

Chapter 5

The microstructure

5.1 Introduction

This chapter gives an explanation of how corpora can be seen as a key to writing better Setswana dictionary articles on the microstructural level. In this chapter, we will discuss three major issues. Firstly, we will cover the importance of corpora in the following areas; as sense distinctions, as a key to writing better dictionary articles, as an aid to retrieve typical collocations and corpora as an aid to select typical and natural examples. Secondly, we will highlight certain microstructural inconsistencies relating to the treatment of verbs and ‘the so-called’ Setswana synonyms. Lastly, the treatment of the Setswana months in the currently available Setswana dictionaries will be critically analysed and evaluated against the background information of the English and the Afrikaans dictionaries. Each section will conclude with suggestions for the improvement of the respective Setswana dictionaries by means of a corpus-based microstructure.

It is argued that, if African-language lexicography is to take its rightful place in the new millennium, the active use of corpora to improve the quality of microstructural elements in the treatment of lemma signs should become an absolute priority. (cf. De Schryver and Prinsloo (2000a and 2000b) and Prinsloo and De Schryver (1999 and 2001). Corpora provide useful evidence of the formal usage of the lexical items, i.e. the associated syntactic structures, phraseological patterns, collocations, contexts of use, etc.

According to Galley (2000:132), the microstructure should include a diverse mass of data, for example; cross-references, paraphrase of meaning, examples, parts of speech, typographical exposition, to mention but a few. The basic aim of the lexicographer is to guide the user in respect of the properties/features, characteristics, use and meaning of a lemma sign. Laufer, (1992:71) formulates this basic aim as follows:

“Knowing a word would ideally imply familiarity with all its properties [...] When a person ‘knows’ a word, he/she knows the following: the word’s pronunciation, its spelling, its morphological components if any, the words that are morphologically related to it, the word’s syntactic behaviour in a sentence, the full range of the word’s meaning, the appropriate situations for using the word, its collocation restrictions, its distribution and the relation between the word and other words within a lexical set”.

The question that arises now is how the utilization of a corpus can help the lexicographer to achieve the ultimate microstructural goal. According to De Schryver and Prinsloo (2000), a large, structured, electronic corpus is the first requirement for corpus-based dictionaries as well as advanced corpus query tools. Such tools must be able to provide at least two basic outputs, namely word-frequency counts and concordance lines as well as the capacity of analysing problematic contexts. Concordance lines culled from living-language sources supplement and support the lexicographer’s (native-speaker) intuition. They take him/her to the heart of the actual usage of word(s) in context, allowing the lexicographer to see up to several dozens of contexts at a glance.

In order to illustrate this interaction between corpus queries and the compilation of a dictionary’s microstructure, the chapter will be structured as follows: First, a brief introduction to corpus queries as an aid to sense distinctions is given with reference to the Setswana homonyms such as, *tshela* and *thari*. This is followed by detailed analysis of corpus lines in combination with frequency counts for the so-called

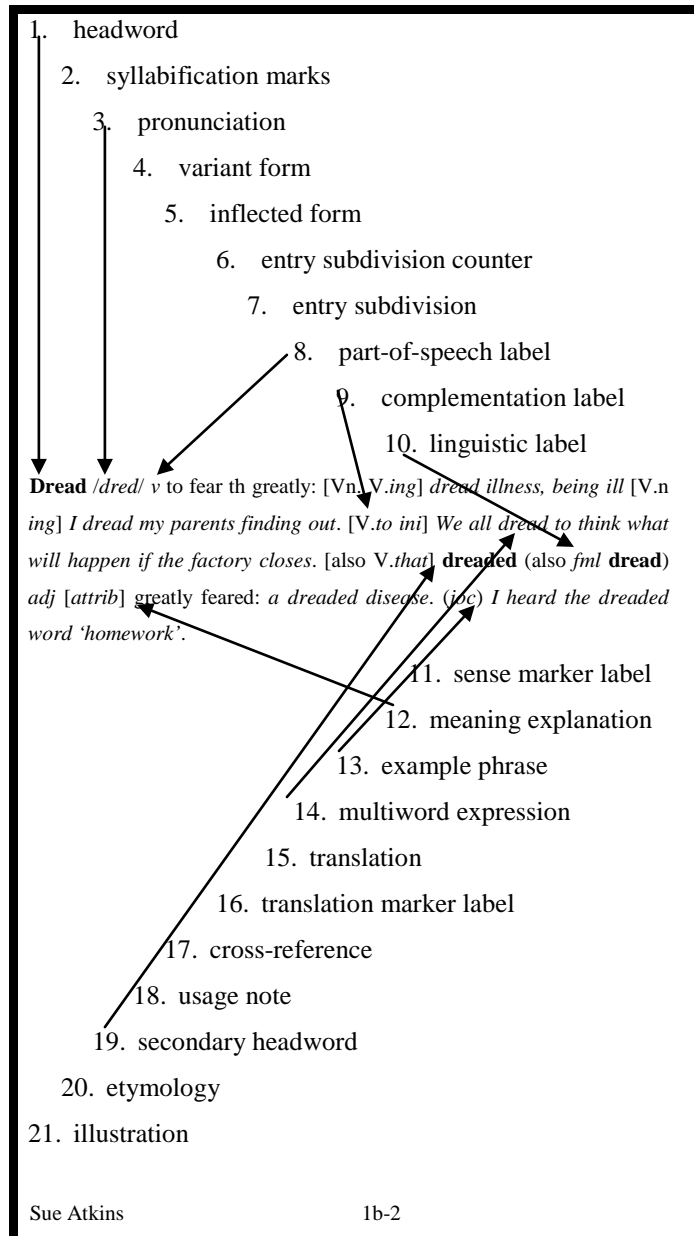
synonyms such as *batla* (to look for) and *senka* (to look for). Finally, inconsistencies of application in the treatment of verbs and the Setswana months within the microstructure of the currently available Setswana dictionaries will be highlighted.

5.2 Inconsistencies of application within the microstructure of the Setswana dictionaries

On the microstructural level, comment on semantics is the most important data type. Gouws (1983:113) states that it is the information type most generally consulted by the target users, most substantial and considered as the central component of the article. A number of important data entries have not been treated satisfactorily in Setswana dictionaries, for example; definitions, translations, sense markers, etymology, to mention but a few. The lexicographer has to decide on a selection of entries to treat in the microstructure of the dictionary.

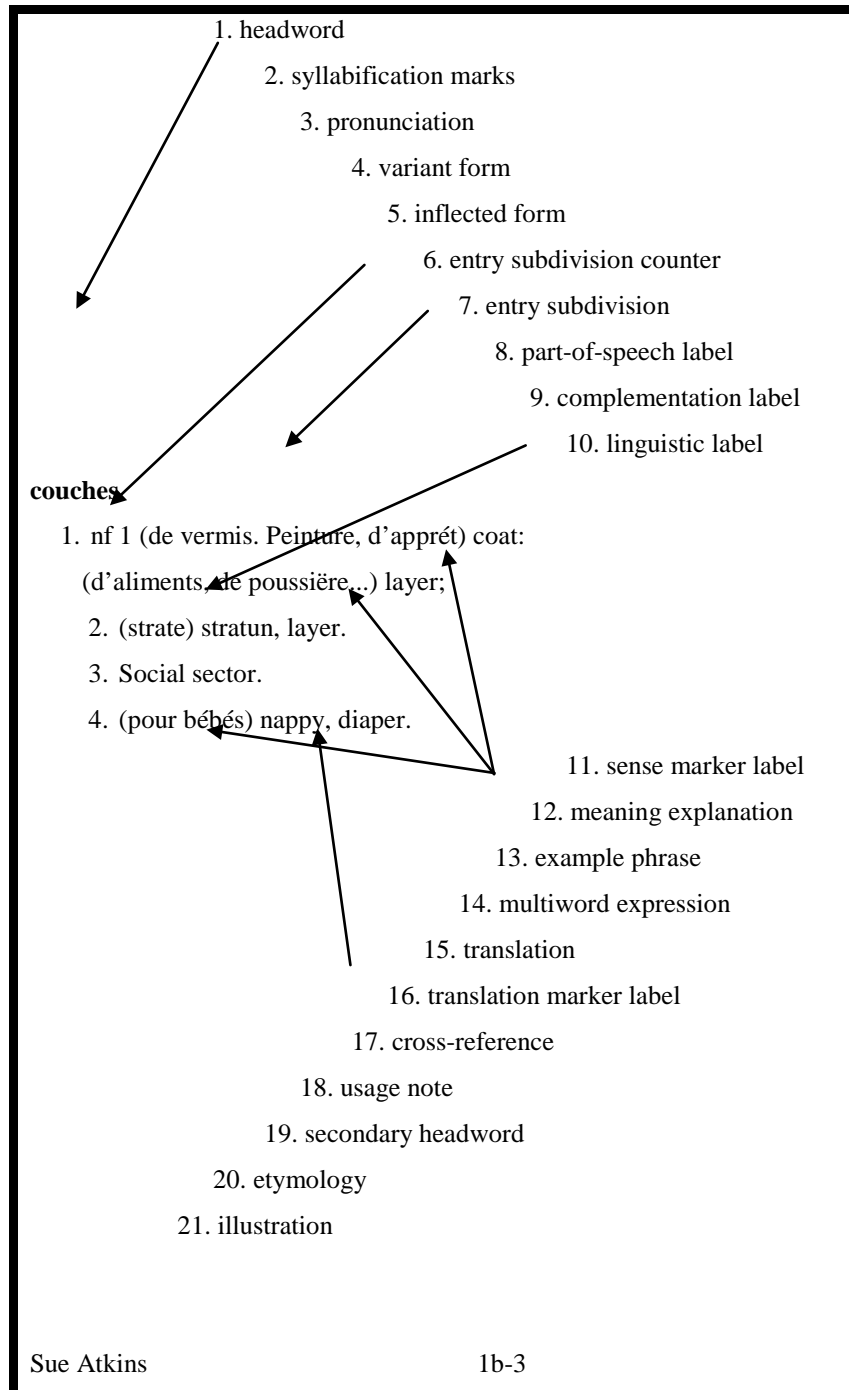
Atkins et al. 1997 give a schematic presentation of typical data types given in comprehensive monolingual and bilingual dictionaries, cf. Tables 27 and 28.

Table 27: Example of data types in a comprehensive monolingual dictionary



(Monolingual dictionary data type: Atkins et al. 1997)

Table 28: Example of data types in a comprehensive bilingual dictionary



(Bilingual dictionary data type: Atkins et al. 1997)

5.3 Corpora as a key to writing better dictionary articles

5.3.1 Corpora as an aid to sense distinctions

According to Prinsloo and De Schryver (2004:4), the lexicographer is always in doubt whether he or she has covered all the relevant senses of a lemma sign in the definition (also called paraphrase of meaning) or translation equivalent paradigm. In terms of Prinsloo and Gouws (1996:43), corpus lines will assist the lexicographer in respect of sense distinction, deciding on translation equivalents, retrieval of typical collocations, pinpointing frequent clusters and the selection of representative, authentic examples to be included in the dictionary.

Prinsloo and De Schryver further state that the lexicographer should be cautious not to regard each corpus line as a different sense but rather learn to ‘see the senses emerge’ from a digestible number of corpus lines studied. It is normally also not the intention to study thousands of corpus lines for each lemma but rather to look at a few hundred lines sorted in this sensible ways, e.g. on the word preceding/following the lemma. These lines help the lexicographer to distinguish various senses of the word. The chances of a dictionary compiler gathering all senses and sub-senses on the basis of intuition are zero. Consider the corpus lines in Table 29 which is an extract from the South African Setswana corpus.

Table 29: Concordance lines for *thari* (210) in the South African Setswana corpus

| | | |
|---|--------------|------------------------------------|
| Mateo wee! maitseo a ile kae? Bana ba | <i>thari</i> | e ntsho, Bana na mmlala wa sebilu, |
| ngwe- nyana a mmolelela gore gatwe | <i>thari</i> | ya ngwana e kwa ga-bona-mogolo |
| go feta baabo abo a ilwa o kabo belege | <i>thari</i> | wa itlhoboga aw ikela le naga ja |
| Morongwenyana a raya Morongwe a re: | <i>thari</i> | ya ngwana e kae?” Morongwe a |
| | | gama |
| Bana ba Afrerika ka bopara, Bana ba | <i>thari</i> | e ntsho dinatla. Ke bone maloba |
| sadi ba baswa le ba ba santseng ba bona | <i>thari</i> | Mo moletlong wa lenyalo kgotsa |

| | | |
|--|--------------|--------------------------------|
| Mokgalajwe le batho ba Madibe ba fitlha | <i>thari</i> | kwa phitlhong mo- rago ga go t |
| Shole o ntete. Bobedi re ke re supe | <i>thari</i> | mmogo, Re supe fa kgole e e le |
| g, gonne ke fano Modimo o mo timile | <i>thari</i> | Jaanong ...” “A re tlogeleng |

The following senses clearly emerge from the concordance lines listed in Table 29:

1. (late coming)

*Goroga ka nako mo tirong. Ga ke a tla **thari**. A gakologelwa gape....*

(Arrive on time at work. I am not late. He remembered

2. (skin used to carry a baby)

*...e ke Mmabatho mmarona, sebelega bana ka **thari** mpeng. Ka yona e kete nka be
ke le thata ka tsaya **thari** ka go belega*

(She is our real mother who always looks after her children. I wish I was strong
enough to carry you with *thari*)

3. (black nation) *thari e ntsho*

*...ee Batswana a re tswaneng re le bana ba **thari** e ntsho. Modisetshaba o re boloke*

(...yes, let's unite as a **black nation**. Foreigners to care for us)

4. (giving birth)

*Modimo o ne o sa tima Motlalepule **thari**, le ene o ka bo a bua monate*

(God did not deprive Motlalepule the opportunity to bear **children**, she should also
be proud of that fact)

Table 30: Concordance lines for *tshela* (1540) in the South African Setswana corpus

| | | |
|--|---|---|
| 16) A fitlha a itulela gone. (17) A go tlhola ka a ne a se mo ntlong. Go oga le rre. Ke eletsa gore a ka be a sa a bothito a ithobe dingalo. Mmatshope a fetsa matsatsi nae, a mpotsa le gore o magalapa a morubisi; Losika lwa bo-Madiba ano a tletse dikwena tea bo- | <i>tshela</i> <i>tshela</i> <i>tshela</i> <i>tshela</i> <i>tshela</i> <i>tshela</i> <i>tshela</i> | ka tsie le dinotshe, (18) Le bofofu ntlheng ya lesaka a du gore a tle a bone maungo a lo metse, a be a a loka ka letsw jang. Maswe a diatla. Morago a -le-baloi. Ntomolele a gana nn -ke-go-garume. A tletse marara |
|--|---|---|

The following senses clearly emerge from the concordance lines listed in Table 30:

1. (survival, to live on)

A fitlha a itulela gone. A tshela ka tsie le dinotshe

(He arrived and settled there. He **survived** on locusts and bees)

2. (alive)

Olga le rre. Ke eletsa gore a ka be a sa tshela gore a tle a bone maungo

(Olga and her father. I wish he was still **alive** to witness the outcomes)

3. (cross)

Madibana a tletse dikwena tla bo-tshela- ke-go- garume. A tletse marara

(The crocodiles are lying in wait to **cross** so that they may attack)

4. (pour)

...a bothito a ithobe dingalo. Mmatshope a tshela metse, a be a a loka ka letswai

(...apply warm water to ease the pain. Mmatshope **pours** water and included salt in)

Corpora can furthermore assist the lexicographer in finding typical collocations and combinations of words as computed with WordSmith Tools.

5.3.2 Corpora as an aid to retrieve typical collocations

In this section we will illustrate how a detailed analysis of corpus lines, in combination with frequency counts at various levels, for the frequent Setswana synonyms *batla* and *senka* (to look for) enables lexicographers to enhance the quality of microstructural elements. The aim here is to find the means of getting the relationships between *batla* and *senka* in terms of various statistics generated by WordSmith Tools as shown in Tables 31 and 34 below.

5.3.2.1 Collocates of *batla* according to the South African Setswana corpus

Table 31: Corpus lines for *batla* (5170) in the South African Setswana corpus

| | | |
|---|--------------|------------------------------------|
| . Fa o pota ka fa o utlwe go twe re | <i>batla</i> | motho yo o falo- tseng materik |
| a a tlhole a tshela. o setse a sule! Ke | <i>batla</i> | go mo utlwisa seo a ntseng a se la |
| ela pelo. "Mme o raya jang? Kana ke | <i>batla</i> | go ya sekolong," a bua a hupe- |
| aana ka ena, mma. A re o bone io, mme o | <i>batla</i> | go ikgolega ka ena. Re romilwe ke |
| , 0 didimaletseng? Motswasele: Kana o | <i>batla</i> | ke bua eng? Modise: Batho ba tsa |
| ona. Yo o mpatlang, o tshwanetse go | <i>batla</i> | fa moraka wa kgomo o leng ten" |
| lhogo. (2) (c) Leubajaaka motho, le | <i>batla</i> | eng mo morafeng? (3) (d) Tiris |
| "Ao! " Maipelo a | <i>batla</i> | a tshwara sengwenyana mo sefatlheg |
| panyi, jk. 44 Tumelo Kganetso Mme o | <i>batla</i> | gore ke nne le mosa- Mme o batla g |
| i Dikwalo ke di | <i>batla</i> | a di huparetse mo legwafeng. K |
| diphatsa lwa bone. A ithaya a re o | <i>batla</i> | go sala mo polasing fa ba tsamaya. |
| Nyaa tsala, wena a re ye gae. Ke | <i>batla</i> | gore o wele makgwafo. Ke bona |
| atshe, tshwene. Motho a fosa a ba a | <i>batla</i> | a tlhoma ka nko moseja ole. Ya re |

The top ten collocates of *batla* occurring immediately to the right of *batla* in the South African Setswana corpus are shown in Table 32 namely *gore* (so that), *mme* (but), *tl*a (will), *kwa* (there), *nna* (me), *itse* (know), *bona* (they), *eng* (what) and *tsa* (for).

Table 32: Collocates of *batla* generated by WordSmith Tools

| collocates (total) | | | | | | | | | | | | | | | |
|--------------------|---------|-------|------|-------|----|----|----|-----|----|----|-----|-----|-----|----|----|
| N | WORD | TOTAL | LEFT | RIGHT | L5 | L4 | L3 | L2 | L1 | * | R1 | R2 | R3 | R4 | R5 |
| 1 | BATLA | 5170 | 66 | 65 | 23 | 19 | 17 | 6 | 1 | 39 | 8 | 0 | 18 | 17 | 22 |
| 2 | GORE | 1020 | 329 | 691 | 40 | 66 | 62 | 157 | 4 | 0 | 382 | 41 | 127 | 83 | 58 |
| 3 | MME | 398 | 156 | 242 | 21 | 27 | 20 | 88 | 0 | 0 | 7 | 63 | 26 | 62 | 84 |
| 4 | TLA | 311 | 211 | 100 | 46 | 29 | 55 | 25 | 56 | 0 | 0 | 13 | 18 | 28 | 41 |
| 5 | KWA | 302 | 110 | 192 | 45 | 31 | 30 | 4 | 0 | 0 | 19 | 28 | 76 | 33 | 36 |
| 6 | NNA | 302 | 148 | 154 | 33 | 24 | 28 | 63 | 0 | 0 | 4 | 68 | 21 | 36 | 25 |
| 7 | ITSE | 258 | 85 | 173 | 26 | 22 | 33 | 4 | 0 | 0 | 0 | 130 | 18 | 16 | 9 |
| 8 | BONA | 238 | 82 | 156 | 25 | 17 | 29 | 9 | 2 | 0 | 0 | 70 | 36 | 21 | 29 |
| 9 | ENG | 238 | 64 | 174 | 7 | 14 | 24 | 17 | 2 | 0 | 108 | 12 | 41 | 7 | 6 |
| 10 | TSA | 224 | 112 | 112 | 25 | 38 | 35 | 3 | 11 | 0 | 1 | 37 | 14 | 35 | 25 |
| 11 | FELA | 200 | 89 | 111 | 14 | 16 | 32 | 27 | 0 | 0 | 24 | 16 | 28 | 22 | 21 |
| 12 | NENG | 189 | 143 | 46 | 10 | 4 | 64 | 65 | 0 | 0 | 0 | 0 | 17 | 17 | 12 |
| 13 | TSE | 178 | 47 | 131 | 25 | 13 | 4 | 5 | 0 | 0 | 8 | 39 | 12 | 40 | 32 |
| 14 | TIRO | 177 | 18 | 159 | 8 | 3 | 4 | 3 | 0 | 0 | 126 | 5 | 11 | 12 | 5 |
| 15 | MOTHO | 149 | 63 | 86 | 4 | 20 | 4 | 35 | 0 | 0 | 35 | 13 | 11 | 8 | 19 |
| 16 | GAGWE | 134 | 81 | 53 | 17 | 15 | 19 | 30 | 0 | 0 | 0 | 0 | 25 | 11 | 17 |
| 17 | NTSE | 134 | 115 | 19 | 15 | 15 | 15 | 69 | 1 | 0 | 0 | 0 | 3 | 4 | 12 |
| 18 | DIRA | 126 | 47 | 79 | 10 | 24 | 11 | 2 | 0 | 0 | 0 | 41 | 14 | 14 | 10 |
| 19 | JAAKONG | 115 | 67 | 48 | 6 | 5 | 20 | 30 | 6 | 0 | 1 | 3 | 6 | 25 | 13 |
| 20 | BUA | 114 | 41 | 73 | 16 | 9 | 16 | 0 | 0 | 0 | 0 | 40 | 16 | 7 | 10 |
| 21 | MOSADI | 110 | 22 | 88 | 6 | 3 | 7 | 6 | 0 | 0 | 42 | 6 | 17 | 12 | 11 |
| 22 | MONNA | 109 | 49 | 60 | 10 | 11 | 15 | 13 | 0 | 0 | 21 | 11 | 9 | 8 | 11 |

According to Leech (1981:17), collocates consists of the associations of meaning in a particular environment. If one instructs the corpus query tool to calculate and list collocates of the verb *batla*, certain useful conclusions can be drawn. For example, the frequent use of *gore* (that), *kwa* (at), *eng* (what), *fela* (just), *nna* (me), *itse* (know) etc which otherwise escaped attention in dictionaries compiled on intuition.

The concordance lines in WordSmith Tools allows one to see the items that are most frequently found to the left and to the right of a search-word as reflected in Table 32. In Table 32 ‘*gore*’ (so that) collocates 386 times with *batla* in the horizon L1-R1. 4 of these collocates occur to the left of *batla*, 382 to the right. The breakdown of occurrences to the right is 382 times R1, 41 times R2, 127 times R3, 83 times R4 and 58 times R5. The second most frequent collocate of *batla* is the word *eng* (what) which collocates 110 times in the horizon L1-R1. 2 of these collocates occur to the left of *batla* and 108 to the right.

Table 33: Top ten collocates of the base *batla* that collocate immediately to the right of *batla* in the South African Setswana corpus

| Base + Collocate | Translation | Frequency |
|-----------------------------|-----------------------|-----------|
| 1. <i>O batla gore</i> | (wants to) | 382 |
| 2. <i>A batla kwa</i> | (seek there) | 19 |
| 3. <i>Ke batla eng</i> | (looking for) | 108 |
| 4. <i>Ba batla motho</i> | (look for someone) | 35 |
| 5. <i>Batla fela</i> | (just look) | 24 |
| 6. <i>Ne a batla tse</i> | (wanted to come) | 8 |
| 7. <i>Ba batla tiro</i> | (seek for job) | 126 |
| 8. <i>Batla mme</i> | (want to know) | 7 |
| 9. <i>Fa o batla mosadi</i> | (looking for a woman) | 42 |
| 10. <i>O batla monna</i> | (looking for a man) | 21 |

It is important to note that there is a recurrent pattern as far as words following *batla* are concerned as indicated in Table 33. One can see demonstratives, conjunctions and the subjectival concord *o* in *o batla go* (he wants to) with the highest frequency followed by *batla gore* (look for), *batla kwa* (look there), *batla eng* (look for what), *batla fela* (look for just) etc, with the lowest frequency. The following is a second example of the corpus lines for the collocates of *senka*.

5.3.2.2 Collocates of *senka* according to the South African Setswana corpus

Table 34: Corpus lines for *senka* (244) in the South African Setswana corpus

| | | |
|---|--------------|------------------------------------|
| e ile." Motsei a re, "A ga ba ka ba | <i>senka</i> | ngaka e nngwe gape?" Bikibiki |
| o a rona a ne a tletsetletse naga re lo | <i>senka</i> | mme re sa bone sepe. Phaladi o bil |
| ke diolo a ntse a apaapa ka dinao a | <i>senka</i> | ka tlhoafalo mo lefifing la bosigo |
| lhelela Keabetswe a ithwala a ya go | <i>senka</i> | tiro. Aitsane le fa 0 ka~ seatla m |
| ao nnyaa Ao, ga se nna! apaapa | <i>senka</i> | mo fifing ka diatla Ke ntse ke |
| Nakedi a nna a apaapa mo lefifing mme a | <i>senka</i> | mo- kgwaro. E rile fa a sena g |
| selekanyo sefe? Ba ya kae? Ba | <i>senka</i> | go le kae ? kaBla supetsa; dir |
| g e re fa boroko bo fedile ke ye go | <i>senka</i> | dikala di le pedi pele ke ya go th |
| phiri sa me, ke tla bo ke ikaletsa. A o | <i>senka</i> | gore ke ipolaye ? Ema, ke mo |
| di tswa mo mometsong. mpampetsa | <i>senka</i> | ka bonolo; itaya ka bonolo |
| lobelo lo logolo ka di ne di gopotse go | <i>senka</i> | kwa metsi a ka bonwang gone. 19 |
| kelong jwa dijo. Ga ba kitla ba re | <i>senka</i> | gona. Ba tla fisa ntlo. Re tla swe |

The following senses clearly emerge from the concordance lines listed in Table 34:

1. (seek)

... sepe ka go boela Gauteng. O ne a ya go *senka* tiro kwa Vereeniging, mme a
ithuta

(... nothing by going back to Gauteng. He went to **seek** job in Vereeniging where he
studied)

2. (find)

... atsaya lobone. O ne a apaapa mo lefifing a *senka* dikgetse tse a di alang

(... he took the lamp. He gropes in the darkness and **find** the sacks to spread out)

3. (search)

Badisa ba tswa ba ya go *senka* dinku, ba di fitlhela di eme

(The shepherds went out to **search** for the sheep, they found them waiting)

4. (want)

... *fa ba ka se ka ba nthusa; ba re ba **senka** barutintshi ba ba boitshoko*

(... if they are not going to assist me; they **want** patient teachers)

Consider the following collocates generated by the WordSmith Tools to identify the co-occurrence to the right and left of *senka*:

Table 35: Collocates of *senka* generated by the WordSmith Tools

| collocates (total) | | | | | | | | | | | | | | | |
|--------------------|--------|-------|------|-------|----|----|----|----|----|-----|----|----|----|----|----|
| N | WORD | TOTAL | LEFT | RIGHT | L5 | L4 | L3 | L2 | L1 | * | R1 | R2 | R3 | R4 | R5 |
| 1 | SENKA | 244 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 244 | 0 | 0 | 0 | 0 | 0 |
| 2 | KWA | 34 | 12 | 22 | 6 | 4 | 2 | 0 | 0 | 0 | 6 | 7 | 5 | 1 | 3 |
| 3 | TSA | 21 | 12 | 9 | 4 | 5 | 3 | 0 | 0 | 0 | 0 | 4 | 0 | 2 | 3 |
| 4 | TIRO | 19 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 1 | 0 | 0 | 1 |
| 5 | TLA | 17 | 10 | 7 | 0 | 1 | 2 | 3 | 4 | 0 | 0 | 2 | 1 | 3 | 1 |
| 6 | MME | 16 | 4 | 12 | 0 | 1 | 0 | 3 | 0 | 0 | 1 | 1 | 2 | 6 | 2 |
| 7 | TSE | 16 | 1 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 7 | 1 | 6 | 0 |
| 8 | GORE | 15 | 2 | 13 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 5 | 3 | 1 | 2 |
| 9 | SENGWE | 15 | 3 | 12 | 1 | 1 | 1 | 0 | 0 | 0 | 10 | 2 | 0 | 0 | 0 |
| 10 | NTSE | 12 | 8 | 4 | 2 | 0 | 1 | 5 | 0 | 0 | 0 | 2 | 1 | 0 | 1 |
| 11 | GAGWE | 9 | 5 | 4 | 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |
| 12 | MONGWE | 9 | 1 | 8 | 0 | 0 | 1 | 0 | 0 | 0 | 5 | 2 | 0 | 0 | 1 |
| 13 | MOTHO | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 0 |
| 14 | BOLO | 7 | 7 | 0 | 2 | 1 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | FITLHA | 7 | 4 | 3 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 0 |
| 16 | MATLHO | 7 | 2 | 5 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 2 | 2 |
| 17 | GAPE | 6 | 1 | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 3 | 0 |
| 18 | TLHOKA | 6 | 3 | 3 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 2 |
| 19 | DIRA | 5 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 20 | LEKA | 5 | 4 | 1 | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 21 | LENG | 5 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 3 |
| 22 | NGAKA | 5 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 |
| 23 | NNA | 5 | 1 | 4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |

In Table 35 item 4, *tiro* (work) collocates 17 times with *senka* in the horizon L1-R1. 17 of these occur to the right and 0 to the left. The breakdown of occurrences to the right is 17 times R1, 1 R2, 0 R3, 0 R4 and 1 R5. When item 8 *gore* (so that) is taken, it collocates 2 times with *senka* in the horizon L1-R1. 2 of these collocates occur 2 times to the right of *senka* and 0 times to the left. The low frequent use of *senka gore* is because the form is more widely used in Botswana than in South Africa.

Table 36: Top ten collocates of the base *senka* that occur immediately to the right of *senka* in the South African Setswana corpus

| Base + Collocate | Translation | Frequency |
|------------------------|------------------|-----------|
| 1. <i>Senka leng</i> | (find when) | 1 |
| 2. <i>Senka tiro</i> | (find the job) | 17 |
| 3. <i>Senka gore</i> | (find that) | 2 |
| 4. <i>Senka motho</i> | (find a person) | 4 |
| 5. <i>Senka tse</i> | (find this) | 1 |
| 6. <i>Senka sengwe</i> | (find something) | 10 |
| 7. <i>Senka mme</i> | (find a mother) | 1 |
| 8. <i>Senka kwa</i> | (find there) | 6 |
| 9. <i>Senka mongwe</i> | (find someone) | 5 |
| 10. <i>Senka ngaka</i> | (find a doctor) | 4 |

Table 37 below will now be used to contrast the presumed Setswana synonyms *batla* and *senka*. When Table 35 and Table 36 are compared using the nodes *gore* and *tiro* to see which items are most frequently found to the right of the search-word *batla* and *senka*, the word '*batla gore*' appears 691 times and *senka gore* appears 13 times. The results suggest that the Setswana users in South Africa are more likely to use *batla* than *senka*.

Table 37: Comparison between *batla* and *senka* in collocation with *gore* and *tiro*

| Word | Left | Right | Total |
|------------------------------------|------|-------|-------|
| <i>batla gore</i> (want to) | 329 | 691 | 1020 |
| <i>senka gore</i> (look for) | 2 | 13 | 15 |
| <i>batla tiro</i> (look for a job) | 18 | 159 | 177 |
| <i>senka tiro</i> (find a job) | 0 | 19 | 19 |

Compare another example which further illustrates the contrast between *batla* and *senka*:

When the first 20 collocates are compared between *batla* and *senka* certain collocates are missing for both the so-called synonyms as indicated by an ‘x’ in Table 38.

Table 38: Comparison of the top ten collocates of *batla* and *senka*

| Collocates | <i>Batla</i> | <i>Senka</i> |
|------------------|--------------|--------------|
| 1) <i>mme</i> | ✓ | ✓ |
| 2) <i>tla</i> | ✓ | ✓ |
| 3) <i>kwa</i> | ✓ | ✓ |
| 4) <i>itse</i> | ✓ | x |
| 5) <i>bona</i> | ✓ | x |
| 6) <i>eng</i> | ✓ | x |
| 7) <i>matlho</i> | x | ✓ |
| 8) <i>sengwe</i> | x | ✓ |
| 9) <i>dira</i> | x | ✓ |
| 10) <i>nna</i> | x | ✓ |

5.3.3 Corpora as an aid to pinpoint clusters

Clusters are words which are found repeatedly in each other's company. There is a recurrent pattern for the words following *batla* and *senka*. In this case the lexicographer can instruct the corpus query tool to calculate and list the frequent clusters. Consider the following:

Table 39: Two-word clusters with *batla* (451) in the initial position in South African Setswana corpus

| 5170 Cluster | | Frequency | Percentage |
|--------------|-------------|-----------|------------|
| <i>batla</i> | <i>gore</i> | 1020 | 19.72 % |
| <i>batla</i> | <i>mme</i> | 398 | 7.69 % |
| <i>batla</i> | <i>tla</i> | 311 | 6.015 % |
| <i>batla</i> | <i>kwa</i> | 302 | 5.84 % |
| <i>batla</i> | <i>nna</i> | 302 | 5.84 % |
| <i>batla</i> | <i>itse</i> | 258 | 4.99 % |
| <i>batla</i> | <i>bona</i> | 238 | 4.60 % |
| <i>batla</i> | <i>eng?</i> | 238 | 4.60 % |
| <i>batla</i> | <i>tsa</i> | 224 | 4.33 % |

Going down one level in respect of words in combination with *batla* as indicated in Table 39, *batla* is followed 1020 times by *gore*, 19.72% of all occurrences of *batla* and by *mme* 398, 7.69% of all occurrences of *batla* and a variety of other clusters ranging from 6,015% to 4, 33%. This means that *batla gore* should definitely be considered for inclusion in the article of *batla* in any Setswana dictionary, which is not the case in existing dictionaries.

Table 40: Three word clusters with *batla* (451) in the initial position in the South African Setswana corpus

| 451 Cluster | Frequency | Percentage |
|----------------------|-----------|------------|
| <i>a batla go</i> | 422 | 93.6 % |
| <i>ke batla go</i> | 406 | 90.02 % |
| <i>ya go batla</i> | 163 | 36.14 % |
| <i>ke batla gore</i> | 161 | 35.7 % |
| <i>ne a batla</i> | 154 | 34.15 % |
| <i>ba batla go</i> | 141 | 31.26 % |
| <i>batla go itse</i> | 134 | 29.71 % |
| <i>o batla gore</i> | 98 | 21.72 % |

From Table 40 moving downwards with words used in combination from left to right, one can see that *batla* appears 422 times between a ~ go and 406 times between ke ~ go. It is important to note that *batla go* preceded by *a* counts 422 or 93,6% out of the possible 451 occurrences of *batla* and *batla go* preceded by *ke* counts 406 or 90,02% of the possible 451 of the occurrences of *batla* with the rest ranging from 6,14% to 21,72%. This means that inclusion of the subject concord *ke* and or *a* as well as the infinitive prefix *go* in the article of *batla* is highly recommended.

It is therefore important for lexicographers to examine synonyms thoroughly before giving the translation equivalents and definitions including cross-reference entries such as *batla* and *senka*.

It is clear that the examples in concordance lines in Table 33 and Table 36 bring about the difference between *batla* and *senka*. True synonyms are rare in Setswana. It is a matter of dialectical preference. For example, *senka* is preferred by the Southern Setswana sub-group, *Setlhaping* and *Setlharo* dialects in the Taung, Vryburg and Kuruman district. *Batla* is preferred by the Eastern Setswana which comprises of the *Sekwena* and *Sekgatla* sub-dialects spoken around Pretoria.

In the next section we will highlight problems relating to the treatment of verbs, polysemy and synonyms in the existing Setswana dictionaries and the Setswana months will be critically analyzed against the background information of the English and the Afrikaans dictionaries.

5.3.4 Corpora as an aid to select typical and natural examples

In this section we will look into the huge potential of combining different corpus query tools, with special reference to the selection of excellent typical and natural examples. According to Fox (1987:138), the terms typical and natural examples can be defined as follows:

“Our first and foremost requirement for examples is typicality: that they should show the way in which people actually use the word they are exemplifying. [...] naturalness [...] is the well-formedness of sentences not in isolation but in text.”

Laufer (1992:72) also stated that lexicographers who are educated native speakers of the language are bound to have correct intuitions about their mother tongue, about the grammaticality of the word, its typical use and its typical environment. These intuitions are necessarily less correct than intuitions of those language users who are represented in the corpus and are therefore not less reliable.

The lexicographer can thus combine the output of different good query tools such as word-frequency counts and concordance line screens. For example,

Table 41: Collocates of the base *tshwanetse* (with horizons L5-R5) in the South African Setswana corpus

| N | WORD | TOTAL | LEFT | RIGHT | L5 | L4 | L3 | L2 | L1 | * | R1 | R2 | R3 | R4 | R5 |
|----|------------|-------|------|-------|----|----|----|-----|----|----|-----|-----|-----|----|----|
| 1 | TSHWANETSE | 3689 | 17 | 23 | 8 | 6 | 0 | 2 | 1 | 49 | 1 | 0 | 4 | 7 | 11 |
| 2 | GORE | 786 | 409 | 377 | 48 | 52 | 59 | 249 | 1 | 0 | 102 | 13 | 147 | 74 | 41 |
| 3 | NNA | 326 | 50 | 276 | 11 | 15 | 6 | 18 | 0 | 0 | 2 | 234 | 13 | 14 | 13 |
| 4 | ITSE | 219 | 103 | 116 | 15 | 19 | 66 | 3 | 0 | 0 | 0 | 58 | 20 | 31 | 7 |
| 5 | KWA | 193 | 65 | 128 | 15 | 16 | 33 | 1 | 0 | 0 | 1 | 2 | 56 | 37 | 32 |
| 6 | MME | 191 | 131 | 60 | 20 | 23 | 24 | 64 | 0 | 0 | 0 | 1 | 14 | 17 | 28 |
| 7 | FELA | 183 | 123 | 60 | 14 | 13 | 13 | 83 | 0 | 0 | 0 | 9 | 20 | 9 | 22 |
| 8 | TSA | 181 | 90 | 91 | 33 | 7 | 50 | 0 | 0 | 0 | 11 | 4 | 6 | 43 | 27 |
| 9 | TSE | 178 | 82 | 96 | 12 | 42 | 6 | 22 | 0 | 0 | 0 | 1 | 11 | 41 | 43 |
| 10 | BONA | 155 | 105 | 50 | 17 | 14 | 45 | 29 | 0 | 0 | 0 | 23 | 4 | 11 | 12 |
| 11 | SENGWE | 136 | 67 | 69 | 3 | 34 | 3 | 27 | 0 | 0 | 0 | 0 | 44 | 5 | 20 |
| 12 | DIRA | 122 | 26 | 96 | 7 | 11 | 5 | 3 | 0 | 0 | 0 | 83 | 6 | 4 | 3 |
| 13 | GAGWE | 111 | 75 | 36 | 7 | 19 | 13 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 36 |
| 14 | MONGWE | 97 | 75 | 22 | 11 | 34 | 4 | 26 | 0 | 0 | 0 | 1 | 4 | 4 | 13 |
| 15 | NENG | 97 | 95 | 2 | 3 | 0 | 1 | 91 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 16 | PELE | 97 | 26 | 71 | 11 | 4 | 5 | 6 | 0 | 0 | 2 | 1 | 23 | 33 | 12 |
| 17 | JWA | 91 | 59 | 32 | 15 | 29 | 15 | 0 | 0 | 0 | 1 | 0 | 2 | 18 | 11 |
| 18 | TENG | 87 | 34 | 53 | 8 | 4 | 6 | 16 | 0 | 0 | 0 | 0 | 11 | 23 | 19 |
| 19 | MOTHO | 84 | 62 | 22 | 13 | 11 | 4 | 34 | 0 | 0 | 0 | 0 | 5 | 10 | 7 |
| 20 | JALO | 83 | 65 | 18 | 13 | 10 | 14 | 27 | 1 | 0 | 0 | 0 | 7 | 5 | 6 |
| 21 | BANA | 80 | 42 | 38 | 5 | 13 | 5 | 19 | 0 | 0 | 4 | 2 | 16 | 13 | 3 |
| 22 | THATA | 80 | 42 | 38 | 15 | 9 | 6 | 12 | 0 | 0 | 0 | 0 | 19 | 5 | 14 |
| 23 | JAAKA | 78 | 43 | 35 | 10 | 3 | 17 | 13 | 0 | 0 | 0 | 0 | 17 | 12 | 6 |
| 24 | JAANONG | 76 | 52 | 24 | 1 | 7 | 6 | 37 | 1 | 0 | 0 | 0 | 13 | 7 | 4 |
| 25 | GAGO | 74 | 53 | 21 | 4 | 13 | 4 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 21 |

In Table 41, a selection of the collocates of the base *tshwanetse* with the horizon L5-R5 is listed. In Table 41 items *gore* (so that) collocates 103 times with *tshwanetse* in the horizon of L1-R1. 1 of these collocates occur on the left of *tshwanetse*, 102 to the right. The second most frequent collocate of *tshwanetse* is *nna* (me) which collocates 252 times in the horizon L2-R2. 18 of these collocates occur on the left of *tshwanetse* and 234 to the right.

Table 42: Three-word clusters with *tshwanetse* (909) in the South African Setswana corpus

| N | cluster | Freq. |
|----|--------------------------|-------|
| 1 | <i>o tshwanetse go</i> | 759 |
| 2 | <i>ba tshwanetse go</i> | 359 |
| 3 | <i>re tshwanetse go</i> | 290 |
| 4 | <i>a tshwanetse go</i> | 284 |
| 5 | <i>ke tshwanetse go</i> | 226 |
| 6 | <i>e tshwanetse go</i> | 215 |
| 7 | <i>tshwanetse go nna</i> | 203 |
| 8 | <i>o tshwanetse wa</i> | 147 |
| 9 | <i>di tshwanetse go</i> | 125 |
| 10 | <i>ne a tshwanetse</i> | 121 |
| 11 | <i>gore o tshwanetse</i> | 114 |
| 12 | <i>re tshwanetse ra</i> | 112 |
| 13 | <i>o tshwanetse a</i> | 110 |
| 14 | <i>tshwanetse wa bo</i> | 106 |
| 15 | <i>o ne a</i> | 102 |

From Table 42 above we see that the most frequent three-word cluster with *tshwanetse* in the South African Setswana corpus is *o tshwanetse go* (he/she is suppose to). The second most frequent cluster is *ba tshwanetse go* (they are suppose to); followed by *re tshwanetse go* (we are suppose to).

Given all these available corpus data, it is now very easy for the lexicographer to select a typical and natural example of usage for inclusion into a dictionary by simply glancing at the output of one or more concordance-line screens.

5.4 Problems related to the treatment of the Setswana verbs

Verbs in Setswana change their forms in order to express, or help express, different perspectives for viewing an action or state, such as the time an event happened, how long it lasted, and the number of different semantic connotations as given for the verb *reka* in Table 22. The base-form focuses on the meaning of a lexical verb without considering its derivations. These verb forms make available some important differences of meanings as reflected in example 35. Consider the following derived forms of the verb *dira* (work, do).

Example 35

| | | | | |
|------------------------------|---------------------|---|-----------------------|------------------------------|
| Applied verb form, | <i>-tsa</i> | > | <i>e.g. diragatsa</i> | (do something for) |
| Causative verb form, | <i>-isa, -ya</i> | > | <i>e.g. dirisa</i> | (cause/let something happen) |
| Neuter verb form, | <i>-ela</i> | > | <i>e.g. direla</i> | (work for) |
| Passive verb form, | <i>-wa, iwa</i> | > | <i>e.g. dirwa</i> | (be doing something) |
| Perfective verb form, | <i>-ile</i> | > | <i>e.g. dirile</i> | (done) |
| Reciprocal verb form, | <i>-na</i> | > | <i>e.g. dirisana</i> | (work together) |
| Reversible verb form, | <i>-ola, -olola</i> | > | <i>e.g. dirolola</i> | (to undo) |

Neither of these forms are clearly defined in the two monolingual Setswana dictionaries as reflected below in the THAN and the THAND. Examples 36 and 37 below are now used to illustrate problems related to the treatment of the verb '*dira*' and its derivatives.

Example 36: THAN

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

Monolingual dictionaries deal with defining equivalents. However, this is not the case with the THAN as indicated in Example 36 above. The dictionary is filled up by morphological and grammatical information. No paraphrase of meaning and examples usage are given. It is important for the lexicographer to pay attention to the subject of explaining and not to giving the morphological form only. Finding the meaning of a word is the primary aim of the Setswana dictionary use.

Example 37: THAND

| |
|--|
| dira tshwara ka diatla go ithusa; baba, dilô tse di tshabêgang Diphôlôgôlô tse di re bolayang ke dira. Êné le nna re dirisantsê thata. Go dirile Modimo. |
|--|

Example 37 defines only two senses i.e. ‘enemies’ referring to both wild animals and people and ‘using of hands to work’ while other senses are excluded. Compare in this regard the treatment of the verb *dira* in the following bilingual dictionaries:

Example 38: SED

Dira, n., pl. of *sera*, A hostile army ; enemies ; war. *Go epela dira*, to make war against ; v.t., pft. *dirile*, work ; do. Same as *diha*.
Dirai, n., pl. of *serai*, Traps.
Dirala, n., pl. of *serala*, Platforms : places for stacking corn, before it is threshed.
vi., pft. *diretse*, happen ; come to pass.

The compilers in example 38 succeeded in giving the translation equivalents of *dira* as enemies, work and do, but often find themselves giving translation equivalents which do not conform to the meaning of the original which often mislead the users. For example, *go epela dira* (to make war against) and pft *dirile* (done).

Example 39: MSED

dira N. CL8 *di-*, PL. OF *sera*, hostile armies; enemies.
dira V. S. SIMP., similar to *bêrêka*, work; make; do; act.
diradira V. S. REP., do repeatedly; do constantly; do a little at a time.

It is clear from the given examples 38 and 39 that more semantic guidance is given compared to the example 37.

In Table 43 below, the article should involve determining the meaning of *dira* in various senses and not only two senses as reflected in examples 37, 38 and 39. Table 43 presents an attempt to improve on typical articles for the verb *dira* and maximally use of corpus data.

Table 43: Corpus lines for *dira* (7074) in the South African Setswana corpus

| | | |
|--|-------------|---|
| tlhoka puo. Sy o o logang maano a go | <i>dira</i> | bosula o tla bidiwa Rra-bolotsa |
| ditlhong. Fa o na le tsholofelo, o | <i>dira</i> | o sa tetesele. Solofela gone ga O |
| O kwena jaaka Banotwa, o ile a ya go | <i>dira</i> | legae ja gagwe Mochudi. Lefatshe ja |
| ladi, monna yo. O nang le maatla go | <i>dira</i> | sengwe le sengwe mo Modimolle. Moth |
| Kana motho yo o latelang kgosi O ka | <i>dira</i> | dira eng se se iseng se dirwe?i '3Ka |
| duele ona molao o o re kganelang go | <i>dira</i> | O a re tshwara. Bosigo o tshw |
| a, o mpolelele, ausi .1 Nka tloga ka go | <i>dira</i> | dilo!" Mooki a tsamaya ka ntlha y |
| ... ;-! ~ 1. (a) Lebota le | <i>dira</i> | eng~ (b) Dit~hare di tla nna mo |
| lhagisiwa: 37 Tumelo Kganetso Ke a | <i>dira</i> | ka jalo ke nna le madi Ga ke dire, |
| e ka mafoko a gago. 4. Thakadu o ne a | <i>dira</i> | dira kae? 5. Maikaelelo a ga , mmagwe M |
| dira, anetse a tlosiwa. Re tshwanetse ra | <i>dira</i> | gore Kgomo a mo kobe mo motseng. Re |
| Jaanong Dafita a ba amogela a ba | <i>dira</i> | ditlhogo tsa masomo a batlhaba |

From 43 the following senses emerge:

1. (enemies)

Mantsho ga tlhwe a utlwa sentle. **Dira** tsa rona di mo boitumelong

(Mantsho is not yet aware. Our **enemies** are rejoicing)

2. (work)

...a **dira** teng mme a tshwanetse go ya go **dira** kwa Taung. Pele a tswa ka goro

(he works there, though he should go and **work** in Taung. Before he left)

3. (make)

*...a tse ba di utlwileng. Batho ba **dira** leratla fa ba bona Moatlhodi a ema*

(what they have heard. People **make** noise when the judge stood up)

4. (do)

*Fa Ramasedi a rata, a ka e **dira** letsatsi le penne*

(when the almighty God wants, he may **do** it during sun shine)

5.5 The treatment of the Setswana months

The Setswana months are not satisfactorily treated in Setswana monolingual dictionaries. The English and Afrikaans dictionaries will be used to compare the treatment of the Setswana months. Many shortcomings exist as far as the presentation of information is concerned. Entries state only the names of the months in their chronological order, and thus no justice has been done to bring about the historical and cultural aspects. These months are defined inadequately, thus they provide no meaning as illustrated in example 42 below. Two Setswana monolingual dictionaries are compared to English and Afrikaans dictionaries.

Example 40: THAN and THAND

| No | THAN | THAND |
|----|---|---|
| 3 | <p><i>Mopitlwe. T.G.T. / mopitlo In / la. bo -. kgwedi ya boraro ya ngwaga.</i></p> <p>March. The third month of the year</p> | <p>3. Mopitlwe kgwedi ya boraro ya ngwaga.</p> <p>March. The third month of the year</p> |
| 4 | <p><i>Moranang T.T.T.T. In / la bo – kgwedi ya botlhano ya ngwaga.</i></p> <p>April. The fourth month of the year</p> | <p>4. Moranang kgwedi ya bone ya ngwaga.</p> <p>April. The fourth month of the year</p> |
| 5 | <p><i>Motsheganong T.T.T.T. In /la. bo – kgwedi ya botlhano ya ngwaga.</i></p> <p>May. The fifth month of the year</p> | <p>5. Motsheganong kgwedi ya botlhano ya ngwaga e mariga a simologang ka yana.</p> <p>May. The fifth month of the year, the beginning of winter.</p> |

It is unfortunate that the two Setswana monolingual dictionaries have not given the cultural meanings of these lemmas i.e. *Mopitlwe, ngwana wa motswana o ja gore a pipitlelwe ka ntlha fa dijo ele ntletse ntletse.* (The Motswana child eats a lot in such a way that he/she constipate, because there is too much food available).

Aspects of culture are also employed to refer to the aesthetic or intellectual quality of a particular language's art, literature and institution. Many of the Setswana concepts with which we operate are culturally bounded, in the sense that their understanding depends upon socially transmitted knowledge. Compare the article (May month) taken from the Oxford English dictionary (OED) and the Oxford School Dictionary (OSD) where copilers make attempts to cover few examples in bringing about definitions and meaning descriptions.

Example 41: OED

May /meɪ/ *n.* **1** the fifth month of the year. **2** (**may**) the hawthorn or its blossom. **3** *poet.* bloom, prime. □ **may-apple** an American herbaceous plant, *Podophyllum peltatum*, bearing a yellow egg-shaped fruit in May. **May-bug** = COCKCHAFFER. **May queen** a girl chosen to preside over celebrations on May Day. **Queen of the May** = *May queen*. [ME *f.* OF *mai* *f.* L *Maius* (*mensis*) (month) of the goddess Maia (see MAIA²), who was worshipped in this month]

may /meɪ/ *v. aux.* (3rd *sing. present* **may**; *past* **might** /maɪt/) **1** (often foll. by *well* for emphasis) expressing possibility (*it may be true*; *I may have been wrong*; *you may well lose your way*). **2** expressing permission (*you may not go*; *may I come in?*). ¶ Both *can* and *may* are used to express permission; in more formal contexts *may* is usual since *can* also denotes capability (*can I move?* = *am I physically able to move?*; *may I move?* = *am I allowed to move?*). **3** expressing a wish (*may he live to regret it*). **4** expressing uncertainty or irony in questions (*who may you be?*; *who are you, may I ask?*). **5** in *purpose* clauses and after *wish*, *fear*, etc. (*take such measures as may avert disaster*; *hope he may succeed*). □ **be that as it may** regardless of whether or not that is so. **may as well** = *might as well* (see MIGHT¹). **that is as may be** that may or may not be so (implying that there are other factors). [OE *mæg* *f.* Gmc, rel. to MAIN¹, MIGHT²]

Example 42: OSD

May the fifth month of the year. □ **May Day** 1 May, kept as a festival with dancing or as an international holiday in honour of workers. **May queen** a girl chosen to preside over festivities on May Day.

may¹ *auxiliary verb* (see also MIGHT²) expressing possibility (*it may be true*), permission (*you may go*), wish (*long may she reign*), or uncertainty (*whoever it may be*).

may² *noun* hawthorn blossom.

Maya (*məˈjɑː*) *noun* **1** (plural **Maya** or

It is important to note that the oral corpus component has brought into light the information that would enrich our dictionary definitions and meaning descriptions.

Consider Table 44 on the Setswana months generated from the South African oral Setswana projects in comparison to the corpus lines generated from the South African written corpus in Table 45.

Table 44: The South African oral Setswana corpus

Project 4 Dikgwedi tsa Setswana

a jewa. Pula yona e tshologa matsorotsoro, re be re re kgwedi ke Tlhakole. Kgwedi ya borara Mopitlwe. Mopitlwe ke kgwedi e e twelang mo setlheng sa letlhafula, kwa masimong, dijalo di kgona go ka jewa. Dijalo di tshwana le merogo, mmidi o motala le ntshwe ya bo e le ntletse-ntletse. Ngwana wa motswana o tla ja mpa go pipitlelwa, ya nna gore kgwedi Mopitlwe, mpa e pipitlelwa ke dijo. Go jewa legwetla, letsatsi letlhaba phakelanyana mme malatsi a maleelenyana. Letsatsi le tlhaba phakelanyana mme masigo a makhutswane, dimela di thunya ka bontsi. Tlhaga e semolola go bona botala jo boša. Mefuta e mentsi ya dimela e a jalwa, kgwedi ya nna Mopitlwe. Kgwedi ya bone Mloranang. Mloranang a dijalo a-nama, e le go tlhalosa kgotsa go bontsha gore go gola ga dijalo go a nama. Go nama go, go tlhalosa go iketla, go gola. Kana motswana fa a ntse fa fatshe a namile maoto, o bontsha gore o iketlile, pula yone e a fokotsega, e ya kwa botelong.

Go tlhagelela mowa o o tsididi, o o kgaolang setlha se. Kgwedi ya botlhano, Mlotsheganong, ke kgwedi e e simololang setlha sa mariga. Dipula di a khutlha, phefo ya borwa e foka ka matla. Ditlhare di foforega matlhare, bojang bo a setlhefala, didiba di fokotsega metsi, metsi a swa dikgapetla, dijalo di a omelela kwa masimong. Binonyane ga di kgone go ka ja dijalo kwa masimong, motswana a be a re, dijalo di tshaga nong, ka jalo di palelwa ke go eja, kgwedi ya nna Mlotsheganong.

Table 44 can now be interpreted as follows:

During this time of the second month of the year, it is heavingly raining, and then the Batswana said: *e tlhakola mogote wa letsatsi*. They enjoy autumn season and the sun rises early and the days are longer and the nights are shorter. The plants and grass are becoming green. Most of the cultivation is taking place during the month of March. As a result the Batswana children are eating over to such an extent that they got constipated. During the month of April, most of the plants are showing their colours. This means that most of the plants are growing well. The Batswana people are relaxing and enjoying and the time of rain is shortening at the end. Then came the cold winds which are separating autumn and winter. The fifth month which is May is the month that starts the winter season. There is no rain; the winds from the south are blowing strong. The trees shade their leaves; the grass becomes greyish, the wells become dry, the water freezes, the plants become dry at the farms. Then the Batswana person says the plants are laughing at the birds and the birds do not eat those plants, now that is the May month.

Table 45: Concordance lines taken from the South African written Setswana corpus

| MOTSHEGANONG: 38 entries (sort: 5L,5L) | | | | | |
|---|---|------------|--------------------|-------------|----------|
| N | Concordance | Set | TagWord No. | File | % |
| 1 | mo) 2. 3. Mopitlo 4. 5. Motsheganong 6. 7. Ph | | 1,120 | us~1.txt | 9 |
| 2 | A re tswelatseng! a. Motsheganong yone, ke | | 1,058 | walo2.txt | 26 |
| 3 | A re tswelatseng! a. Motsheganong yone, ke | | 2,931 | walo2.txt | 75 |
| 4 | 96, Port Elizabeth. 10 Motsheganong, 1965. K | | 23,152 | tlha~1.txt | 91 |
| 5 | Port Elizabeth. 10 Motsheganong, 1965. | | 22,583 | tlha~2.txt | 90 |
| 6 | e hard sorghum grains Motsheganong Motshe | | 31,411 | csstt~1.txt | 58 |
| 7 | lepong ka kgwedi ya Motsheganong. E tlhaga | | 43,536 | ojimst.txt | 65 |
| 8 | a dijalo go di segaka. Motsheganong o tlaa go | | 3,770 | lofe~1.txt | 43 |
| 9 | go tloga ka kgwedi ya Motsheganong go fitlha | | 5,678 | wemst.txt | 15 |
| 10 | nong Motsheganong Motsheganong motshe | | 31,413 | csstt~1.txt | 58 |
| 11 | e di welang. Maloba fa Motsheganong e thulam | | 22,432 | om~1.txt | 48 |
| 12 | ga mang? Botsenwa! Motsheganong a tihola | | 11,267 | gwa~1.txt | 96 |
| 13 | joo. Mo kgwedding ya Motsheganong Mafoko | | 9,493 | dile~1.txt | 48 |
| 14 | otshediswamelelwane Motsheganong birds are | | 31,402 | csstt~1.txt | 58 |
| 15 | edi o ntlele ka botshelo! Motsheganong ke ole o | | 1,985 | oko~1.txt | 38 |
| 16 | a kgwedi, ka la di 31 Motsheganong, Khwela | | 14,125 | tsw~2.txt | 24 |
| 17 | a kgwedi, ka la di 31 Motsheganong, Khwela | | 15,098 | tsw~1.txt | 27 |
| 18 | phirima ka kgwedi ya Motsheganong. sele | | 41,912 | anodi.txt | 78 |
| 19 | mabele ka kgwedi ya Motsheganong." Mah | | 23,711 | swabil.txt | 65 |
| 20 | - .; nwe ke bonno. Motsheganong: Ke kgw | | 7,635 | ana~1.txt | 96 |
| 21 | habula, Fa a omelela Motsheganong a fetoga. | | 4,277 | tsa~1.txt | 41 |
| 22 | M~tseno a ~ago _Motsheganong, Phuk | | 9,083 | tsa~1.txt | 88 |
| 23 | gopolo e e phoso Mei Motsheganong meiyar | | 25,057 | csstt~1.txt | 46 |
| 24 | mologong ya kgwedi ya Motsheganong ka Lama | | 18,961 | wemst.txt | 49 |
| 25 | beilweng botsetse. Motsheganong kgwedi y | | 32,421 | anodi.txt | 60 |
| 26 | -wa-godimo motshedi Motsheganong motshe | | 19,717 | tcotel.txt | 61 |
| 27 | l. E ne e le kgwedi ya Motsheganong. Serame | | 45,429 | ulemst.txt | 90 |
| 28 | ne ra 0 tlakaula mmidi Motsheganong - Ra s | | 6,736 | \tsaya.txt | 21 |
| 29 | tlele ka kgwedi leina, Motsheganong e le leina | | 3,123 | ana~1.txt | 39 |
| 30 | babetsa dinala serame Motsheganong: Mme | | 6,689 | \tsaya.txt | 21 |

When Tables 44 and 45 are compared, one notices the importance of gathering more information from the spoken corpus to add more value to the existing Setswana monolingual dictionaries.

5.6 Towards a sound lexicographical treatment of the Setswana months

Setswana months are named after nature and the change of the seasons. Every month falls within a season and it correlates with the historical events of that particular month as indicated above. According to the Batswana people, the first month of the year starts in August (*Phatwe*) and not January (*Ferikgong*) since they start with the plough and the first rain (*Kgogolammoko*) starts to fall. According to them, *Phukwi* (July) is regarded as the last month of the year and not December since all the work of the plough in the fields is completed. The meaning attached to these months is deduced from a specific cultural context. An illustration of that matter is given as follows:

Table 46: Treatment of the Setswana months

| | | |
|----|--|---|
| 1 | <p><i>Ferikgong</i></p> <p>January</p> | <p><i>Go a fisa gore logong lwa mofiri lo robege bonolo. Ka jalo kgwedi ya nna Ferikgong, mm eke kgwedi ya bofelo ya setlha sa selemo.</i></p> <p>Eng. The weather is very hot, drying up trees, especially those that produce firewood. It becomes easier to break firewood from such trees. It is the last month of Summer).</p> |
| 2. | <p><i>Tlhakole</i></p> | <p><i>Pula e na matsorotsoro, ka jalo e tlhakola mogote wa Ferikgong</i></p> |

| | | |
|----|---|--|
| 3. | <p>February</p> <p>Mopitlwe</p> <p>March</p> | <p>Eng. During this month, very heavy rains are experienced, and those rains help to cool the hot January-weather off. It is the frost month of Autumn.</p> <p><i>Dijalo le mmidi ke nletsentletse. Ngwana wa motswana o tla ja mpa gore a pipitlelwe, ka jalo ya nna Mopitlwe.</i></p> <p>Eng. During this month, it is green everywhere, the fields and the veld. There is also more than enough food for everyone. (<i>o tla ja mpa gore o pipitlelwe</i>)</p> <p><i>Go bonala go gola ga dijalo go anama. Go nama go iketla go namile maoto. Moranang a dijalo a nama.</i></p> <p>Eng. Plant growth slows down, some trees start losing leaves or their leaves start browning.</p> <p><i>Go tsididi, phefo e foka ka maatla, ditlhare di foforegi matlhare. Dinonyane ga di kgone go ka ja dijalo kwa masimong, ka jalo Motswana a be a re dijalo di tshega nong ka jalo kgwedi e bidiwa Motsheganong.</i></p> <p>Eng. The weather starts getting cooler and windy, the trees/ plants starts losing leaves. Birds can no longer feed on the seeds easily in the fields, thus the Batswana mock (tshega) these birds frustration. It is the first month of winter.</p> |
| 4. | <p>Moranang</p> | |
| 5. | <p>Motsheganong</p> <p>May</p> | |

| | | |
|----|--------------------|---|
| 6. | Seetebosigo | <i>Kgwedi e, e kganela go eta bosigo ka o tla tlhela batho dikobo. Ka jalo Seete bosigo.</i> |
| | June | Eng. During this month, people are discouraged from visiting, especially for overnight/sleepover, because it is cold and guests might be an inconvenience if you do not have enough blankets to share with them (guests). |
| 7. | Phukwi | <i>Ke kgwedi ya bofelo ya setlha sa mariga. Serame se a laela gore ke a feta. Go fitlha pula ya kgozolammoko. Motswana a bo a re Phukwa! A ngwaga o wele kwa! go tle o moša re tle re simolole botshelo seša. Go jewa dijo tse di bolelo le go apara diaparo tse di bokete.</i> |
| | July | Eng. The winter season comes to an end. It is also the end of the Batswana year. The first rains are expected in anticipation of the new year |
| 8. | Phatwe | <i>Kgwedi e simolola setlha sa dikgakologo. Serame se a gakolologa. Ke kgwedi ya ntlha ya motswana.</i> |
| | August | Eng. This is the first month of the new year, and the first for the Spring season. The weather is slightly warmer and dusty but pleasant. |
| 9. | Lwetse | <i>Ke kgwedi ya dikgakologo, botshelo jo boša bo a simolola, bo tla ka malwetse bo batho le diphologolo, ya nna kgwedi ya lwetse. Malwetse a mantsi a itemogelwa ka kgwedi e.</i> |

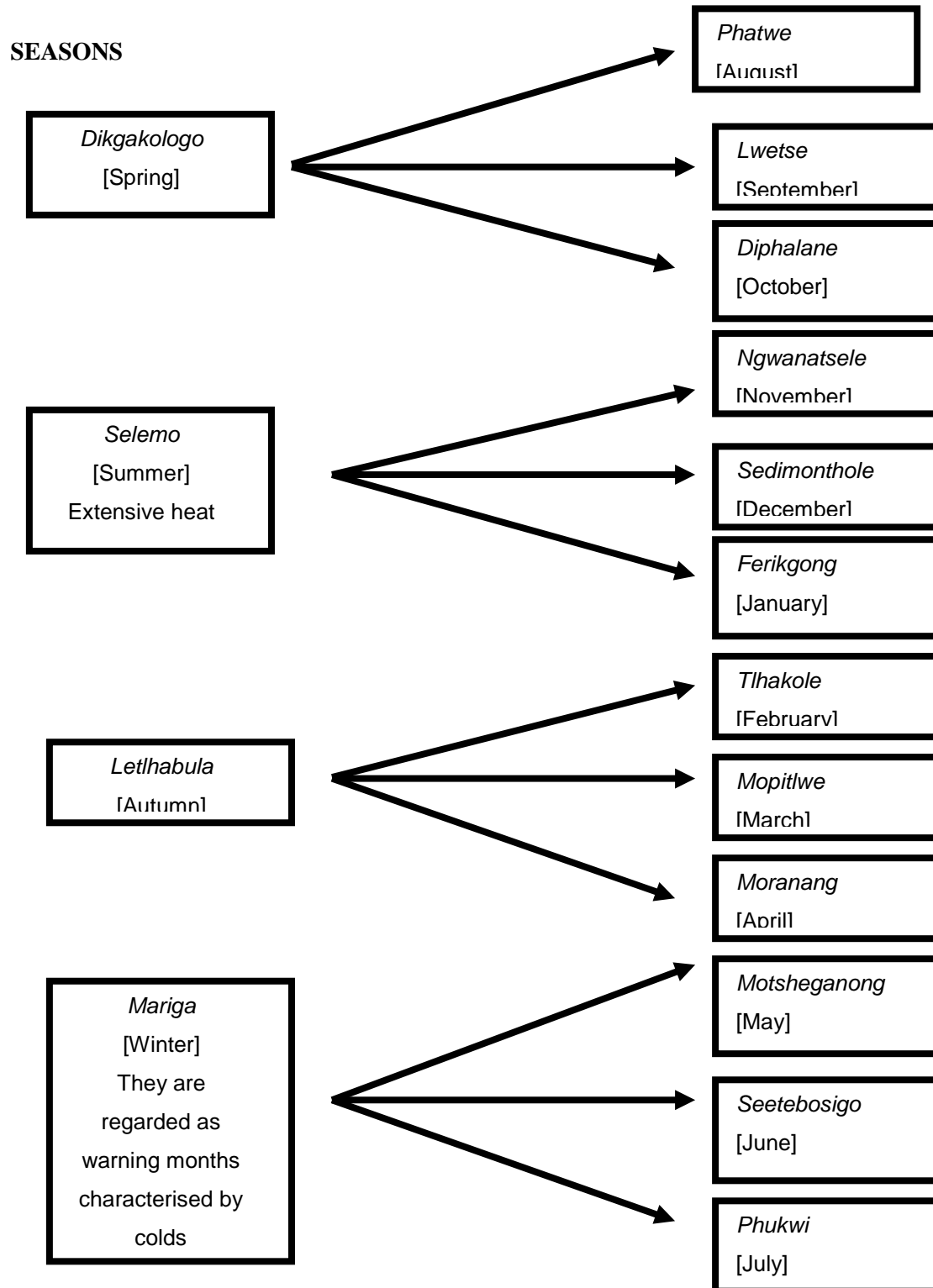
| | | |
|-----|--|--|
| 10. | <p>September</p> <p><i>Diphalane</i></p> | <p>Eng. This is a month for new beginnings, e. g. plants begin to blossom, most animals are now out of herbanation, etc. Also, because of the blossoming of plants, there are lots of pollen and allergens in the atmosphere, hence most diseases are spread and or experienced during this month.</p> <p><i>Ke kgwedi ya dikgakologo. Bojang bo tlhogile, naga e talafetse, go utlwala melodi ya dononyane tse di itumeletseng bontle jwa naga ka bophara. Kgwedi eo e nne jaakadiphala mo tsebeng ya monna wa motswana.</i></p> |
| 11. | <p>October</p> <p><i>Ngwanatsele</i></p> | <p>Eng. This is the last month of Spring. It is green everywhere and birds are singing all over, the music is sweet to listen to.</p> <p><i>Ke tshimologo ya setlha sa selemo. Dophologolo do simolola go baya kgotsa go tsala. Kgwedi ya nna ngwanatsele. Bana ba itsela magapu le maugo. Ke ka moo go twang ngwana itseele.</i></p> |
| 12. | <p>November</p> <p><i>Sedimonthole</i></p> | <p>Eng. This is the beginning of summer. Most animals give birth to their young ones, most fruit are ripe, and thus no one will stop you from having any fruit you want (ngwana itseele)</p> <p><i>Maungo mo kgwedi e, a mantsi, merogo e simolola e jewa thata. Fa o ile sekgweng kgotsa masimong, o tla o rwele o imelwa ke dijo, o sa kgone le go 'ithola' morwalo. Ka jalo ra re 'sedi nthole ke palelwa ke go ithola' ka jalo kgwedi ya bidiwa</i></p> |

| | | |
|--|----------|---|
| | December | <i>Sedimonthole.</i> Eng. There are lots and lots of fruit and vegetables everywhere. For those who are reaping manually, using baskets would be so full, one won't be able to remove the basket from one's head by him. You might even wish that there was a bigger power to help you out. |
|--|----------|---|

According to the research, the names for Setswana months were shifted according to the historical events and the changes of seasons. It is important to state that a descriptive monolingual dictionary should be characterized by this kind of approach that will give true status of a language in question and THAN should not be an exception. This is about balance between the normal dictionary information and, where necessary, encyclopaedic information. The lexicographer should take a balanced approach in respect of the inclusion of encyclopaedic information in paper dictionaries. Due to the limitation on space encyclopaedic information should be limited to instances, e.g. crucial cultural information where the standard treatment in the article is insufficient. Paper Setswana dictionaries should therefore not be a combination of dictionary and encyclopaedia per se.

Table 47 below gives an overview representation of these Setswana months:

Table 47: Schematic representation of the Setswana months



5.7 Conclusion

In this chapter we have demonstrated the value of corpora, how it can be used to improve the quality of the microstructural elements in the treatment of lemma signs. The impact of corpora as a key to writing better dictionary articles, as sense distinctions for writing better definitions in monolingual dictionaries and setting up translation equivalents in bilingual dictionaries and as an aid to pinpoint frequent clusters was emphasized. We have also illustrated how a detailed analysis of corpus lines, in combination with frequency counts, enables lexicographers to tremendously enhance the quality of microstructural elements. Furthermore, typical collocations in addressing typical microstructural inconsistencies existing in the currently available Setswana dictionaries have been highlighted. Problems relating to the treatment of verbs, polysemy and the so-called synonyms in the Setswana dictionaries have been addressed. The study has also shown how the lexicographer with good query tools at his/her disposal can combine the output of different tools such as word-frequency counts and concordance lines screens.

We have also seen that the intuition of even a trained native-speaker lexicographer cannot compare to the accuracy of corpus-based queries. As a result corpus lexicography has the potential to result in much sounder and more user-friendly dictionaries than those compiled during the era of the so-called traditional manual lexicography. The treatment of the Setswana months in currently available Setswana dictionaries was critically analysed and evaluated against the background of the English and the Afrikaans dictionaries. The problem of what is regarded as the first month of the year for Setswana was outlined through the use of examples and diagrams. Suggestions for the improvement by means of a corpus-based microstructure has been identified and discussed in detail.

Chapter 6

Conclusion

We have entered the corpus era in the dictionary compilation and this study gave a comprehensive discussion of how future Setswana dictionaries should be compiled maximally using corpora, corpus query tools and advanced tools such as a ruler and block system. In chapter 2, the extensive historical background of the Batswana as a group (and how they are divided in both South Africa and Botswana), and Setswana as a language with diverse dialects was discussed in detail. It was also noted that Setswana, like any other language, is influenced by other languages spoken within or around its environment, unless where the environment is homogenously Setswana speaking where the influence is likely to be minimal or non-existent.

Just like other languages, Setswana is growing and this growth dates back to the arrival of the missionaries in the 1820's to date. The contribution made by the missionaries is valuable and worth noting. Various tables, maps, illustrations, etc, have been used to demonstrate the Batswana geographical locations, their language Setswana, dialects and their geographical distribution in both South Africa and Botswana, including the statistical representations of the Batswana and their dialects.

Language planning was also discussed in-depth, with emphasis on the three main sub-dimensions, characterizing language planning. Language is not stagnant, and thus language change also affects Setswana as a language. Factors like new development, technology, etc, are bound to affect language (particularly spoken language) and leads to the creation of new concepts and terms in relation to the new developments or for effective communication. New concepts usually originate from borrowing (from other languages) and analogical implications on some words frequently used. Just like many

progressive languages, Setswana has grown and developed to be a language in its own right, through aspects such as the writing system orthography; which elaborated in-depth about the Setswana grammar, including the grammatical rules applicable. Various examples were given to illustrate and to support the discussions given.

The education system in South Africa, right from Bantu Education to date, greatly impacted on Setswana, particularly written or academic Setswana. Education systems around the world are prescriptive as to how language should be taught or learned – thus some words might be modified or disallowed completely, and some new (unfamiliar) words might be included. Language Boards also play a role in prescribing the parameters within which a language can operate, and some concepts might not be approved for use, as is the role of the PANSALB.

As the language develops, it becomes necessary to have some form of reference to keep up with the development, leading to the writing of dictionaries. In spite of the weakness of most of the Setswana monolingual dictionaries in comparison to English dictionaries, they (dictionaries) provide valuable information and serve as the basis for further language development. However, suggestions to deal with the weakness were given so that these dictionaries can be as useful as they were meant to be.

In chapter 3, the actual compilation of Setswana corpora was discussed in detail. Extremely useful theoretical insights of the strategies employed by COBUILD and LDOCE which are corpus-based are used as excellent examples for the compilation of African language dictionaries; in particular Setswana. Projects were used to illustrate the compilation of the Setswana oral corpus, and these beautifully illustrated the influence on Setswana language by various environmental aspects, such as the rural v/s the urban areas; level of literacy or education; attitudes to or against language changes; the origin of certain Setswana words (like Setswana months); and so forth.

The oral corpus serves as the foundation for the written Setswana, and the effect or role of spoken language has been highlighted effectively, covering various issues, e.g. keyness (both positive written corpus and negative corpus). The corpus query program

was also used to give vital statistical information on the corpus. The statistical analysis between the oral corpus and the written corpus was highlighted. These analytical tests are applied in order to detect systematic differences between the text categories. Illustrations to indicate the statistical analyses were taken from WordSmith Tools, which is a very useful tool for distinguishing word types, tokens and the written corpus. The significance of frequency counts as an extremely useful tool in the compilation of a lemma list for a new dictionary was also emphasized.

The desire to redress the imbalance in favour of the unprejudiced investigation of the spoken and written language to justify its correctness was done by comparing the oral and written Setswana corpus in terms of the Keyness function using WordSmith Tools. Another very important comparison is given, between the South African corpus and the Botswana corpus. In this regard, words have been sorted by consistency, rank and overall frequency. However, the South African corpus is larger than the Botswana corpus as indicated through the illustrations used, and this difference is also evident with the statistical analysis. Words that are regarded as frequently used in the South African corpus are not necessarily regarded as frequently used in the Botswana corpus.

In chapter 4, much emphasis was placed on the urgent need for the utilization of electronic corpora on the macrostructural level. The macrostructural inconsistencies existing in Setswana dictionaries have been critically evaluated and analyzed. The use of frequency lists have been emphasized ensuring that frequently used words are not accidentally omitted and also that the dictionary space is not occupied by articles of lemmas unlikely to be looked for by the target users, particularly in the two Setswana monolingual dictionaries, i.e. THAN and THAND.

The limitations of the Setswana dictionaries are such that a huge number of words are entered without semantic information, thus the user has to frequently consult the front matter. This is a very time consuming and confusing process. Lexicographers therefore have to adopt a holistic approach when compiling dictionaries. This will also eliminate the possibilities of imbalances with regards to the alphabetical stretches,

which are usually the result of the lexicographers adding new words as they come across them, disregarding frequency counts. Sometimes lemma-signs are either under or over-treated. The reasons for this have been explored and graphical illustrations using the two well-known Setswana monolingual dictionaries were used.

The use of the multi-dimensional ruler to correct or balance the inconsistencies has been demonstrated. Lexicographers should negotiate a complex interplay and overlap between lemmatization strategies, approaches, lexicographic traditions, and verbal structures and conjunctiveness. Each of these were discussed in detail and typical examples provided or highlighted.

The rules for the lemmatization of nouns, both singular and plural, were looked into, once more referring to the two Setswana monolingual dictionaries. The problems experienced by users of these dictionaries or language users, in terms of the inconsistencies of some words converting to plural forms, have also been indicated and discussed to interpret the differences or inconsistencies for users familiar with the language and those that are not.

This chapter also opened new doors for the lemmatization strategies of nouns and verbs based on the frequency of use, using the available corpus data. The feasibility was exemplified through the macrostructural inconsistencies that exist in Setswana dictionaries. The problems inherent in the lemmatisation of the Setswana homonyms and tonal indication are real. These studies were performed against the background of the user- perspective.

Chapter 5 illustrates the contribution of corpora during the microstructural compilation of dictionary articles. Detailed analysis of corpus lines in combination with frequency counts at various levels was outlined through suitable examples taken from the Afrikaans and English dictionaries. The most significant corpus query output to generate the concordance lines and the occurrences of a word or phrase extracted from the corpus was emphasized. Such corpus lines will assist the lexicographer in respect of sense distinction, deciding on translation equivalents, retrieval of typical

collocations, pinpointing frequent clusters and in the authentic examples to be included in a dictionary. Corpora in this instance can furthermore assist the lexicographer in finding typical collocations generated by WordSmith Tools and combination of words for writing a better definition or choosing better equivalents. Inconsistencies regarding the treatment of nouns, synonyms and cross-referencing as an interconnection of the microstructural component through the use of a reference marker were reviewed.