating their significance at the macrostructural and microstructural level. This is followed by a discussion of the design of a modern corpus-based tool, a true Ruler, to measure and regulate the compilation of Sesotho sa Leboa dictionaries. Corpus-based compilations are then briefly contrasted to the intuition that characterised lexicography until recently.

Word-frequency counts and the creation of corpora for the African languages – the early days

It stands to reason that a large balanced and representative collection of Sesotho sa Leboa texts (a corpus), is a prerequisite for assessment of word frequency. Collecting such texts on computer only became a possibility in the mid-1980s, with the advent of the personal computer (PC). Secondly, typing millions of Sesotho sa Leboa words manually into the computer was not a realistic option, given the fact that only 360 assistant hours were available per year for the entire project. In the early-1990s this problem was resolved by the introduction of Optical Character Recognition (OCR) software. This basically entails the scanning of written texts into computer memory. Initially, this new technology had several shortcomings, at least as far as the creation of a Sesotho sa Leboa corpus is concerned. The equipment was very expensive and the project could not afford to buy it and had to work on a beg-and-borrow basis. Furthermore, only predetermined sets of characters could be recognised (such as *Arial 10* or *Courier 12*) by first-generation OCR hard- and software, and just a small percentage of typed / printed Sesotho sa Leboa text that happened to match one of the built-in character sets could be recognised. Also, the initial scanners could only operate on loose single pages that literally moved through the body of the machine, similar to sheets of paper going through a modern-day printer. Especially problematic for Sesotho sa Leboa was that all occurrences of *š* were read as *s* and corrections had to be effected manually.

Fortunately, OCR hard- and software gradually improved: flatbed scanning hardware with automatic document feeder and programs such as OmniPage or Recognita, which even offer training facilities for enhanced recognition, were soon introduced. Trainable software solved, among other things, the recognition

problem regarding *s* versus *š*. The corpus became known as the *Pretoria Sesotho sa Leboa Corpus* (PSC) and gradually grew from 156 000 running words or 'tokens' in 1990 (Prinsloo, 1991) to 5.8 million words a decade later (De Schryver & Prinsloo, 2001). The availability of a corpus opened doors to a variety of new research possibilities for and insights into lexicography, linguistics, translation studies, etc. In due course, corpora for *all* official South African languages were built at the University of Pretoria, with sizes averaging several million tokens per language.

The final but crucial missing link in the quest for frequency counts and the ability to study multiple occurrences of a single word in context was a computer program capable of actually sorting and counting words and capable of producing concordance lines. Such a program, based on Microsoft Access, was designed at the University of Pretoria and used up to 1999, when it was replaced by the commercial program *WordSmith Tools* (Scott, 1999). The availability of a Sesotho sa Leboa corpus and corpus query programs in the 1990s marked the real beginning of corpus-based African-language dictionaries.

The role of word-frequency counts and concordance lines at macrostructural and microstructural level

At this stage it was possible to generate the three crucial outputs for use in the dictionary project, namely (a) overall frequency counts, (b) distributions of those counts across the various sources or sub-corpora, and (c) concordance lines. The latter are also known as KWIC [keyword in context] lines, as they list numerous occurrences of a specific 'node' (being a word, part of a word, or even an entire phrase – including wild cards and Boolean operators) in context with 'co-text'. Not only could the intuitive H, M and L as well as H-M and M-L labels of the early Dictionary Committee days now be replaced by sound corpus thresholds, the entire treatment of each article could also be approached with much sounder tools.

Indeed, when it comes to concordance lines, it can be argued that their consideration by the modern lexicographer is indispensable in compiling a better microstructure. Consider Table 1, which shows a sample of KWIC lines generated from PSC for *ntšha* [take out].

Table 1: Sample of concordance lines for the Sesotho sa Leboa word *ntšha* [take out]

#	Left co-text	Node	Right co-text
1	*bo bonala bo sa le gona. Ba mo tšea ba mo*	***ntšha***	*moleteng. Ge ba mo lebelediša ka seetša*
2	*ka gana ge e le yona ka re ke noko. O ile a*	***ntšha***	***mphaka*** *wa gagwe wa bogale. Ka wona a*
3	*seatla sa yona. Ge ba goroga ke ge* ***letšatši le***	***ntšha***	***nko****. Ba gorogile ka mogobo wa nngalaba,*
4	*Ba tliše!" Phukubje ya tsena ka mphomeng ya*	***ntšha***	*tawana e tee ya e iša go mmagoyona, ya e*
5	*a ba Jabese wa ga Gileada. Bjale* ***phuthego ya***	***ntšha***	***banna*** *ba dikete tše lesome le metšo e*
6	*Ka lebaka la bodiidi le la bohumi kgadi yeo ya*	***ntšha***	*lentšu ya re: "Re tlo bona ge o ka tla wa*
7	*ke utswitšego thekethe ya Seila ya dipere ka yo*	***ntšha***	***tšhelete*** *ka yona. O be a romilwe ke*

A single glance at the concordance lines for *ntšha* is sufficient to highlight possible senses and sub-senses such as the basic sense 'to take out something', for instance a knife in Line 2, 'to pay / earn money' (literally 'to take out money') in Line 7, or idiomatic uses in the sense of 'the sun taking out its nose' (that is 'rise') in Line 3, or 'the congregation taking out men' (that is 'nominate / identify') in Line 5. The indispensability of KWIC lines for compiling dictionary articles and the usefulness of such lines as a tool for writing definitions, as well as selecting translation equivalents and typical examples of usage, is described in detail in De Schryver and Prinsloo (2000b). Currently, concordance lines constitute the microstructural backbone of the compilation of the first comprehensive monolingual dictionary for Sesotho sa Leboa. Such lines were unfortunately not available to the early pioneers or to the Groenewald initiative.

In order to briefly study the power of a corpus at the macrostructural level, one can consider Table 2, which reflects frequency counts in four, relatively small, randomly selected different sub-corpora from PSC. Each of these sub-corpora consists of ten texts (varying in length), labelled Sub-corpus 1 to Sub-corpus 4. In Column 2 the total count or overall number of occurrences of each sample word in all four sub-corpora is given. Words such as *thuto* [lesson], *polelo* [language], *ngwala* [write], and so on, all occur with a relatively high count in each of the four sub-corpora. A high total count and general distribution across different sub-corpora is normally a strong recommendation for the inclusion of a particular word in a dictionary, be it as a lemma sign or as part of a microstructural treatment (in other words, inside a dictionary article).

Table 2: Random sample of corpus words *cum* frequency counts and their distributions across four different sub-corpora

Word	Total Count	Count Sub-corpus 1	Count Sub-corpus 2	Count Sub-corpus 3	Count Sub-corpus 4
dingwe	12	1	1	3	7
gabotse	20	3	12	4	1
gomme	7	3	2	1	1
khwaere	12	—	—	12	—
kutu	36	—	—	14	22
lediri	44	—	—	18	26
maleme	24	6	1	10	7
ngwala	23	1	8	5	9
pedi	13	2	6	—	5
phuthego	12	—	—	12	—
polelo	30	3	2	18	7
potšišo	41	—	30	10	1
sefela	12	—	—	12	—
sekaseka	13	1	8	4	—
thuto	64	1	3	25	35

Since *thuto* is fairly generally used in all contexts, its occurrence in all the sub-corpora is not unexpected. Its relatively higher occurrence in Sub-corpora 3 and 4 is also acceptable since these sub-corpora contain a larger proportion of academic texts. This is clearly evident for grammatical terms such as *lediri* [verb] and *kutu* [stem] for which the total counts of 44 and 36 respectively lie entirely within Sub-corpora 3 and 4. *Khwaere* [choir], *phuthego* [congregation] and *sefela* [hymn] have a high total count in this selection of relatively small corpora, but occur only in Sub-corpus 3, which includes religious texts.

The macrostructures of the initial bilingual project launched by P.S. Groenewald, and of dictionaries such as the *New Sepedi Dictionary* (Prinsloo & Sathekge, 1996), the *Popular Northern Sotho Dictionary* (Kriel *et al.*, 1997[4]) or *SeDiPro 1.0* (Prinsloo & De Schryver, 2000), were all based on such frequency counts. In the latter three dictionaries, words with overall counts of at least eight, five and four respectively in the 'growing' PSC (when PSC stood at 0.85 million, 1.83 million and 3.66 million words respectively; cf. Prinsloo & De Schryver, 2001:108), were considered for inclusion. All these dictionaries thus have macrostructures that were clearly dominated by word-frequency considerations.

In selecting the lemma-sign list of a dictionary on the basis of frequency of use, the lexicographer avoids a number of typical inconsistencies characteristic of dictionaries compiled in the traditional way. Compare, for example, the typical situation of the unequal treatment of 'verb root *cum* verbal extensions' in African-language dictionaries, which results from a lemmatisation approach where

lexicographers simply add the words, in this case those verbal derivations, which 'happen to cross the compilers' way'. In Table 3, four randomly selected verb stems that occur with a relatively high frequency in PSC, namely *dira* [make; do], *kwa* [hear; feel], *tseba* [know] and *fihla* [arrive; hide], are listed in respect of a number of derivations (Column 1) of these verb stems, together with an indication of their frequency counts in the 5.8-million-word PSC.

Table 3: Frequency counts of *dira*, *kwa*, *tseba* and *fihla*, and some of their verbal derivations, in the 5.8-million-word PSC

Frequency		13 983		12 187		11 140		8 070	45 380
Root → **↓ Derivation**	*dira*	8 130	*kwa*	6 207	*tseba*	8 888	*fihla*	5 698	28 923
+ perfectum	***DIRILE***	**1 255**	*kwele*	1 689	***TSEBILE***	**326**	*fihlile*	1 093	4 363
+ causative	*diriša*	1 186	***KWEŠA***	**143**	*tsebiša*	554	*fihliša*	171	2 054
+ passive	*dirwa*	1 039	***KWEWA***	**131**	***tsejwa*** ***TSEBJA***	**0** **651**	*FIHLWA*	17	1 838
+ applicative	*direla*	641	***KWELA***	**271**	*TSEBELA*	74	*fihlela*	841	1 827
+ reciprocal	*dirana*	8	*kwana*	1 169	*tsebana*	160		0	1 337
+ neutro-passive	*direga*	866	*KWEGA*	18	*tsebega*	244	*FIHLEGA*	1	1 129
+ neutro-active		0	*kwala*	1 052		0		0	1 052
+ causative + causative		0	***KWEŠIŠA*** *KWIŠIŠA*	**918** 86	*tsebišiša*	29		0	1 033
+ causative + passive	***DIRIŠWA***	**306**	*KWEŠWA*	33	***TSEBIŠWA***	**117**	*FIHLIŠWA*	10	466
+ intensive neutro-active	*diragala*	66	*kwagala*	288	*tsebagala*	2		0	356
+ perfectum + passive	***DIRILWE***	**222**	*KWELWE*	38	*TSEBILWE*	14	*FIHLILWE*	14	288
+ applicative + applicative	*DIRELELA*	21	*KWELELA*	33		0	*fihlelela*	135	189
+ applicative + perfectum	*diretše*	80	*KWETŠE*	42	*TSEBETŠE*	1	*fihletše*	39	162
+ applicative + passive	*DIRELWA*	63	*KWELWA*	17	*TSEBELWA*	1	*FIHLELWA*	20	101
+ causative + reciprocal	*dirišana*	49	*KWEŠANA*	4	*TSEBIŠANA*	7	*fihlišana*	17	77
+ neutro-active + neutro-passive		0	*KWALEGA*	1	*tsebalega*	63		0	64
+ neutro-active + causative		0	*KWATŠA*	46	*tsebatša*	9		0	55

+ applicative + perfectum + passive	*diretšwe*	38		0		0	*FIHLETŠWE* 14	52
+ transitive reversive	*dirolla*	11		0		0	0	11
+ causative + applicative	*DIRIŠETŠA*	2	*KWEŠETŠA*	1	***tsebišetša***	**0**	0	3
+ denominative + neutro-active + causative		0		0	***tsebafatša***	**0**	0	0
+ causative + intensive neutro-active		0		0	***tsebišagala***	**0**	0	0
+ neutro-active + intensive neutro-active		0	***kwalagala***	**0**		0	0	0

Words given in lowercase were either entered as lemma signs or treated in the microstructure in *The New Sesotho – English Dictionary* (Kriel, 1950), while those in uppercase were not. The first row reflects the total frequency counts of each stem and its verbal derivations. The final column indicates the total frequency counts for each type of verbal derivation. Both have been sorted in order of decreasing frequency.

When studying Table 3 as a whole, it is hard to explain why highly-used derivations, especially those with frequency counts higher than 100 (indicated in uppercase bold), such as *dirile* (1 255), *kwešiša* (918), *tsebja* (651), and so on, were omitted, and this at the expense of rather rare derivations, especially those lacking even a single occurrence in PSC (indicated in lowercase bold), such as *tsejwa* (0), *kwalagala* (0) or *tsebišetša* (0).

In a similar vein, Ziervogel and Mokgokong's (1975) *Comprehensive Northern Sotho Dictionary* could be criticised for including only five of the seven days of the week, missing out on *Mošupologo* [Monday] and *Mokibelo* [Saturday]. Ironically, these two days are among the three most-frequently-used days; and even belong to the top-3 000 words of the Sesotho sa Leboa lexicon (cf. De Schryver & Lepota, 2001:6, 37). A number of similar macrostructural inconsistencies in existing African-language dictionaries are discussed in great detail in De Schryver and Prinsloo (2000a:293-297). The recurrent theme is that words likely to be looked for were omitted, whilst precious dictionary space was utilised for words unlikely to be looked for by the target users, or thus that all these instances concern the lemma-sign list in terms of inclusion versus omission.

Focusing once more on Table 3, one could say that the omission of certain items such as *dirile*, *kwešiša* or *tsebja* versus the inclusion of other items such as *tsejwa*, *kwalagala* or *tsebišetša*, although occurring at a macrostructural level, actually corresponds to micro-elements of the macrostructure in the sense that these are dispersed *ad hoc* lemma signs which should be added or left out throughout the dictionary. Consequently, the problem of inconsistency also has a *macro*-macrostructural dimension, namely inconsistencies observed when zooming out and examining the dictionary as a whole. Such inconsistencies are normally detected when random sections of a dictionary are compared with each other. A typical inconsistency found for African-language dictionaries in this regard is the tendency to over-treat the initial sections or alphabetical categories of a dictionary and to under-treat the final ones. Such imbalances for Sesotho sa Leboa have, for instance, been reported in De Schryver and Prinsloo (2001:377-378) for Kriel's (1983[3]) *Pukuntšu woordeboek*. In this dictionary, it is clear that Kriel started off with great enthusiasm, lumping verbal and nominal derivations of a particular stem together, giving rather detailed grammatical guidance, and treating expressions and collocations extensively. The number of articles on page 2, a random page in the alphabetical category **A**, is 22. However, towards the end of the dictionary, Kriel not only changed his lemmatisation approach from lumping to splitting, he also limited the treatment per article to an absolute minimum. The number of articles on page 281, for example, a random page in the alphabetical category **S**, is 75 – thus three times more than for **A**.

It is interesting to note that the *opposite tendency* regarding this type of macro-macrostructural inconsistency is found in dictionary compilation for Indo-European dictionaries (Sue Atkins, *personal communication*, 29 January 2003), presumably as a result of the fact that lexicographers can gain confidence as they proceed. Be this as it may, the need for a measurement and prediction instrument, or Ruler, at the macro-level of macrostructural compilation is evident. Ideally, such a Ruler should make both measurements and predictions possible when it comes to (a) the number of lemma signs and/or the number of pages per alphabetical category, (b) the relative length of articles (expressed as the number of articles per page or as the number of column-lines per article), and even (c) the time spent on or needed for compiling each alphabetical category.

Crafting a multidimensional Ruler for the compilation of Sesotho sa Leboa dictionaries

Work on such a Ruler for Sesotho sa Leboa already started in 1999-2000. The need was quickly felt, however, to test the concept on languages with an established dictionary culture first, which led to the design of so-called 'multidimensional lexicographic Rulers' for English and Afrikaans (cf. Prinsloo & De Schryver 2002; 2003). Conversely, a Ruler was also designed for isiNdebele, for which not a single

general-language dictionary exists with isiNdebele as the source language (cf. De Schryver 2003). The undertaken research showed that a sound general-language Ruler for a particular language can successfully be built from an average of corpus counts on the one hand and measurements of page allocations in existing dictionaries on the other. The research further indicated that, wherever either corpora or either dictionaries are not available, a Ruler could also be built from just one of these two components.

Keeping in mind that all Sesotho sa Leboa dictionaries from the early days lack a solid lexicographic tradition, it does not seem appropriate to include existing dictionary data in the design of a 'Sesotho sa Leboa Ruler'. Hence, unlike for example the Afrikaans Ruler which is built from both corpus and existing dictionary data, the Sesotho sa Leboa Ruler will be built using corpus data only. Furthermore, unlike the isiNdebele Ruler for which corpus lemmatisation is absolutely crucial, owing to the disjunctive orthography of Sesotho sa Leboa combined to the fact that a user-friendly Sesotho sa Leboa dictionary is word-based,[3] in crafting the Sesotho sa Leboa Ruler counts derived from an unlemmatised corpus may successfully be consulted, rather than lemmatised counts.

In hindsight, the design of the Sesotho sa Leboa Ruler is rather straightforward indeed, and should be considered a special case of the greater theoretical framework – much as Isaac Newton's equations turned out to be special cases of Albert Einstein's theory of relativity. In simple terms this means that below the surface of undemanding word-frequency counts, a solidly-tested method for the design of multidimensional lexicographic Rulers resides.

Nonetheless, one could say that in crafting a multidimensional Sesotho sa Leboa Ruler, calculations regarding the frequency of occurrence of Sesotho sa Leboa words are employed in a more advanced dimension of corpus-based dictionary compilation. Intuitively, the function of such a Ruler can be understood as determining the percentage that should be allocated to each alphabetical category. It would, of course, be foolish to simply allocate equal space to each of the alphabetical categories in a Sesotho sa Leboa dictionary, since it is well known that the alphabetical category **M**, for example, containing among others all nouns from classes 1, 3, 4, 6 and 18, should be allocated more space than many other alphabetical categories taken together. In Column 2 of Table 4 the actual breakdown of the different words or 'types' per alphabetical category in the 5.8-

[3] In a user-friendly African-language dictionary, nouns are entered under both their singular *and* plural (as they are found in different alphabetical categories), each derivation of a verb is entered as a *separate* article (and thus not crammed into the article of the root), adjectives are entered under *each* of their forms (to be found throughout the alphabet), etc. – with the appropriate cross-references.

million-word PSC is shown. When the number of types per alphabetical category (Column 2) is expressed as a percentage (Column 3), one arrives at the sought 'Sesotho sa Leboa Ruler'. From it one can for instance deduce that the category **M** comprises as much as 16.92% of the entire dictionary, whereas, say, **A** only takes up 2.47%.

Also in Table 4, this Ruler is compared to the page-breakdown of the Sesotho sa Leboa to Afrikaans and Sesotho sa Leboa to Afrikaans/English sections in two dictionaries for which a straightforward word-lemmatisation approach was followed, namely Kriel's *Pukuntšu woordeboek* (1983[3]) and Lombard *et al.*'s *Sediba* (1992) respectively. Note that Ruler percentages are compared with dictionary-page percentages, a sound methodology, implicit in the theoretical Ruler framework. Kriel's category **M**, for instance, takes up 20.60% of the entire Sesotho sa Leboa to Afrikaans section, which is 3.67% larger in absolute terms (and 21.71% larger in relative terms) than the Ruler suggestion. Similarly, the next category, **N**, is under-treated by 0.96% in absolute (and 16.18% in relative) terms. Here one immediately sees the power of a Ruler: an existing dictionary can be 'measured' and alphabetical sections that deviate too much can be 'rectified' in future editions.

Table 4: The 'Sesotho sa Leboa Ruler' compared to two straightforward word-based Sesotho sa Leboa dictionaries

	RULER (PSC)		**Ruler vs. Kriel 1983[3]**		**Kriel 1983[3]**		**Ruler vs. Lombard *et al.* 1992**		**Lombard *et al.* 1992**		
	types	**%**	**abs. %**	**rel. %**	**pp.**	**%**	**abs. %**	**rel. %**	**pp.**	**%**	
A	3 638	2.47	-1.00	-40.32	4.8	1.47	-0.82	-33.04	1.0	1.65	**A**
B	13 984	9.49	+0.92	+9.65	33.9	10.41	+1.75	+18.45	6.8	11.24	**B**
D	9 964	6.76	-1.02	-15.11	18.7	5.74	+0.18	+2.68	4.2	6.94	**D**
E	2 338	1.59	-0.88	-55.50	2.3	0.71	+0.56	+35.44	1.3	2.15	**E**
F	3 645	2.47	-0.57	-23.06	6.2	1.90	-0.32	-13.12	1.3	2.15	**F**
G	5 397	3.66	-1.21	-32.95	8.0	2.46	+0.80	+21.86	2.7	4.46	**G**
H	5 549	3.77	-0.33	-8.70	11.2	3.44	-0.46	-12.21	2.0	3.31	**H**
I	6 074	4.12	-0.62	-15.10	11.4	3.50	-2.14	-51.88	1.2	1.98	**I**
J	798	0.54	-0.33	-60.32	0.7	0.21	-0.21	-38.95	0.2	0.33	**J**
K	9 404	6.38	+3.81	+59.69	33.2	10.19	-0.10	-1.57	3.8	6.28	**K**
L	9 137	6.20	+1.50	+24.26	25.1	7.70	+2.23	+35.96	5.1	8.43	**L**
M	24 937	16.92	+3.67	+21.71	67.1	20.60	+2.09	+12.33	11.5	19.01	**M**
N	8 742	5.93	-0.96	-16.18	16.2	4.97	-0.81	-13.62	3.1	5.12	**N**
O	1 995	1.35	-0.80	-59.19	1.8	0.55	-0.53	-38.95	0.5	0.83	**O**

P	6 854	4.65	+1.18	+25.39	19.0	5.83	-0.68	-14.71	2.4	3.97	**P**
R	4 566	3.10	-1.26	-40.56	6.0	1.84	+0.37	+12.03	2.1	3.47	**R**
S	12 887	8.74	-1.69	-19.27	23.0	7.06	-0.48	-5.49	5.0	8.26	**S**
T	14 907	10.12	+0.81	+8.02	35.6	10.93	-1.02	-10.13	5.5	9.09	**T**
U	826	0.56	-0.38	-67.14	0.6	0.18	-0.40	-70.51	0.1	0.17	**U**
V	346	0.23	-0.20	-86.93	0.1	0.03	-0.23	-100.00	0.0	0.00	**V**
W	814	0.55	-0.40	-72.22	0.5	0.15	-0.22	-40.15	0.2	0.33	**W**
Y	459	0.31	-0.22	-70.44	0.3	0.09	+0.51	+165.34	0.5	0.83	**Y**
Z	108	0.07	-0.04	-58.12	0.1	0.03	-0.07	-100.00	0.0	0.00	**Z**
	147 369	100.00	***r* = 0.972**		325.8	100.00	***r* = 0.979**		60.5	100.00	

On the whole, Kriel's dictionary under discussion here did rather well, as the correlation coefficient *r* with the Ruler is as high as 0.972. In this case, if Kriel's dictionary were to be revised within the same lemmatisation framework, then special attention should be given to the under-treatment of the categories **E** and **O**, and the over-treatment of **K**, which are relatively large alphabetical categories for which the relative deviation is rather high.[4] Compared to Kriel's dictionary, Lombard *et al.*'s fares even better when measured against the Ruler, as the correlation climbs to 0.979, and only the category **I** is seriously under-treated.

It should be borne in mind that this Sesotho sa Leboa Ruler was designed with straightforward, word-based dictionaries for Sesotho sa Leboa in mind, as it was assumed that future dictionaries for Sesotho sa Leboa will also follow this lemmatisation approach. When Van Wyk revised Kriel's *Pukuntšu woordeboek* (1983[3]), for instance, although still adhering to a word-based approach, he consistently did away with plural nouns and lumped many verbal derivations together. Rules in the dictionary's front matter are supposed to provide enough guidance to the user. As Van Wyk also added numerous new words, especially under the category **M**, direct comparisons are not possible. The point is thus that the Sesotho sa Leboa Ruler as designed cannot be successfully used for dictionaries other than those with a straightforward, word-based lemmatisation. This is illustrated in Table 5, where one can see that the correlation between Van Wyk's revision and the Ruler goes down to 0.932.

In simple terms this means that another Sesotho sa Leboa Ruler must be designed for dictionaries compiled within Van Wyk's framework. Likewise, yet another Ruler must be designed for stem-based Sesotho sa Leboa dictionaries, as a comparison between Ziervogel and Mokgokong's (1975) *Comprehensive Northern*

[4] Note that when the relative deviation is high for small alphabetical categories, there is of course less reason for concern.

Sotho Dictionary and the Ruler in Table 5 clearly indicates. As expected, with an *r*-value of 0.497, there is no longer any correlation whatsoever.

Table 5: The 'Sesotho sa Leboa Ruler' compared to a rule-driven, word-based Sesotho sa Leboa dictionary and a stem-based Sesotho sa Leboa dictionary respectively

	RULER (PSC)		Ruler vs. "Van Wyk" Kriel *et al.* 1989[4]		"Van Wyk" Kriel *et al.* 1989[4]		Ruler vs. Ziervogel & Mokgokong 1975		Ziervogel & Mokgokong 1975		
	types	%	abs. %	rel. %	pp.	%	abs. %	rel. %	pp.	%	
A	3 638	2.47	-1.32	-53.65	3.2	1.14	+0.42	+16.98	43.8	2.89	**A**
B	13 984	9.49	-0.29	-3.05	25.9	9.20	-2.82	-29.70	101.1	6.67	**B**
D	9 964	6.76	-4.92	-72.79	5.2	1.84	-4.93	-72.97	27.7	1.83	**D**
E	2 338	1.59	-0.58	-36.27	2.8	1.01	-0.22	-13.85	20.7	1.37	**E**
F	3 645	2.47	-0.65	-26.41	5.1	1.82	+2.34	+94.48	72.9	4.81	**F**
G	5 397	3.66	-1.19	-32.39	7.0	2.48	+2.41	+65.74	92.0	6.07	**G**
H	5 549	3.77	-1.15	-30.57	7.4	2.61	+2.32	+61.51	92.2	6.08	**H**
I	6 074	4.12	-2.89	-70.20	3.5	1.23	-3.12	-75.71	15.2	1.00	**I**
J	798	0.54	-0.12	-21.62	1.2	0.42	-0.22	-39.99	4.9	0.32	**J**
K	9 404	6.38	+4.79	+75.00	31.4	11.17	+10.57	+165.71	257.0	16.96	**K**
L	9 137	6.20	+2.89	+46.63	25.6	9.09	-2.92	-47.04	49.8	3.28	**L**
M	24 937	16.92	+5.77	+34.09	63.9	22.69	-13.63	-80.55	49.9	3.29	**M**
N	8 742	5.93	-1.79	-30.23	11.7	4.14	-0.76	-12.87	78.3	5.17	**N**
O	1 995	1.35	-0.87	-64.29	1.4	0.48	-0.15	-10.80	18.3	1.21	**O**
P	6 854	4.65	+2.24	+48.16	19.4	6.89	+4.94	+106.11	145.3	9.59	**P**
R	4 566	3.10	-0.82	-26.44	6.4	2.28	+1.59	+51.37	71.1	4.69	**R**
S	12 887	8.74	-0.05	-0.61	24.5	8.69	-2.72	-31.13	91.3	6.02	**S**
T	14 907	10.12	+2.07	+20.45	34.3	12.18	+7.69	+76.03	269.9	17.81	**T**
U	826	0.56	-0.39	-69.20	0.5	0.17	-0.10	-18.18	7.0	0.46	**U**
V	346	0.23	-0.20	-83.21	0.1	0.04	-0.23	-100.00		0.00	**V**
W	814	0.55	-0.30	-53.58	0.7	0.26	-0.28	-51.32	4.1	0.27	**W**
Y	459	0.31	-0.15	-49.37	0.4	0.16	-0.10	-33.26	3.2	0.21	**Y**
Z	108	0.07	-0.07	-100.00		0.00	-0.06	-84.24	0.2	0.01	**Z**
	147 369	100.00	***r* = 0.932**		281.6	100.00	***r* = 0.497**		1 515.5	100.00	

Given that the compilation of mostly extremely user-friendly dictionaries for Sesotho sa Leboa is envisaged in the years ahead, or thus dictionaries with a straightforward lemmatisation approach for which corpus types roughly correspond with dictionary-citation forms, other Sesotho sa Leboa Rulers will not be crafted at his stage.

The percentage breakdown in Column 3 of Tables 4 and 5 can now be plotted as a *physical ruler*, as has been done in Figure 1. Such a physical ruler as shown in Figure 1 can also actually be held against the flank of a dictionary, and space allocation per alphabetical category can be measured and compared.

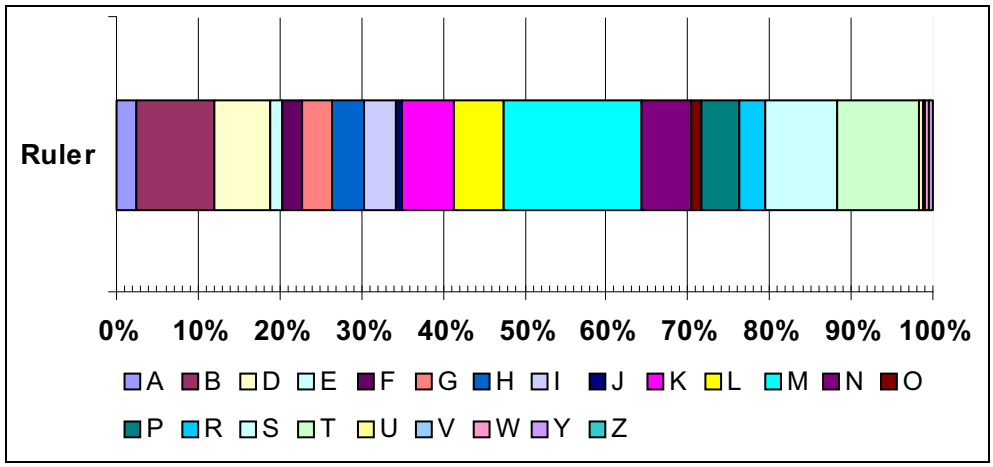

Figure 1: The multidimensional Sesotho sa Leboa Ruler as a physical ruler

Where dictionary pages contain a running and coloured thumb index (one of the so-called 'rapid outer access structure' devices), the 'running ruler' visible on the flank of the dictionary can even directly be placed against a physical ruler such as the one depicted in Figure 1.

Furthermore, where exact lemma-sign counts are available, one can compare the Ruler with the breakdown of those counts. Exact counts are for example available for SeDiPro, the project for which P.S. Groenewald laid the basis. In De Schryver and Prinsloo (2001:384-385) the number of lemma signs in an earlier version of the SeDiPro database is compared to the number of types in the 5.8-million-word PSC, and this for each alphabetical category, or thus with the Ruler. This comparison, for which the correlation is 0.964, is shown in Figure 2.

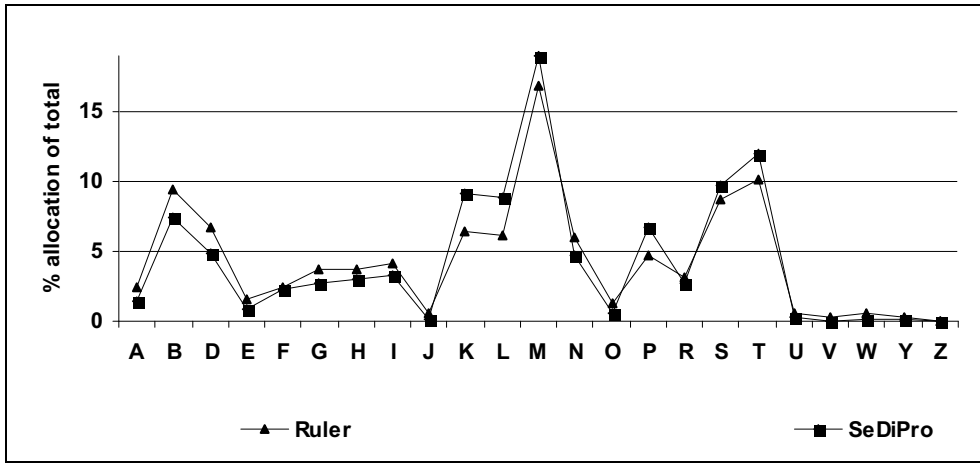

Figure 2: Lemma-sign counts per alphabetical category in the SeDiPro database versus the Sesotho sa Leboa Ruler

For the purposes of practical dictionary compilation any Ruler can of course be broken down in an arbitrary number of so-called 'Thorndike blocks' (cf. Landau, 2001:360-362), meaning, blocks reflecting smaller equal-size chunks. Such a breakdown for Sesotho sa Leboa into 100 blocks in terms of PSC is given in Table 6.

Table 6: A '100-block Ruler' for the compilation of Sesotho sa Leboa dictionaries

% Marker	% Marker	% Marker	% Marker	% Marker
1 ALAF	21 FAHL	41 KUKE	61 MONO	81 SEET
2 AROG	22 FETL	42 LAMO	62 MOŠA	82 SEJA
3 BAFE	23 FOŠW	43 LEDI	63 MOTO	83 SEMA
4 BANK	24 GALA	44 LEKA	64 MPHO	84 SERE
5 BEAB	25 GAYA	45 LEPO	65 NASW	85 SETS
6 BITL	26 GOLE	46 LETŠ	66 NGWE	86 SITE
7 BOGE	27 HAHA	47 LOGW	67 NKOK	87 STEF
8 BOKO	28 HLAH	48 MABE	68 NTEB	88 SWEL
9 BOMM	29 HLOG	49 MAGA	69 NTSE	89 TEKE
10 BOPU	30 HOSE	50 MAKA	70 NYAK	90 THAT
11 BOTL	31 IHLO	51 MAMO	71 OLEL	91 THOM
12 BUWA	32 ILWA	52 MARA	72 PANK	92 TIKR
13 DIAP	33 IPIT	53 MATH	73 PHAK	93 TLHA
14 DIIP	34 ITLH	54 MEAG	74 PHET	94 TONA
15 DIKU	35 JESU	55 MELO	75 PIPA	95 TSEN
16 DIPE	36 KATO	56 MIDI	76 PŠHA	96 TŠHI

17 DITE	37 KGAN	57 MMAS	77 RANG	97 TSOL
18 DITO	38 KGOH	58 MOBO	78 RETA	98 TUME
19 DUDI	39 KGWA	59 MOHL	79 RRAG	99 WABO
20 EMAE	40 KLAS	60 MOKO	80 SATH	100 ZOUN

From such a 100-block Ruler, page, lemma sign and time dimensions can be deduced. Indeed, the first block (which goes down to ALAF) represents 1% of the dictionary in more than one dimension. At the level of page allocation it simply means 1% of the total number of central-section pages normally allocated and predetermined by the publisher. Say, for example, it is required that the central section of a dictionary should not exceed 550 pages, then each block should not be longer than five and a half pages. Consider, secondly, the relative length of each individual article. Imagine that the lexicographer decides to include only those words that occur more than three times in the corpus. At that point the number of words with a frequency greater than three for one block in PSC can be taken, and the average length of each article is calculated as 5.5 pages divided by this number. If the time allocated to the project is for example five years, completing two blocks per month will be good progress, and the dictionary will be completed on time – a rare phenomenon in dictionary compilation.

The more comprehensive the dictionary, the more useful such a Ruler determining page, article length and time intervals will be. It is virtually indispensable for multi-volume projects such as those to be completed by the National Lexicography Units for all nine official African languages of South Africa. Such a multidimensional lexicographic Ruler in fact dictates the total factor of *effective progress* in the compilation of a dictionary and represents a truly advanced application of word-frequency counts obtained from a corpus.

Basing both macro- and microstructural aspects as well as a Ruler on frequency is exciting, but one should not lose perspective on a sound balance between frequency assessments on the one hand, and the intuition and skills of the lexicographer on the other. This article therefore concludes with an effort to restore such a balance.

Harmonising intuition and frequency

Is there still room for intuition when concordance lines are available for every single word in Sesotho sa Leboa, so that lexicographers can avail themselves of hundreds or even thousands of occurrences of a word in context with co-text, and when a multitude of basic and advanced frequency assessments and statistical Rulers can also guide, monitor and predict effective progress? Does it mean that the traditional approach has now been superseded and replaced by invariably advanced, yet in the end dumb, computer calculations and frequency cut-off

points? The answer is: 'Not at all'. In the following paragraphs it is briefly argued that the lexicographer's expertise is timeless and of great value, and that the ideal situation will be a sensible harmonisation of human skills and machine capabilities.

Firstly, at the *microstructural* level, a good illustration is the approach towards the nature of examples of usage in dictionaries. In this respect the followers of the tradition of invented or made-up examples are opposed by those lexicographers who believe that examples should be taken verbatim from corpora – both parties claiming superiority of approach. Their conflicting standpoints are analysed in great detail by Prinsloo and Gouws (2000). Such an ideological conflict is unfortunate and unnecessary, since it is possible to harmonise these extremes. The lexicographer can use corpus examples as a point of departure and then edit these, thus combining the better of two worlds: the advantages of authentic use and the lexicographer's experience and intuition. Exactly the same approach should be taken in the writing of definitions, in other words, starting with what is offered by concordance lines – as in the oversimplified case of *ntšha* in Table 1 – and to supplement this with the lexicographer's own knowledge and experience as a mother-tongue speaker.

Secondly, the same holds true at the *macrostructural* level. So, for example, De Schryver and Prinsloo (2001) have shown that a well-planned combination of a variety of existing lists that were assembled without the advantage of having a corpus results in a lemma-sign list with a remarkable internal consistency. Whilst having the advantage of numerous basic and advanced statistical outputs, one should not overreact and assume that alternative methods for the creation of a dictionary's macrostructure have no virtues, or that different approaches are in principle and per definition marred by inconsistencies.

What is thus called for is both a macro- and a microstructural perspective on *corpus*-based activities versus *intuition*-based compilations by lexicographers. Today, the Groenewald dream is carried forward by young, enthusiastic lexicographers such as M.P. Mogodi, M.C. Mphahlele and M.R. Selokela, who combine their lexicographic skills with the power of the corpus. This section and in fact the entire article can therefore be concluded in terms of Martin *et al.*'s (1983:87) proposal that 'we have reached a stage where co-operation between man and machine is useful and perhaps indispensable in making better dictionaries'.

References

De Schryver, Gilles-Maurice. 2003. Drawing up the macrostructure of a Nguni dictionary, with special reference to isiNdebele. *South African Journal of African Languages* 23(1): 11-25.

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

De Schryver, Gilles-Maurice (ed.), B. Lepota, M.P. Mogodi and M.C. Mphahlele (lexicographers). 2001. *Pukuntšutlhaloši ya Sesotho sa Leboa 1.0 (PyaSsaL's First Parallel Dictionary)*. Pretoria: (SF)^2 Press.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000a. 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. 2000b. 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.

Gouws, Rufus H. 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. In Reinhard R.K. Hartmann (ed.). *Lexicography in Africa, Progress reports from the Dictionary Research Centre Workshop at Exeter, 24-25 March 1989*: 52-65. (Exeter Linguistic Studies 15.) Exeter: University of Exeter Press.

Gove, Philip B. (ed.). 1961^3. *Webster's Third New International Dictionary of the English Language*. Springfield: Merriam-Webster.

Hartmann, Reinhard R.K. (ed.). 1983. *Lexicography: Principles and Practice* (Applied Language Studies 5.) London: Academic Press.

Kriel, Theunis J. 1950. *The New Sesotho – English Dictionary.* Johannesburg: Afrikaanse Pers-Boekhandel.

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., D.J. Prinsloo and Bethuel P. Sathekge. 1997^4. *Popular Northern Sotho Dictionary, Northern Sotho – English, English – Northern Sotho*. Cape Town: Pharos.

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.

Landau, Sidney I. 2001. *Dictionaries: The Art and Craft of Lexicography (2nd edition)*. Cambridge: Cambridge University Press.

Lombard, Daniel P., Rietta Barnard and Gerhardus M.M. Grobler. 1992. *Sediba, Practical List of Words and Expressions in Northern Sotho, Northern Sotho – Afrikaans – English, English – Northern Sotho / Praktiese lys van woorde en uitdrukkings in Noord-Sotho, Noord-Sotho – Afrikaans – Engels, Afrikaans – Noord-Sotho.* Pretoria: Via Afrika.

Martin, Willy J.R., Bernard P.F. Al and Piet J.G. van Sterkenburg. 1983. On the Processing of a Text Corpus, From textual data to lexicographical information. In Reinhard R.K. Hartmann (ed.): 77-87.

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. 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. In Anna Braasch and Claus Povlsen (eds). *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002*: 483-494. Copenhagen: Center for Sprogteknologi, Københavns Universitet.

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 [Effective Progress with the *Woordeboek van die Afrikaanse Taal* as Measured in Terms of a Multidimensional Ruler]. In Willem Botha (ed.). *'n Man wat beur: Huldigingsbundel vir Dirk van Schalkwyk*: 106-126. Stellenbosch: Bureau of the WAT.

Prinsloo, D.J. and Gilles-Maurice de Schryver (eds), Pieter S. Groenewald *et al*. (dictionary committee). 2000. *SeDiPro 1.0, First Parallel Dictionary Sepêdi – English*. Pretoria: University of Pretoria.

Prinsloo, D.J. and Rufus H. Gouws. 2000. The Use of Examples in Polyfunctional Dictionaries. *Lexikos* 10 (AFRILEX-reeks/series 10: 2000): 138-156.

Prinsloo, D.J. and Bethuel P. Sathekge. 1996. *New Sepedi Dictionary, English – Sepedi (Northern Sotho), Sepedi (Northern Sotho) – English.* Pietermaritzburg: Shuter & Shooter.

Scott, Mike. 1999. *WordSmith Tools* version 3. Oxford: Oxford University Press. See for this software also <http://www.lexically.net/wordsmith/index.html>.

Snyman, Jannie W. (ed.). 1990. *Dikišinare ya Setswana – English – Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.

Tomaszczyk, Jerzy. 1983. On Bilingual Dictionaries. The case for bilingual dictionaries for foreign language learners. In Reinhard R.K. Hartmann (ed.): 41-51.

WAT. 1926–. *Woordeboek van die Afrikaanse Taal*. Stellenbosch: Bureau of the WAT. See also <http://www.sun.ac.za/wat/index.htm>.

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

Sesotho sa Leboa Dictionaries in a Historical Perspective

I Bilingual / Trilingual Dictionaries

I.1. Early Endeavours

I.1.1. German

- **Endemann, Karl**. 1911. *Wörterbuch der Sotho-Sprache (Süd-Afrika)*. Hamburg: L. Friederichsen & Co.

I.1.2. Afrikaans

- **Endemann, Theodor M.H.** 1939. Sotho-Woordelys met die Afrikaanse ekwiwalente versamel uit bestaande Sotho-Literatuur, Behoort by die Handleiding by die aanleer van Transvaal-Sotho (Sepedi). Pretoria: J.L. van Schaik.

I.2. Dictionary Pioneer: Kriel, Theunis J.

I.2.1. Afrikaans (Pukuntšu)

- **Kriel, Theunis J.** 1942. *Sotho – Afrikaanse Woordeboek*. Pretoria: J.L. van Schaik.
- **Kriel, Theunis J.** 1966. *Pukantšu, Noordsotho – Afrikaans, Afrikaans – Noordsotho*. Pretoria: Dibukeng.
- **Kriel, Theunis J.** 1977[2]. *Pukuntšu woordeboek, Noord-Sotho – Afrikaans, Afrikaans – Noord-Sotho*. Pretoria: J.L. van Schaik.
- **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.

I.2.2. English (New)

- **Kriel, Theunis J.** 1950. *The New Sesotho – English Dictionary*. Johannesburg: Afrikaanse Pers-Boekhandel.
- **Kriel, Theunis J.** 1958[2]. *The New English – Sesotho Dictionary*. Johannesburg: Afrikaanse Pers-Boekhandel.
- **Kriel, Theunis J.** s.d.[3]. *The New English – Sesotho Dictionary*. Johannesburg: Afrikaanse Pers-Boekhandel.
- **Kriel, Theunis J.** 1976[4]. *The New English – Northern Sotho Dictionary, English – Northern Sotho, Northern Sotho – English*. Johannesburg: Educum Publishers.

I.2.3. English (Popular)

- **Kriel, Theunis J.** 1971. Popular N.Sotho Pocket Dictionary, N'Sotho – English, English – N'Sotho. Pretoria: Dibukeng.
- **Kriel, Theunis J.** 1976[2]. Popular Northern Sotho Dictionary, N.Sotho – English, English – N.Sotho. Pretoria: J.L. van Schaik.
- **Kriel, Theunis J.** 1988[3]. Popular Northern Sotho Dictionary, N.Sotho – English, English – N.Sotho. Pretoria: J.L. van Schaik.
- **Kriel, Theunis J.**, **D.J. Prinsloo** and **Bethuel P. Sathekge**. 1997[4]. Popular Northern Sotho Dictionary, Northern Sotho – English, English – Northern Sotho. Cape Town: Pharos.

I.3. Linguistic Approach: Ziervogel, Dirk

I.3.1. Afrikaans (Woordelys)

- **Ziervogel, Dirk**. 1949. Woordelys. In Dirk Ziervogel. Noord-Sotho-Leerboek, Met oefeninge en vertalings benewens leesstukke en 'n woordelys. Pretoria: J.L. van Schaik.
- **Ziervogel, Dirk**. 1953[2]. Woordelys. In Dirk Ziervogel. Noord-Sotho-Leerboek, Met oefeninge en vertalings benewens leesstukke en 'n woordelys: 124-155. Pretoria: J.L. van Schaik.

I.3.2. Afrikaans / English (Klein)

- **Ziervogel, Dirk** and **Pothinus C. Mokgokong**. 1961. Klein Noord-Sotho woordeboek, N.-Sotho – Afrikaans – English, Afrikaans – N.-Sotho, English – N.-Sotho. Pretoria: J.L. van Schaik.
- **Ziervogel, Dirk** and **Pothinus C. Mokgokong**. 1965[2]. Klein Noord-Sotho woordeboek, N.-Sotho – Afrikaans – English, Afrikaans – N.-Sotho, English – N.-Sotho. Pretoria: J.L. van Schaik.
- **Ziervogel, Dirk** and **Pothinus C. Mokgokong**. 1979[3]. Klein Noord-Sotho woordeboek, N.-Sotho – Afrikaans – English, Afrikaans – N.-Sotho, English – N.-Sotho. Pretoria: J.L. van Schaik.
- **Ziervogel, Dirk** and **Pothinus C. Mokgokong**. 1988[4]. Klein Noord-Sotho woordeboek, N.-Sotho – Afrikaans – English, Afrikaans – N.-Sotho, English – N.-Sotho. Pretoria: J.L. van Schaik.

I.3.3. Afrikaans / English (Groot)

- **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.

- **Ziervogel, Dirk** and **Pothinus C. Mokgokong**. 1985[2]. 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.

I.3.4. Afrikaans / English (Oudiovista / Tegnidisc)

- **Ziervogel, Dirk** and **Pothinus C. Mokgokong**. *s.d.* [1981]. N. Sotho woordeboek / N. Sotho Dictionary. Goodwood: Oudiovista-Produksies.

I.3.5. English (Van Schaik)

- **Anon**. [**Ziervogel, Dirk**]. 1962[4]. English – Northern Sotho Vocabulary, Northern Sotho – English Vocabulary. In Anon [Dirk Ziervogel]. *Van Schaik's Northern Sotho Phrase Book with Vocabulary For Use in the Transvaal*. Pretoria: J.L. van Schaik.
- **Anon**. [**Ziervogel, Dirk**]. 1990[5]. English – Northern Sotho Vocabulary, Northern Sotho – English Vocabulary. In Anon [Dirk Ziervogel]. *Van Schaik's Northern Sotho Phrase Book with Vocabulary*: 52-84. Pretoria: J.L. van Schaik.

I.4. Terminology / Special Purpose

I.4.1. Terminology

- **Department of Native Affairs**. 1957. Bantu Education, Sotho (N.Sotho, S.Sotho, Tswana), Terminology and Orthography No. 1. Pretoria: Government Printer.
- **Department of Bantu Education**. 1962[2]. Northern Sotho Terminology and Orthography No. 2 / Noord-Sotho terminologie en spelreëls No. 2. Pretoria: Government Printer.
- **Departmental Northern Sotho Language Committee**. 1972[3]. Northern Sotho Terminology and Orthography No. 3 / Noord-Sotho terminologie en spelreëls No. 3. Pretoria: Government Printer.
- **Departmental Northern Sotho Language Board**. 1988[4]. Northern Sotho Terminology and Orthography No. 4 / Noord-Sotho terminologie en spelreëls No. 4 / Sesotho sa Leboa mareo le mongwalo No. 4. Pretoria: Government Printer.

I.4.2. Special Purpose

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

I.5. Miscellaneous

I.5.1. Afrikaans

- **Kotzé, Nico J.** 1957. Noord-Sotho – Afrikaans, Afrikaans – Noord-Sotho Woordelys, Met 'n byvoegsel van Sotho-vakterminologie vir gebruik in Bantoeskole. Johannesburg: Voorwaarts.

- **Gerber, Hendri H.** 2000. Woordeboek Afrikaans – Noord-Sotho / Pukuntšu Seburu – Sesotho sa Leboa. Eldoraigne: Arbeidsprestasie BK.

I.5.2. Afrikaans / English

- **Joubert, P.J.** and **Matome J. Mangokoane**. 1975. Verklarende woordelys Afrikaans – Engels – Noord-Sotho, Volume 1, A-C / Lenanentšu-tlhalosi Seafrikaans – Seisemane – Sesotho sa Leboa, Bolumo 1, A-C. Johannesburg: Suid-Afrikaanse Uitsaaikorporasie.
- **Joubert, P.J.** and **Matome J. Mangokoane**. s.d. [1975-78]. Verklarende woordelys Afrikaans – Engels – Noord-Sotho, Volume 2, D-J / Lenanentšu-tlhalosi Seafrikaans – Seisemane – Sesotho sa Leboa, Bolumo 2, D-J. Johannesburg: Suid-Afrikaanse Uitsaaikorporasie.
- **Radio Bantu**. s.d. Lenanentšu la Seafrikaans – Seisemane – Sesotho sa Leboa / Afrikaans – Engels – Noord-Sothowoordelys. Johannesburg: Suid-Afrikaanse Uitsaaikorporasie.
- **Grobler, Gerhardus M.M.** 1991. The Concise Trilingual Pocket Dictionary, English – Northern Sotho – Afrikaans / Die Kort Drietalige Sakwoordeboek, Afrikaans – Noord-Sotho – English. Parklands: Ad Donker Publisher/Uitgewer.
- **Lombard, Daniel P.**, **Rietta Barnard** and **Gerhardus M.M. Grobler**. 1992. Sediba, Practical List of Words and Expressions in Northern Sotho, Northern Sotho – Afrikaans – English, English – Northern Sotho / Praktiese lys van woorde en uitdrukkings in Noord-Sotho, Noord-Sotho – Afrikaans – Engels, Afrikaans – Noord-Sotho. Pretoria: Via Afrika.

I.5.3. English

- **Hartshorne, Kenneth B.**, **J.H.A. Swart** and **Edgar Posselt**. 1984. *Dictionary of Basic English – N.Sotho Across the Curriculum*. Johannesburg: Educum Publishers.
- **Anon**. 1985. *Learner's English – N/Sotho Dictionary*. Alberton: Librarius Felicitas.
- **Wilken, Pam** and **Isaac S. Masola**. 1994. *Understanding Everyday Northern Sotho, A vocabulary and reference book / Puku ya tlotlontšu le tšupetšo*. Cape Town: Maskew Miller Longman.

I.6. Frequency Approach

I.6.1. English (New Sepedi)

- **Prinsloo, D.J.** and **Bethuel P. Sathekge**. 1996. New Sepedi Dictionary, English – Sepedi (Northern Sotho), Sepedi (Northern Sotho) – English. Pietermaritzburg: Shuter & Shooter.

I.6.2. Afrikaans (Nuwe Sepedi)

- **Prinsloo, D.J.**, **Bethuel P. Sathekge** and **Lizeth Kapp**. 1997. *Nuwe Sepedi Woordeboek, Afrikaans – Sepedi (Noord Sotho), Sepedi (Noord Sotho) – Afrikaans.* Pietermaritzburg: Shuter & Shooter.

II Multilingual Dictionaries

II.1. English, Afrikaans, Sesotho sa Leboa, Sesotho, Setswana, isiXhosa, isiZulu

- **Reynierse, Cecile**. (ed.). 1991. South African Multi-Language Dictionary and Phrase Book: English, Afrikaans, Northern Sotho, Sesotho, Tswana, Xhosa, Zulu. Cape Town: The Reader's Digest Association South Africa.

II.2. English, isiXhosa, isiZulu, Sesotho sa Leboa, Sesotho, Setswana, Afrikaans

- **Jennings, Lionel E.**, **Petrus C. Taljaard**, **Gerhardus M.M. Grobler**, **Rosemary H. Moeketsi** and **J.C. le Roux**. 1995. The Concise Multilingual Dictionary: English, Xhosa, Zulu, Northern Sotho, Southern Sotho, Tswana, Afrikaans / Die kort veeltalige woordeboek: Afrikaans, Xhosa, Zoeloe, Noord-Sotho, Suid-Sotho, Tswana, Engels. Johannesburg: Ad Donker Publisher/Uitgewer.

III Frequency-based Pilot Dictionaries

III.1. English

- **Prinsloo, D.J.** and **Gilles-Maurice de Schryver** (eds), **Pieter S. Groenewald** *et al*. (dictionary committee). 2000. *SeDiPro 1.0, First Parallel Dictionary Sepêdi – English*. Pretoria: University of Pretoria.

III.2. Monolingual

- **De Schryver, Gilles-Maurice** (ed.), **B. Lepota**, **M.P. Mogodi** and **M.C. Mphahlele** (lexicographers). 2001. *Pukuntšutlhaloši ya Sesotho sa Leboa 1.0 (PyaSsaL's First Parallel Dictionary)*. Pretoria: $(SF)^2$ Press.

IV Terminology with Monolingual Definitions

IV.1. Monolingual

- **Serudu, Maje S.** 1989. Koketšatsebo: mongwalo, mareongwalo, tsebokakaretšo. Pretoria: De Jager-HAUM.

IV.2. English, Sesotho sa Leboa, Sesotho

- **Diale, Rose M.** and **Sipho J. Neke**. 2001. Eskom Glossary of Energy Terms, English – Sepedi – Sesotho. Sandton: Eskom.

V Monolingual Dictionaries

- 'The Road Ahead'