A Critical Evaluation of Three Sesotho Dictionaries[1]

Mmasibidi Setaka, *South African Centre for Digital Language Resources,*
South Africa and University of Pretoria, Pretoria, South Africa
(mmasibidi.setaka@nwu.ac.za)
and
D.J. Prinsloo, *Department of African Languages, University of Pretoria,*
Pretoria, South Africa (danie.prinsloo@up.ac.za)

Abstract: This article gives a perspective on Sesotho lexicography and a critical analysis of the macrostructures and microstructures of three selected Sesotho dictionaries. The monolingual paper dictionary *Sethantšo sa Sesotho*, the bilingual paper dictionary *Southern Sotho–English Dictionary* and the *Sesotho online Bukantswe v.3* are evaluated. Their virtues and shortcomings as reference works will be viewed against dictionaries of high lexicographic achievement in order to establish to what extent they fulfil the most basic requirements of macrostructures and microstructures. The inconsistencies addressed in this article reflect the need for Sesotho lexicographers to use corpora in dictionary compilation in order to enhance the quality of entries on both microstructural and macrostructural levels. It will be argued that much more research and description of lexicographic issues is required to bring Sesotho lexicography on a par with its sister languages, Sepedi and Setswana and with good dictionaries for major languages of the world. After decades in existence, currently available Sesotho dictionaries are in dire need for revision and new dictionaries aimed at specific target users should be compiled.

Keywords: LEXICOGRAPHY, DICTIONARIES, SESOTHO, CORPORA, LEMMATISATION, LEMMA TREATMENT, CONCORDANCES, MICROSTRUCTURE, MACROSTRUCTURE

Opsomming: 'n Kritiese evaluering van drie Sesotho woordeboeke. Hierdie artikel gee 'n perspektief op Sesotho-leksikografie en 'n kritiese ontleding van die makrostrukture en mikrostrukture van drie geselekteerde Sesotho woordeboeke. Die eentalige papierwoordeboek *Sethantšo sa Sesotho*, die tweetalige papierwoordeboek *Southern Sotho–English Dictionary* en die *Sesotho online Bukantswe v.3* word geëvalueer. Hulle deugde en tekortkominge as naslaanwerke sal beskou word teenoor woordeboeke van hoë leksikografiese gehalte om vas te stel in watter mate hulle aan die mees basiese vereistes van makrostrukture en mikrostrukture voldoen. Die teenstrydighede wat in hierdie artikel aangespreek word, weerspieël die noodsaaklikheid dat Sesotho leksikograwe korpora in woordeboeksamestelling gebruik om die gehalte van inskrywings op mikrostrukturele sowel as makrostrukturele vlak te verhoog. Daar sal geargumenteer word dat baie meer navorsing en beskrywing van leksikografiese kwessies nodig is om die leksikografie van Sesotho op gelyke voet te bring met die sustertale Sepedi en Setswana asook met goeie woordeboeke van wêreldtale. Na dekades van gebruik, moet die Sesotho woordeboeke wat tans beskikbaar is dringend hersien word en nuwe woordeboeke saamgestel word wat op spesifieke teikengebruikers gerig is.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEKE, SESOTHO, KORPORA, LEMMA-TISERING, LEMMABEWERKING, KONKORDANSIES, MIKROSTRUKTUUR, MAKROSTRUK-TUUR

1.   Introduction

Sesotho lexicography receives very little attention in the literature, compared to e.g. its sister languages Sepedi and Setswana. For Sepedi, numerous studies have been done on problematic macrostructural and microstructural aspects such as lemma selection, treatment of lemmas and the utilisation of electronic corpora to enhance lexicographic quality. On the macrostructural level, most modern Sepedi and Setswana dictionaries utilise frequency counts from corpora as an aid to decide on inclusion or omission of lemmas for newly compiled or revised dictionaries. On the microstructural level, concordance lines culled from corpora contribute to quality enhancement in the writing of definitions, selection of translation equivalents, selection of examples, etc. De Schryver and Prinsloo (2000a) give a detailed discussion of the shortcomings in African language dictionaries on the macrostructural level due to inadequate lemma offerings, mostly as a result of including lemmas in the dictionary "as they cross the compiler's mind" rather than by means of a specific selection strategy such as frequency lists from corpora. African language dictionaries also do not perform well at the microstructural level due to inadequate treatment of the lemmas, as will be shown in more detail below. See also De Schryver and Prinsloo (2000b), Gouws and Prinsloo (2005) and Otlogetswe (2009a and b, 2012, 2013), for detailed discussions for Sepedi and Setswana. With the exception of Moleleki (1999), Prinsloo (2013), and Motjope-Mokhali (2016), no in depth lexicographic research has been recorded for Sesotho. A work of merit is Motjope-Mokhali's (2016) critical comparison of the *Sesuto–English dictionary* and *Sethantšo sa Sesotho* with reference to lexical entries and dictionary design.

In this article we will focus on Sesotho lexicography and will explore and critically analyse one monolingual dictionary, *Sethantšo sa Sesotho* (henceforth referred to as *Sethantšo*), one bilingual, *Southern Sotho–English Dictionary,* (*SSED*) and an electronic dictionary, *Sesotho Online Bukantswe v.3* (*Bukantswe*) in terms of their macrostructural and microstructural characteristics. Benchmarking the quality of these Sesotho dictionaries will be done against dictionaries of high lexicographic achievement on the basis of a number of basic requirements of macrostructures and microstructures.

2.   Criteria for the evaluation of Sesotho dictionaries

Prinsloo and Taljard (2017: 428-430) give a detailed discussion of the problematic aspects of evaluation of macrostructures and microstructures of dictionaries for African languages. They state that any comparison between African language dictionaries with dictionaries deeply rooted in a long and rich lexicographical tradition is somewhat unfair. Ideally, specific types of dictionaries with narrowly defined target users should be compared with each other in terms of the specific quality criteria applicable to that specific type of dictionary. So, for example, a Sesotho monolingual dictionary for advanced learners should be compared with English monolingual dictionaries for advanced learners. None of the three Sesotho dictionaries evaluated in this article, however, has narrowly defined target users. They are all aimed at undefined target users and have to serve all possible user profiles. The researcher has no other option than to revert to an evaluation based upon the most basic criteria for the judgment of lexicographic quality of Sesotho dictionaries. Gouws (1990: 52) emphasizes that good dictionaries, as containers of knowledge, are characterised by their offering of a variety of information types. *Macmillan Dictionary* (*MD*) gives a concise summary of such basic requirements of a dictionary:

> A dictionary is a description of the vocabulary of a language. It explains what words mean, and shows how they work together to form sentences. http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/

Gouws and Prinsloo (2005: 144) say that "the word must be defined in such a way that the users will get all the answers to the questions that made him or her consult the dictionary".

Finding the word he/she is looking for in the dictionary and finding sufficient information about the meaning and use of the word relates to the compilation of the lemma list (as part of the macrostructure) and treatment of the lemma (data types in the microstructure).

3.   A survey of Sesotho dictionaries

Sesotho dictionaries known to the authors are listed with brief descriptions of author, type and size, where available, in table 1. This survey of dictionaries is focused on general dictionaries which according to Nkomo (2010: 372) "… have a very important role to play in the development, acquisition and use of indigenous African languages". The sizes and scope of these dictionaries are different, as some are mere word lists while others provide more comprehensive treatment of the lemma.

Table 1:   Sesotho dictionaries

Name	Date of publication First published/ latest revision	Author	Type and size
Paper dictionaries			
English–Southern Sesotho dictionary	2015	Motsapi J.	Mono-directional bilingual 408 pages
English–Southern Sesotho dictionary	1965	Hamel, L. (OMI)	Bilingual. 6 volumes
English–Sesotho vocabulary	1905	A. Casalis	Bilingual: English lemma list with basic Sesotho translations
English–Sotho, Sotho–English pocket dictionary	1960	S. Christeller	Bilingual Pocket Dictionary 144 pages
English–Se-Suto vocabulary	(Unknown)	Ellenburger, D.F.	(Unknown)
Sethantšo sa Sesotho	2005	Hlalele, B.	Monolingual 325 pages
Sesuto–English dictionary	1937	Mabille, A. and Dieterlen, H.	Bilingual 445 pages
Bukantswe ya maleme-pedi, Sesotho–Seafrikanse	1974	J.A. du Plessis, J.G. Gildenhuys and J.J. Moiloa	269 pages
English–Sotho, Sotho–English pocket dictionary.	1960	(Unknown)	Bilingual
Sehlalosi: Sesotho cultural dictionary	1994	Matšela, F.Z.A.	74 pages
Southern Sotho–English dictionary	1988	Paroz, R.A.	598 pages
Khetsi ea Sesotho	1997	Pitso, T.T.E.	Monolingual 323 pages
Se-Sotho–English vocabulary	(Unknown)	Verdier, T.	(Unknown)
New South Sotho dictionary	1997	Chaphole, S.R.	Bilingual 103 pages
Oxford First Bilingual dictionary	2007	Paizee, D.	Bilingual 64 pages
English Sesotho Official Foundation Phase CAPS Picture Dictionary	2017	Sesotho National Lexicography unit	Bilingual dictionary 148 pages
Online Sesotho dictionaries			
Bukantswe http://bukantswe.sesotho.org/		Olivier J.A.K.	Bilingual dictionary 10,075 entries
Sesotho dictionary-Bilingo https://www.bilingo.co.za/sesotho-dictionary-2/			(Not available at time of consultation (9 Jan 2020))
Free English–Sesotho dictionary and translator-FREELANG https://www.freelang.net/dictionary/sesotho.php			Bilingual Sesotho > English: 9,980 words, English > Sesotho: 6,638 words
English–Southern Sotho Dictionary, Glosbe https://glosbe.com/en/st			Bilingual
Sesotho–English Dictionary — Apps on Google Play https://play.google.com/store/apps/details?id=alldictdict.alldict.sten&hl=en_US			Bilingual

4.   Lemma selection and treatment

In this section the aim is to evaluate the macrostructures and microstructures of *Sethantšo*, *SSED* and *Bukantswe*. The lemma lists and the treatment of lemmas in these dictionaries were evaluated based on introspection as well as through comparison with Sesotho corpora. The focus will be on the merits and contributions made to the knowledge of Sesotho, but a number of shortcomings on macrostructural and microstructural level will also be highlighted and briefly discussed. These presumed shortcomings revolve around (a) insufficient basic information in respect of meaning and translation equivalents, (b) alphabetical ordering, (c) morphological information, (d) pronunciation guidance, (e) examples of usage, (f) inconsistencies in the presentation and treatment of lemmas and (g) inadequate search functions.

4.1   *Sethantšo*

This is a monolingual dictionary written in Lesotho Sesotho orthography and first published in 2005. It contains approximately 10,000 lemmas and the typical information given in the articles of each lemma includes paraphrase of meaning (a definition), part of speech, noun class indication and prefix of the plural form in the case of nouns, past tense derivation in the case of verbs and etymology as in figure 1. Basic orthographic differences between South African and Lesotho orthographies are SA *di* > Lesotho *li*, *kg* > *kh*, *tjh* > *ch*, *fj* > *fsh*, etc. See https://en.wikipedia.org/wiki/Sesotho_orthography for a detailed discussion and typical examples.

kalatis.a(.itse) */kutu-ketso/* ho pepesa makhabane kapa seo motho a se fihletseng 'me a ithorisa ka tsona; *ho ikalatisa*: ho ipolelisa; ho ipatla ka ho ipolela le ho ipolelisa makhabane le liketso tsa bohaka; ho se ikopanye le ba bang ka hore ba tlase, ka ho iphahamisa.

kale */kutu-tlhakiso/* e boholo bo bokae? e kholo bo bokana.

kaliana(li.) */lereho 9/* lijo tse jeoang mohla bale ba tsoang ba tebukang; lijo tsa bohlokoa tsa mekete e khethehileng.

kalim.a (.ile & .me) */kutu-ketso/* ho lumella tšebeliso ea seng sa hao se tla khutlisoa; ho nehana ka seng sa hao se sebelisoa ke e mong nakoana e itseng se tla khutlela ho uena ha tšebeliso eo e felile.

kamo */sere/* le supa ho robeha ha ntho e khaohang ho sa lebelloa, haholo e entsoeng ka letsopa kapa ka ho hong ha e le motopo.

kamoh.a(.ile) */kutu-ketso/* ho robeha ka boeona ho sa lebelloa; ho pomeha ha ntho ka boeona. (<kamo)

kamol.a(.tse) */kutu-ketso/* ho poma ntho e motopo; ho khaola e entsoeng ka letsopa kapa ho roba seqo kotoana se itseng. (<kamo)

kamore(li.) */lereho 9/* phapusi ea ntlo; karolo e khaotsoeng ka lerako kahar'a ntlo. (<A)

kampel.a(.tse) */kutu-ketso/* ho koalla ka motero; ho thibeletsa ka motero k.h.r. terata. (<kampo: A & E)

kampo */sehokelo/* kapa; afa; empa.

Figure 1:   An extract from *Sethantšo*

In the treatment of *kamore* 'room' the plural form is indicated in brackets as the final part of the lemma as *(li.)*, i.e. the plural form is *likamore* 'rooms'. This is followed by part of speech indication given in italics between forward slashes indicating that it is a noun from class 9 followed by a definition and that it is borrowed from Afrikaans. The articles in this dictionary are relatively short — on average 32 articles per page in double columns. No examples of usage, collocations, pronunciation, etc. are given. In particular, examples of usage could be valuable in illustrating the different types of rooms and related terms, as has been successfully done in the *English–Sesotho Dictionary* (*ESD*) for the lemma *room* in figure 2.

room *n. (~ in a building)* kamore, phaposi;
  (sitting ~ in a house) foreisi, kamore ya
  boikgathollo, phaposi ya boikgathollo;
  (space) sebaka; *(waiting ~)* kamohelong,
  kotoise

Figure 2:   '*Room*' in the *English–Sesotho Dictionary*

The *English–Sesotho Dictionary* did well in giving a clear scope of different types of rooms by means of labelling them in brackets but still being economical, using only five lines of dictionary column space. Adding examples of use would assist users in text production and should be considered for future revisions of this dictionary. Consider the corpus lines for *kamore* in table 2.

Table 2:   Corpus lines for *kamore* 'room' in Sesotho

Tjhopho o ile a re 59 'Tjhopho said. 59'	*kamore* [… room]	*ya ho robala. Karabo: Tjho* 'the bedroom. Karabo: Tjho'
sona, o ne o ka tiisa hore 'it, you would be certain that'	*kamore* [… room]	*ya habo ya ho phomola ke s* 'the sittingroom is s'
Ka ho ya kaBarbara, 'According to Barbara,'	*kamore* [… room]	*ya bonaya ho robala, ke mo* 'their 'the bedroom, is here'
e tsohle tsa bona di na le 'all of them have'	*kamore* [… room]	*e ikgethang ya ho hlapela* 'a special bathroom'

The concordance lines in table 2 show the different types of rooms which ought to be put in the article to, help the user to understand the meaning of the word, its range of application and the opportunity to extract authentic examples of usage from the corpus as emphasized by De Schryver and Prinsloo (2000b). Such examples bring the meaning of different types of rooms to the fore and can be regarded as a natural extension of the definition. Future compilers of

Sesotho dictionaries are advised to consult Sesotho corpora in the compilation of dictionaries to enhance the quality of their micro- and macrostructural compilations. Consider the value added by an example of usage of *room* in the *Oxford Bilingual School Dictionary: Northern Sotho and English* (*ONSD*) in figure 3.

room *** *noun* (*pl.* **rooms**) ▦ phapoši ♦ Paledi's house has four **rooms**. *Ntlo ya Paledi e na le diphapoši tše nne.*

Figure 3:   Phapoši *'room'* in *ONSD*

In figure 3 the three-star markup (***) is a valuable indication to the user that *room* is a highly frequent word and such markup is often wisely perceived by learners as an implicit recommendation to not only find its meaning but also to learn such a word to extend their vocabulary of the language. The structural markers ■ and ♦ help to demarcate the different information categories, translation equivalents and examples, respectively. These markers also contribute to a user-friendly layout and are appealing to the eye.

Collocations could economically be indicated as part of examples to give an indication of words which more often than chance predicts co-occur with *kamore* 'room', thus giving users a clearer picture of its meaning.

One has to keep in mind that space is limited in single-volume paper dictionaries. Compilers often have to strike a balance between the number of entries that can be accommodated versus the exhaustiveness of the treatment. So, for example, Prinsloo (2009: 162) says the compiler is caught up in a triangulation of number of lemmas versus exhaustiveness of treatment versus price.

> In principle, these limitations leave the compiler with two basic options: the inclusion of a large number (e.g., 20,000–30,000) of lemmas with limited (e.g., 1–2 lines double column) treatment, or a limited number (e.g., 10,000) of lemmas with more exhaustive (e.g., 5–7 line) treatment.

This is also true for Sesotho dictionaries.

Sesotho monolinguals should nevertheless strive towards giving a more detailed treatment of lemmas in order to meet the basic requirements as stated above in terms of *MD* and Gouws and Prinsloo (2005).

Consider example (1) as an attempt at a model entry for *mala* 'entrails'.

(1)
mala¹ (ma-la) 5/**6** *[mala]*
Setho sa mmele se fumanwang ka mpeng ya motho kapa ka mpeng ya
    phoofolo se jarang dijo, ho di tsamaisa le ho di kenya maro a itseng
    a tsoang mabopong a ona. *Dikgoho di na le mala a masesane. Batho
    bohle ba na le mala. Ke ja mala le mohodu. Mala a kgomo a maholo.*

mala²
Lefu le tshwarang motho ha a dubehile ka mpeng, a jele dijo tse senyehileng kapa dijo di sa dula hantle ka mpeng. *Ntate o tshwerwe ke mala. Mala a mmangwane a bohloko, ebile o a tsholla. Nthabiseng o sebedisitse mala. Ke tshwerwe ke mala. Mme o mathiswa ke mala.*

mala³
Lela la pene. *Pene ya Tshepo e na lela le le lelele. Dipene di na le mala a fapaneng. Enke e ka hara lela lena e omme.*

In example (1) the three homonym distinctions are separated and indicated by superscript numbers. Syllable division, noun class singular/plural is given as well as a specific class indication, class 6 in this case, and a phonetic transcription. The class and class pair indication is done in a very economical way showing the relation 5/6 with the relevant one in this entry given in boldface.

Consider also the treatment of *mala* in figure 4, in the paper version for Sepedi, a sister language of Sesotho in *Pukuntšutlhaloši ya Sesotho sa Leboa* (*PTLH*) which is a reflection of a well-compiled article.

mala¹ *(leina ka bontši) (5/6)* kanale
ye teleletelele ye boleta yeo e
phuthaganego ka teng, e hlotla le
go sepetša dijo tšeo di šitšwego ke
mogodu go di ntšhetša ntle : *Ge go
hlabilwe kgogo, bana ba fiwa mala,
hlogwana le menatlana*

Figure 4:   Treatment of the first sense of *mala* in *PTLH*

The treatment of *mala* in example (1) and figure 4 is more appropriate: it distinguishes between the different senses and it gives usage examples. It also indicates the different data categories — someone who does not understand Sesotho will perhaps not be able to distinguish the different categories. The absence of frequency indication in (1) indicates relatively low frequency. Examples of usage are clearly illustrated in bold and paraphrased in such a way that the user is able to understand the true meaning of *mala.*

On the level of the macrostructure the lemma offering of *Sethantšho* for the alphabetical stretch "L" was compared with words occurring more than 200 times in a Sesotho corpus of approximately 1.5 million words. Only non-derived words were taken into consideration as it cannot necessarily be expected from a dictionary to lemmatise nominal and verbal derivations. For example, a dictionary cannot be criticized if it lemmatises only the non-derived forms of frequently used words, (frequencies given in brackets), such as the

verb stem *lahla* (729) 'lose, throw away' and the noun *lebenkele* (799) 'shop' and not any of their derivations. Typical derivations are verb stems containing verbal suffixes e.g. the perfect, applicative, passive and relative forms, or nouns occurring with locative, diminutive or augmentative suffixes. Consider, for example, the derived verb stems *lahlile* (215) (perfect) 'lost' and *latelwang* (251) (applicative + passive + relative) 'which is followed' as well as the locative derivation *lebenkeleng* (906) 'at the shop'. It would be user friendly if a dictionary did lemmatise frequently used derivations as has been done in *ONSD*. Lemmatisation of derivations, however, cannot be put as a requirement in this evaluation because the editorial policy of the dictionary could simply be *not* to lemmatise certain regularly derived forms of verbs and nouns. This is typically the approach for passive, perfect and locative forms irrespective of the frequency of the specific derivation, as in *Pukuntšu* (*PUKU*). *Sethantšo* did, however, lemmatise a number of frequently derived words, e.g. *lapile* (281) 'hungry'. Consider table 3 for an edited list of these words compared to the lemma list of *Sethantšo*.

Table 3: Top frequencies in the Sesotho corpus for the alphabetical stretch "L" compared to the lemma list of *Sethantšo*

Corpus word	Freq	Sethant.						
			LEBOYA	1748	NO	LAHLA	729	YES
LE	496927	NO	LEHAE	1682	YES	LAELA	673	NO
LA	79427	YES	LAOLA	1668	YES	LEBATOWA	610	NO
LEHA	12217	YES	LELOKO	1571	NO	LEHE	607	YES
LEBITSO	6783	YES	LEBESE	1543	NO	LEEME	605	NO
LEBAKA	5725	NO	LAKATSA	1272	YES	LEMA	567	NO
LELAPA	5443	NO	LEKALA	1266	YES	LEHLAKORE	500	YES
LEFU	5287	YES	LEBA	1257	NO	LEBE	490	YES
LEKA	4833	YES	LEKANA	1160	YES	LAESENSE	480	NO
LEFATSHE	4799	NO	LEIHLO	1079	YES	LABORARO	474	NO
LATELA	4245	NO	LEKE	1026	YES	LEBO	429	NO
LEFAPHA	4173	NO	LEBALE	1024	NO	LATA	425	YES
LEANO	4149	NO	LEBELLA	939	NO	LEBOHELA	416	NO
LEFA	3914	NO	LEHOLO	878	YES	LEFIFI	410	YES
LEHLOHONOLO	3884	YES	LEKOLA	874	NO	LEFEELA	408	NO
LEBALA	2009	YES	LEETO	814	NO	LEKANYA	407	YES
LEFELLA	1837	NO	LEBENKELE	799	NO	LEE	404	NO
LEBOHA	1762	YES	LEBELO	790	NO	LEFUBA	394	YES

LEHLABULA	387	YES	LABOBEDI	315	NO	LEHANO	228	YES
LELEME	380	NO	LEHLATSIPA	311	YES	LEHOLIMO	224	NO
LABOHLANO	379	NO	LEA	304	NO	LATOLA	221	YES
LELEKA	365	YES	LEBOLLO	297	YES	LEHARE	214	YES
LEBOKOSE	358	NO	LAPILE	281	YES	LELEIRI	214	NO
LABONE	343	NO	LABALABELA	243	YES	LEETSI	204	NO
LEKUNUTU	320	NO	LEHLABATHE	231	YES			
LAKALETSA	316	YES	LEFUTSO	229	NO			

The lemma list of *Sethantšo* does not compare well with the Sesotho words which occur more than 200 times in the Sesotho corpus. The dictionary lemmatised and treated only 35 of the 75 i.e. 47% of these top frequencies that can be assumed to be words likely to be looked up especially by learners of the language. Common words such as *lebaka* (5,725) 'reason', *Laboraro* (474) 'Wednesday' and *leano* (4,149) 'a plan', are obviously missing. De Schryver and Prinsloo (2000c) suggest that compilers should do much better in the compilation of lemma lists on intuition. Compilers should at least be able to capture the most frequently used words in a language even without the help of frequency lists culled from corpora.

4.2   *SSED*

The *Sesuto–English Dictionary* was published in 1876 in Lesotho. Several editions followed. The 4th edition, enlarged by Dieterlin was published in 1911. In 1959 a revised and enlarged edition, the *Southern Sotho–English Dictionary* by Paroz was published. This edition was presented in the South African orthography (Moleleki 1999: 243). The edition under discussion in this article is the 8th edition of 1961.

SSED is a classic example of dictionaries that were not compiled through the use of corpora but on introspection over time by Mabille, Dieterlin and Paroz. Moleleki (1999: 243) regards *SSED* as "the most useful and consulted work for Sesotho". He, however, bluntly states that "the work is not user-friendly. It is structured by the presupposition that the user is very conversant with the structure of Sesotho". He is of the opinion that the dictionary is not meant for learners but for those who already have a sound competence in the language. Narrowing the target users down to users who have a sound basic knowledge of the grammar of Sesotho is fine in itself, and ideally dictionaries should be aimed at clearly defined target users (Gouws and Prinsloo 2005: 3). However, if a dictionary is the only significant available reference work for a specific language it has to serve by necessity the needs of the broader Sesotho-speaking community. This includes learners of the language who are mother-tongue

speakers as well as non-mother-tongue speakers. It will be briefly argued in terms of Van Wyk (1995) below that this dictionary unnecessarily excludes users who do not have the required grammatical knowledge of Sesotho simply because it opted for bad choices in lemmatisation strategy and alphabetical ordering.

Consider the following typical examples of articles from *SSED* in figure 5.

se ·**baka** (*di* ·) n., place; distance, space; opportunity, time, occasion, chance; *ho hloka* –, to have no time; *ho*	*mo* ·**hla** (*me* ·) n., day, date; season; *le ka* – *o le mong*, never; *mehla e a ja, e a roba*, time leaves its trail of disabilities; loc. *mohla(ng)*, *me*-
*mo*l**êthêma**, n.cl.2, length; adj., long. *bo*l**êthô**, n., *hlonepho* substitute for *bosiu*, night. [cp. *bo*§*lôthô*].	*mo*·**phata** (*me*·) n., cracked pot; big clay pot. *ma***phata**, n.plur., tapeworm.
le·**kgabunyane** (*ma*·) n., homeless person taken as a servant; female servant;	*lĕ***bèlè** (*ma*–) n., grain or plant of corn;

Figure 5:   Selected entries in *SSED*

The entry s*ebaka* in this extract is indicated as a noun with the translation equivalent paradigm 'place, distance, space, opportunity, time, occasion, chance'. The plural form is indicated in brackets as (*di·*). This is followed by several examples of usage e.g. *ho hloka sebaka* 'to have no time'. Nouns are lemmatised according to the first letter of the noun stem, thus the word *lebele* is lemmatised in the alphabetical stretch B on its stem form *-bele* and *mohla* under *-hla*. Likewise *lekgabunyane* is lemmatised under *-kga*, *molethema* and *boletho* under *-le, and mophata* and *maphate* under *-pha*. If the user wants to look up *mohla* 'day' in the *SSED* he/she won't find it under M because the user is supposed to know that, since a stem lemmatisation strategy is followed, the prefix *mo-* has to be removed and that it should be looked up on the stem under H, i.e. *-hla*. However, the lemma is not found under H in a normal alphabetical order. The alphabetical stretch H runs from stems beginning with *ha-*, *hi-*, *ho-*, … *hwi-* but words starting with *hl-* are not listed — they are given under a following main stretch HL as mo.**hla**.

Van Wyk (1995), supported by Prinsloo and De Schryver (1999) and Prinsloo and Theletsane (2018), strongly argue against the use of stem lemmatisation for disjunctively written languages such as Sesotho. They argue that stem lemmatisation is unnecessary and unwanted for disjunctively written languages and that future compilers of paper dictionaries should stick to word lemmatisation. It simply means that instead of expecting the user to identify the stems *-bele, -lethema, -kgabunyane*, etc. as a prerequisite for look-up, the lexicographer could simply have lemmatised *lebele* and *lekgabunyane* under L, *boletho* under B and *mohla*, *molethema*, *mophata* and *maphata* under M. Van Wyk (1995) also dismisses all claims that stem lemmatisation is superior, more scientific and more economical than word lemmatisation.

Thus, we believe that for *SSED* changing from a word to a stem tradition in the 7th edition was a step in the wrong direction. In the front matter the motivation for changing to stem lemmatisation was the belief that the noun prefix is a mobile and exchangeable element (*SSED*: xii). Even this belief was refuted by Van Wyk (1995: 86) saying that "this assumption is, however, wrong; the morphology of the noun differs in crucial ways from that of the verb. The noun prefix is not mobile or freely exchangeable as Paroz claims".

SSED motivates its viewpoint with the example that *motho* 'a person' should be lemmatised on its stem form *-tho* in order to bring together words "which are similar in origin and related in meaning and of showing better the relative place of a given word in the language". This would hold true for *motho* ʼpersonʼ, *batho* 'peopleʼ, *setho* 'culture' and *botho* 'mankind'. The resulting entry is given in figure 6 as two columns.

Figure 6:   The entry for *motho* in *SSED*

Since *setho*, *botho*, *bothohadi*, etc. are treated separately no significant space saving is achieved. The only real advantage is that no separate entry is required for the plural form *batho*, thus saving less than a single line in the dictionary article. Providing plural forms and their treatment as separate articles does require duplication of the treatment of the singular form. However, plural forms are handled in a very economic and effective way in *ONSD* by sacrificing one line for the lemma and a skeleton treatment thereof, with a cross-reference to the singular as in example (2).

(2) **batho** *** *pl. noun* 1/2 See sg. MOTHO

So, in *ONSD* the lemmas *motho*, *batho*, *setho* and *botho* can all be found alphabetically under their first letters by even inexperienced users who do not have any grammatical knowledge of the language.

A second drawback is the phonetic ordering of lemmas in *SSED* instead of an ordinary alphabetical ordering. Digraphs and trigraphs are treated as single letters in their own right instead of as two or three individual letters for alphabetical ordering, i.e. as *a, b, c, (ch), d, e, f, fj, g, h, hl, i, j, k, kg, kh, l, m, n, nc, ng, nq, nx, ny, o, p, ph, pj, pjh, q, qh, r, s, sh, t, th, tj, tjh, tl, tlh, ts, tsh, u, v, w, x, (xh), y, z.* This means that ordinary alphabetical categories are divided into different subsections. For example, T is split up into no less than eight categories e.g. *t, th, tj, tjh, tl, tlh, ts* and *tsh*. For example, *th* in a word such as *thaba* 'mountain' comes after *tetetsa* 'bruise, beat' separated by 24 dictionary pages where in the *New South Sotho dictionary* (*NSSD*) *tetetsa* and *thaba* are separated by a single entry. To the ordinary user who does not have in-depth phonetic knowledge, this arrangement is user-unfriendly, difficult to comprehend and simply does not make sense. This affects the alphabetical ordering of bigrams and trigrams inside words as well.

It can be concluded in terms of Van Wyk (1995), Prinsloo and De Schryver (1999) and Prinsloo and Theletsane (2018) that stem lemmatisation brings no gain but imposes an unnecessary burden on the user, — it makes it very difficult for especially the inexperienced user to find words. Exactly the same goes for a phonetic instead of an ordinary alphabetical ordering. When stem lemmatisation and phonetic ordering are combined it is even worse and even experienced users struggle to look up words in such dictionaries. They often have to revert to a guidance page, if provided, or can even incorrectly conclude that the word is not in the dictionary. Compilers of future Sesotho dictionaries should seriously consider abandoning both stem lemmatisation and a phonetic alphabetical ordering.

On the level of the macrostructure, the lemma list of *SSED* compares well with top frequencies in the Sesotho corpus as indicated in table 4.

Table 4:   Top frequencies in the Sesotho corpus compared to the lemma list of *SSED*

Corpus word	Freq	SSED						
			LEFATSHE	4799	YES	LEBOHA	1762	YES
LE	496927	YES	LAPENG	4632	YES	LEBOYA	1748	YES
LA	79427	YES	LATELA	4245	YES	LEHAE	1682	YES
LEHA	12217	YES	LEFAPHA	4173	YES	LAOLA	1668	YES
LEBITSO	6783	YES	LEANO	4149	YES	LELOKO	1571	NO
LEKGOTLA	5749	YES	LEFA	3914	YES	LEBESE	1543	YES
LEBAKA	5725	YES	LEHLOHONOLO	3884	YES	LAKATSA	1272	YES
LELAPA	5443	YES	LEBALA	2009	YES	LEKALA	1266	YES
LEFU	5287	YES	LEFELLA	1837	NO	LEBA	1257	YES
LEKA	4833	YES	LEKGETHO	1787	YES	LEKANA	1160	YES

LEIHLO	1079	YES	LABORARO	474	YES	LAKALETSA	316	YES
LEKE	1026	YES	LEKGOLO	435	YES	LABOBEDI	315	YES
LEBELLA	939	NO	LEBO	429	YES	LEBETSE	312	YES
LEHOLO	878	YES	LATA	425	YES	LEHLATSIPA	311	YES
LEKOLA	874	YES	LEBOHELA	416	YES	LEA	304	YES
LEHODIMO	871	NO	LEFIFI	410	YES	LEBOLLO	297	YES
LEETO	814	YES	LEFEELA	408	YES	LEKGOWA	295	NO
LEBENKELE	799	YES	LEKANYA	407	YES	LAPILE	281	YES
LEBELO	790	YES	LEE	404	YES	LABALABELA	243	YES
LAHLA	729	YES	LEFUBA	394	YES	LEHWATATA	240	YES
LEKGETLO	703	YES	LEHLABULA	387	YES	LEHLABATHE	231	YES
LAELA	673	YES	LELEME	380	YES	LEFUTSO	229	YES
LEBATOWA	610	NO	LABOHLANO	379	YES	LEHANO	228	YES
LEHE	607	YES	LELEKA	365	YES	LATOLA	221	YES
LEEME	605	NO	LEBOKOSE	358	YES	LEHARE	214	YES
LEMA	567	YES	LABONE	343	YES	LELEIRI	214	NO
LEHLAKORE	500	YES	LEKGOBA	343	YES	LEETSI	204	YES
LEBE	490	YES	LEHLOKWA	323	YES			
LAESENSE	480	NO	LEKUNUTU	320	YES			

SSED lemmatised and treated 75 of the 84 words considered, i.e. 88%. In this case the compilers did well in the selection of top frequencies.

4.3   *Bukantswe*

It is generally believed that electronic dictionaries made a slow start, but will eventually supersede paper dictionaries in many ways.

> Though 'electronic lexicography' — the use of digital media for delivering dictionary data — dates back at least as far as 1990, the pace of change has picked up dramatically in the last five years, after a leisurely start. (Rundell 2012: 72)

Good electronic dictionaries are characterised by the utilisation of electronic features enabled by computer technology and utilisation of virtually unlimited space on the internet. The interested reader is referred to De Schryver (2003), and Prinsloo (2019a) for a more detailed discussion of such features and to Bothma, Prinsloo and Heid (2018), Prinsloo, Prinsloo and Prinsloo (2018), Prinsloo (2019a), Prinsloo and Bothma (2020) and Prinsloo and Taljard (2019) for

detailed discussions on user support tools in electronic dictionaries.

Bukantswe has more than 10,000 Sesotho entries with their English equivalents available from http://bukantswe.sesotho.org/. Searches can be done in English and Sesotho. In its self-description it is stated that *Bukantswe* is a "Bilingual English–Sesotho dictionary, [the] dataset represents a basic Sesotho dictionary compiled in the creation of a Sesotho language resource". The dictionary was developed by Jako Olivier and is "based on an on line word list published and revised since 1996" (https://repo.sadilar.org/handle/20.500. 12185/419).

At first glance when a user opens *Bukantswe* a search box is presented and users can start right away by typing in the word they are looking for. Consider figure 7 for the search word *lapeng*.

Figure 7:   *Bukantswe's* opening screen and results for *lapeng*

The user is informed about the size of the dictionary — it contains 10,075 entries. The current screen layout has been changed from the previous one which offered a clickable A–Z alphabetical option as in figure 8.

Figure 8:   A–Z sequence of alphabetical stretches from *Bukantswe*

Bukantswe gives users translation equivalents of words and not any pronunciation guidance or examples of usage.

It is a virtue of the dictionary that *noun classes* and persons are indicated as in (3a–3d). It indicates singular forms of nouns with an "s.", followed by the class number in brackets and the same for plural forms with "pl." as in (3a). Class numbers and indication of first and second person singular and plural are also given for pronouns as in (3b-3d), but are missing in cases such as (3e) and (3f) where class indication as (s.9) and (pl.02) respectively should be given.

(3)
a. **agente (s.9) diagente (pl.10)**   agent
b. **Ana** demonstrative pronoun (06)   these
c. **lohle**   all (05) (quantitative pronoun)
d. **wena**   you (singular) (absolute pronoun)
e. **kakaretso [1]**   abstract
f. **badimo [1]**   ancestors

Part of speech (POS) is indicated for nouns (n.), verbs (v.), adjectives (adj.), pronouns (e.g. quantitative pronoun), etc. as in example (4)

(4)
karabo (s.9) dikarabo (pl.10) [1]   answer (n.)
baleha   flee (v.)
a matonana   huge (06) (adj.)
ohle   all (06) (quantitative pronoun)
bja   slap (ideophone)

However, POS indication is not done consistently — consider the missing POS indication for nouns in (5a) and (5b) and missing indications for verbs in (5c) and (5d).

(5)
a. **avenyu (s.9) diavenyu (pl.10)**   avenue
b. **ketso (s.9) diketso (pl.10) [1]**   action
c. **bipetsana**   suffocate
d. **ntjhafatsa**   modify

The dictionary indicates etymology by means of a convention "(<="">" as in (6a). In this case it indicates that *saena* is derived from the English word sign. Nothing is inserted between the two double quotation marks. If, however, the origin is from another language, e.g. Afrikaans the original word or language is indicated between these two double quotation marks followed by a closed bracket as in (6b–6f). In (6b), for example, it is indicated that *kalaka* is borrowed

from 'kalk'. The placement of the closed bracket directly after the Afrikaans word is unclear.

(6)
a. **saena**   sign (v.) (<="">
b. **kalaka**   lime (<="" kalk)="">
c. **bora**   drill (v.) (<="" boor)="">
d. **amen**   amen (<="" afr)="">
e. **ankere (s.9) diankere (pl.10)**   anchor (<="" anker)="">
f. **borashe ba terata**   wire brush (<="" draad)="" afr="">

Another virtue of *Bukantswe* is that *homonyms* are distinguished. *Amohetse* has three unrelated translations, 'accepted', 'accommodated' and 'adopted'. Homonym distinction is made by homonym numbers in square brackets following the lemma as in (7), indicating three unrelated meanings for *amohetse*. Homonym numbers are even supplied for translation equivalents as in (7b) but in such cases it merely looks like synonym paradigms, i.e. *bjara* and *bjaratsa* as translation equivalents of *crush.*

(7)
a. **amohetse** [1] accepted (v.)
   amohetse [2] accommodated (v.)
   amohetse [3] adopted (v.)
b. **bjara** crush (v.) [1]
   bjaratsa crush (v.) [2]

Scientific and domain labels are used throughout the dictionary. Consider (8a) and (8b) for natural elements and (8d) for domain indication.

(8)
a. **Aluminiamo**   Aluminium (Al) [Element]
b. **Argone**   Argon (Ar) [Element]
c. **bela**   boil (liquids) (v.)
d. **Amose**   Amos (Biblical Name)

In (8) *Aluminiamo* and *Argone* are labelled as [Element] i.e. belonging to the periodic table of natural elements and *Amose* as a Biblical name in the religion domain. Such labels are valuable to the user to distinguish between words belonging to the general language versus words occurring in specific domains. So, for example, *solution* refers to the solving of a problem in the general language but is domain specific if referring to a chemical solution.

A number of shortcomings were, however, noticed.

Users who want to look up frequently used words, e.g. *mosadi* (5,140) 'woman' and *monna* (7,191) 'man' find no results. In the case of *monna* the *full*

string "monna (s.1) banna (pl.2) [2]" as search node is required; even searching for "monna (s.1) banna (pl.2)", i.e. without "[2]" renders no results. Exactly the same holds true for *mosadi*, i.e. searching for the full string "*mosadi (s.1) basadi (pl.2) [3]*" is required. This is a serious problem which existed at the time of consultation (December 2019–January 2020) and needs to be corrected urgently. As it stands the user would simply conclude that these top frequency items, *monna* and *mosadi* are not in the dictionary.

Another major problem is that the results obtained from the look-up often reflect partial-matches, i.e. mere blind data base hits instead of reflecting a dictionary article. For example, the result for the search of the second-most frequently used word in the Sesotho corpus, *le* (426,927) returns no less than 270 results of which (9) is an extract. It is clear that all words containing '*le*' (boldfaced in (9)), either in Sesotho or English have simply been blindly extracted from the database.

(9)

tenehi**le**	irritated
teotsa ya pense**le**	pencil sharpener
themperetjhara ya mme**le**	body temperature
thomello dikerafike ho tswa ka nt**le**	importing graphics
Thuto (**le**bitso) [1]	**Le**sson (female name)
tsamaisa [2]	**le**ad (v.)
tsamaisa [3]	**le**t go (v.)
tsamaisa [4]	**le**t someone/something walk (v.)

The user who wants to know the meaning of the word *le* has to read through 270 unwanted entries and ironically, the most basic meanings of *le* 'and, with, also' are not given, with *hammoho le* 'together with' as the closest match. This probably represents the worst case of information overload — something that is frequently cautioned against in the literature, cf. Gouws and Tarp (2017).

Consider also the results for *motho* (23,052) 'person' in (10):

(10)

ha ho motho	nobody
tidima ya tse jang motho	pathologist

Not indicating the basic meaning 'human being, person' but giving 'pathologist' is completely illogical and misinforms the user.

From all of these examples it is clear that even for the correctly treated lemmas there are not much microstructural information or data types offered, as discussed in terms of Gouws and Prinsloo (2005) above. The information types are limited to a translation equivalent and hardly fulfil the basic requirements of a dictionary. The dictionary does not give any synonyms where applicable, no related words, no examples of usage, no pronunciation guidance, etc. Pronuncia-

tion guidance can be very effectively given in electronic dictionaries by means of clickable icons; see figure 10. These entries in *Bukantswe* are examples of what Prinsloo and Taljard (2017: 431) call "ontoereikende bewerking" (insufficient treatment) of the lemmas. Consider the information given for *ja* 'eat' in (11) compared to corpus extracts in table 5, *SSED* and *ONSD* in figure 9 and *MD*, figure 10 respectively. Insufficient treatment is unacceptable in electronic dictionaries because virtually unlimited space and true electronic features enabled by the computer are available, cf. Prinsloo (2019b) for a detailed discussion.

(11)
ja [1]   eat (v.)

Prinsloo (2015) indicates the value of even a very limited corpus in the compilation of dictionaries. Corpus lines suggest a number of senses that can be distinguished for *-ja*. The boldfaced words in the final column in table 5 describe the sense in which the lemma *-ja* 'eat' has been used in the different lines.

Table 5:   Concordance lines and senses of *ja*

0 etsetswang Jehova; le tla	ja	mahobe a se nang ditomoso ka	**to eat**
anyetsa, a hopola hare 0 tla	ja	mofufutso wa bohlale ba hae	**to succeed**
ka ba Modimo wa rona. Le tla	ja	menono ya ditjhaba, le ithor	**to enjoy**
ka, o ne a nahana hore o tla	ja	mohlapeng? O ne a nahana ho	**to use what is near**
etsang. 35 Ba bohlale ba tla	ja	lefa la hlompho; ho e le dit	**to respect**
la di sebetsa." "Monna a tla	ja	batho." "Ke ile ka tshoha ha	**to deceive**
di tla be di 0 dumella 0 tla	ja	monate. Morwetsana 0 habile	**to live**
e tla nyopa, mme mollo 0 tla	ja	moAHo wa ya kgopo. 35 O emol	**to burn down**
hapanyetsa taba ena, empa ya	ja	setsi sekolong mona. Eitse	**to happen frequently**
a, ba re ho yena: O ke ke wa	ja	lefa la ntlo ya ntata rona	**to take the inheritance**
ba 0 romiIeng? 22 0 se ke wa	ja	mofutsana ka hoba e le mofut	**to rob a poor person**
a bolela, feel a 0 se ke wa	ja	tlwaye ere ha ke 0 roma, 0 b	**get used to**
ba molemo haholo ha 0 ka wa	ja	masapo a hlooho hore ka moso	**to think**
marao a tshabile disale. Tsa	ja	mothinya wa ho qetela di nts	**to run away**
leleme; ba le ratANg ba tla	ja	tholwana tsa lona. 22 Ya fum	**to reap rewards of**

Such corpus lines are invaluable to the lexicographer to distinguish the different senses of *eat* for consideration for inclusion in the dictionary. It often happens that lexicographers are alerted to senses that they might have missed if they had to rely on intuition only. It has to be realised, however, that not every single concordance line represents a different sense — it is the task of the lexicographer to decide on the number of senses to be distinguished.

The entry for *ja* in *SSED*, although not compiled using a corpus, as well as in *ONSD* is richer because it captures a number of senses through translation equivalents such as 'to eat' 'to despoil', 'to cost', 'to cause pain', 'to ache', etc.

SSED	*ONSD*
ja³ (*jelē*, imper. *eja*, *jaa*, pres., past *eja*, pass. *jewa*, *jowa*, perf. *jelwē*, *jewē*, *jowē*, obs. perf. *llē*, obs. pass. perf. *hlewē*) v.t., to eat; fig., to despoil; to cost; to cause pain, to ache; to affect (a disease); to put on; to bump; *ho ja*, food; act of eating;	**ja** *** *verb* ■ eat ♦ Ke ja nama ya kgomo. *I eat beef.* ◇**ja bogale** ■ be angry ◇**ja fase** ■ fail; run away ◇**ja hlogo** ■ think; be deep in thought ◇**ja leonyane** ■ stalk ◇**tša go ja** ■ something to eat

Figure 9: *-ja* in *SSED* and *ONSD*

The entry for *eat* in *MD* also indicates lexicographic richness of treatment.
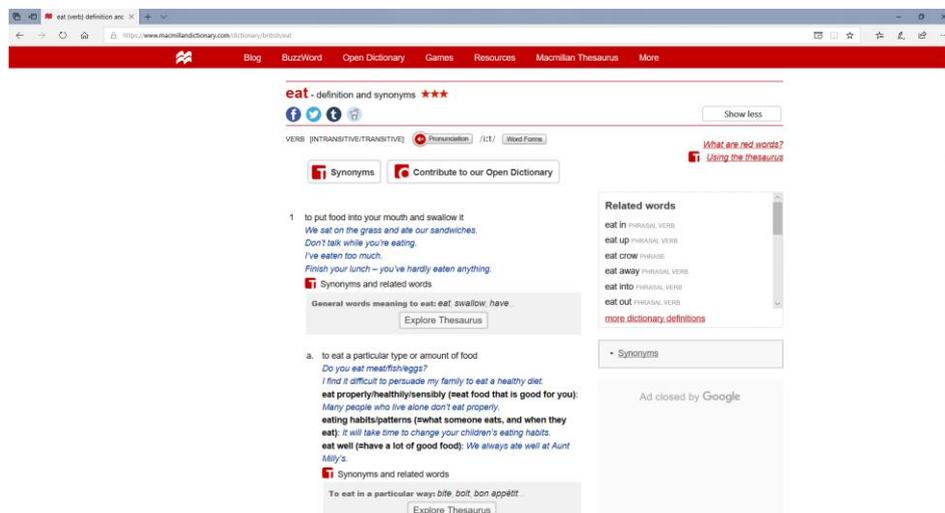
Figure 10: *eat* in *MD*
(https://www.macmillandictionary.com/dictionary/british/eat)

In figure 10 a wealth of information types are given such as different senses, translation equivalents, examples of usage, frequency indication, pronunciation, word forms, definitions, etc. This is a good example of what future Sesotho electronic dictionaries should look like.

A further shortcoming in *Bukantswe* is the lack of even a very basic user's guide to the dictionary.

On the level of the macrostructure, as indicated in table 6, the lemma list of *Bukantswe* does not compare well with top frequencies in the Sesotho corpus. This is aggravated by the presumed technical detection problem described above, i.e. cases where the dictionary gives look-up results only if the full string

is entered. It simply means searches for hundreds, if not thousands of espe-
cially nouns will not render any results.

Table 6:   Top frequencies in the Sesotho corpus compared to the lemma list of
Bukantswe

Corpus word	Freq	Bukantswe						
LE	496927	NO	LEKANA	1160	NO	LEE	404	YES
LA	79427	NO	LEIHLO	1079	NO	LEFUBA	394	NO
LEHA	12217	NO	LEKE	1026	NO	LEHLABULA	387	NO
LEBITSO	6783	NO	LEBALE	1024	NO	LELEME	380	NO
LEKGOTLA	5749	NO	LEBELLA	939	YES	LABOHLANO	379	YES
LEBAKA	5725	NO	LEHOLO	878	NO	LELEKA	365	NO
LELAPA	5443	NO	LEKOLA	874	NO	LEBOKOSE	358	NO
LEFU	5287	NO	LEHODIMO	871	NO	LABONE	343	YES
LEKA	4833	NO	LEETO	814	YES	LEKGOBA	343	NO
LEFATSHE	4799	YES	LEBENKELE	799	NO	LEHLOKWA	323	NO
LAPENG	4632	YES	LEBELO	790	YES	LEKUNUTU	320	YES
LATELA	4245	NO	LAHLA	729	NO	LAKALETSA	316	YES
LEFAPHA	4173	NO	LEKGETLO	703	NO	LABOBEDI	315	YES
LEANO	4149	NO	LAELA	673	NO	LEBETSE	312	YES
LEFA	3914	YES	LEBATOWA	610	NO	LEHLATSIPA	311	NO
LEHLOHONOLO	3884	NO	LEHE	607	NO	LEA	304	NO
LEBALA	2009	YES	LEEME	605	NO	LEBOLLO	297	YES
LEFELLA	1837	NO	LEMA	567	NO	LEKGOWA	295	NO
LEKGETHO	1787	NO	LEHLAKORE	500	NO	LABALABELA	243	NO
LEBOHA	1762	YES	LEBE	490	NO	LEHWATATA	240	NO
LEBOYA	1748	YES	LAESENSE	480	NO	LEHLABATHE	231	NO
LEHAE	1682	NO	LABORARO	474	YES	LEFUTSO	229	NO
LAOLA	1668	NO	LEKGOLO	435	NO	LEHANO	228	NO
LELOKO	1571	NO	LEBO	429	NO	LATOLA	221	NO
LEBESE	1543	NO	LATA	425	NO	LEHARE	214	NO
LAKATSA	1272	NO	LEBOHELA	416	YES	LELEIRI	214	NO
LEKALA	1266	NO	LEFIFI	410	NO	LEETSI	204	NO
LEBA	1257	NO	LEFEELA	408	NO			
			LEKANYA	407	NO			

Bukantswe lemmatised and treated 19 of the 84 words considered i.e. 23% which reflects insufficient coverage of the selected top frequencies. Even without a corpus, the lexicographer is expected to capture a greater percentage of the most frequently used verbs on intuition as mentioned above in terms of De Schryver and Prinsloo (2000c).

5.   Conclusion and future work

Sesotho lexicography is in a developing phase and much more research and description of lexicographic issues is required to bring the body of knowledge for this language on a par with its sister languages Sepedi and Setswana and lexicographic achievement of dictionaries for major languages of the world such as English, French, German, etc. The virtues and shortcomings raised in this article in respect of *Sethantšo*, *SSED* and *Bukantswe* are a true reflection of most African language dictionaries. Gouws and Prinsloo (2005: 9) say that "the publication of any dictionary should not only be the result of the preceding compilation activities but it has to be regarded as the culmination of a much more comprehensive set of activities, the so-called lexicographic process". Furthermore, the inconsistencies addressed in this article reflect the need for Sesotho lexicographers to use corpora in dictionary compilation in order to enhance the quality of entries on both microstructural and macrostructural levels. Corpus utilisation will also enable compilers to indicate frequencies of words in the dictionary as has been done in *ONSD* by means of a 3-star rating system, as in figure 9 above. After decades in existence, currently available Sesotho dictionaries are in dire need for revision and new dictionaries aimed at specific target users should be compiled. A language cannot be served by only a few dictionaries compiled as a one-size-fits-all for user needs. Gouws and Prinsloo (2010: 505) state that no single dictionary can be everything for everyone. There is also a strong need for community involvement in the compilation of Sesotho dictionaries in a true Afro-centric approach where more mother-tongue speakers of Sesotho take the initiative to compile good Sesotho dictionaries, cf. Prinsloo (2017) and Prinsloo (2019b).

Acknowledgements

Endnote

1. The term 'Bantu' got stigmatised during the Apartheid Era in South Africa. Therefore the term 'African' will be used in this article even in reference to what is internationally referred to as 'Bantu languages'.

Bibliography

Dictionaries

(*Bukantswe*) *Sesotho Online, Bukantswe v.3.* Available at http://bukantswe.sesotho.org/. [accessed: January 7, 2020].
(ESD) Motsapi, M. (Ed.). 2015. *English–Sesotho Dictionary*. Cape Town: South African Heritage Publishers.
(MD) Macmillan Dictionary. Available at http://www.macmillandictionaries.com/ [accessed: May 5, 2019].
(NSSD) Chaphole, S.R. 1997. *New South Sotho Dictionary. English–South Sotho, South Sotho–English.* Pietermaritzburg: Shuter & Shooter.
(*ONSD*) De Schryver, G.-M. (Ed.). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English.* Cape Town: OUP Southern Africa.
(PTLH) Mojela, M.V. (Ed.). 2007. *Pukuntšutlhaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.
(*PUKU*) Kriel, T.J. and E.B. van Wyk. 1989. *Pukuntšu*. Pretoria: J.L. van Schaik.
(*Sethantšo*) Hlalele, B. 2005. *Longman Sethantšo sa Sesotho*. Maseru: Longman Lesotho.
(*SSED*) Mabille, A. and H. Dieterlen. 1988. *Southern Sotho–English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.

Other

Bothma, T.J.D., D.J. Prinsloo and U. Heid. 2018. A Taxonomy of User Guidance Devices for e-Lexicography. *Lexicographica. International Annual for Lexicography* 33: 391-422.
De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age *International Journal of Lexicography* 16(2): 143-199.
De Schryver, G.-M. 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, G.-M. 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, G.-M. and D.J. Prinsloo. 2000c. (In)consistencies and the Miraculous Consistency Ratio of '(x 1.25)4 = x 2.44', A Perspective on Corpus-based versus Non-corpus-based Lemma-sign Lists. Paper presented at the Fifth International Conference of the African Association for Lexicography (AFRILEX), University of Stellenbosch, Stellenbosch, 3–5 July 2000.

Gouws, R.H. 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Ed.). 1990. *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.

Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: AFRICAN SUN MeDIA.

Gouws, R.H. and D.J. Prinsloo. 2010. Surrogaatekwivalensie in tweetalige woordeboeke met spesifieke verwysing na zero-ekwivalensie in Afrikataalwoordeboeke. [Surrogate Equivalence in Bilingual Dictionaries with Special Reference to Zero-equivalence in Dictionaries for African Languages]. *Tydskrif vir Geesteswetenskappe* 50(4): 502-519.

Gouws, R.H. and S. Tarp. 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.

Moleleki, M.A. 1999. The State of Lexicography in Sesotho. *Lexikos.* 9: 241-247.

Motjope-Mokhali, Tankiso Lucia. 2016. *A Comparative Analysis of* Sesuto–English Dictionary *and* Sethantso sa SeSotho *with Reference to Lexical Entries and Dictionary Design.* Unpublished doctoral thesis. Pretoria: UNISA. Available at URI: http://hdl.handle.net/10500/22205.

Nkomo, D. 2010. Affirming a Role for Specialised Dictionaries in Indigenous African Languages. *Lexikos*. 20: 371-389.

Otlogetswe, T.J. 2009a. *English–Setswana Dictionary*. Second edition. Pentagon Publishers: Gaborone.

Otlogetswe, T.J. (Ed.). 2009b. *MLA Kgasa: A Pioneer Setswana Lexicographer*. CASAS Book series 64. Cape Town. CASAS.

Otlogetswe, T.J. 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.

Otlogetswe, T.J. 2013. Oxford English–Setswana, Setswana–English School Dictionary. Oxford: OUP.

Prinsloo, D.J. 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.

Prinsloo, D.J. 2013. Lexicography of the Sotho Languages. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.).. 2013. *Dictionaries: An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography:* 929-946. Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science. HSK Vol. 5.4. Berlin/Boston: Mouton de Gruyter.

Prinsloo, D.J. 2015. Corpus-based Lexicography for Under-resourced Languages — Maximizing the Limited Corpus. Paper read at AELINCO 2015, 7th Conference on Corpus Linguistics, Valladolid, Spain, 5–7 March 2015.

Prinsloo, D.J. 2017. Analyzing Words as a Social Enterprise: Lexicography in Africa with Specific Reference to South Africa. Miller, J. (Ed.). 2017. *Analysing Words as a Social Enterprise: Celebrating 40 Years of the 1975 Helsinki Declaration on Lexicography. Collected Papers from AustraLex 2015:* 42-59. https://www.adelaide.edu.au/australex/publications/.

Prinsloo, D.J. 2019a Detection and Lexicographic Treatment of Salient Features in e-Dictionaries for African Languages. *International Journal of lexicography*. ecz031, https://doi.org/10.1093/ijl/ecz031. Published: 27 November 2019.

Prinsloo, D.J. 2019b A Perspective on the Past, Present and Future of Lexicography with Specific Reference to Africa. Gürlek, Mehmet, Ahmet Naim Çiçekler and Yasin Taşdemir. 2019. *Proceedings of the 13th International Conference of the Asian Association for Lexicography, ASIA-*