

Introducing *Tlhalosi ya Medi ya Setswana*: The Design and Compilation of a Monolingual Setswana Dictionary

Thapelo J. Otlogetswe, *Faculty of Humanities, University of Botswana, Gaborone, Botswana (otlogets@mopipi.ub.bw)*

Abstract: This paper presents the design elements of a recently published monolingual Setswana dictionary, *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), the fourth main Setswana monolingual dictionary to appear. The paper situates the dictionary within a recent, but growing Setswana monolingual dictionary tradition whose roots may be traced to 1976 when M.L.A. Kgasa's *Thanodi ya Setswana ya Dikole* was published. However, before this date, the development of Setswana lexicography lies wholly in a bilingual dictionary tradition which dates back to the 1800s. The paper also discusses the different features of *Tlhalosi ya Medi ya Setswana* demonstrating that it possesses some dictionary features which are not found in previous Setswana dictionaries such as frequency information, phonemic transcription and extensive cross-referencing.

Keywords: TLHALOSI, THANODI, DICTIONARY, SETSWANA, MONOLINGUAL

Opsomming: Bekendstelling van *Tlhalosi ya Medi ya Setswana*: Die ontwerp en samestelling van 'n eentalige Setswanawoordeboek. Hierdie artikel bied die ontwerpbeginsels aan van 'n onlangs gepubliseerde eentalige Setswanawoordeboek, *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), die vierde belangrike Setswana eentalige woordeboek wat verskyn het. Die artikel plaas die woordeboek binne 'n onlangse, maar groeiende Setswana eentalige woordeboektradisie waarvan die wortels teruggevoer kan word tot 1976 toe M.L.A. Kgasa se *Thanodi ya Setswana ya Dikole* gepubliseer is. Voor hierdie datum egter lê die ontwikkeling van die Setswana-leksikografie geheel en al in 'n tweetalige woordeboektradisie wat terugdateer tot die 1800's. Die artikel bespreek ook die verskillende eienskappe van *Tlhalosi ya Medi ya Setswana* deur aan te toon dat dit 'n aantal woordeboekeienskappe besit wat nie in vorige Setswanawoordeboeke aangetref word nie soos frekwensie-inligting, fonemiese transkripsie en uitvoerige kruisverwysings.

Sleutelwoorde: TLHALOSI, THANODI, WOORDEBOEK, SETSWANA, EENTALIG

1. Introduction

The Setswana language has a fairly long lexicographic tradition characterised by low dictionary production. The tradition began with a bilingual Setswana-English dictionary by the London Missionary Society missionary, John Brown,

that was first published in 1875, enlarged and revised in 1895, and revised again in 1925 by John Tom Brown. For over 130 years this dictionary, which was updated by Matumo (1993) has been the only general Setswana–English dictionary. The Setswana monolingual dictionary tradition started in 1976 with *Thanodi ya Setswana ya dikole* (Kgasa 1976). And for about 20 years this was the only monolingual Setswana dictionary, although Setswana was taught in both Botswanan and South African schools (cf. Volz 2003). Much of this lexicographic tradition is dominated not by monolingual but by bilingual dictionaries. Jones (in Matumo 1993: vii) traces the origin of Setswana lexicography to John Brown's bilingual dictionary (1875) and to Robert Moffat's Setswana version of the Gospel of St Luke (1830), which has definitions of difficult words in its final back pages. Jones observes:

In 1830 Robert Moffat published a Setswana version of the gospel of St Luke, and at the back offered two pages of explanations of the more "difficult" words. Is it fanciful to regard this as the first small germ of a dictionary? ... but the first published dictionary of which the Botswana Book Centre has record is that of John Brown in 1875.

Alternatively, Cole (1955: xxviii) dates Setswana lexicographic research to later years to the plant names compilation of Miller (1936) and the list of kinship terms of Van Warmelo (1931). However, lexicographic research in Setswana may be traced to an earlier period, much earlier than Moffat's 1830 writings that Jones refers to, and certainly far earlier than Cole's botanical and kinship references. Research demonstrates that Lichtenstein in the second volume of *Travels in Southern Africa in the Years 1803, 1804, 1805, and 1806* had as an Appendix a list of about 270 Setswana words and phrases. The original book in German appeared in 1811. Therefore the earliest lexicographic activity, at least of a headword list with its English equivalents, can be traced to 1803–1806, in Lichtenstein's works. In 1815, Campbell in his *Travels in South Africa* gave a list of 80 "Bootchuana" words at the end of chapter 19 in the book. Salt (1814) in *A Voyage to Abyssinia* presents a list of 20 "Mutshuana" words and their English equivalents. Lexicographic work in Setswana, regardless of its size and detail, therefore existed before the work of Moffat, who only came to Southern Africa in 1816.

The first Setswana bilingual dictionary, *Lokwalo loa Mahúku a Secwana le Seeneles*, was compiled by Brown (1875) of the London Missionary Society. An enlarged and revised version appeared in 1895 and was reprinted in 1914 and 1921. In 1925 Brown produced the third edition of this dictionary with the aid of A.J. Wookey's research (Peters and Tabane 1982: xxiv). However, since the 1925 edition of Brown's dictionary to the mid-1970s, no Setswana dictionary was compiled. That represents a fifty year gap. It was not until 1976 that Kgasa published his 134-page monolingual dictionary, *Thanodi ya Setswana ya Dikole* (The Setswana Dictionary for Schools), whose main target users were primary school pupils. Kgasa's dictionary, which took ten years to compile (Rasebotsa

2008), is significant for it marks the start of the Setswana monolingual dictionary tradition. A smaller, but detailed, trilingual dictionary — Setswana, English and Afrikaans — was compiled by Snyman et al. (1990), *Dikišinare ya Setswana–English–Afrikaans Dictionary/Woordeboek*, whose target is secondary school and university users. This dictionary filled a gap that existed in the South African market where both English and Afrikaans were used widely by Setswana speakers.

The Compact Setswana Dictionary by Dent (1992) is an abridged 300-page dictionary "intended for those people who find more comprehensive dictionaries too cumbersome or too detailed for their needs" (Dent 1992: introduction). It has a large headword list, though it lacks detailed lexicographic metalinguistic matter.

Matumo (1993) revised Brown's 1925 dictionary into what is now the *Setswana–English–Setswana Dictionary*. Prinsloo (2004) revisits this dictionary suggesting how it may be revised by the design and use of a multidimensional Ruler and Block System for the measurement and balancing of alphabetical stretches in terms of time, average length of articles and number of pages per alphabetical category.

Cole (1995) compiled a dictionary of plants and animals called *Setswana — Animals and Plants (Setswana — Ditshedi le Ditlhare)*. While in the foreword L.W. Lanham notes that "the author ... eschews the label 'dictionary' for it, preferring to identify it as a 'lesser listing of vocabulary'" (Cole 1995: ix), it is a bilingual dictionary, Setswana to English and English to Setswana, with some of the entries included with their Latin names.

In 1998, in collaboration with Tsonope, Kgasa compiled the second monolingual dictionary *Thanodi ya Setswana* which until now remains the definitive Setswana monolingual dictionary used widely in Botswana schools and other domains both in Botswana and South Africa. It has been prescribed in Botswana junior and senior secondary schools for over fifteen years since its publication as a reference text for Setswana language and literature classes.

Créissels and Chebanne's *Dictionnaire Français–Setswana Thanodi Sefora–Setswana*, published in 2000, is the only French/Setswana bilingual dictionary. Its primary target group is students of French at secondary and university level. It is the first Setswana dictionary with Setswana phonemic transcriptions, though this part of the dictionary is limited. In 2007 Otlogetswe published a smaller bilingual dictionary, the *English–Setswana Dictionary*, for use as a reference work in Botswana primary schools.

In 2008 a fairly large Setswana dictionary of about 600 pages was compiled by Mareme with the aid of the Setswana National Lexicographic Unit (NLU) based in Mafikeng. Like Kgasa and Tsonope's dictionary, it is called *Thanodi ya Setswana*. This dictionary has not been popular in Botswana and is also not widely used in South Africa.

Cole and Moncho-Warren (2012) have published a Setswana–English, English–Setswana dictionary of nearly 1200 pages entitled *Macmillan Setswana*

and *English Illustrated Dictionary*. The dictionary will be of interest to language specialists, especially at the university, because, as the back cover states, the dictionary is "intended for use by scholars and teachers at tertiary level, and in libraries". G.B. Mareme, writing in the introduction of the dictionary, is probably right that "learners in the new South African education system, especially those studying life sciences, physics and chemistry, will find this dictionary invaluable".

The discussion above has outlined the Setswana dictionary landscape dating to the 1800s. It demonstrates that *Tlhalosi ya Medi ya Setswana* comes at a time when for the past 20 years there has been some exciting Setswana lexicographic projects and a growing Setswana monolingual dictionary tradition.

What follows is a description of the compilation and different features of *Tlhalosi ya Medi ya Setswana*.

2. Headword collection for the *Tlhalosi ya Medi ya Setswana*

The craft of lexicography demands not only the ability to collect data, but also the ability to make sense of it. The headword list collection for the compilation of *Tlhalosi ya Medi ya Setswana* was achieved through three strategies to capture the broad lexical representation of the Setswana language. The word list collection used a corpus, semantic domains and the consultation of other publications such as dictionaries, grammar books and language documentations.

2.1 Corpus contribution

Developments in corpus linguistics and corpus lexicography of the past 30 years, particularly in English lexicography, have resulted in unique techniques of lexical analysis for lexicography. The use of statistical analysis through the use of Corpus Querying Software (CQS) such as the *Oxford WordSmith Tools* (Scott 2004) and *SketchEngine* (Kilgarriff et al. 2004) has meant that it is now possible to analyse large amounts of text running into millions of words. *Tlhalosi ya Medi ya Setswana* was based on a 16 million token Setswana corpus (Otlogetswe 2008). The corpus is broad in its composition, with 90% of the text coming from written texts and 10% being transcribed speech. The corpus comprises novels, science texts, political texts, grammar books, poetry texts, religious texts, prose texts, spoken texts recorded from a variety of contexts such as funerals, Setswana classroom and family meetings. The corpus was analysed using the *Oxford WordSmith Tools* version 5 (Scott 2004). Frequency analyses were conducted to generate a dictionary headword list while word concordances were extracted to study how different words behave in context to aid sense discrimination. Some congrams were also run to identify word clusters of certain Setswana terms. The corpus was further used as a source of illustrative examples.

The contribution of a corpus to the dictionary making process has been discussed extensively in lexicographic literature by Béjoint (2000: 97), and Sinclair (1987, 1991) who demonstrate the crucial nature of a corpus and frequency lists for the dictionary compilation process. A corpus has increasingly become useful in providing definitive dictionary information for many modern dictionary compilations. However, claims and judgments on the definitiveness of the data abstracted from a corpus cannot be made independent of the claims and judgments on the quality and constitution of the corpus from which the data was abstracted (Otlogetswe 2008). Put differently: if the corpus data is skewed and badly structured, the data abstracted from the corpus will also reflect this — one cannot take from a corpus what has not been put into it. Focus must increasingly be directed to the structure and content of the corpus and not just what is abstracted from it (Biber 1993). The corpus should attempt to be as broad as possible in its coverage of the language varieties. Such a broad spectrum corpus would include novels, newspapers, magazines, spoken language, news, meeting transcriptions, health literature etc (Biber 1994). Since about 90% of the corpus used for this dictionary was written text and 10% was transcribed text from spoken language, semantic domains (Moe 2001) were used as a supportive measure to add words to the dictionary.

Many corpora in African languages are insufficient as sources of dictionary material since their text coverage is limited. Most corpora comprise published works which are a poor representation of the linguistic variation of a speech community. There are numerous oral domains in which African languages are used such as law, business, farming and health, which are largely excluded from a corpus of written texts. To fill this gap, other strategies of data collection were used.

2.2 Semantic domains

Apart from the corpora collection approach discussed above, Moe (2001) proposes a method of semantic domains to be used for the collection of words. He argues that the methodology is particularly attractive for minority languages, most of which have none or few written texts, or no corpora. His argument is that the methodology is 100 times faster than collecting words without a structure. He argues that 12,000 words have been collected in a few weeks through what is effectively a simple methodology but one which is able to produce a massive classified dictionary and thesaurus.

Moe analysed the domain classification of words as suggested by Murdock et al. (1987), Roget (1958 and 1985) and Louw and Nida (1989) and found them inadequate for eliciting vocabulary. What Moe (2001: 151) attempts to compile is "a universal list of semantic domains" that field lexicographers could use to prompt native speakers to think of words in their language. However, semantic domains have a greater relevance than mere elicitation of mother tongue speakers' words: "It could be used to collect words, it could serve to

classify a dictionary, and it could aid in semantic investigation" (Moe 2001: 152). Underlying this system is a mental approach to the lexicon: words are all linked together in the mind in a gigantic multi-dimensional web of relationships which cluster around a central nexus (Moe 2001: 4). The mental lexicon is not alphabetical but words congregate around key concepts which Moe calls semantic domains (Moe 2003: 216). His argument is therefore that related words should be collected at the same time. To guide field workers, Moe phrases domains as questions such as the following for the domain "sing":

What words refer to singing? sing, serenade, warble, yodel, burst into song
What words refer to singing without using words? hum, whistle

These series of questions are central to what Moe calls the Dictionary Development Process (DDP) which he used in Uganda in training lexicographers in collecting words for the Bantu language Lunyole. The DDP has 1,700 domains each with 8–10 questions which could elicit over 10 words per domain which means that the dictionary would have at least 17,000 entries.

This method was used to obtain specialised information which is difficult to find in a corpus such as the names of stars, types of cows, meat cuts, animal colours, terminology relating to traditional beer making processes etc.

2.3 Other literature

Setswana has a number of relevant literatures which were consulted to develop the headword list of the dictionary. Some are dictionaries such as Kgasa (1976), Kgasa and Tsonope (1995) and Mareme (2008). Others are grammar books such as Cole (1955) and Mogapi (1984). The headword list of *Tlhalosi ya Medi ya Setswana* was compared against that of Kgasa and Tsonope (1995). The grammatical analysis of Cole (1955) and Mogapi (1984) provided a useful resource for closed word class members such as demonstratives and pronouns. We also benefited from Cole (1995) from which a wealth of animal and plant names were culled.

3. Choosing *tlhalosi* over *thanodi*

Since this is the first Setswana dictionary called *tlhalosi*, it is important to provide a justification for the choice of the term *tlhalosi* over the more established Setswana term *thanodi* to refer to a Setswana monolingual dictionary. The first Setswana dictionary to carry the title *thanodi* was the monolingual dictionary Kgasa (1976). It was published about 100 years after the first bilingual Setswana–English dictionary (Brown 1875) appeared. Therefore, when Kgasa published the first monolingual dictionary, the term *dictionary* was very much synonymous with a bilingual dictionary, that is, a dictionary that offered translations from one language to another. This observation is important since

it may proffer an explanation why the term *thanodi* may have been chosen to mean "dictionary" in Setswana.

The word *thanodi* is derived from the Setswana verb *ranola* through the linguistic process of strengthening. It means "to unravel the signification of a hidden tongue or opaque semantics". However, in current usage, the word *ranola* means "translate". The term *thanodi* is therefore better suited to refer to bilingual or multilingual dictionaries such as the Matumo (1993) and Snyman et al. (1990) dictionaries, since they deal with translation and equivalents. The term *tllhalosi* on the other hand is derived from the Setswana verb *tllhalosa* which means "to explain". Since a monolingual dictionary explains words in a single language, the term *tllhalosi* is preferred as a term reserved for a monolingual dictionary. To use the term *thanodi* for *tllhalosi* would therefore be erroneous. That is why in part the current monolingual dictionary is termed *tllhalosi* and not *thanodi*, since it does not translate words from one language to another but it explains Setswana words using Setswana vocabulary. It has also been discovered that a number of Sotho-Tswana languages use the word *tllhalosi* to refer to either monolingual dictionaries or dictionaries in general.

For instance, the monolingual *Pukuntšutllhaloši ya Sesotho sa Leboa* (Mojela 2007) uses the word *tllhaloši* and not *thanodi* in its title. Other compilers use the term *tllhalosi* for dictionaries in general. An example is *Lenanentswe-hlalosi: Seafrikaans–Seisemane–Sesotho sa Leboa* (Joubert and Mangokoane 1975). In the Lozi language the word for *dictionary* is *sitalusa-manzwi* (*setllhalosa mantswwe* in the Setswana orthography), "explainer of words". The word *manzwi* or *mantswwe* means "words".

Compared to *thanodi*, the word *tllhalosi* is also appropriate since it is derived from *tllhalosa* a verb which means "to explain". A word's explanation goes beyond its meaning. It includes its use in a sentence, its part of speech and noun class, its stylistic use, that is, whether it is formal or informal. It also includes a word's etymology, pronunciation and usage frequency. All these characteristics of a word are not translating but form part of explaining a word. This is why the word *tllhalosi* was given preference above *thanodi*.

4. Dictionary target users

The dictionary was designed with a broad spectrum of monolingual users in mind. In a community which has a vibrant and established dictionary culture, it is easier to divide target users into smaller specialised and focused groups. However, in a less vibrant dictionary environment like the one we are dealing with, it is important to consider a fairly broad spectrum of users for the dictionary, instead of restricting its scope of usage. This is significant because since 1976, *Tllhalosi ya Medi ya Setswana* is the forth Setswana monolingual dictionary to be compiled. It was therefore written with the following users in mind:

- (a) It was meant for use by Setswana language classes from secondary school level up to tertiary levels. Phonemic transcriptions with tonal markings were therefore included to cater for the needs of advanced language classes. Synonyms and etymological information were added to fulfil the needs of both secondary and tertiary level classes.
- (b) The dictionary is suitable for utilisation by Setswana writers and authors. It contains information on dialectal variation and stylistic markings which will aid writers in their work. These could be professional writers or students writing their essays at school or university.
- (c) The dictionary was compiled for Setswana speakers in general wherever they might live. Setswana has mother-tongue speakers in at least four countries: South Africa, Botswana, Namibia and Zimbabwe. The largest number of speakers is found in South Africa (over 4 million speakers, about 8% of the population) where Setswana is one of the eleven official languages. Zimbabwe has an estimated 29,000 Setswana speakers and Namibia has approximately 6,000. In Botswana, Setswana is used by circa 1.5 million speakers (70-90% of the population) as their mother tongue (Andersson and Janson 1997). Selolwane (2004: 4) observes that "the SeTswana language is the most dominant of all the language groups found in Botswana, with at least 70% of the population identifying it as a mother tongue and another 20% using it as a second language". In Botswana Setswana is the national language. It is studied as a compulsory subject by all Botswana students from primary up to senior secondary school. The dictionary is therefore written to instil pride in the Setswana language amongst Setswana speakers. This is important since Setswana is an important intangible heritage of its speakers.
- (d) The dictionary was written for Setswana language teachers and lecturers. It is designed to aid them in their instructional classes as well as in the development of the Setswana language.

5. *Tlhalosi ya Medi ya Setswana's* key features

The following are the microstructural features of *Tlhalosi ya Medi ya Setswana*.

5.1 Sense presentation

Each word has at least one core meaning, to which a number of polysemous senses may be attached. Core meanings represent typical, central uses of the word in question in current Setswana, based on corpus analysis and established research. The core meaning also represents the most literal sense that the word has in ordinary modern usage. This is not necessarily the same as the oldest meaning of the word, because word meaning change over time. In our analysis, the core meaning is based on frequency and an analysis of concordance lines. For instance, the entry for the word *bona* "see" is the following:

bôna¹ **** /bónà/ •ldr• 1^o go dirisa matlho go leba *Fa a bona kolo e tla e tabogile a fapoga tsela ka bofefo go iphemela*
2^o amogela sengwe *O bona madi a mantsi ka go rekisa phane*
3^o go tihaloganyana sengwe *Jaanong ke a bona gore ga a nthate* 4^o go kopana le mongwe *Ba rile ba tlaa re bona teng*

Some related words have multiple parts of speech. The decision of which part of speech appears first is made according to the most common usage. In the dictionary, the core sense acts as a gateway to other related polysemous senses.

The word *kampa* "camp" can be used as noun as well as verb in sense 3.

kampa¹ /kámpá/ •In. 9. n-, *10. din-• 1^o lefelo le go nniwang mo go lone ka nakwana bogolo jang le le kgakala le motse = BOTHIBÊLÊLÔ 2^o lefelo la bonno la masole, mapodisi kgotsa batlhokomedi ba diphologolo 3^o •ldr• go nna mo lefelong lengwe ka nakwana, bogolo jang le le mo sekgweng ⇐ SeE: camp

Other words have figurative meaning extension of the core sense. The core sense of the word *mokwatla* is "back part of the body" and the figurative meaning extension is "support of a system or organization".

5.2 Phonemic presentation

The dictionary is the first Setswana dictionary to have comprehensive phonemic transcriptions for each lexical entry. Since Setswana is a tonal language, each transcription has been given tonal markings. Phonemic transcriptions are important since they help users, language learners, linguists and teachers, to have a pronunciation reference point. The tonal markings in particular are important for disambiguating words which, though orthographically written the same, have different tonal distributions. These are words such as *mosimanyana* (mòsimànjàná) "small boy" and *mosimanyana* (mòsimànjàná) "small hole". Every headword in the dictionary is accompanied by phonemic transcriptions with the exception of multiword expressions such as idioms and proverbs. For instance:

gôpane o letsa kwadi •seane• go thata
Re lekile go thusa ka ditsela tsothe go agisanya bobedi jo, mme go supa-fala gore gopane o letsa kwadi fela багаетшо

gôpane wa moikapari •leele• mohumanegi yo mogolo *Ke ne ke tlaa mo duela ka eng ke le gopane wa moikapari jaana?* = MOHUMANEGI, MODIDI YO O NTA E TLHÔGÔ E MOTÔPÔ

gôpanyane /χópánànì/ •In. 1a. Ø, *2a. bo-• ngotlo ya lefoko "gopane"

gopê **** /χúpè/ •lbd• lebadi la maina a setlhopha 15 le le tlhobosang bole-ngteng jwa sengwe *Ke badile pego e yotlhe mme ga go na gope fa a tlhalosang gore monna yo o dirile molato teng*

gopêlêla /χúpélélà/ •ldr• go neela

mongwe magopelo <i>Re tlaa gopelela</i> <i>modisa le fa a ne a seyo fa re bua</i> <i>kgomo ya gagwe</i>	gore di wele mo go one, o khuru- metswe ka bojang le dikala <i>Batlho-</i> <i>komedu ba diphologolo tsa naga ba</i> <i>fulhetse go epilwe gopo mo gare ga</i> <i>sekgwa</i>
gopô /χòpò/ •ln. 9. n-, *10. din-• mosima o o epilweng go thaisa diphologolo	

5.3 Setswana dialectal and stylistic varieties

The dictionary is compiled to capture all the Setswana lexical varieties such as Sengwaketse, Sekgatla, Sengwato and others. For instance, the word for *tin* is lexicalised differently in a number of Setswana dialects. It is *mmolopita* in Sekgatla, *tsiri* in Sengwaketse and *sebagabiki* in Sengwato. In the dictionary, all these words are cross-referenced and indicated by a dialectal marker. The dictionary therefore captures high levels of synonymy. It has over 4,400 cross-references of synonyms from a variety of dialects. Synonyms appear at the end of head words and are preceded by the equal sign [=]. The headwords bear the following dialectal marks *AfBo.* (South African), *Bots.* (Botswana), *Ngwk.* (Sengwaketse), *Lete.* (Selete), *Kgat.* (Sekgatla), *Ngwt.* (Sengwato), *Rolo.* (Serolong), and *Kwen.* (Sekwena). It will be observed that some of the markers are broad, such as *AfBo.* and *Bots.*, while others, such as *Ngwk.* and *Kwen.*, are narrow. This is to capture both the dialectal variations associated with different Tswana ethnic groups while at the same time documenting the developing regional varieties of Botswana and South Africa used by the different ethnic groups found in these regions. As more research goes into Setswana dialectal studies, a more detailed characterisation of the Setswana lexical wealth will be revealed. Currently *Tlhalosi ya Medi ya Setswana* has 371 headwords with dialectal markings. The dictionary also has stylistic markings for *informal* entries and terms which may be considered *offensive* by the users and speakers of the language. Informal entries are marked by the replacement character [◆], while the offensive entries are marked by two exclamation marks [!!].

5.4 Frequency information

Tlhalosi ya Medi ya Setswana is the first Setswana dictionary to mark frequency. The most frequent 4000 headwords have been marked with the use of stars immediately following the headword. A frequency list was generated from a Setswana corpus of about 16 million tokens. The most frequent 1000 words were given four stars (****), words ranked between 1001 and 2000 got three stars (***) , words ranked between 2001 and 3000 got two stars (**), words ranked between 3001 and 4000 got one star (*), while words ranked 4001 and below got no star. Kilgarriff (1997: 135) notes: "A central fact about a word is how common it is. The more common it is, the more important it is to know it." Additionally, highly frequent words are more likely to be polysemous.

5.5 Parts of speech presentation

Like any general language dictionary, *Tlhalosi ya Medi ya Setswana* has an elaborate mark-up for parts of speech. These, amongst others, include *lethusi* (lths), *sekaledi* (skld), *lebopi* (lbp), *lebadi* (lbd), *lediri* (ldr), *lerui* (ler), *leetsisi* (leet), *legokedi* (lgkd), *lekopanyi* (lkpn), *lelatihelwa* (lltl), *leemedi* (lmd), *lesupi* (lsp), *lesoboki* (lsbk) and *letlhaodi* (ltlh). An elaborate mark-up in the dictionary is that of nouns since Setswana, like other Bantu languages, has a complex noun prefix system. Singular noun entries include mark-ups for both the singular noun class prefix and for the plural form of the noun. The structure of the singular entry therefore includes:

[In.] (*leina* "noun" abbreviation) [singular noun class]. [singular noun class prefix], [a star], [plural noun class of the headword]. [plural noun class prefix of the headword]

This translates to the following part of speech label:

•*ln. 9. n-*, **10. din-*• which means that a word is a noun *ln* of noun class 9 whose prefix is *n-* and that its plural belongs to noun class 10 whose noun class prefix is *din-*.

This may correspond to a real entry like *hōsetêlê* "hotel" below:

hōsetêlê /hōsítêlê/ •*ln. 9. n-*, **10. din-*• matlo a borobalo a a agetsweng badiri kana bana ba sekolo *Go agiwa hosetele e nngwe ya basetsana kwa sekolong sa rona* ≡ SeE: hostel

What makes this format attractive is that it marks not just the singular noun class of the headword and its class prefix, the plural noun class of the headword and its class prefix are also included in the mark-up. This gives users not just the grammatical information of the headword but also the formation of its plural form. As this dictionary is written primarily for monolingual speakers of the language, the above design was considered appropriate and sufficient. Nouns which have no plural form, carry only the mark-up for the singular form. For instance, the headword *bonana* "youthfulness" below:

bonana *** /bònánà/ •*ln. 14. bo-*• mokgwa kana seemo sa motho yo o mo dingwageng tse di mo magareng ga bongwana le bogolo *Ga twe o akole bonana jwa gago fa o sa tshedile*

5.6 Etymology and borrowing

Setswana has for a long time been in contact with Afrikaans and English and therefore has borrowed extensively from these two languages. Other languages from which borrowings have taken place, include Zulu and Kalanga

(cf. Mathangwane 2008). Of the over 15,000 headwords in *Tlhalosi ya Medi ya Setswana*, 1,179 contain etymological information. Headwords with etymology information therefore constitute about 7.6% of the dictionary. The dictionary traces a word to the language of its origin as well as to its etymon (a word in the source language). Three strategies are used in the dictionary etymology mark-up. First, the source language as well as the etymon is given. For instance:

jê^hmê /dʒémè/ •*ln.* 9. *n-*, *10. *din-*• sejo se se borekereke se se botshe se se dirilweng ka maungo se se tshasiwang mo borothong *Ke rata botoro fela ka gonne jeme yone e na le sukiri e ntsi* ≡ SeE: jam

This entry *jê^hmê* is a borrowing from English (SeE) *jam*.

The second strategy used is an elaborate descriptive one for words borrowed from other Setswana words. This is especially perceptible in the etymology of months. For instance:

Ferikgong /firiq^hòŋ/ •*ln.* 1a. Ø, *2a. *bo-*• kgwedi ya ntlha ya ngwaga ≡ Kgwedi e e reilwe ka lephoi la mofiri kgotsa kofiri le le a bong le sela dikgonnyana, le aga sentlhaga, go tla le simolola go baya mae. Ka go nna jalo leina le le tswa mo go mofiri le dikgong

The entry *Ferikgong* "February" is traced to two Setswana words *kofiri* "dove" and *dikgong* "pieces of wood".

The final strategy is non-elaborate where the same word is traced to a Setswana word as shown in the examples below:

gôpane ** /χópáni/ •*ln.* 1a. Ø, *2a. *bo-*• segagabi se setona se se tshwanang le kwena e nnye = KGWATHÊ ≡ gôpa²

mabarebare /màbàribàri/ •*ln.* 6. *ma-*• dikgang tse di sa tlhoma-misegeng sentle *Ke utlwile mabarebare a gore o a nyala* = MAGATWE ≡ ba re

The entry *gopane* "iguana" is traced to the verb *gopa* "crawl on the belly", while the word *mabarebare* "rumours or hearsays" is traced to the expression *ba re* "they say".

5.7 Idiom and proverb presentation

The dictionary comprises 1,538 proverbs and idioms. This is about 10% of the dictionary's entries. The inclusion of these entries is important since it captures the wealth of the language's idiomatic expressions. Historically it has been argued that Setswana idioms (*maele*) begin with *go* (Mogapi 1984). This classifi-

cation of Setswana idioms is erroneous, since the *go* that is usually found at the beginning of many Setswana idioms is only the infinitive marker of Setswana verbs. Therefore in Setswana idioms it performs the similar function of the infinitive marker of a following verb. Defining Setswana idioms in this matter excludes other Setswana idioms which do not fit this syntactic characterisation such as:

- | | |
|----------------------------------|------------------------------|
| (a) mosamarea yo o molemô | (i) naga e ntse tshêtlha |
| (b) mosêkêla mpeng | (j) nakô kgolo ke eno |
| (c) mosi o tswa ka sekhurumêlô | (k) ntwá e e matlhô mahibidu |
| (d) motho a sa kgwe mathe | (l) ntwá ya tiitii pôtê |
| (e) motsetô o fapogile marago | (m) sebe sa phiri |
| (f) motsetsi o ja ka letsôgô | (n) motsetsi o ja ka letsôgô |
| (g) naga e jelê bore jwa thamaga | (o) tselapedi a tseye e fe |
| (h) naga e ntse kgologolo | |

In *Tlhalosi ya Medi ya Setswana* the *maele* label for Setswana idiomatic constructions is not restricted to those beginning with the infinitive marker *go* as in *go ja motho setshego* "to laugh at somebody". Setswana idioms have been characterised according to their figurative meaning and the fact that they fracture principles of semantic compositionality (Moon 1998). This is important, since the idiom label is not just syntactically determined, but most importantly it is semantically determined. This decision has a bearing on how the idioms are entered in a dictionary. In *Tlhalosi ya Medi ya Setswana* all idioms and proverbs are entered as separate headwords and not subsumed as subentries under a main entry. This is because, as they are semantically non-compositional, they are separate entities, semantically, syntactically and morphologically. For instance, *tsaya boipuso*, *tsaya dikgang*, *tsaya dikobô ka dikgoka* and others are not entered as subentries of *tsaya* but as independent headwords as shown below.

tsaya **** /tsájà/ •ldr• **1**• go ntsha sengwe fa se teng o se isa go sele
Tsaya ngwana o mo robotse mo ntlong **2**• go tsholetsá sengwe
3• go dirá gore sengwe se nne ka fa tlase ga taolo ya gago
Ditsouludi di ne tsa tsaya puso ka go tlhankgola tautona = gapa
4• go nyala mosadi *Mosimane yo a re o batla go tsaya ngwana wa moagisani wa rona*

tsaya boagedi •leele• go nna moagi wa lefatshe le lengwe

tsaya boipuso •leele• go simolola go ipusa ga lefatshe le sa busiwe ke lefatshe le lengwe *Ka 1966 ke fa lefatshe la Botswana le tsaya boipuso*

tsaya botshelô •leele• go fedisa botshelo kgotsa go bolaya *Bolwetse bo itshokile jwa bo jwa tsaya botshelo jwa ngwana*

tsaya dikgang •leele• **1**• go tlotla *Basadi ba fitlhile ba ema fa*

kgorong ya jarata ba tsaya dikgang ka lobaka 2° go ya mo lefelong motho a bo a gatisa ditiragalo tse di kgathisang mo lefelong leo, a bo a di isa kwa seromamoeng kgotsa kwa dikoranteng Tiro ya go tsaya dikgang e monate

tsaya dikobô ka dikgoka •leele• go tlhakanela dikobo le motho ntleng ga go rata ga gagwe *Mosimane yo o lebisitswe molato wa go tsaya dikobo ka dikgoka* = thubetsa, betêlêla

The decision to enter idioms and proverbs as separate headwords was taken because it is believed that such a strategy is linguistically motivated and especially user-friendly. The user is not left to guess under which headword the idiom or proverb is subsumed.

5.8 Middle section

The dictionary also comprises a rich middle section full of educationally and culturally important material. Amongst these are pages that explain punctuation marks, a detailed presentation of the etymology of Setswana months, kinship terminology, and the presentation of cattle colours, ear markings, cattle horn shapes and others in full colour.

6. Conclusion

This paper has presented Setswana's fourth monolingual dictionary *Tlhalosi ya Medi ya Setswana*. It is the first Setswana dictionary to carry the title *tlhalosi*. The dictionary is situated within a broad Setswana lexicographic tradition which dates to the 1800s. It has discussed some of its key features such as frequency mark-up and dialectal representation. In its compilation a 16 million word corpus, Moe's semantic classification strategy and the use of already published material in the Setswana language have been used. A Setswana monolingual dictionary tradition is still developing. However, there is much to be learnt from developments in lexicography elsewhere. Developments in English lexicography and developments in lexicography in the African languages such as Shona and Swahili can be instructive to the development of Setswana lexicography.

References

- Andersson, L.-G. and T. Janson. 1997. *Languages in Botswana; Language Ecology in Southern Africa*. Gaborone: Longman Botswana.
- Béjoint, H. 2000. *Modern Lexicography: An Introduction*. Oxford/New York: Oxford University Press.
- Biber, D. 1993. Using Register-diversified Corpora for General Language Studies. *Computational Linguistics* 19(2): 219-241.

- Biber, D.** 1994. An Analytical Framework for Register Studies. Biber, D. and E. Finegan (Eds.). 1994. *Sociolinguistic Perspectives on Register*: 31-56. Oxford: Oxford University Press.
- Brown, J.T.** 1875. *Lokwalo loa Mahúké a Secwana le Seeneles*. Johannesburg: Yates and Alexander, for the London Missionary Society.
- Brown, J.T.** 1925. *English Dictionary*. Johannesburg: Pula Press.
- Campbell, J.** 1815. *Travels in South Africa, Undertaken at the Request of the Mission Society*. London: Black, Parry and Co.
- Cole, D.T.** 1955⁹. *An Introduction to Tswana Grammar*. Cape Town: Longman.
- Cole, D.T.** 1995. *Setswana — Animals and Plants (Setswana — Ditschedi le Dithhare)*. Gaborone: The Botswana Society.
- Cole, D.T. and L. Moncho-Warren.** 2012. *Macmillan Setswana and English Illustrated Dictionary*. Johannesburg: Macmillan South Africa.
- Créissels, J. and A.M. Chebanne.** 2000. *Dictionnaire Française–Setswana, Thanodi Sefora–Setswana*. Mogoditshane: Tasalls Publishing.
- Dent, G.R.** 1992. *Compact Setswana Dictionary: English–Setswana, Setswana–English*. Pietermaritzburg: Shuter and Shooter.
- Joubert, P.J. and M.J. Mangokoane.** [1975–1978]. *Lenanentšú-tlhalosi: Seafrikaans–Seisemane–Sesotho sa Leboa*. Johannesburg: Suid-Afrikaanse Uitsaaikorporasie.
- Kgasa, M.L.A. (Ed.).** 1976. *Thanodi ya Setswana ya Dikole*. Cape Town: Longman Penguin Southern Africa.
- Kgasa, M.L.A. and J. Tsonope.** 1998. *Thanodi ya Setswana*. Gaborone: Longman.
- Kilgarriff, A.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Kilgarriff, A., P. Richly, P. Smrz and D. Tugwell.** 2004. The Sketch Engine. Williams, J. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 105-116. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud. Available online at: <http://www.lexmasterclass.com/people/akcv.htm>.
- Lichtenstein, H.** 1815. *Travels in Southern Africa, in the Years 1803, 1804, 1805 and 1806. Volume II*. Translated from the original German by Anne Plumptre. London: Henry Colburn.
- Louw, J.P. and E.A. Nida.** 1989. *Greek–English Lexicon of the New Testament: Based on Semantic Domains. Volume 1*. New York: United Bible Societies/Kaapstad: Bybelgenootskap van Suid-Afrika.
- Mareme, G.B.** 2008. *Thanodi ya Setswana*. Pietermaritzburg: Nutrend Publishers.
- Mathangwane, J.T.** 2008. The Treatment of Loanwords in Setswana Dictionaries. Otlogetswe, T.J. and M.L.A. Kgasa. 2008. *A Pioneer Setswana Lexicographer*: 133-146. Cape Town: Centre for Advanced Studies of African Society (CASAS).
- Matumo, Z.I.** 1993. *Setswana–English–Setswana Dictionary*. Gaborone: Botswana Book Centre and Macmillan Botswana Publishers.
- Miller, O.B.** 1936. *Check-lists of the Forest Trees and Shrubs of the British Empire: No. 6. Bechuanaland Protectorate*. Oxford: Imperial Forestry Institute.
- Moe, R.** 2001. Lexicography and Mass Production. *Notes on Linguistics* 4(3): 150-156.
- Moe, R.** 2003. Compiling Dictionaries Using Semantic Domains. *Lexikos* 13: 215-223.
- Mogapi, K.** 1984. *Thutapuo ya Setswana*. Gaborone: Longman.

- Mojela, V.** 2007. *Pukuntšuthaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend Publishers.
- Murdock, G.P. et al.** 1987⁵. *Outline of Cultural Materials*. New Haven: Human Relations Area Files.
- Otlogetswe, T.J.** 2007. *English–Setswana Dictionary*. Gaborone: Pentagon.
- Otlogetswe, T.J.** 2008. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Otlogetswe, T.J.** 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.
- Peters, M.A. and M.M. Tabane.** 1982. *Bibliography of the Tswana Language/Bibiliokerafi ya fine ya Setswana*. Pretoria: State Library.
- Prinsloo, D.J.** 2004. Revising Matumo's *Setswana–English–Setswana Dictionary*. *Lexikos* 14: 158-172.
- Rasebotsa, N.** 2008. M.L.A. Kgasa; The Legacy of my Father. Otlogetswe, T.J. and M.L.A. Kgasa. 2008. *A Pioneer Setswana Lexicographer*: 26-39. Cape Town: Centre for Advanced Studies of African Society (CASAS).
- Roget, P.M.** 1958, 1985. *Roget's Thesaurus*. Harmondsworth: Penguin Books.
- Salt, H.** 1814. *A Voyage to Abyssinia and Travels into the Interior of that Country*. Philadelphia: M. Carey, and Wells and Lilly.
- Scott, M.** 2004–2006. *Oxford WordSmith Tools. Version 4.0*. Oxford: Oxford University Press.
- Selolwane, G.D.** 2004. Ethnic Structure, Inequality and Governance of the Public Sector in Botswana. *UNRISD Project on Ethnic Structure, Inequality and Governance of the Public Sector*.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.M. (Ed.)**. 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins ELT.
- Snyman, J.W., J.S. Shole and J.C. le Roux.** 1990. *Dikišinare ya Setswana–English–Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.
- Van Warmelo, N.J.** 1931 *Kinship Terminology of the South African Bantu*. Ethnological Publications. Volume II. Pretoria: Government Printer.
- Volz, S.** 2003. European Missionaries and Tswana Identity in the 19th Century. *Pula: Botswana Journal of African Studies* 17(1): 3-19.