Chapter 4

The macrostructure

4.1   Introduction

The aim of this chapter is firstly to critically evaluate and analyse the macrostructure of existing Setswana dictionaries with a focus on the compilation of and the deficiencies in the lemma lists. Typical macrostructural inconsistencies existing in Setswana dictionaries will be highlighted. The extent of the inconsistencies will generally show how the respective Setswana dictionaries succeed or fail to treat the most commonly used words. The focus will mainly be on the following macrostructural aspects: inconsistencies regarding the lemmatization of nouns and verbal derivations, imbalances regarding the alphabetical stretches, lemmatization approaches, lemmatization strategies, lemmatization traditions and grammatical aspects regarding the lemmatization problems of the noun prefixes of classes 5 and 11; the absence of noun inflections and the inconsistencies regarding the lemmatization of homonyms and the absence of tonal indication.

Secondly, the chapter will demonstrate how corpus query tools can be used to generate alphabetical word lists and frequency lists reflecting the overall counts or specific words or words in context. Thereafter follows the plotting of data to indicate the relationship between rank and the frequency of tokens. Thirdly, the chapter will demonstrate how the Setswana dictionaries handle the treatment of dialectical words. Finally, each section dealing with the inconsistencies will conclude with suggestions for the improvement of the respective Setswana dictionaries by means of a corpus-based macrostructure.

According to Prinsloo and Gouws (1996:103), the lexicographer for African languages must find lemmatization strategies that result in a user-friendly end product. It is important for the lexicographer to find a sound balance in terms of the selection of lemmata for words likely to be looked up by the target users. Martin et al. (1983:81-82, 87) state that:

> "The decision what to include in the dictionary still has to be made by the lexicographer himself, however, and this depends in turn upon the nature and size of the dictionary and its intended users. In this respect lemmatized frequency-lists can be a further help… We have reached a stage where co-operation between man and machine is useful and perhaps indispensable in making better dictionaries"

When examining the macrostructure of the existing Setswana dictionaries in comparison to word lists culled from a Setswana corpus, one can easily determine the typical macrostructural inconsistencies that need to be rectified through the electronic corpora. Gouws (1990:55) states:

> "Lexicographical activities on the various indigenous African Languages […have] resulted in a wide range of dictionaries. Unfortunately, the majority of the dictionaries are the products of limited efforts not reflecting a high standard of lexicographical achievement."

4.2 Typical macrostructural inconsistencies existing in Setswana dictionaries

There is no dictionary that can be comprehensive enough to give a balanced account of a representative selection of a lexicon. However selection of lexical items to be included as lemmas cannot be done in an arbitrary way, but have to comply with the lexicographical standards rooted in a sound theory (cf. Prinsloo and Gouws 1995:1). The analysis of dictionaries in randomly selected alphabetical stretches or sections of

alphabetical stretches in Table 15 reveals the importance of the utilization of corpora during the creation of a dictionary's lemma-sign list.

Table 15: Comparison of the macrostructure between the fixed points *rabbit* and *rally* in various dictionaries

S.A. Oxford School dictionary. (Oxford university Press 2004:352-357)	Major Dictionary. (Eksteen, 1997:1238-1241)	Setswana-English Dictionary (Brown, 1964:512-513)	Dikišinari ya Setswana. (Snyman et. al, 1990:313)	Setswana-English-Setswana dictionary. (Matumo, 1993:590)	English- Sepedi Kriel 1976
rabbit	rabbit	rabbit	rabbit	rabbit	rabbit
race	race	race	race	race	race
racialism	_____	____	____	____	____
racism	_____	____	____	____	racism
racist	racist	____	____	____	_____
rack	rack	____	____	rack	rack
radial	radial	____	____	____	_____
radiate	radiate	radiate	radiate	radiate	radiate
radiator	radiator	____	____	____	_____
radical	radical	____	____	____	_____
radio	radio	____	____	____	radio
radiographer	radiographer	____	____	____	____
radiology	____	____	____	____	radiology
radish	radish	radish	____	____	radish
radius	radius	____	____	____	radius
raffle	raffle	____	radius	____	____
rag	rag	rag	rag	rag	rag
raid	raid	raid	____	raid	raid
rail	rail	rail	rail	rail	rail
rain	rain	rain	rain	rain	rain
raise	raise	raise	raise	raise	raise
rake	rake	rake	____	rake	rake
rally	rally	____	____	rally	rally

In Table 15, six dictionaries are compared and viewed over 23 items in the alphabetical stretch *rabbit* to *rally*. The two English dictionaries Oxford University Press (2004) and Eksteen (1997) are used as a base to indicate the macrostructural inconsistencies existing in Setswana dictionaries. Brown (1964) treats 10 items, Snyman et al (1990) treat 8 items, Matumo (1993) treats 11 and Kriel (1976) treats 16 items. It is important to note that words most likely to be consulted by the target users have been left out in most of the Setswana dictionaries as reflected in Table 15. The number of the lexical items and the absence of the treatment of most commonly used words like *racial, racist, radial, radiator, radical, radiology, radius* etc. in Setswana dictionaries, prove the urgent need of corpus utilization of word frequency counts to compile a lemmatized frequency list.

For the revision of existing Setswana dictionaries, frequency lists can play a vital role in ascertaining that frequently used words are not accidentally omitted, and, on the other hand, that dictionary space is not occupied by articles of lemmas unlikely to be looked for by the target users.

Consider the second example where inconsistency regarding the entering of derivations is evident in Setswana monolingual dictionaries.

Example 3: *Reka (buy)* (672)

rekegela (36), *rekela* (125), **rekelana (0), rekelane (0), rekelwa (8),** *reketswe* (15), *rekile* (150), *rekileng* (26), *rekilwe* (34), *rekisa* (280), *rekisang* (32), **rekiseditswe (2),** *rekisetsa* (40), *rekisitswe* (14), *rekisiwa* (43), **rekiwa (5),** *rekolola* (35), *rekwa* (69), <u>*reketse*</u> (35), <u>*rekang*</u> (37), <u>*reke*</u> (79), <u>*rekegela*</u> (36), <u>*rekegele*</u> (35), <u>*rekele*</u> (20), <u>*rekise*</u> (30), <u>*rekileng*</u> (26), <u>*rekisang*</u> (32), <u>*rekisitse*</u> (21), <u>*rekwe*</u> (19).

From this example, the inconsistency regarding the entering of derivations in THAN is evident. It is difficult to justify the inclusion and the exclusion of the derivations of the verb *reka* (buy). Frequencies are given between brackets in example 3. Highly used underlined derivations such as <u>*reke*</u> (79), <u>*reketse*</u> (35),

rekang (37), *rekise* (30), *rekisang* (32) etc. have been omitted while less frequently used derivations (bolded) such as **rekelana** (0), **rekelane** (0), **rekelwa** (8), **rekiseditswe** (2), **rekiwa** (5) etc. are entered.

Consider the third example of inconsistency in the THAN where highly frequently used derivatives of the verb *dira* are omitted or not lemmatised while less frequently used ones are lemmatized:

Example 4: THAN

```
dira  TT   tpt.  -ile.  tsêna mo tirong
nngwe: bêrêka
dirafala  TTTT   tpt.  -itse.  >dira+afala;
tôta lefoko le ka diragala ka gore le tswa
mo go dirêga
diragadiwa  TTTTT   tpt.  -itse.
>dira+agala+iwa
diragala  TTTT   tpt.  -itse.  >dira+agala
diragalang  TTTTG   tpt.
>dira+agala+ng
diragalêlang  TTTTTG   tpt.
>dira+agala+ela+ng
diragaletse  TTTTT   tpt.
>dira+agala+itse
diragaletswe  TTTTT   tpt.
>dira+agala+itse+iwa
diragatsa  TTTT   tpt.  -itse.
>dira+ega+isa
diragatsang  TTTTG   tpt.
>dira+ega+isa+ng
diragetseng  TGGTG   tpt.

>dira+ega+itse+ng
di•rala  TGT   ln./8.  O  bontsi jwa
serala
dirala  TTT   tpt.  -itse.  >dira+ala
diralang  TTTG   tpt.  >dira+ala+ng
dirana  TTT   tpt.  -ile.  >dira+ana
dirang  TTG   tpt.  >dira+ng
dirêga  TTT   tpt.  -ile.  >dira+ega
diregang  TTTG   tpt.  >dira+ega+ng
diregile  TTTT   tpt.  >dira+ega+ile
diregileng  TGGTG   tpt.
>dira+ega+ile+ng
dirêla  TTT   tpt.  -itse.  >dira+ela
go dirêla motho e se nama = go dira
ditirô tsa motho yo mongwe. mme ênê a
sa dire sepê
dirêlang  TTTG   tpt.  >dira+ela+ng
dirêlwa  TTT   tpt.  -itse.
>dira+ela+iwa
dirêlwang  TTTG   tpt.
>dira+ela+iwa+ng
di•retlo  TTT  diretlwa  ln./8.  O
mateng a a apeilweng a
tlhakatlhakantswe a seruiwa. gantsi a
```

```
schutshane
diretswe  TTT   tpt.
>dira+ela+iwa+itse
diretsweng  TGTG   tpt.
>dira+ela+iwa+itse+ng
dirile  TTT   tpt.  >dira+ile
dirileng  TGTG   tpt.  >dira+ile+ng
dirilwe  TTT   tpt.  >dira+iwa+ile
dirilweng  TGTG   tpt.
>dira+iwa+ile+ng
dirisa  TTT   tpt.  -itse.  1. >dira+isa 2
thusa yo mongwe mo tirong nngwe
dirisana  TTTT   tpt.  -itse.
>dira+isa+ana
dirisang  TTTG   tpt.  >dira+isa+ng
dirisantsê  TTTTT   tpt.
>dira+isa+ana+itse
dirisediwa  TTTTT   tpt.  -itse.
>dira+isa+ela+iwa
dirisediwang  TTTTTG   tpt.
>dira+isa+ela+iwa+ng
dirisitse  TTTT   tpt.  >dira+isa+itse
dirisitseng  TGGTG   tpt.
>dira+isa+itse+ng
dirisitswe  TTTT   tpt.
>dira+isa+itse+iwa
dirisitsweng  TGGTG   tpt.
>dira+isa+itse+iwa+ng
dirisiwa  TTTT   tpt.  >dira+isa+iwa
dirisiwang  TTTTG   tpt.
>dira+isa+iwa+ng
diriswa  TTT   tpt.  >dira+isa+iwa
diriwa  TTT   tpt.  -ilwe.  >dira+iwa
dirolola  TTTT   tpt.  -itse.  >dira+olola
dirololwa  TTTT   tpt.
>dira+olola+iwa
dirwa  TT   tpt.  >dira+iwa
dirwang  TTG   tpt.  >dira+iwa+ng
```

omitted while less frequently used words such as **direlang** (1). **dirisitswe** (2)

dirisantse (1), *dirileng* (1), to mention but few, have been lemmatized. Consider Table 16 below:

Table 16: Overall frequency counts for the derivatives of the verb *dira* (7053) in WordSmith Tools

Rank	Word	Frequency		Rank	Word	Frequency
21286	dire	2,201		21454	dirisetswa	77
21426	dirisa	1,974		21167	diragale	76
21245	dirang	1,337		21136	dirafatsa	72
21401	dirile	1,098		21171	diragalelang	72
21322	direla	955		21170	diragalela	69
21473	dirisiwa	705		21302	diregile	58
21473	dirisiwa	705		21375	diretswe	52
21617	dirwa	654		21455	dirisetswang	51
21162	diragala	537		21479	dirisiwe	46
21163	diragalang	305		21300	direge	42
21476	dirisiwang	255		21334	direleng	33
21625	dirwang	253		21466	dirisitseng	33
21413	dirilwe	250		21376	diretsweng	30
21328	direle	229		21338	direlwe	29
21441	dirise	224		21303	diregileng	26
21193	diragetse	211		21157	diragadiwa	25
21414	dirilweng	185		21337	direlwang	23
21626	dirwe	181		21499	diriwang	23
21433	dirisang	174		21125	dirafalela	22
21190	diragatsa	164		21145	dirafetse	17
21297	direga	156		21173	diragalela	9
21470	dirisitswe	154		21182	diragaletseng	5
21371	diretse	146		21443	diriseditsweng	2
21464	dirisitse	143		21402	dirileng	1
21372	diretseng	107		21322	direlang	1
21472	dirisitsweng	105		21435	dirisantse	1
21340	direng	104				
21336	direlwa	101				
21119	dirafala	81				

It is noticed from the above examples that Kgasa and Tsonope (1995) apparently lemmatized these derivations without considering frequency of use offered by a large corpus, although they indicate that a corpus was consulted. For this reason,

the dictionary does not address the needs of the users since derivations likely to be looked up by users are left out. Ideally one needs the overall frequencies as well as information on the spreading of these words across the different sources for the compilation of the lemmalist.

4.2.1   Physical limitation on the volume

According to Prinsloo (1994:94), limitations mostly on the number of pages or amount of entries that can be accommodated in a specific dictionary or sub-dictionary, has a far greater impact on the lemmatization of African languages than one would expect. Busane (1990:30) states that:

> "One of the basic problems of lexicography is to decide what to put
> in the dictionary and what to exclude"
> "…the problem remains as to whether all the lexical units that are
> likely to be derived from the main entry or the stem should be
> entered in the dictionary"

This implies that the need to select what to include or exclude in a dictionary proved to be a major concern for the compilers of the Setswana dictionaries.

Taking into consideration the limitations on the volume of Setswana dictionaries, the THAN contains 330 pages with approximately 16500 entries. This dictionary contains a huge number of derived forms presented as lemmas with elaborate comments on the form or morphological information without semantic information as reflected in example 4 above. This can lead to frustration and uncertainty about the exact meaning of the derivations and it is time consuming to try to find their meanings by artificially adding on the meaning of each derivation. (See *dikgakollišano* in Table 23 below).

4.2.2   Imbalances regarding the alphabetical stretches

The second macrostructural problem apart from inclusion versus omission is the balancing out of the entire alphabetical categories of the Setswana dictionaries as a whole. This simply means that dictionary compilers often starts off treating the first few alphabetic categories exhaustively, but then grow tired towards the end of the alphabet. Compare (Prinsloo 2000a and Prinsloo 2000b) in reference to Kriel's dictionary. Landau 2001:398 states:

> "Nothing is more difficult to predict or control than a dictionary
> begun from scratch"

This remark is equally applicable to dictionaries that were compiled without the availability of a corpus. (See De Schryver and Prinsloo (2000) and Prinsloo and De Schryver (2003) for numerous examples of inconsistencies regarding over and under treatment in terms of alphabetical categories). Consider the following example below:

Example 5

PUKU1

> **aka**, *a.ka.* (*-ile*, *-etše*), lieg, leuens vertel, jok, onwaarheid spreek (dial. kyk: *aketša*).
> **aka**, *a.ka*, inhaak, vashaak, haak, aanhaak, soen, omarm, lieg, liefkoos; *akwa*, gehaak/ingehaak word; *akêla*, haak vir; *akelana*, mekaar liefkoos, vriendskaplik verkeer; *akelwa*, ingehaak word vir; *akiwa*, ingehaak word; *ake*, *ga*, *sa*, nie (in)haak nie; *akê*, mag/moet haak of inhaak; *moaki*, haker; *baaki*, hakers.
> **akalala**, *a ka la.la*, sweef, hang oor, oorhang; *akalalêla*, sweef vir/oor; *akalatša*, laat sweef, vlerke oopsprei om te sweef; *akaladitše*, het laat sweef; *se bone nong go -*, *go wa fase ke ga lona*, hoogmoed kom tot 'n val; *akalatšwa*, genoodsaak om te sweef; *akalalwa* gesweef word; *akalêla*, hang/sweef oor, wydsbeen staan oor; *akaletše*, het gesweef oor; *moakaladi*, persoon wat sweef.
> **akama**, *a ka.ma*, verwonder/verbaas wees; *akamela*, inlaat (bemoei) met; *akametša*, (laat) verbaas, verbasing wek, aangaap, toeroep; *akametšwa*, verbaas/aangegaap word, toegeroep word.
> **akere**, *'a kê.'rê*, akker.
> **aketša**, *a ke.tša*, leuen vertel, lieg, jok; *akeditše*, het (gelieg) 'n leuen vertel; *sa aketše*, nie lieg nie.
> **akga**, *a.kga*, werp, gooi, slinger, swaai, beweeg; *akgaakga*, heen en weer beweeg (soos branders), slinger, skommel; *akgaakgwa*, heen en weer geslinger word; - *diatla*, arms swaai, met leë hande loop; - *dinao*, voet in die wind slaan; *akgwa*, beweeg/geslinger word; - *akgêga*, skommel, swaai; - *akgêla*, slinger, swaai, werp; *akgêla*, slinger na/vir, tou om die horings gooi, met 'n vangtou vang, uitkrap, soos kole uit 'n vuur; *akgelwa*, geslinger word, gevang word met 'n tou; - *dikobo*, klere uitpluk.
>
> **tsirikana,** *'tsi'ri ka.na*, klink.
> **tsirima**, *'tsi'ri.ma*, klink, lui, uitspuit, vorentoe spring.
> **tsirimetša**, *'tsi'ri me.tša*, laat klink, vasbyt, laat lui, styf vasbind.
> **tsirinya**, *'tsi'ri.nya*, laat klink, lui.
> **tširoga**, *'tši ro.ga*, wakker skrik, senuweeagtig word, opskrik, moedeloos word.
> **tširogo** *'tši ro.gô*, impuls.
> **tširoša** *'tši ro.ša*, wek, skrikmaak.

It is clear from example 5 that the first alphabetical words like *aka – akga* have been exhaustively treated while words towards the end of the alphabet like *tsirikana – tširoša* have received less attention.

A multi-dimensional Setswana Ruler will now be introduced to study imbalances in Setswana dictionaries and to suggest a norm.

4.2.3 Building and applying a multi-dimensional Lexicographic Ruler

Prinsloo (2004) defines a Ruler as a practical instrument for measurement of the relative length of alphabetical stretches in alphabetically ordered dictionaries. Rulers are designed according to the generally accepted principle that alphabetical categories in any given language do not contain an equal number of words. Rulers are based upon the percentages of types per alphabetical category in corpora.

According to Prinsloo (2004:9) the real value of the Ruler lies in the fact that it focuses the attention of the compiler on potential ill-balanced areas, therefore the aim of the multi-dimensional Lexicographic Ruler for Setswana should be to eliminate the imbalances as reflected in Table 17 and Table 18 below.

Consider the Ruler for Setswana in Figure 23, based on the average of the percentage breakdown of types in a Setswana corpus.
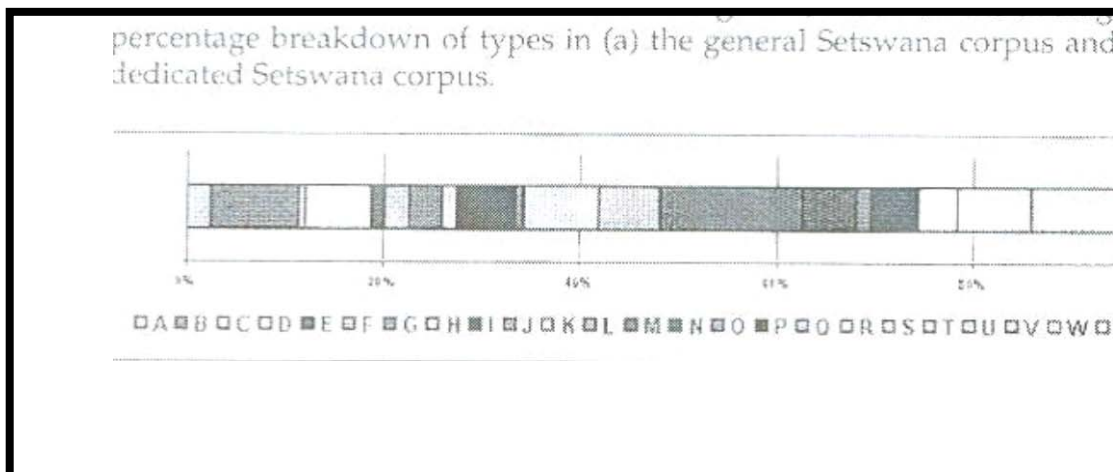
(Prinsloo, 2004:8)

Figure 24: A Ruler for Setswana

This ruler can also be expressed in terms of a percentage breakdown, i.e. divided into 100 blocks as a so-called block system.

Table 17: A block system for Setswana

1.	ALAF		21.	FELE		41.	KOUS		61.	MOTL		81.	SELE
2.	AROG		22.	FOLO		42.	LAEL		62.	MPHE		82.	SERA
3.	BADI		23.	GAGW		43.	LEBO		63.	NATE		83.	SETO
4.	BANN		24.	GATS		44.	LEKI		64.	NGWA		84.	SIMO
5.	BATW		25.	GOLO		45.	LERI		65.	NKUK		85.	SUAS
6.	BIRO		26.	GWET		46.	LETS		66.	NTEM		86.	TALE
7.	BOGA		27.	HUBE		47.	LOKO		67.	NTSH		87.	THAA
8.	BOLA		28.	IJES		48.	MAAD		68.	NYOR		88.	THIB
9.	BONK		29.	IKGO		49.	MAHA		69.	OOMA		89.	THWE
10.	BORU		30.	INOL		50.	MALE		70.	PANT		90.	TLAM
11.	BOUT		31.	IPUS		51.	MARA		71.	PHAK		91.	TLHA
12.	DAAM		32.	ITIS		52.	MATL		72.	PHIM		92.	TLHO
13.	DIFA		33.	ITSH		53.	MEFA		73.	PITL		93.	TLWA
14.	DIKG		34.	JOKO		54.	MESU		74.	PUDU		94.	TSAP
15.	DINK		35.	KANY		55.	MMAL		75.	RAMO		95.	TSHE
16.	DIRA		36.	KERO		56.	MMOL		76.	RENG		96.	TSHW
17.	DITH		37.	KGAR		57.	MOFI		77.	ROKG		97.	TSUN
18.	DITU		38.	KGOM		58.	MOKG		78.	RURU		98.	UBAU
19.	EGEP		39.	KHAN		59.	MONG		79.	SEBA		99.	WABO
20.	ETLH		40.	KODU		60.	MORW		80.	SEHI		100.	ZIMB

According to Prinsloo (2004:8), the Block System opens the door to a number of very practical applications. It gives a clear guidance in terms of page allocation, average length of articles, progress in terms of time and even remuneration intervals for part-time compilers.

From the above given statistics, it means that for a dictionary like THAN which contains 330 pages. 3,3 pages should roughly correlate with each block/percentage block. Consider now a comparison between MSED and the Ruler in terms of lemma percentage and page percentage.

Table 18: MSED versus the Ruler in terms of page allocation and the number of lemmas

	MSED:	MSED:	**Setswana**	MSED lemmas	**MSED Page %**
	Lemmas %	Pages %	**Ruler**	vs. the Ruler	vs. Ruler
A	1.2	1.3	2.6	-1.4	-1.3
B	4.7	4.6	9	-4.3	-4.4
C	0	0	0.6	-0.6	-0.6
D	6	6.4	6.6	-0.6	-0.2
E	1.2	1.3	1.4	-0.2	-0.1
F	3.7	3.3	2.4	1.3	0.9
G	5.2	5.3	3.4	1.8	1.9
H	0.9	0.9	1.5	-0.6	-0.6
I	5.3	4.9	5.9	-0.6	-1
J	0.7	0.7	0.8	-0.1	-0.1
K	12.2	11.9	7.7	4.5	4.2
L	6.7	6.8	6.1	0.6	0.7
M	12.5	13.7	14.6	-2.1	-0.9
N	4	4	5.5	-1.5	-1.5
O	1.3	1.3	1.6	-0.3	-0.3
P	5.9	6	4.6	1.3	1.4
Q	0	0.2	0.2	-0.2	0
R	3.9	3.5	3.9	0	-0.4
S	8.5	8.6	7.5	1	1.1
T	15.4	14.1	12.2	3.2	1.9
U	0.5	0.4	0.6	-0.1	-0.2
V	0	0	0.3	-0.3	-0.3
W	0.1	0.2	0.4	-0.3	-0.2
X	0	0.2	0.1	-0.1	0.1
Y	0.1	0	0.3	-0.2	-0.3
Z	0	0	0.2	-0.2	-0.2
		99.8	100		-0.2

The two Setswana dictionaries, i.e. THAN and THAND will now be compared in terms of pages utilized per alphabetical stretch and the number of the lemmas respectively.

Table 19: Comparison of the alphabetical stretch in THAN and THAND

(THAN) Tsonope (330 pages)			(THAND) Kgasa and Tsonope (126 pages)		
Category	Number of pages	Percentage	Category	Number of pages	Percentage
A	8	2,4	A	2	1,6
B	25	7,58	B	13	10,32
C	1	0,30	C	-	-
D	4	1,21	D	1	0,79
E	2	0,61	E	1	0,79
F	7	2,12	F	1	0,79
G	10	3,0	G	5	3,97
H	1	0,30	H	1	0,79
I	14	4,24	I	1	0,79
J	1	0,30	J	1	0,79
K	29	8,79	K	10	7,94
L	31	9,39	L	10	7,94
M	43	13,03	M	19	19,09
N	13	3,94	N	4	4,8
O	3	0,91	O	1	0,79
P	19	5,8	P	7	5,56
Q	1	00,30	Q	-	-
R	9	2,73	R	4	3,8
S	31	9,38	S	14	11,11
T	56	16,97	T	19	15,07
U	1	0,30	U	1	0,79
V	-		V	1	0,79
W	1	0,30	W	1	0,79
X	1	0,30	X	-	0,79
Y	1	0,30	Y	1	0,79
Z	-		Z	-	-

Table 19 indicates that most Setswana dictionaries have an alphabetical stretch of many pages for the lemmas B, K, L, M, P, S, and T. For a dictionary like the THAN, the stretches M and T fill a high number of pages 13, 03% and 16, 97% respectively. The same situation prevails for THAND with M and T at 19, 09% and 15, 07% respectively.

The difference in size for the alphabetical stretch M between these two dictionaries namely THAN and THAND could have been caused by the lexicographer's addition of new words and the ignorance of frequency counts, thus not taking a holistic approach. The alphabetical stretches for B, K, L, M, S and T in THAN occupy 65, 15% of the total number of 330 pages, and 67, 4% of the 126 number of pages for THAND.

Table 19 will now be graphically represented as indicated in Figures 25 and 26. The vertical axis indicates number of pages, the horizontal axis indicates category of the alphabetical stretches.

4.2.4 Graphical representation of the macrostructure of the two monolingual Setswana dictionaries
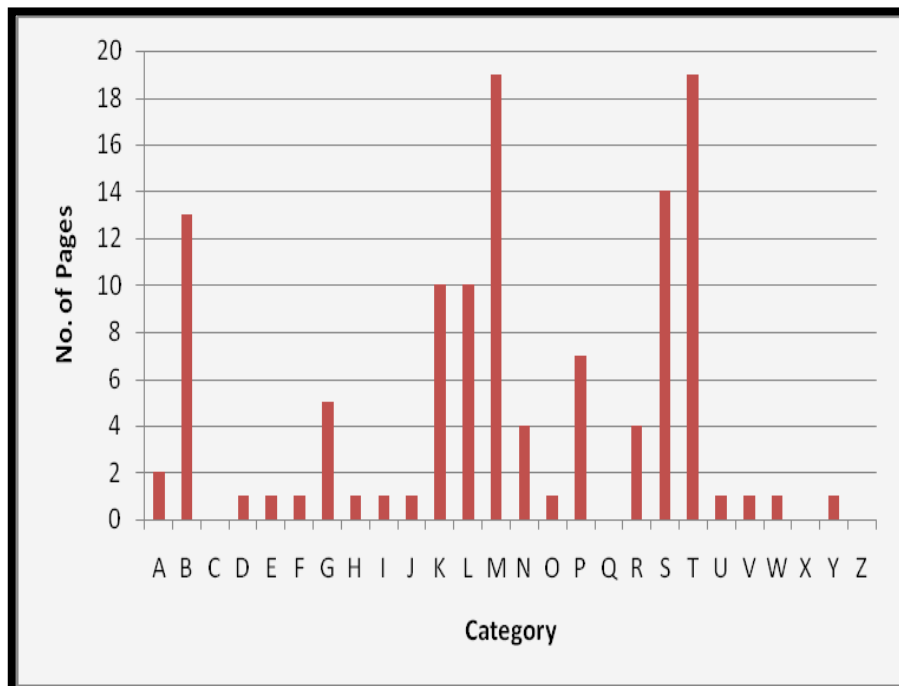
Figure 25: Graphical representation of THAND

It is clear from figure 25 that the alphabetical stretches for B, M, S, and T are relatively big and in particular contains large numbers of lemmas. The alphabetical stretches B, M and S contain the plural class prefix *ba-*, *me-* and *ma-* and the singular class prefix *se-* while the alphabetical stretches for T contains the majority of the Setswana verbs and few nouns.
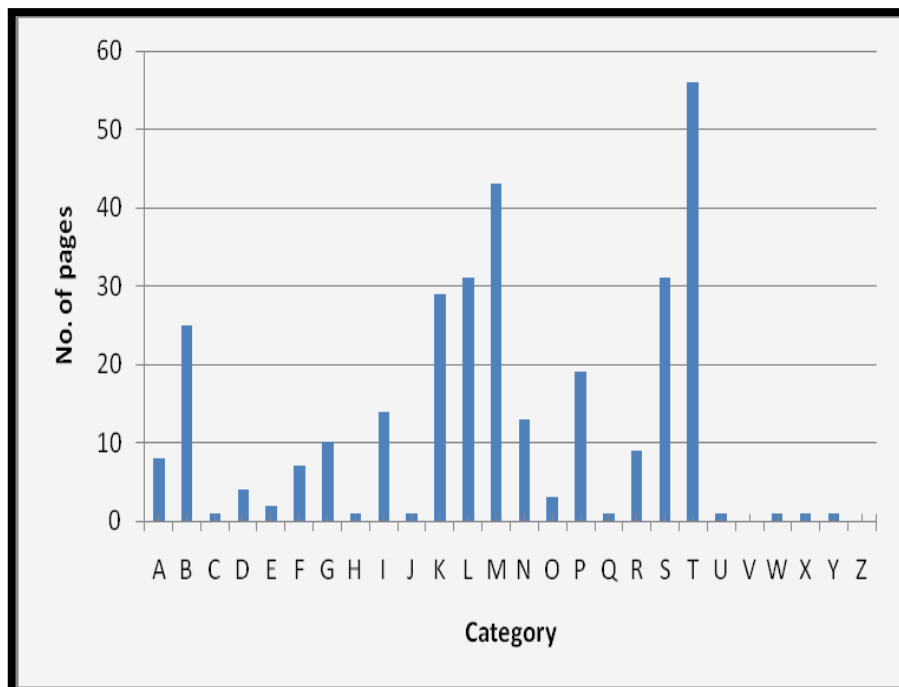
Figure 26: Graphical representation of THAN

Figure 26 contains the alphabetical stretches for M and T which are relatively big. The alphabetical stretch M and T contain the singular class prefixes *mo-*, and the plural class prefixes *ma-* and *me-* while the alphabetical stretches for T contains the majority of the Setswana verbs and nouns.

It is clear from both dictionaries that both the pages and the number for the lemma signs B, K, L, M, S and T occupy a huge number of pages and deserves exhausted

treatment and E, H, J, Q, W, X and Y occupy a smaller number of pages while V and Z in THAN and C, Q and Z in THAND are empty, because the Setswana language does not contain words which start with V and Z.

4.3   Lemmatisation approaches, strategies and traditions

According to Prinsloo and Gouws (2005:85), it is important for the   lexicographer when dealing with lemmatization in African languages to negotiate a complex interplay and overlap between (a) lemmatization approaches, (b) lemmatization strategies, (c) lexicographic traditions, (d) nominal and verbal structures and (e) conjunctiveness versus disjunctiveness. Compare the most relevant relations categorically in terms of columns A-E and   rows 1-5.

Table 20: Lemmatization approaches, strategies, traditions, etc.

	A	**B**	**C**	**D**	**E**
	Lemmatization approaches	**Lemmatization strategies**	**Lexicographic traditions**	**Nominal and verbal structures**	**Conjuctiveness versus disjuctiveness**
1	Traditional	Stem	Word	Verbal prefixes	Conjunctive orthography
2	Paradigms	Singular and plural	stem	Verbal suffixes	Disjunctive orthography
3	Rule orientated	Singular only			
4	frequency	Left-expanded			
5		First or third letter			

In terms of Table 20 a complex set of 1-1 relations as given in Table 21 exists and has to be negotiated in any discussion of the lemmatization of nouns and verbs in African languages.

Table 21: Complex set of 1-1 relations in the Lemmatization's of nouns and verbs in African languages

A1:B1	A1:C1	A1:D1	A1:E1	A1:B2	A1:C2	A1:D2	A1:E2	A1:B3	A1:D3	A2:B1
;	;	;	;	;	;	;	;	;	;	;
A2:C1	A2:D1	A2:E1	A2:B2	A2:C2	A2:D2	A2:E2	A2:B3	A2:D3	A3:B1	A3:C1
;	;	;	;	;	;	;	;	;	;	;
A3:D1	A3:E1	A3:B2	A3:C2	A3:D2	A3:E2	A3:B3	A3:D3	A4:B1	A4:C1	A4:D1
;	;	;	;	;	;	;	;	;	;	;
A4:E1	A4:B2	A4:C2	A4:D2	A4:E2	A4:B3	A4:D3	B1:C1	B1:D1	B1:E1	B1:C2
;	;	;	;	;	;	;	;	;	;	;
B1:D2	B1:E2	B1:D3	B2:C1	B2:D1	B2:E1	B2:C2	B2:D2	B2:E2	B2:D3	B3:C1
;	;	;	;	;	;	;	;	;	;	;
B3:D1	B3:E1	B3:C2	B3:D2	B3:E2	B3:D3	B4:C1	B4:D1	B4:E1	B4:C2	B4:D2
;	;	;	;	;	;	;	;	;	;	;
B4:E2	B4:D3	B5:C1	B5:D1	B5:E1	B5:C2	B5:D2	B5:E2	B5:D3	C1:D1	C1:E1
;	;	;	;	;	;	;	;	;	;	;
C1:D2	C1:E2	C1:D3	C2:D1	C2:E1	C2:D2	C2:E2	C2:D3	D1:E1	D1:E2	D2:E1
;	;	;	;	;	;	;	;	;	;	;
D2:E2	D3:E1	D3:E2								
;	;	;								

A brief overview of verbs will now be given to serve as a basis of the interpretation of the different lemmatisation strategies, approaches and traditions, as given in Table 21.

In the case of verbs numerous derivations of a single verb stem exist, consisting of the root plus one or more prefix(es) and or suffix(es) as is clearly indicated in Table 22 for the verb stem *reka* 'buy' which is structurally analysed in terms of 18 numbers.

Table 22: Derivative of reka (buy)

01	Root + standard modifications	VR	*reka*
		VRPre	*rekile*
		VRPas	*rekwa*
		VRPerPas	*rekilwe*
02 ANA	Root + reciprocal + standard modifications	VRRec	*rekana*
		VRRecPer	*rekane*
		VRRecPas	*rekangwa*
		VRRecPas	*rekangwe*
03 ANTSHA	Root + reciprocal + causative + standard modifications	VRRecCau	*rekantsha*
		VRRecCauPer	*rekantshitse*
		VRRecCauPas	*rekantshwa*
		VRRecCauPerPas	*rekantshitswe*
04 ANYA	root + alt. causative + standard modifications	VRAlt-Cau	*rekanya*
		VRAlt-CauPer	*rekantsha*
		VRAlt-CauPas	*rekangwa*
		VRAlt-CauPerPas	*rekangwe*
05 EGA	Root + neutron passive + standard modifications	VRNeu-Pas	*rekega*
		VRNeu-PasPer	*rekegile*
		VRPas	
		VRPerPas	
06 ELA	Root + applicative + standard modifications	VRApp	*rekela*
		VRAppPer	

		VRAppPas	
		VRAppPerPas	*reketswe*
07 ELANA	Root + applicative + reciprocal + standard modifications	VRAppRec	*rekelana*
		VRAppRecPer	*rekelane*
		VRAppRecPas	*rekelangwa*
		VRAppRecPerPas	*rekelangwe*
08 ISA	Root + causative + standard modifications	VRCau	*rekisa*
		VRCauPer	*rekisitse*
		VRCauPas	*rekisiwa*
		VRCauPerPas	*rekisitswe*
09 ISANA	Root + causative + reciprocal +standard modifications	VRCauRec	*rekisana*
		VRCauRecPer	*rekisane*
		VRCauRecPas	*rekisangwa*
		VRCauRecPerPas	*rekisangwe*
10 ISEGA	Root + causative + neutron passive + standard modifications	VRCauNpas	*rekisega*
		VRCauNpasPer	*rekisegile*
11 ISETSA	Root + causative + applicative + standard modifications	VRCappu	*rekisetsa*
		VRCauAppPer	*rekiseditse*
		VRCauAppPas	*rekisetswa*
		VRCauAppPerPas	*rekiseditswe*
12 ISETSANA	Root + causative + applicative + reciprocal + standard modifications	VRCauAppRec	*rekisetsana*
		VRCauAppRecPer	*rekisetsane*
		VRCauAppRecPas	*rekisetsanwa*
		VRCauAppRecPerPas	*rekisetsanwe*

13 OLOLA	Root + reversive transitive + reciprocal + standard modification	VRRevt	rekolola
		VRRevtPer	rekolotse
		VRRevtPas	rekololwa
		VRRevtPerPas	rekolotswe
14 OLOLANA	Root + reversive transitive + reciprocal + standard modifications	VRRevtApp	rekololana
		VRRevtRecPer	rekololane
		VRRevtRecPas	rekololanwa
		VRRevtRecPerPas	rekololanwe
15 OLOLELA	Root + reversive transitive + applicative + standard modifications	VRRevtApp	rekololela
		VRRevtAppPer	rekololetse
		VRRevtAppPas	rekololelwa
		VRRetAppPerPas	rekololetswe
16 OLOLELANA	Root + reversive transitive + applicative + reciprocal + standard modifications	VRRevtApp	rekololelana
		VRRevtAppPer	rekololelane
		VRRevtAppPas	rekololelanwa
		VRRevtAppPerPas	rekololelanwe
17 OLODISA	Root + reversive transitive + causative + standard modifications	VRRevtCau	rekolodisa
		VRRevtCauPer	rekolodisitse
		VRRevtCauPerPas	rekolodiswa
18 OLODISANA	Root + reversive transitive + causative + reciprocal + standard modifications	VRRevtCauRec	rekolodisana
		VRRevtCauRecPer	rekolodisane
		VRRevtCauRecPas	rekolodisangwa

		VRRevtCauRecPas	*rekolodisangwe*

From Table 22, it is clear that the lexicographer has to consider a huge number of derivations per verb.

4.3.1 Lemmatization approaches

4.3.1.1 Traditional approach

This approach highlights a scenario whereby dictionary compilers seem to be unaware of the need to reduce the number of entries for a specific verb. Prinsloo (2004) defines the traditional approach as the worst situation where dictionary compilers fail to employ relevant or appropriate selection strategies and are even unaware of the problem of what to include in and what to omit from the dictionary. In case of nouns and verbs the compilers would e.g. conveniently ignore the need to reduce the number of derivations, which resulted in the compiler randomly adding words to the dictionary until the publication deadline.

In the preface to the SEAD, the compilers honestly admit:

> "The dictionary team is aware of the fact that common and even
> essential words may easily be omitted during the compilation of a
> dictionary. This can take place simply because the lexicographer had
> not encountered such words."

The traditional approach emphasizes the problem of essential words being accidentally excluded and a lot of rare words which are unlikely to be looked up by the target users are included. A typical example of this kind of an approach was discussed in Example 3 and Example 4 of this chapter where the imbalances arise from the traditional approach of the different derivations of verbs as highlighted. The issue whether frequency of use should be a determining factor for the inclusion versus

omission of lemmas in dictionaries is often debated. It is relatively easy to define frequency as a guideline for inclusion/omission for bilingual dictionaries in the South African language lexicography context. Publishers normally limit the compiler to 5,000 lemmas in each side of the dictionary where target users include learners of the language. Given these requirements most of the 5,000 lemmas have to be selected on frequency of use. The situation becomes more problematic in monolingual dictionaries. If the monolingual dictionary is intended for mother-tongue speakers one could perhaps argue that they will not be looking for common frequently used words. Firstly, it is accepted that a mother-tongue speaker might not be inclined to look up the words such as *table* and *chair* for their meaning but for other purposes e.g. idiomatic use. Secondly monolingual dictionaries for Setswana include non-mother-tongue learners as target users who are likely to mainly look up frequently used words. Compilers of monolingual dictionaries are encouraged to compile rather comprehensive dictionaries, even as the first attempts for the language to include frequently used words but also a substantial number of infrequent words to solve the problem.

4.3.1.2 Completing paradigms

Prinsloo (1994:97) also calls this an 'enter-them-all' approach. In the THAN attempts were made to enter all nominal and verbal derivations to such an extent that mother-tongue speakers doubt whether many of these derivations are actually and actively used. Compare a section of the article of *aga* (build) in THAN in this regard:

Example 6: THAN

adisa

adisa TTT *tpt.* -itse. 1. >ala+isa 2. go thusa go ala
aêga GGG ¦sekêga, seêga *tpt.* -ile. baya sengwe se itshegeditse ka se sengwe
aêgolola GGGTT *tpt.* -itse. >aêga+olola
aêgwa GGG *tpt.* -ile. >aêga+iwa
aena GGG *tpt.* -ile. tlosa mangwana a seaparo ka sengwe se se bolelô
aene GTG *ln./9.* di-. tshipi e e gatisang diaparô go tlosa matsutsuba
-afala TTT ¦-agala *tiregi.* mogatlana o o supang phetogô ya seêmô kgotsa boleng jwa sengwe: *Kgwafo* fala .
afe GT *td.* ka ga tlhaolô ya maina a setlhôpha sa boraturo: *O raya mariga* afe *?*
Aforika GTGG ¦Afrika, *ln./1a.* bó-. nngwe ya dikarolo tse tlhano tsa kgolokwe ya lefatshe tse di bipilweng ke mmu
Aforika Borwa GTGGTG *ln./1a.* bó-. lengwe la mafatshe a Aforika
aga GT *ltls.* go dira sengwe kgapetsakgapetsa; tlhôla: *Mmê o aga a nkômanya.*
aga GT *tpt.* -ile. 1. dira bonnô kana sedirisiwa 2. dira lelapa ga monna le mosadi 3. tswa monni wa lefelô lengwe
-agala TTT ¦-afala *tiregi.* mogatlana o o supang phetogô ya seêmô kgotsa boleng jwa sengwe: *Bon* agala . BONA -ala; -êga; -afala; -êsêga
agang GTG *ltls.* -ile. >aga+ng: *Ke wêna yo o* agang *o re tshwenya.*
agang TTG *tpt.* -ile. >aga+ng
agêga GGG *tpt.* -ile. >aga+êga
agêgile GGGT *tpt.* >aga+êga+ile
agêla GGG *tpt.* -itse. >aga+êla
agêlêla GGGT *tpt.* -itse. aga ka go dikologa kgotsa ka go thekeletsa
agêletse GGGT *tpt.* >agêlêla+itse
agêletswe GGGT *tpt.* >agêlêla+iwa+itse
ageletsweng GGGGG *tpt.* >agêlêla+iwa+itse+ng
agêlwa GGG *tpt.* -itse. >aga+êla+iwa
agetse GGG *tpt.* >aga+itse
agetswe GGG *tpt.* >aga+iwa+itse
agetsweng GGGG *tpt.* >aga+iwa+itse+ng

agile GGG *ltls.* >aga+ile; dirile sengwe ka go se boeletsa
agile GGG *tpt.* >aga+ile
agileng GGGG *tpt.* >aga+ile+ng
agilwe GGG *tpt.* >aga+iwa+ile
agilweng GGGG *tpt.* >aga+iwa+ile+ng
agisa GGG *tpt.* -itse. 1.>aga+isa 2. thusa sengwe go aga
agisana GGGT *tpt.* -ile. 1.>aga+isa+ana; thusana mo tirong ya kagô 2. tshela mmôgô ka kutlwanô
agisane GGGT *tpt.* >agisana+ile
agisantse GGGGT *tpt.* >aga+isa+ana+itse
agisanya GGGT *tpt.* -itse. >aga+isa+ana+anya; jaaka go dira gore batho ba tshedisanê ka kagisô
agisanyang GGGTG *tpt.* -itse. >aga+isa+ana+anya+ng
agisitse GGGT *tpt.* >aga+isa+itse
agiwa GGG *tpt.* -ile. >aga+iwa
agologa GGGT *tpt.* -ile. >aga+ologa
agolola GGGT *tpt.* -itse. >aga+olola
agololwa GGGT *tpt.* -itse. >aga+olola+iwa
agwa GT *tpt.* -ile. >aga+iwa
agwang GGG *tpt.* -ile. >aga+iwa+ng
ahaa! GGG *lltl.* lelatlhêlwa le le supang tumêlô kgotsa kgakologêlô
ahee! GGG *lltl.* lelatlhêlwa le le supang tumêlô
aila GGG *tpd.* -itse tsamaya go sena maikaêlêlô kgotsa maikemisetso
aitsane TGTG ¦aitse *lkpn.* lekôpanyi le le ngôkang theetso ya mmuisiwa ka go tlhagisa sengwe se sesa mo puisanong; le tswa mo tsetleng- o a itse
aitse TGT ¦waitse *lkpn.* BONA aitsane
aiyo! GGG *lltl.* lelatlhêlwa le le kayang maikutlô a motho fa sengwe se diragala se sa solofêlwa
ajwa TT *tpt.* >aba+iwa
ajwang TTG *tpt.* >aba+iwa+ng
-aka TT *tiraka.* mogatlana o o supang go êtêgêla kgotsa go fetêlêla ga se se dirwang: *Rog* aka .
aka TT *tpt.* -ile. go bua se eseng nnete
akabaditse TTTTT *tpt.*

2

It is clear from the above given example that users have to struggle through numerous columns of fine print to find the meaning of words such as *ageletsweng*, *agetsweng*, *agisanyang* etc. as illustrated above. The problem is clearly illustrated by Gouws and Prinsloo (2005:73). Consider the following example in terms of accessibility and unambiguous retrieval of the information from the perspective of an inexperienced learner of Sesotho sa Leboa. The user wants to look up the word *dikagollišano*. (S)he firstly has to strip the suffixes in order to find the verb stem and then to 'add' the

semantic connotations in a cumulative way in order to find the meaning – thus up to 12 steps in total as given in Table 23 below.

Table 23: Accessibility and information retrieval process for *dikagollišano* in NSDN

1.	*dikagollišano*	↓	plural deverbative consisting of root + reversive transitive + causative + reciprocal + ending
2.	*kagollišano*	↓	singular deverbative consisting of root + reversive transitive + causative + reciprocal + ending
3.	*agollišana*	↓	verb root + reversive transitive + causative + reciprocal + ending
4.	*agolliša*	↓	verb root + reversive transitive + causative + ending
5.	agolla	↓	verb root + reversive transitive + ending
6.	**aga**	↓	**verb** (stem)
7.	**build**	↓	**meaning of the verb**
8.	break down	↓	reverse or opposite meaning 'un-build'
9.	cause to break down	↓	add causative sense of 'let/force'
10.	cause each other to break down	↓	add reciprocal sense of 'each other'
11.	the process of causing each other to break down	↓	change 'the process of …' to the plural
12.	**the process of causing each other to break down**		

(Gouws and Prinsloo 2005:40)

4.3.1.3 Rule-orientated approach

According to Prinsloo (1994:98), this approach presents a scenario where dictionary compilers still in principle, strive to complete paradigms and still attempt to make Provision for 'all' verbs and nouns and their derivatives. The compiler does not 'enter them all', but makes use of certain rules or guidelines which should be followed, if a word cannot be directly looked up in the dictionary. The target user is expected to interpret or reverse regularly derived derivations by means of a set of rules given in the non-alphabetical section of the dictionary as well as within the dictionary itself. In this regard the emphasis is on limiting the number of lemmas or sub lemmas in a dictionary. For example, (a) lemmatizing only the singular forms of nouns and only

the basic verbal stems and (b) giving sets of rules in the user's guide for the user to strip suffixes and add meaning components. In principle, it still reflects the urge to 'enter-all' although it is quite economical in terms of dictionary space and it is also user-unfriendly.The strategy of lemmatizing singular forms of nouns as described in terms of THAND in 4.3.2 does however imply plural-to-singular guidance rules. A typical example is Pukuntšu (1989).

Example 7

PUKU 2 (Kriel and Van Wyk 1989: Preface)

Perfecta

-dile: *-la,*   e.g.   *badile*   under   *bala*

-ditše: *-tša,*   e.g.   *biditše*   under   *bitša*

-etše: *-ela*,   e.g.   *rapetše*   under   *rapela*

          -ala,   e.g.   *robeštše*   under   *robala*

-tše: *-as*,   e.g.   *bešite*   under   *beša*

          -tšha,   e.g.   *bontšhitše*   under   *bontšha*

          -sa,   e.g.   *lesitše*   under   *lesa*

          -tswa,   e.g.   *hlatswitše*   under   *hlatšwa*   e.t.c

Applicative:

-etša: *-ša,*   e.g.   *bešetsa*   under   *tloša*

          -tšha,   e.g.   *tlhakantšhetša* under   *ntšhetša*

          -ša,   e.g.   *lesetša*   under   *leša*

          -tšwa,   e.g.   *hlatswetša*   under   *hlatšwa*

          -nya,   e.g.   *bofanya*   under   *bofa*

-letša: *-tša,*   e.g.   *biletša*   under   *bitša*

According to Prinsloo (2004:94), this approach on the other hand runs into serious difficulty with regard to practicality and user friendliness. Busane (1990:28) states its underlying disadvantages as follows:

"---- many introductory pages are usually allocated to grammatical sketches of the language concerned without the knowledge of which is deemed hazardous to use the dictionary successfully. We believe, however, that these sections and introductory explanations are not sufficient provisions for a user friendly product. Dictionary users are known to allocate little time to the study of these prefatory matters."

In this regard the compiler of a dictionary based on frequency of use can easily capitalise on the virtues of the rule-orientated approach.

4.3.1.4   Frequency-based approach

One could summarize the discussion thus far by stating that the corpus era for African languages introduced by Prinsloo (1991), opened new doors for the lemmatization of nouns and verbs namely lemmatization based on the frequency of use. According to Prinsloo (2004:93), using corpus data, the lexicographer can ensure that frequently used words are not accidentally omitted and, on the other hand, that precious dictionary space is not taken up by articles of lemmas which are unlikely to be looked up by the target users. Following this approach, the lexicographer can sensibly and drastically reduce the number of lemmas for a specific verb such as *rema* (chop) on frequency of use. The best point of departure is a frequency list of actual occurrences of the verb taken from a Setswana corpus.

Example 8

rema (34), remegang (5), reme (2), remege (1), remiremi (1), remile (35), remileng (77), remisa (1), remisang (100), remise (3), remiseditseng (2), remisetsa (1), remisetswe (1), remiseditsweng (4), remiso (1), remiswa (3), remiswang (4), remiswe (5).

The lexicographer can now lemmatize and reduce this list on the basis of the frequency of occurrence in the corpus. The lexicographer can for example decide not to lemmatize derivations containing the relative suffixes *-ng* and not to include infrequent derivations thus reducing the list to *rema, remile, remisa, remisitse, remetswe, remiswa*.

4.3.2 Lemmatization strategies

4.3.2.1 Lemmatization of nouns

According to Prinsloo (2004:95) lemmatization approaches are illustrated in a number of specific strategies such as lemmatizing: (a) both singular and plural, (b) only singular forms (c) noun stems, (d) on first or third letter and (e) using left expanded article structures.

Table 24: Noun classes and examples of Setswana

Class	Prefix	Example	Translation
1	*mo-*	*monna*	man
2	*ba-*	*banna*	men
1a	*Ø*	*rrangwane*	uncle
2a	*bo+*	*borrangwane*	uncles
3	*mo-*	*monwana*	finger
4	*me-*	*menwana*	fingers
5	*le-*	*lesogana*	young man
6	*ma-*	*masogana*	young men
7	*se-*	*selepe*	axe
8	*di-*	*dilepe*	axes
9	*N-/Ø*	*nku*	sheep
10	*di+*	*dinku*	sheep

11	*lo-*	*lobota*	wall
12	*di-*	*dipota*	walls
13	*bo-*	*bogobe*	porridge
14	*ma-*	*magobe*	different kinds of porridge
15	*go*	*go bona*	to see
16	*fa-*	*fase*	below
17	*go-*	*godimo*	above
18	*mo-*	*morago*	behind

(a) Lemmatizing both singular and plural noun forms

Prinsloo (2004:95) highlights that lemmatizing both singular and plural nouns is an extremely user-friendly strategy and very popular among inexperienced users and learners of a language. The user does not necessarily require previous knowledge. Unfortunately, the redundancy factor in terms of dictionary space is almost 80% high and has to be weighed up against the advantages in terms of the approach's user-friendliness and practicality. The compiler for MSED opted for lemmatising both singular and plural forms of nouns as suggested by Prinsloo and De Schryver (1999) and Gouws and Prinsloo (2005a:84-85). Compare example 9 in this regard.

Example 9: MSED

a	b
morara N. CL.3 *mo-,* SING. OF *merara,* ivy. **morutwa** N. CL.1 *mo-,* SING. OF *barutwa,* DER. F *rutwa,* same as *murutiwa* and *morutwi,* a Learner; disciple.	**merara** N. CL.4 *me-,* PL. OF *morara,* creeping plants; climbers, including Gymnema sylvestre and Secamone parvifolio; when used as a REl. s, it signifies branching in all directions. **barutwa** N. CL.2 *BA-,* PL. OF *morutwa,* students;

Note that in (9a and b) it is suggested that the treatment be given for the most frequent member of the singular form/plural pair. In the case of *morutwa* and *merara* full

treatment is given while only basic information in *morara* and *barutwa* is given. This approach is in line with the more radical approach suggested by Gouws and Prinsloo (2005a), giving the treatment at the more frequently used member of the pair. For example, for *morara* and *morutwa* versus *merara* and *barutwa* in example 9, treatment is given to the singular form *morutwa* and the plural form *merara*. Consider example 10 where compilers of the SED opted to lemmatise both singular forms and plural forms giving equal treatment to both singular forms and plural forms.

Example 10: SED

a	b
Morutwa, N. CL.1 mo-, SING OF barutwa, DER. F. rutwa, same as morutiwa and morutwi, a learner; disciple. **Morara, n.,** A species of tree- climbing plant ; a wild vine.	**Barutwa,** N. CL.2 BA- PL. OF morutwa, students; scholars. **Merara,** n., pl. of morara, creeping plants; adj., Branching out in aal directions. *Ditshika tse di merara*. These veins spread or branch in all directions.

It is important to note in (10) that treatment is given to both members of the singular form/plural pair. Redundancy is increased to a great extend, although very user-friendly.

(b) Lemmatizing only singular noun forms

In terms of Prinsloo (2004:96), lemmatizing only singular noun forms, is a sound lexicographic strategy. It is not expecting too much from e.g. an advanced learner to know the regular productive rules of the language governing the formation of singular and plural forms, as illustrated in Table 25 which is an adaptation of the Sepedi rules for Setswana.

Table 25: Rules for looking up plural forms in Setswana

Rule		Examples	
word starts with	look word up under	word start with	look word up under
ba-	*mo-*	*basadi*	*mosadi*
	ngw-	*bana*	*ngwana*
bab-	*mm-*	*babetli*	*mmetli*
bo-	*(the stem)*	*bomalome*	*malome*
di-	*se-*	*dilepe*	*selepe*
	n-	*dinku*	*nku*
	(the stem)	*dikgomo*	*kgomo*
	lo-	*dinthe*	*lore*
ma-	*le-*	*maleme*	*leleme*
	bo-	*magobe*	*bogobe*
maj	*ma-*	*majang*	*bojang*
me-	*mo-*	*megoma*	*mogoma*
	me	*metsi*	*metsi*
meb-	*mm-*	*mebutla*	*mmutla*
mengw-	*ngw-*	*mengwaga*	*ngwaga*

A typical example of this kind of approach is found in THAN where the lemmatization strategy of singular nouns is followed. However inconsistencies regarding a number of words on the typical plural forms of the nouns *ba-* , *di-* and *me-* are noticed.

Although the editorial policy of dictionaries such as THAN is to lemmatize the singular forms only, plurals such as *badimo* (gods), *barwa* (sons), *baša* (youths) *and Bararo* (the three) are lemmatized. It is unfortunate that it is not always easy for users to look up plural forms under their singular forms because in most cases, from a user's point of view, there is no straightforward one to one correlation between some of the singular/plural class pairs first, *di–* (class 8 and class 10) has a one to two correlation. For example, the user who does not know the meaning of the word *dinku* (sheep) has to look it up under *senku or nku*). Secondly, a one to three correlation also exists in the case of *ma-* (class 6) and a one to four correlation in the case of *me–* (class 4). For example if the user is confronted with the word *metsi* (water). The rule in Table 25

states that *me-* should be looked up under *mo-* in class 4 and according to Table 25 it should be looked up under **motsi*, which is ungrammatical and the user has been misled in this regard*.* There are also irregular nouns that change their roots when taking the plural forms, e.g. singular *lore* (wood), plural *dinthe* (woods) and singular *lorako* (wall), plural *dithako* (walls*) in class 10.* Again the user who is not familiar with the language may get lost by simply adding the prefix (di-) * to the root *-rako* * *dirako* instead of (*dithako*) and *dinthe*.

(c) Lemmatizing nouns on the first or the third letter

Lemmatizing nouns on the first or the third letter is a method used by Snyman et al (1990) in their SEAD. Consider the following example:

Example 11: SEAD

First letter	Third letter
kwáno (here)	*kwalo, le- ma-* (handwriting, orthography)
kwatla (a cut of meat from the back of a pig)	*kwapa, bo- le- ma-* (flat scale)
kwena (become strong and sturdy)	*kwé (kwê), le-* (Vaal river)
Laboraro (Wednesday)	*lapa, le- ma-* (traditional courtyard, home)

This approach has certain advantages for the inexperienced learner of Setswana, but can be frustrating to the user, because there are always two options to choose from when looking for the nouns. Redundancy is avoided by not having separate articles for lemmas such as *lekwalo* (letter) and *makwalo* (letters), etc.

(d) Lemmatizing noun stems

Lemmatizing nouns on their stems is a choice between the two conflicting lexicographic traditions namely, word versus stem followed in the lemmatization of African languages. This kind of an approach is not found in Setswana dictionaries. Consider the following examples of stem lemmatization taken from the CNSD and the SZD:

Example 12: CNSD

xample 13: SZD

gesticulation.

-bhekwa (v) be watched.
   ukubhekwa yisambane — to meet with misfortune.

-bhele (i- ama-) (n) bear; bale (e.g. grass fodder).

-bhelebhele (i- ama-) (n) edible fruit of 'ingotsha' shrub.

-bhelenja (um- imi-) (n) loin skin worn by girls.

-bhelihodi (i- ama-) (n) saddle belt which goes under stomach of horse.

-bhelu (isi- izi-) (n) tail of rabbit; commotion; riot.

-bhelu (u-) (n) Afrikander beast.

-bhema (v) smoke; take snuff.
   ukubhema ukholwe — to have enough of situation.

-bhembuluka (v) flee.

-bhememe (u-) (n) famine; commotion.

-bhemisa (v) give snuff; tobacco.

-bhembetheka (v) drink thirstily; pierce.

-bhena (v) curve in (of the back).

-bhendi (i- ama-) (n) band (instrumental).

bhengce (ideo) of evasiveness.

-bhengceza (v) be evasive; prevaricate.

-bhengele (i- ama-) (n) armlet; bangle.

-bhengeza (v) spread a story.

bhengu (ideo) of whirling (e.g. wind); of commotion.

-bhengu (isi- izi-) (n) whirlwind; commotion.

-bhengubhengu (isi- izi-) (n) strong wind; hurricane; commotion.

-bhengubhengu (ubu-) (n) of being blown away by strong wind; commotion.

-bhenguza (v) blow violently (of wind); get into a temper.

-bhensa (v) have bent back; be insolent; be wasteful.

-bhentshi (i- ama-) (n) bench.

-bhense (isi- izi-) (n) person with curved in back.

-bhenzini (i-) (n) benzine.

-bheqeza (i- ama-) (n) large flap; broad collar or cap.

-bherethi (i- ama-) (n) beret.

-bheseni (u- o-) (n) basin.

-bheshu (i- ama-) (n) skin buttock-covering worn by men; men who still wear Zulu traditional dress.

-bheshwana (u- o-) (n) small flaps of beads worn by boys and girls over buttocks, at sides and in front.

-bhesi (i-) (n) bass; bass voice.

-bhetri (i- ama-) (n) battery.

-bheva (isi- izi-) (n) fiery tempered person.

-bheyila (v) bail out; pay bail; come out on bail.

-bheyili (i-) (n) bail.

-bhibha (v) spread as festering wound; fly (as flag).
   ukubhibha kohlanga — winter time.

-bhibhi (ubu-) (n) meerkat.

-bhibhidla (v) learn to do a thing; bubble.

bhibi (ideo) of striking with back of hand on mouth.

-bhibiza (v) hit on the mouth with back of hand.

bhici (ideo) of something soft spreading out because of pressure (as stepping on a tomato); of oozing.

-bhici (isi- izi-) (n) something overripe (as fruit); rotten stuff.

-bhici (u- izim-) (n) discharge from eyes.

-bhicika (v) become squashed; ooze.

-bhicongo (isi- izi-) (n) disaster; devastation.

-bhida (v) bid (as at a sale).

-bhidana (v) get confused.

-bhidakala (v) get spoilt; quarrel.

-bhidi (i- ama-) (n) object with many colours; variegated.

-bhidi (um- aba-) (n) leader of marriage dancing group; conductor.

-bhidi (izi-) (n) sediment.

-bhidilisha (v) act in an uncertain manner; act as a learner.

-bhidisi (um- aba-) (n) conductor.

-bhidiliza (v) act in an uncertain manner; act as learner.

-bhidisa (v) conduct a choir.

-bhidla (v) rain heavily.

bhidli (ideo) of falling apart); of collapsing (as house); of appearing in large numbers (e.g. sores on body).

-bhidli (isi- izi-) (n) something in abundance; be large (e.g. a town with numerous houses).

-bhidlika (v) fall down; collapse (as a building); come out in large numbers (as sores).

-bhidlibhidli (imbidlimbidli izim-) (n) clumsy stout person.

-bhidliza (v) demolish.

-bhido (um-) (n) edible wild herbs.

-bhija (v) sway the body; twist arm; catch red-handed.

In (12) the Sepedi nouns *lebopa* and *mabopa* are lemmatized on their nominal stem form *–bopa*. The same holds true for this isiZulu nouns *ibhendi* and *amabhendi* lemmatised under *-bhendi*.

Lemmatising nouns stems is not user-friendly especially for inexperienced users and learners of the language and it introduces unnecessary problems in respect of stem identification. Central to these traditions stands the issue of conjunctivism versus disjunctivism. Consider an example where Setswana (disjuntivism) is compared with isiZulu (conjuctivism).

4.3.3   Conjunctiveness versus disjunctiveness

Table 26: Conjunctivism versus disjunctivism

Setswana		isiZulu	
Re a leboga	(We are thankful)	*Siyabonga*	(We are thankful)
Ke ne ke opela	(I was singing)	*Bengicula*	(I was singing)
A re kopaneng	(Let us meet)	*Masihlangane*	(Let us meet)

For words in case of 'We are thankful', 'I was singing' in Setswana are written as four separate words while in isiZulu are written as a single word. Van Wyk (1995) describes conjunctivism versus disjunctivism as purely a matter of orthographical convention. The stem tradition has mostly been followed for dictionaries for the conjunctively written languages namely isiZulu, isiXhosa, isiNdebele and Siswati (Nguni languages) and the word tradition for the disjunctively written ones, Sesotho sa Leboa, Setswana, Sesotho, Tshivenda and Xitsonga.

4.3.4   Lexicographic Traditions

4.3.4.1   Stem versus word tradition in respect of verbs

Van Wyk (1995) states that it is important to note the difference between nouns and verbs when it comes to affixes (prefixes and suffixes). According to Prinsloo (2009:6), lemmatising stems of verbs in particular makes sense for the conjunctively written languages. A huge number of prefixes, up to more than 4000 per verb, combine freely and productively with verbs, such as subject concords, object concords, negative morphemes, the progressive, the potential, future, etc. Lexicographers of conjunctively as well as disjunctively written languages agree that stem lemmatisation is the best option. It would also be totally redundant to lemmatise each verb stem plus prefixes separately. For example, *ngiyabonga* (I am thankful) and *masihlangane* (let us meet) etc. in isiZulu are all lemmatised under their stem *-bonga* (thankful) and *hlangane* (meet). According to Prinsloo (2009:9), in case of verbal suffixes in particular, verb stems plus suffixes should be lemmatised separately to avoid very long articles where treatment of the numerous derivations is attempted under a single stem form, for example, as in the Popular Northern Sotho Dictionary (POP) in contrast to the above.

Example 14: **POP**

badiša cause to read/count

bala   read; count, reckon: include

balêga be counted

balêgê go se ~ innumerable

balêla read/count for…

balola recount…

*balw***a** be read, counted, ~ *le* included

4.4 Problems regarding the lemmatization of nouns with prefixes *le-* and *lo-* in Setswana dictionaries

In Setswana, particularly the South African written variant, use is made of the singular class prefixes *lo-/le-* with their accompanying plural forms *di-/ma-* respectively. It has to be clarified that even in dictionary entries, *lo-*, *le-*, *di-* and *ma-* should be captured differently to avoid a situation where one excludes another. For example, in Setswana sometimes we speak of *lofuka* (wing) whose plural form is *diphuka* (wings), *lofofa* (wing) whose plural form is *diphofa* (wings) while in other instances we will use *lefofa* (wing) with the plural *mafofa* (wings). While the two versions are intelligible in all instances, *diphofa* includes references to big feathers and a multiplicity of feathers while *mafofa* simply refers to plurality. It has to be borne in mind that, while class prefixes *lo-* and *le-* can be used interchangeably in some instances, their usage becomes absolute in others. For example, there are instances where *lo-* cannot be replaced by *le-* as in *lore* (wood) whose plural is *dinthe* (woods), and similarly, *le-* cannot be substituted for *lo-* as in *lesepa* (mess) whose plural form is *masepa* (mess). Also in terms of concordial forms, *lo-* and *le-* used variably after *lona* as in '*lona lo bona lo le batho*' (you think you are people) and '*lena le bona le le batho'*. (you think you are people). *Lo-* in this regard serves as a variant of *le-* although the two are used interchangeably. It is recommended that the written variant be used consistently when chosen, that is, if *lo-* is used, it has to be used in the whole text. The same applies to *le-*. Be that as it may, *lo-* is used consistently in some editions of the Bible and it has acquired a sense of religiosity and respect and accords these qualities to the addresses.

A number of nouns of class 5 *le-* are normally used in the plural only. The *lo-* class prefix belongs to the noun class 11 and takes the plural prefix *diN-* and it is typically found in the central and southern dialects of Setswana.

According to Cole (1975:91), this class has been partially or completely absorbed into class 5 *le-*. The singular form of class 6 has the singular prefix *le-*, but retains

the plural prefix diN-, though frequently replaced by the prefix *ma-*. It is indicated that class 11 *lo-* is gradually becoming obsolete in Setswana.

From the above given examples, it is clear that there is no fixed law or principles of dealing with the lemmatization of the noun prefixes *le-* and *lo-*. Consider the following examples 15, 16, and 17 below:

Example 15: MSED

(a) Singular *le-*	Plural
Lerapo	***Marapo***
lerapô *la mokwatla*, N. PL. a lumbar vertebra; *la molala*, cervical vertebra. lerapô N. CL. 5 *le-*, SING. OF *marapô*, a trap; *kgori e bôna mae lerapô ga e lo bone*. lerapô N. CL. 5 *le-*, SING. OF *marapô*, bone.	marapô N. CL. 6 *ma-*, PL. OF *lerapô*, or *lesapô*, same as *masapô*, bones; used as a REL. S., it means strength; power, or forcefulness. *O marapô*, he is strong. marara N. CL. 6 *ma* riddle; complications, also
Leratla	***Maratla***
leratla N. CL. 5 *le-*, SING. OF *maratla*, a great clatter; the noise of a crash; a crashing noise.	maratla N. CL. 6 *ma-*, PL. OF *leratla*, noises.
Lerako	***Marako***
lerakô N. CL. 5 *le-*, SING. OF *marakô*, a stone wall.	marakô N. CL. 6 *ma-*, PL. OF *lerakô*, ruins.
(b) Singular *lo-*	Plural
Lorapo	***Marapo***
lorapô N. CL. 11 *lo-*, SING. OF *dithapô*, a string; a snare. PROV. *kgori e bôna mae, lorapô ga e lo bone*. lorapô N. CL. 11 *lo-*, SING. OF *marapô*, a bone.	marapô N. CL. 6 *ma-*, PL. OF *lerapô*, or *lesapô*, same as *masapô*, bones; used as a REL. S., it means strength; power, or forcefulness. *O marapô*, he is strong. marara N. CL. 6 *ma* riddle; complications, also
Loratla	***Maratla***
loratla N. CL. 11 *lo-*, SING. OF *dithatla*, a loud, or startling noise.	maratla N. CL. 6 *ma-*, PL. OF *leratla*, noises. maratô N. CL. 6 *ma* PL. OF *leratô* DER. E *rata*
Lorako	***Dithako***
lorago N. CL. 11 *lo-*, SING. OF *marago*, the buttock. lorakô N. CL. 11 *lo-*, SING. OF *dithakô*, a stone wall; a stone dyke.	persons; dextrous. dithakô N. CL. 10 *din-*, PL. OF *lorakô*, stone walls.

The compilers of the MSED opted to lemmatise both the singular and the plural forms of the nouns of class prefixes *le-* and *lo-* separately and give treatment to both pairs. However, redundancy becomes a factor in dictionaries that offer treatment of both the singular and plural forms as indicated in example 16 above.

Consider another example where compilers opted to lemmatise only the most frequent forms of both the singular and the plural forms of the prefixes *le-* and *lo*.
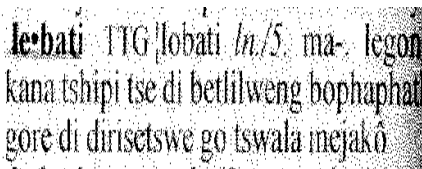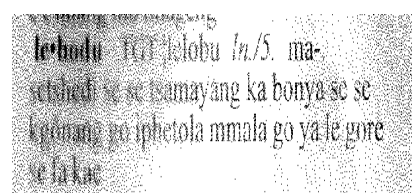
Example 16: THAND

Le-	*Lo-*
leofa tlola molaô wa Modimo; dira bosula *Rotlhe rea* leofa. *Rotlhe re* leofile. **leoto(ma)** lekoto; se setshedi se tsamayang ka sônê kgotsa koloi, baesekele, jalo, jalo. **leotwana(ma)** mopakô o o tshegetsang tlhomeso. **lepa** tlhôkômêla tiragalô ka lobaka *Go* lepa *dilô go ruta thata*. **lepai(ma)** kobô ya Sekgowa e e aparwang fa godimo. **lepê** Ga go lepê (lefoko) — ga go na se se buiwang. **lêpêlêla akgêga** ; ikutlwa tlala thata **lephoi(ma)** nônyane ya naga e e jang mabêlê; leeba. **lephutshe(ma)** sengwe sa dijalo se se monate; se ratile go nna sesweu ka fa ntlê mme mo teng sehibidu; se dithôtsê ditshweu. **leradu(ma)** kgomo e e gangwang mo gae. **leragô(ma)** nama ya moragô e go nnwang ka yônê. **lerama(ma)** nama ya lotlhaa. **leratla** modumô o mogolo jaaka wa go lela ga tlou gongwe batho ba omanêla kwa godimo. **lere** isa diatleng tsa yo mongwe; se go palangwang lomôta ka sônê Lere lere *ele ke palamê*. **lerêmêla(ma)** phatsanyana e e dirwang ke marêmô a selêpê. **lerinini(ma)** nama e e tshwereng mênô. **leritiba** motho yo o tlhôkômêlang motse kgotsa dijô. **lerôbôrôbô** bolwetse bo bo bolayang batho ba le bantsi thata. **leroo(ma)** kgatô ya sebatana — katse, ntša, tau, nkwê, jalo, jalo. **lerope(ma)** fa go kileng ga bo go agilwe teng gongwe go lengwa teng. **lerophi(ma)** kokobalô ya letlalô la seatla e e dirwang ke go tshwara tirô e motho a sa e tlwaêlang.	**lôpô** mokwatla wa tlhôbôlô o lerumô le tswang ka ônê. **lopôlô** poo ya tshêphê. **lora** nna le ditshwantshô mo tlhaloganyong mo borokong. **loraba** mongala, bophadiphadi. **loradu** kgakologêlô e e lotobo motho a sa le monnye. **lorakô(dithakô)** kagô ya majê e e thekeletsang. **loratla** modumô o motona thata. **loratô** mowa wa go rata; mowa wa go baya pelo mo mothong yo mongwe go mo dirêla gotlhe mo go siameng; mowa o o fenyang sengwe le sengwe; le dilô di ka ratwa. **lore(dinthe)** thôbane e motsofe o ikôkôtlêlang ka yônê. **loriba(ma)** sekôpô se setona sa tshwene. **lorole(dithole)** mmu o mosesane o o tsosiwang ke motsamaô wa sengwe gongwe ke phefô. **lorôlô** dithôlô tse dintsi. **loruô** mo gontsi ga diruiwa mo motho o nang nagô. **losa** tlhabantsha; lwa le phôlôgôlô; dira gore diphôlôgôlô di lwê *Mogale o* losa *nkwê*. *Mosimane o* losa *dintša*. **losaisai** bôna lesaisai. **lošalaba** bôna lešalaba. **losea(ma)** ngwana yo o tshotsweng. **losêlô(ditshêlô)** selô se go fêfêrwang mabêlê kana mmidi o o tlhobotsweng ka sônê. **losiê** phôlôgôlô e e mogatla o serutha, o e itshereletsang ka ônê ka go o tshamekisa e o isa kwa le kwa fa e tlhasetswe. **losika(ditshika)** mo go êlêlang madi; batho ba ba tshwaraganang ka go tsalwa; mo go rokang matlal **losire(ma)** bôna lesire. **loso(dintsho)** phêlô ya botshelô; tlogô ya mowa mo nameng. Loso *ga lo tlwaelwe*.

This approach reduces redundancy, but is heavily dependent on previous knowledge of the language.

Consider the third example taken from the THAN where the class prefixes *le-* and *lo-* are both lemmatised, and instead of treating them, they supplied a cross-reference to the prefix *le-*. Consider for example, nouns such as *lonao*, *longana*, *loologa*, *lookwane*, *looto*, to mention but few.

Example 17: THAN

Lebati	*Lobati*
le•bati TTG |lobati *ln./5.* ma-. legon kana tshipi tse di betlilweng bophaphad gore di dirisetswe go tswala mejakô	
Lebodu	*Lobodu*
le•bodu TTG |lobu *ln./5.* ma-. ...ayang ka bonya se se ...phetola mmala go ya le gore	
Lebone	*Lobone*
Lebota	*Lobota*

This strategy increases redundancy to a great extent and a number of cross-references cause a great deal of page turning which is time consuming. A further

complication arises from the fact that the relations between singular and plural are not always regular and not satisfactorily treated in dictionaries such as MSED and THAND. Consider the following example in this regard:

Example 18

Singular		Plural
Letsatsi (sun / day) [Matsatsi]	>	*Malatsi*
Letsogo; Lebogo (hand) [Mabogo]	>	*Matsogo*
Letsele; Lebele (breast) [Matsele]	>	*Mabele*

From the above example it is clear that preference is given *to malatsi*, *matsogo* and *mabele* that are irregular forms. The latter is further supported by an example of a proverb *mabogo dinku a a thebana* (it is important to assist one another).

The above mentioned nouns, i.e. *matsatsi* and *malatsi; matsogo* and *mabogo* etc. should be treated as variants and lemmatized as single entries. An example should also be given to help identify subtle differences between similar words as indicated in brackets *[malatsi], [matsogo] and [matsele]* which are preferred to *matsatsi, mabogo* and *mabele.*

Furthermore, both the THAN and the THAND lemmatize singular nouns only and give the plural prefix as part of the treatment. This can be confusing to the user. Suppose a foreign user is confronted with the word *malats*i. (S)he will look for the word under **lelatsi*. In this instance the user will get lost since the plural form has undergone a morphological change. It is recommended that both singular and plural forms of nouns be lemmatized to render a user-friendly

product. This approach should also be based on a frequency count as outlined previously in this chapter.

4.5   The absence of nominal derivations

Dictionary information on nominal derivations such as diminutives are not provided by Setswana lexicographers. It is unfortunate that the majority of Setswana dictionaries provide inadequate treatment of these lexical items. It is interesting to note that diminutives can have different meanings (senses) when nouns referring to various body organs or human beings are used. Diminutives associated with body organs express behavioural patterns, i.e. gossip and idling around.

Example 19

(a) Diminitives associated with body organs

O tota o le loleme	> (You really have a tongue)	'you talk too much'
O tota o le lolengwana	> (You really have little tongue)	'you gossip too much'
O lenao	> (You are a foot)	'you are loitering'
O lenaonyana	> (You are a little foot)	'you are idling or gallivanting'

The suffixes *-ngwana* and *-nyana in lolengwana and lenaonyana* justify treatment in their own right.

(b) Human beings

   i. *Bona **mosadinyana** yoo*. (Look at that little woman).
  ii. *Ao **mosadinyana** wa me*. (Oh my little woman) Oh my adorable woman.

In the above given example in b(i), the sentence can refer to the following:

> ❖ Look at that adorable woman
>
> ❖ Look at that little woman
>
> ❖ Look at that little woman showing disrespect but may also indicate an expression of love or passion

Example b (ii) indicates an expression of love or passion.

The above-mentioned assertion underlines the need to be included and treated in the dictionary. The absence of the oral corpus and the exclusion of the diminutive can deprive the user of certain information, especially if there is a high frequency of usage.

Consider example 20 below where diminutives have undergone morphological change:

Example 20

Word			Diminutive - form
Tsebe	(ear)	>	*Tse**jwana*** (small ear)
Moriri	(hair)	>	*Mori**tshana*** (small hair)
Legodu	(thief)	>	*Lego**tswana*** (small thief)
Kgomo	(cow)	>	*Kgo**ngwana*** (calf)
Ngwana	(child)	>	*Ngwa**nyana*** (little baby)

From the above given examples it can be seen that the diminutives have undergone morphological change. The user who is not familiar with the language may easily be confused and will be unable to look them up under their non-derived forms.

Although the THAN is directed at the native speakers of Setswana who have a strong command and knowledge of their language, users are also expected to have sufficient knowledge of morphology when it comes to nouns suffixes. This is often not the case. It is disappointing that noun inflections such as diminutives are omitted. Once again the question can be asked, on which grounds the items were selected.

4.6   Inconsistency regarding the lemmatization of homonyms

Homonyms are described as two or more unrelated meanings associated with the same form, for example, *bank* (of a river) and *bank* (financial entity). The treatment of homonyms in Setswana dictionaries also reveals some shortcomings. The question pertaining to the selection of entries is when to consider two occurrences of a word as having related or unrelated meanings. Compare the following examples 21 and 22 from (THAN and SEAD) and (THAN and THAND) respectively.

Example 21: tshela

THAN	SEAD

siameng mo botshelong; dira gore go nnê le lesegô
 	
tshegofatsô GGGT *ln./9.* Ø tirô ya go eleletsa dilô tse di siameng mo botshelong
 	
tshegofatsong GGGTT *ltls.* >tshegofatsô+ng
 	
tshekatshekô GGGT *ln./9.* di-. tirô ya go tlhatlhoba sengwe; kanokô; patlisisô
 	
tshekatshekong GGGTT *ltls.* >tshekatshekô+ng
 	
tshekêga GGG *tpt.* -ile. 1. baya ka letlhakore 2. borapalalô kgotsa tshekamô ya selekanyô sa 45°
 	
tshêkêlô TTT *ln./9.* di-. lefelô le go sêkêlwang teng
 	
tshêkgê! TT *ltss.* modumô wa fa go tlhabiwa kgotsa go segiwa mo nameng ka sengwe se se bogalê; rasu!
 	
tshêkgênngwa TTTT *tpt.* >tshêkgênya+iwa
 	
tshêkgêntsê TTTT *tpt.* >tshêkgênya+itse
 	
tshêkgênya TTT *tpt.* -itse. sega kgotsa tlhaba mo nameng ka sengwe se se bogale go dira nthônyana; rasunya
 	
tshekisanyô TTTT *ln./9.* di-. >tshekisô+anya
 	
tshekisô TTT *ln./9.* di-. tirô ya go botsolotsa yo o dirileng molato le go mo atlhola
 	
tshêkô TT *ln./9.* di-. tirô ya go bôna gore molato o dirilwe ke mang le go baya katlholô
 	
tshêkong TTT *ltls.* >tshêkô+ng
tshela GT *lts.* tlwaêla go dira sengwe; aga; tlhôla
 	
tshela GT *tpt.* -ile./-itse. 1. seêmô sa go kgôna go itirisa ga dilô tsa tlhôlêgô mo go êmisiwang ke loso 2. mokgwa wa botshelô ♠ *go tshela sa ntša le phiri = go ilana thata*
 	
tshela GT *tpt.* -ile./-itse. ralala sengwe, jaaka metsi, molelwane, jj.
 	
tshêla TT ¦thêla *tpt.* -itse. 1. tsenya sengwe se se seng pope mo go se sengwe 2. ntsha mantlê a a metsi thata; tšhwêga ♠ *go tshêla motho mmu mo matlhong = go tsietsa yo mongwe*
 	
tshelaganya GGGT *tpt.* -itse. >tshela+ega+anya; feta sengwe ka mokgwa wa go se ralala
	'n kookhouer oor die vuur te hou), drie-poot
	tshêgô, mo- *dev* < *tshêga,* manner of laughing // manier van lag; se- di-, laughter // lag, gelag
	tshêgô *rel.* lucky, fortunate // geseënd, gelukkig, voorspoedig, eg *Mosadi yo o tshego,* A lucky woman // 'n Gelukkige vrou
	tshêgófádïwa *in Go tshegofadiwa ka thari,* To be blessed with children // Om geseënd te wees met kinders
	tshêgófala *den* < *lesegô,* become blessed *or* lucky // geseënd *of* gelukkig raak
	tshêgófátsa *caus* // *kous* < *tshegofala,* bless, make lucky // seën, gelukkig maak
	tshêgófátsô *dev* < *tshegofatsu,* blessing, grace, luck // seëning, genade, geluk
	tshêka, 1. *n* bo-, north // noorde; 2. *adv* bo- *or* // *of in kwa botsheka,* north // noord, *eg Ba ile (kwa) botsheka,* They have gone north // Hulle het noord(waarts) gegaan
	tshêkágányô *dev* < *sekaganya,* diagonal, chiasmus *or* crossparallelism (literary) // diagonaal, oorhoekse lyn, chiasmus *of* kruisparallellisme (lettk.)
	tshékêgá, le-, side // kant
	tshékge (tshêkgê) *id. vide tshêkgênya,* denoting nicking *or* the making of small incisions // wat kerf *of* die maak van klein snytjies aandui, *eg Ngaka ya mo re tshekge! ka logare,* The medicineman nicked him with a blade // Die medisyneman het hom 'n klein snytjie met 'n lemmetjie gegee
	tshêkgênngwa *pass* < *tshêkgênya,* be nicked *or* incised // word gekerf *of* gesny
	tshêkgêntsê *perf* < *tshêkgênya,* (have *or* has) nicked *or* made a small incision // het gekerf *of* 'n klein snytjie gemaak
	tshêkgêntsha *caus* // *kous* < *tshêkgênya,* cause *or* help to nick *or* make a small incision // laat *of* help kerf, 'n klein snytjie laat *of* help maak
	tshêkgênya, nick, make a small incision // kerf, 'n klein snytjie maak
	tshêkgênyêtsa *appl* < *tshêkgênya,* nick *or* make a small incision for *or* on behalf of *or* at // kerf *of* 'n klein snytjie maak vir *of* namens *of* by
	tshêkisô *dev* < *sêkisa,* prosecution, hearing (a case), trial // vervolging, verhoor
	tshêkô, di- *dev* < *sêka,* court case, legal suit // hofsaak, regsgeding
	tshéla, be alive, live (*ie* exist) // lewe (*ie* in lewe wees), leef (*ie* bestaan voer)
	tshéla, cross water (*eg* a stream *or* an ocean) // oor water gaan *of* water kruis (*eg* 'n stroom *of* 'n oseaan)
	tshéla, pour // giet, (in)skink, ingooi; *Go tshela bola,* To throw bones for divination // Om dolosse te gooi; *Go tshela lešalaba (lošalaba),* To cheer loudly // Om luidkeels toe te juig
	tshélakgábo, mo- me-, pole *or* tree trunk serving as a bridge over a stream

From example 21 above, THAN provides treatment of homonyms *tshela* as *tshela* (used to), *tshêla* (pour), *tshela* (cross) and *tshela* (to live) while SEAD provides only

three homonyms i.e. *tshela* (to live), *tshela* (to cross) and *tshêla* (pour). Consider also the second example of *thari* taken from the THAN and the THAND:

Example 22: thari

THAN	THAND
thankgola latlhêla sengwe kgakala. **thanolô** tlhalosô e e tletseng; tshinosó *Morutwana o tlhôka thanolô mo moruting wa gagwê.* **thanthanyêga** ša ka go ntsha ditlhase. **thanthologa** tswa mo morutsheng jaaka tlhale. *Tlhale e* thanthologile. **thantholola** ntsha mo morutseng; thatholola *Ke e* thantholotsê. **thapa** nna le motho yo o dirang ka tuêlô *Re tshela ka go* thapa *ba bangwe.* *Ke* thapile *batho ba.* **thapêlô(di)** kopô mo Modimong; se motho a se lebogêlang Modimo gongwe yo mongwe; puisanyô e tona e e tlhwaafetseng e e kopang kagisô kgotsa boitshwarêlô. **thapô** tirô ya go thapa; se se mo kungweng e le peo ya yônê; thudi, se se bôfang. **tharabololô** tlhalosô gongwe thanolô ya se se ne se le thata go tlhaloganngwa. **thari(di)** letlalô le bogologolo go ne go bêlêgwa ngwana ka lônê. Gompieno go dirisiwa letsela la "kaki" gongwe la "matalanyane". **tharo** raro, pedi le nngwe. **thata** nonofô, maropô; tlhôka kutlwêlôbotlhoko ya go thusa; bokete; sa tobetsegeng *Ke na le* thata *ya go go lêlêka.* *Gôpane ke motho yo o* thata *go thusa ba bangwe.* *Malatsi ano a* thata *ruri.* *Go* thata *go êpa foo.* **thatafatsa** dira gore go nnê thata; nonotsha; ketefatsa pelo *Faro o ne a* thatafatsa *pelo ya gagwê.* *Ke ne ke* thatafadiwa *ke mafoko a gagwê.* *Rre o* thatafaditse *pelo ya gagwê.*	**thari** TG *ln./9.* di-. 1. letlalô kana letsela le go bêlêgwang ngwana ka lônê 2. mo go tswang moragô ga ngwana a sena go tsalwa; motlhana **tharing** TGT *ltls/* >thari+ng **tharo** GT *td.* bobedi bo tlhakane le bongwe **thata** GT *lmn.* mo go palêlang go dirwa sengwe **thata** GT *ln./9.* di-. 1. itekanêlô mo mmeleng 2. tshwanêlô go dira ka thatô ♣ *ga ke thata ke le nosi, ke thata ka lentswe = kgakololô ya batho ba ba ntsi e gaisa ya motho a le nosi* **thata** GT *ltls.* mo go fetileng selekanyô **thatafaditse** GGGTT *tpt.* dirile thata **thatafadiwa** GGGTT *tpt.* >thata+afala+iwa **thatafala** GGGT *tpt.* -itse. >thata+afala **thatafalêla** GGGTT *tpt.* -itse. >thata+afala+ela **thatafaletse** GGGTT *tpt.* >thata+afala+ela+itse **thatafaletswe** GGGTT *tpt.* >thata+afala+ela+iwa+itse **thatafatsa** GGGT *tpt.* -itse. dira thata **thatafatsang** GGGTG *tpt.* dirang thata **thatafatsô** GGGT *ln./9.* di-. tirô ya gore go nnê thata kana go palêgê **thatafetse** GGGT *tpt.* nnile thata **thataro** GGG *td.* boraro bo tlhakane le boraro

From example 22 above, both the THAN and the THAND managed to provide only one sense of the homonym *thari* as (skin used to carry a baby) while other senses are excluded.

From examples 21 and 22 for *tshela* and *thari*, it is clear that frequency counts and user-friendliness were not taken into consideration. These are words which are frequently used. The problem with frequency counts by corpus query programs such as WordSmith Tools is that a single total count is given for the orthographic form shared by both homonyms. The lexicographer has to determine the frequencies manually. It is important to emphasize that the distinction between the homonyms be determined on the basis of frequency counts and that the structural markers be employed to indicate the different contexts in which the lemma signs *tshela* and *thari* can occur. The above mentioned statement is supported by Hausman and Wiegand (1989:356) where they emphasize the use of the structural markers to demonstrate the different meanings of homonyms and maintain that these should not be presented haphazardly but according to a predetermined set of criteria accounted for in the front matter text. Consider the following example in this regard:

Example 23

Tshêla[1]	(to put in)
Tshéla[2]	(to cross)
Tshéla[3]	(to live)

Compare also the following examples where Afrikaans and English dictionaries are consistently managing to lemmatize the homonyms:

Example 24: Verklarende Afrikaanse WoordeBoek (VAW)

> **Graaf¹,** (s), **grawe**. **1**. Adellike titel. **2.** Iem. Met so 'n titel.
>
> **Graaf²,** (s), **grawe. 1**. Spitwerking. **2**. Lengte van 'n   graafblad. **3**. Hoeveelheid wat op 'n keer met 'n graaf geskep kan word; ~ steel.

Example 25: TSAOSD

> **lag¹** *verb* (**lagged, lagging**) go too slowly and fail to keep up with others.
>
> **lag²** *noun* lagging; a delay.
>
> **lag³** *verb* (**lagged, lagging**) wrap pipes or boilers etc. in insulating material to keep them warm.

4.7 The absence of tonal indication

Tonal indication is of crucial importance in Setswana for it is used to make both semantic and grammatical distinction. As far as tonal indication is concerned, it is important to note that dictionaries employ different conventions. Some dictionaries indicate high as well as low tones while others indicate only high tones. For example, Ziervogel enters head words or sub-head words in capital letters with indication of high tone e.g. *RÉKA*. The word is repeated in brackets with the circumflexes indicated on *e* or *o*: *rêka*. For such entries he uses lower case: *lerêko*, *ma- rêko*. A second option is to indicate the tonal pattern separately by means of the upper case character **L** for low tone and **H** for high tone. For example, Van Wyk in Pukuntšu enters head words with the circumflexes indicated, as follows: *boikêtsiso* and indicates the tonal pattern further down in the article as, for example, LLHHL.

As far as the Setswana dictionaries are concerned, it is unfortunate to state that only SEAD succeeded in showing tonal indication. In this dictionary, a distinction is made between a high, a low and a falling tone which are indicated by means of two tonal diacritics i.e. (′) high tone, (¯) a falling tone while the low tone is left unmarked. In the dictionary the mid-low vowels [ɛ] and [Φ] are marked with a circumflex, for example *ê* in *êma* and *ô* in *ôma*. One gets the impression that compilers of the SEAD regarded tonal information of equal importance to lexical information. It is disappoinding that other Setswana dictionaries such as THAN, THAND, MSED and SED excluded tonal distinction. The example below will now be used to demonstrate the importance of tonal indication in Setswana dictionaries.

Example 26

Item **240** in the Setswana lemmatized frequency list is *tshela*. Without tonal indication, this form could mean any of the three possibilities listed below:

1. *tshéla[1]*   [verb]   'to jump'
2. *tshêla[2]*   [verb]   'to pour'
3. *tshéla[3]*   [verb]   'to cross'

It is suggested that tonal indication be included in the forthcoming Setswana dictionaries to render a more user-friendly Setswana dictionary. Compare now an example extracted from SEAD where compilers consistently strive to make use of tonal indication.

Example 27: SEAD

katosa *caus* // *kous* < **katoga**, move farther away from // verder weg beweeg van
katsa, become muscle-bound // styf raak (spiere wat ooreis word)
katsa, be out of rhythm *or* off beat *or* out of step (lit. and fig.) // uit die maat *of* uit die pas wees (lett. en fig.)
katsakuba, di-, bulge (*eg* as a result of an injury) // knop (*eg* as gevolg van 'n besering)
kátsê, di- < *Afr*, cat // kat
katsé, mo- me- *Dicerocaryum zanguebarium* (167), boot-protector // beesdubbeltjie, duiwelsdis
kátsê, mo- me-, a hybrid between a water-melon and a wild water-melon // 'n baster tussen 'n waatlemoen en 'n makataan
katsea, be bloated, be hard (*eg* an inflated ball *or* a swelling) // opgeblaas wees, hard wees (*eg* 'n opgeblaasde bal *of* 'n swelsel)
kátsô, di- *dev* < **atswa**, alms, a tip // liefdegawe, fooitjie, aalmoes
kátšwa *pass* < **kapa**, be caught *or* grabbed (something in mid-air) // word gevang *of* gegryp (iets wat gegooi word)
katúne < *Afr katoen*, cotton plant // katoenplant
káu, 1. *n* **bo-**, young-manhood // jongelingskap; **le- ma-**, unmarried young man, smartly dressed man // ongetroude jong man, netjies geklede man; **2.** *adv* **se-**, like a young man // soos 'n jong man, *eg Ntate o sa ntse a tsamaya sekau*, My father still walks like a young man // My pa loop nog soos 'n jong man
káulêngwe, bo-, brotherhood // broederskap; **mo- ba-**, friend, brother, fellow member, associate // vriend, broer, medelid, deelgenoot; **se-**, brotherhood, friendship // broederskap, vriendskap
kaútu, mo- me-, big strongly built person // groot frisgeboude persoon
káwa, mo- me- *Coccinea sp?*, a species of wild cucumber // 'n wildekomkommerspesie
kawanyana, mo- *rel*, a few // 'n paar, *eg Batho ba ba mokawanyana*, A few people // 'n Paar mense
káwédisa (kawêdisa) *caus* // *kous* < **kawêla**, cause to bloat // laat opblaas
káwélã (kawêla) *intr*, bloat // opblaas
káwétsê (kawêtsê) *perf* < **kawêla**, (have *or* has) bloated // het opgeblaas
kaya, point out, mean, imply, allege, regard as // aanwys, bedoel, beteken, beskou as, wys, beweer
káya, strap (*eg* a cow to be milked) // span (*eg* 'n koei om gemelk te word)
ke, *sc 1st p. sing* // ek, *eg Ke a bua*, I am talking // Ek praat
ke *aux* // *hulpww*, never // nooit, *eg* (i) *Ga re ke re bua jalo*, We never say that // Ons sê nooit so nie; (ii) *O ka se ke . . .*, You can never . . . // Jy kan nooit . . .
kê *aux* // *hulpww*, please do // asseblief tog, *eg A o ke o mphe dijo*, Please do give me food // Gee my asseblief tog kos
ke, (i) *descr cop of 1st p. sing* // bep kop van 1ste p. sing, I am // ek is, *eg Ke bo-*

tlhale, I am intelligent // Ek is intelligent, *or* // of *Ga ke botlhale*, I am not intelligent // Ek is nie intelligent nie; (ii) *id cop of 1st p. sing* // id kop van 1ste p. sing, I am // ek is, *eg Ke Motswana*, I am a Motswana // Ek is 'n Motswana, *or* // of *Ga ke Motswana*, I am not a Motswana // Ek is nie 'n Motswana nie
ké, *id cop of 3rd p.* // id kop van 3de p., he *or* she *or* it is, they are // hy *of* sy *of* dit *of* hulle is, *eg* (i) *Ena ke Mokwena*, He is a Mokwena // Hy is 'n Mokwena, *or* // of *Ena ga se Mokwena*, He is not a Mokwena // Hy is nie 'n Mokwena nie; (ii) *Ke jaaka ba bolela*, It is as they say // Dit is soos hulle sê
ké *adv pref*, by // deur, *eg Re romilwe ke ntate*, We have been sent by my father // Ons is deur my pa gestuur
kebéke (kêbêkê), se- di-, thug // boef, skurk
kedikilwê (kedikilwê), bó- (bô-) *Francolinus africanus and* // en *Francolinus levaillantoides*, grey-winged francolin, Orange River francolin // bergpatrys, Vrystaatse patrys (Kalahari-patrys)
kéétâne (kêêtane), di- < *Afr ketting*, chain // ketting
kéi (kêi), se- di- < *Afr jukskei*, jukskei // ketting
kêka *intr*, spread unobtrusively over a large area (*eg* a fire *or* the infection of a wound) // onopsigtelik versprei oor 'n groot oppervlakte (*eg* 'n vuur *of* die infeksie van 'n wond)
kêkê, le- ma-, carrier ant (termite) // stokkiesdraer (termiet)
kêkêla *intr*, spread unobtrusively over a large area (*eg* a fire *or* the infection of a wound) // onopsigtelik versprei oor 'n groot oppervlakte (*eg* 'n vuur *of* die infeksie van 'n wond)
kêkêma, le-, side // kant
kékétã (kêkêta), gnaw, chip *or* notch repeatedly // knaag, herhaaldelik laat happe kry, herhaaldelik (in)keep
kékéte (kêkêtê), 1. *n* **le- ma-** *dev* < **kêkêta**, gnaw-mark, notch *or* chipmark on an edge // knaagmerk, 'n hap *of* 'n keep in 'n rand; **2.** *rel* **ma-**, gnawed, serrated, chipped // geknaag, getand, happerig, *eg Nkgwana e e makekete*, The chipped clay pot // Die happerige kleipot
kêkêtse *perf* < **kêkêla**, (have *or* has) spread unobtrusively // het onopsigtelik versprei
kêkêtsha *caus* // *kous* < **kêkêla**, cause to spread unobtrusively // onopsigtelik laat versprei
kêkologa, avoid, approach deviously, digress // vermy, met 'n draai benader, met 'n draai nader, langdradig praat
kêlêdi, di- < **êlêla**, tear // traan
kélékã (kêlêka), look *or* examine carefully // deeglik kyk *of* ondersoek
kélékêtla (kêlêkêtla), flow out profusely (*eg* blood *or* tears) // uitstroom (*eg* bloed *of* trane)
kélékêtlo (kêlêkêtlô) *dev* < **kêlêkêtla**, continuous flow // aanhoudende vloei
kelerwá (kêlêrwa), di-, divining-bone, dice, lot // dolos, dobbelsteen, lot
kêlêtsô, di- *dev* < **êlêtsa**, envy, desire,

wish // benyding, begeerte, wens, *eg keletso ya dijo*, appetite // eetlus
kêlô, di- *dev* < **êla**, capacity (volume) // kapasiteit, inhoud (volume)
kelotlhóko (kêlôtlhôkô), 1. *n*, attention, care // aandag, versigtigheid; **2.** *rel*, attentive, careful // oplettend, versigtig, *eg Moithuti yo o kelotlhoko*, The attentive student // Die oplettende student
kêma, di- *dev* < **êma**, a tuft of hair left unshaven on the head // 'n klossie hare wat ongeskeer gelaat word op die kop
kémélélano (kêmêlêlanô) *dev* < **emêlêlana**, quarrelling // getwis
kémo (kêmô) *dev* < **êma**, standing, stance // stand, houding
kémonôsî (kêmônôsi), di-, political independence // politieke onafhanklikheid
kênêkênê, se-, porridge prepared from fermented sorghum husks, soup prepared from ground and whole locusts // pap wat uit gegiste sorghumsemels voorberei word, sop wat uit gemaalde en heel sprinkane voorberei word
kêpu, di- *dev* < **êpa**, digging-stick, crowbar // graafstok, koevoet
kéta (kêta), play the "hole-and-pebble game" // die "gat-en-klipspel" speel
kêtapele, lobolo beast presented by the uncle of a bridegroom // *lobolo*-bees deur 'n bruidegom se oom geskenk
keté *aux* // *hulpww* in e kete, used in expressing a wish *or* "it seems as if . . ." // wat gebruik word om 'n wens *of* "dit lyk asof . . ." uit te druk, *eg E kete o ka wa*, May you fall // Mag jy val; *E kete ga o bone*, It seems as if you do not see // Dit lyk asof jy nie sien nie
kétê (kêtê), bo-, such-and-such a place // so en so 'n plek; **mo-**, such-and-such a person, so and so // so en so 'n persoon, so en so; **se-**, such-and-such a thing // so en so 'n ding
kete, 1. *n* **bo-**, weight, heaviness, difficulty // gewig, swaarheid, moeilikheid; **se- di-**, thousand // duisend; **2.** *rel* **bo-**, difficult // moeilik, *eg Tiro e e bokete*, The difficult work // Die moeilike werk
kétê, mo- me-, feast // fees
ketefala *den* < **bokete**, become heavy *or* difficult // swaar *of* moeilik word
ketefatsa *caus* // *kous* < **ketefala**, make heavy *or* difficult // swaar *of* moeilik maak
kétékã, celebrate // vier
kétékô (ketekô) *dev* < **keteka**, celebration // viering
kêtêlêlapele, foreword, preface, introduction // voorwoord, inleiding
kêtêlêlôpele, manner of leading the way, leadership // wyse van leiding gee, leierskap
kêtêlêtsôpele, foreword, preface, introduction // voorwoord, inleiding
kétla (kêtla) *tr*, chip off (chippings) // spaanders afkap
kétlo (kêtlô), 1. *n* **le- ma-** *dev* < **kêtloga**, chip // hap, skaar (*eg* in 'n lem); **2.** *rel* **le- ma-**, chipped // happerig, *eg Boroto e e leketlo*, The chipped plate // Die happerige bord

In example 27 different meanings of *kete* are clearly distinguished on the basis of tonal indication. If such distinctions were not given the user would find it difficult to retrieve the correct information.

4.8 Dialect words as lemmas in the dictionary are evaluated against the existing Setswana dictionaries

The Setswana language has eight dialects. Dictionaries such as THAN, THAND, MSED and SED do not cover these regional varieties. They are only limited to one dialect i.e. the Serolong dialect which is considered as the standard language.

Wikipedia (2007) defines a dialect as a variety of a language used by people from a particular geographical area. Anthropological linguists on the other hand define a dialect as a specific form of a language used by a speech community. For example, in Setswana the sound represented by *f is* unknown to the Batlhaping tribe but very common among the Barolong tribe. The **h** in the word *le**h**atshe* (earth) is pronounced as **f** in Serolong as *le**f**atshe*. The glottal sound **g** of the southern Batswana tribe loses its glottal sound among the more northern tribes and becomes softened into the sound **h** and the sound **sh** becomes a pure **s**. Compare the following examples in this regard:

Example 28

Alternates *f*, **g** and **h**

English	Setlhaping	Serolong	Standard Setswana
Earth	*le**h**atshe*	*le**f**atshe*	*le**f**atshe*
Cow	*k**h**omo*	*k**g**omo*	*k**g**omo*
Fat	*ma**h**ura*	*ma**h**ura*	*ma**f**ura*
Scares	*tlhoka**h**ala*	*tlhoka**g**ala*	*tlhoka**f**ala*
Rich	***f**uma/**h**uma*	***g**uma*	***h**uma*
Found	***f**umane*	***f**itlhela*	*bone*

Compare other examples below where the vowel *o* in *lo-* is frequently replaced by *e* in *le-* with a possessive *la*.

Example 29

Le*riba* **la** *noka*   **Lo***riba* **lwa** *noka*   (bank of the river)

*Le*silo **la** *mosimane*   *Lo*silo **lwa** *mosimane*   (a stupid boy)

*Le*sogodi **la** *ditsie*   *Lo*sogodi **lwa** *ditsie*   (a swarm of locusts)

Consider other examples taken from a Botswana Television conversation:

Example 30

English	Botswana T.V.	Sekgatla	Sekwena	Standard
beat	***betsa***	*itaya/šapa*	*otla*	*betsa*
school	***sekwele***	*sekolo*	*sekolo*	*sekolo*
week	***biki***	*beke*	*beke*	*beke*
huge	***setona***	*setona/segolo*	*segolo*	*segolo*

If one compares the Botswana Television conversation with the standard language, one realises that we do not speak of dialects but alternates patterns since the dialectical boundries are broken up by the population shifts, urbanized, political re-organisation and technology.

Consider another example of alternates below:

Example 31

Alternates tš, ts, tšh and tšhwa

English	Serolong	Sekgatla	Sekwena	Standard
dog	*ntsha*	*mpša*	*ntšwa/ntšha*	*ntšhwa*
ostrich	*mpshe*	*mpše*	*ntšhe*	*ntšhwe*
youths	*batšha*	*bašwa*	*baša*	*baswa*

The lexicographer will have to justify the alternants in deciding whether to include or exclude in the macrostructure of the dictionary. Consider cojuctions as other examples of alternants:

Example 32

Conjuctions

English	Serolong	Sengwaketse	Setlhaping	Standard
though	*f*ela	*g*ela	*h*ela	*f*ela
if	*f*a	*g*e	*h*a	*f*a

It is important to note that (*f*) is pronounced as (*g*) in Sengwaketse and as (*h*) in Setlhaping. Consider the following example sentences:

> *Ke tla bua **fela fa** o ka ntetla* (I will speak only if you allow me)
> *Ke tla bua **gela ge** o ka ntetla* (I will speak only if you allow me)
> *Ke tla bua **hela ha** o ka ntetla.* (I will speak only if you allow me)

According to the research, the *f* in Setswana is favoured more than the *h*. Singled it can be justifiable. The *h* is a matter of Southern Sotho influence which is also affected by the geographical regions. Consider other examples taken from the Sekwena dialect:

Example 33

English	Sekwena	Standard
I don't know	*kgitse*	*ga ke itse*
I have been	*kgebolo*	*ga ke bolo*
I don't want	*kgebatle*	*ga ke battle*
Here	*kweno*	*kwano*

If one compares the Sekwena dialects with the standard language from example 33, it is important to note that the negative morpheme *ga ke* (I don't) has been coined to the verb stem itse, *bolo* and *battle*. Consder the following:

Kg + itse   *> kgitse*

Kg + bolo   *> kgebolo*

Kg + battle   *> kgebatle*

Consider other examples taken from selected forms of the Setswana drama books and novels:

Example 34

Selected passages from dramas and novels

Motimedi: D.P. Moloto

> Molatlhegi a nna jalo molomo o atlhame, mathe a tsutsurutla, a bile a elela fa fatshe. Metlhagare e opa, e lapile; mme le ha go ntse jalo keledi yona a se ka a ba a e rothisa. Tumelo o na a rata ha e re a otla ngwana a lele; ha a ka se ka a lla o na galefa thata. Jalo ge, ha a bona Molatlhegi a sa ledisiwe ke dipetso tsa gagwe, a mo isa kwa ntle a mo emisa ka lekoto le le lengwe, mme a mo rwesa maje a mabedi a magolo mo letsogong le lengwe le le lengwe. A mo tlogela gore a eme jalo go fitlhela sekolo se tshameka. Ha sekolo se tshameka a ya kwa go Mola-tlhegi a fitlha a mo nwa ka tshetlha (letlhaka), a re tsholeletsa maje go dimo, le ha go ntse jalo Molatlhegi a se ka a lela.

From the above example, it is clear that the author used different dialects and languages as alternates. For example, the use of '*ha*' instead of '*ga*' which is the direct influence from the Setlhaping dialect; the use of '*otla*' (beat) instead of *betsa* or *itaya* which is the direct influence from the Sekwena dialect and the use of the word *lekoto* (leg) which derives from the Sepedi language.

Moratho o montsho: S.S. Tshetlho

Ka ikutlwa ke otlwa ke letswalo. Ke tsamaiwa ke phefo e e maruru mo mokwatleng ke le mosadi. Ya re a laela, modumo wa tlhatloga le go feta.

From the above given example, the author makes use of the word '*otlwa*' (beaten) instead of *betswa* or *itawa* which is the direct influence from the Sekwena dialect.

Consider another example below:

Ke ne ke bapile le Ntsie. Ra nna ra boga bontle jwa 'tiro tsa Modimo, re sa lemoge gore bosigo bo ntse bo totoba bo ya pele. Legofi la nna la otlwa. Bangwe re le kwa mmakarakaputla-a-bonoko, re segetswe mo lefureng ka gonne e se gantsi re ya go itlhabisa phefo.

"Ntsie, a o gopola 'tsatsi la tshirololo ya letlapa la ga nkgolo?"
Ntsie a se tlhole a botsa gore a naare ga a dumedise pele phakela. A utlwa seseseedi se mo huduga tlhogo. A gopola gore o itlhomile a weditse melato.

In this example the author makes use of the word '*nkgolo*' (grandfather) instead of *rremogolo* which derives from the Southern Sotho language and uses the sound '*h*' instead of '*g*' which is the direct influence from the Serolong dialect.

From the above given examples, one notices the inconsistencies of the contemporary author's orthography i.e. authors do not spell the words in the same way which indicate that there is no standard writing in the field.

It is important to note that lexicographers should look at a dialect as a bundle of characteristics peculiar to a language in a specific social environment. The lexicographer should keep in mind that a dictionary should be fully descriptive. It should record objectively various dialects and different styles. However this is not the case with the Setswana dictionaries. It is important to state that future Setswana dictionaries should pay more attention to dialectical forms. Although the whole issue regarding Setswana dialects is complicated, at least frequently used dialectical forms should be given and appropriately labelled.

4.9   Conclusion

In this chapter the extent to which the respective Setswana dictionaries succeeded or failed to treat the most commonly used words was indicated. We have critically analysed and evaluated the typical macrostructural inconsistencies existing in Setswana dictionaries. Macrostructural aspects relating to the inclusion versus the omission of individual lemmata was dealt with. In addition it was also stressed how corpora can be put to good use in revising and improving the macrostructure of the existing Setswana dictionaries. The unequal treatment of derived forms of verbs which results from a lemmatisation approach where lexicographers simply add words as they come across them was also illustrated. Key components of the revision strategy

including the design and the use of a multi-dimensional Ruler and Block System for the measurement and balancing of the alphabetical stretches in terms of number of pages per alphabetical category was also highlighted. In addition, the importance of the relationship existing between frequencies and dictionaries was also emphasized. It is also important to state why the focus is on the high frequencies and not on the low frequencies. For example, for bilinguals, given the restriction on the number of pages top frequencies can hardly be covered and that our tests have proven that the top 10, 000 frequencies cover more than 95% of Setswana texts. As for monolingual we try to include lower frequencies of special relevance such as cultural terms but given the users from small children to adults we cannot dare to omit the top frequencies. In this chapter, we have also provided a perspective on how the South African Bantu language lexicography reflects a complex interplay of lemmatisation traditions, lemmatisation strategies and lemmatisation approaches.

The importance of tonal indication in Setswana where tones are used to make both semantic and grammatical distinctions was illustrated by means of suitable examples taken from the two monolingual Setswana dictionaries i.e. THAN and THAND. Problems regarding the lemmatization of the noun prefixes *le-* and *lo-*, the absence of nominal derivations and the absence of the treatment of homonyms in the Setswana dictionaries were highlighted and critically analysed. The evaluation of dialect words as lemmas in Setswana dictionaries was demonstrated and critically analysed to determine whether the Setswana is standardised or not.

It is also important to note that the dictionary situation in Setswana is such that we do not have the luxury of compiling dictionaries for narrowly defined target users, separate dictionaries for productive and receptive use or dictionaries for the different dialects. To date the Setswana lexicographer was forced to compile general dictionaries for use by everyone and to include the most relevant dialectical forms. On the question, do Setswana dictionaries need a descriptive dictionary or a normative one, one could say in principle descriptive but also normative aspects since the language is not fully standardized.