

Treatment of Spelling Variants in Setswana Monolingual Dictionaries

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

Abstract: This paper argues that the Setswana language is characterised by spelling variants which are a consequence of multiple factors. It considers spelling variants found amongst individual words as well as those found in multi-word expressions (MWEs). It argues that spelling variation may be a result of historical fissions and amalgamations of the Batswana groups as well as borrowings from adjacent languages such as Afrikaans and English. The paper considers how three monolingual Setswana dictionaries of the past twenty years, *Thanodi ya Setswana* (Kgasa and Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012) have lemmatised spelling variants. The paper argues that spelling variants must be included in a general monolingual dictionary and that how such variants are handled must be informed by frequency information from corpus data. The paper concludes by proposing three strategies for addressing variation in MWEs where a difference between the two or more MWEs occurs because of a single word in the MWE or where variation is caused by the presence or absence of a word in a MWE. The third solution applies to cases where the variants differ in too many places such that it would be much more elegant to treat them as separate entries.

Keywords: SPELLING VARIATION, DIALECT, SETSWANA CORPUS, MULTI-WORD EXPRESSION, BORROWING, HISTORY, MONOLINGUAL DICTIONARY, LEMMATISATION, CROSS-REFERENCING

Opsomming: Die hantering van spellingvariante in eentalige Setswana woordeboeke. Hierdie artikel argumenteer dat die Setswanataal gekenmerk word deur spellingvariante wat die gevolg is van verskeie faktore. Oorweging word geskenk aan spellingvariante wat gevind word by individuele woorde sowel as dié wat in meerwoordige uitdrukkinge gevind word (MWU's). Daar word geargumenteer dat spellingvariasie die gevolg kan wees van historiese verdelings en vermengings van die Batswanagroepe, sowel as ontlenings aan aanliggende tale soos Afrikaans en Engels. Daar word gekyk na die lemmatisering van spellingvariante in drie eentalige Setswanawoordeboeke van die afgelope twintig jaar, *Thanodi ya Setswana* (Kgasa en Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) en *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). Hierdie artikel argumenteer dat spellingvariante in 'n algemene, eentalige woordeboek ingesluit moet word en hoe hierdie variante hanteer moet word deur gebruik te maak van frekwensie-inligting uit korpusdata. As slot word drie strategieë aangebied wat variasie in MWU's aanspreek, waar daar 'n verskil tussen die twee of meer MWU's bestaan as gevolg van 'n enkelwoord in die MWU of waar variasie veroorsaak word deur die teenwoordigheid of afwesigheid van 'n woord in 'n MWU. Die derde strategie geld vir gevalle waar die variante in soveel opsigte verskil dat dit 'n beter opsie sal wees om hulle as aparte inskrywings te hanteer.

Sleutelwoorde: SPELLINGVARIASIE, DIALEK, SETSWANA KORPUS, MEERWOORDIGE UITDRUKKING, ONTLENING, GESKIEDENIS, EENTALIGE WOORDEBOEK, LEMMATISERING, KRUISVERWYSING

Introduction

Natural languages are characterized by numerous variants. The variants may be pronunciation variants, as those found, for instance, in words such as *data* (dɛɪtə or dɑ:tə), *potato* (pətɛɪtəʊ or pətɛɪtəʊ) and *either* (aɪðə or i:ðɜːr) (Wells 2000). There are also spelling variants such as distinctions between American and British spelling peculiarities; found in terms such as *color* and *colour* and *behavior* and *behaviour*. The Setswana language like all natural languages is characterised by variants of various kinds. Some of the variants are dialectal (Sutherland and Katamba 1996: 565) while others are spelling variants. By dialects we refer to features in an individual's speech that are associated with pronunciation, vocabulary, grammar and idiom (Honey 1997: 3; Crystal and Davy 1969: 67). A dialect may also be defined by other factors, such as social class or education. This paper restricts itself to the investigation of Setswana spelling variants most of which are a consequence of phonological features of different dialects. The paper argues that Setswana is characterized by multiple variants that pose a challenge to Setswana lexicography. The challenge that confronts a lexicographer is whether all variants of a form should be entered in a dictionary and if they were to be entered, how they would be treated in a monolingual Setswana dictionary.

First, we outline historical events which have had a bearing on spelling variation. We then demonstrate, with much evidence, that spelling variation is common in Setswana texts. The paper then considers how three monolingual Setswana dictionaries of the past twenty years have dealt with spelling variation. The dictionaries used are *Thanodi ya Setswana* (Kgasa and Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). We then measure the frequency of variants in a twenty million-word Setswana corpus to determine which variant is to be the primary form in the dictionary. Variants of multi-word expressions are discussed and three strategies of how such variants could be lemmatized proposed. This paper therefore answers a specific question: How should variants be treated in Setswana dictionaries? In answering this specific question the paper proposes different strategies of how variants should be handled in a general Setswana dictionary.

Sources of variation in Setswana

Spelling variation in Setswana words and expressions is pervasive and is a consequence of numerous causes. Some of the spelling variation in Setswana has historical roots. It may be traced to the historical splits that occurred hundreds of years ago amongst the various Batswana groups. The civil wars, inter-

nal dissension and drought amongst the Batswana resulted in a permanent splitting of Batswana tribes which over the years led to dialectal variation (Otlogetswe 2014).

Below we discuss briefly how the separations between different Batswana groups gave way to dialectal variation amongst them.

Historical conditions for dialects

Dialects are caused largely by isolation and distance. Speakers of the same language who are separated from one another by a border, geographical distance, a hill, or a political border fence will end up developing dialects of the same language. If through separation, speakers come into contact with speakers of other languages, their language will continue to change over time. For Setswana, this is evident in the development of the Sekhurutshe dialect (which has developed from Sehurutshe), in northern Botswana which has been influenced extensively by Kalanga. Setawana has also been influenced by Seyei and languages in the Maun/Chobe areas of Botswana. Setswana in general has borrowed words extensively from English and Afrikaans.

Amongst the Batswana, civil wars, internal dissension and drought resulted in a permanent splitting of tribes. There is even a widely-held theory that the name Batswana is derived from the reciprocal verb stem *-tšwana*, (come or go out from one another, to separate) suggesting that its meaning is "(the offshoots or separatists), reference being made either to the separation of the Batswana from the main Bantu (or Sotho) stock to which they originally belonged, or to the separation from one another of the various tribes which we know today" (Cole 1955: xxi). While there is no compelling linguistic evidence to support this argument, there is however sufficient evidence that secession was common amongst the Batswana (Schapera 1963: 164). It is no wonder that Tlou and Campbell (1997: 96) characterise the history of the Batswana thus: "This is the history of the Batswana: groups of people splitting up and then other groups joining together." Such splits were a consequence of many factors, amongst these droughts which made subsistence difficult. Family and tribal feuds also contributed to more splits and separations. Since separation creates boundaries between people and boundaries between people create dialects of the same language, we now discuss the separation that occurred amongst the Batswana that has given way to dialectal variation.

The splits amongst the Batswana will be expounded through a discussion of the historical splits amongst the Bahurutshe. A comprehensive discussion of Batswana splits and their linguistic implications is found in Otlogetswe (2014). Much of the discussion that follows relies heavily on Ngcongco (1979) and Tlou and Campbell (1997).

Historically the Bahurutshe and Bakwena used to be a single group termed the Baphofu Confederacy (Tlou and Campbell 1997: 97). This group included the ancestors of the Bahurutshe, Bakwena (including Bangwaketse

and Bangwato), Batlharo, Bakgatla and some Bapedi. Towards the end of the 15th century the Confederacy began to disintegrate. First, the Batlharo separated and later a group led by Mokgatla seceded moving northwards to live with the Bapedi and later established an independent settlement.

Around 1500, the remaining Baphofu under Malope were living in upper Limpopo near Mabjanamatshwana (near current Pretoria). It is believed that Malope had heirs in two houses: the first born child in Malope's senior house was a daughter, Mohurutshe, while the first born child in the second house was a son, Kwena. A dispute arose as to whether the chiefdom should be ceded into the hands of the eldest child in the senior house despite her being female, or whether the leadership should be kept male by appointing the senior son of the second house as chief. Following Malope's death, the dispute caused a split around 1475 to 1505. The followers of Mohurutshe were forced to leave Mabjanamatshwana and moved south as a separate group, with a separate totem (the baboon).

Between five and seven generations after the separation of the Bahurutshe and the Bakwena (around 1625–1655), while Mogopa was still ruling the Bakwena, a terrible famine which was famously termed '*tlala e e boitshegang*' scattered and dispersed the Bakwena clans far and wide. As a result of this famine, many Bakwena clans migrated south of the Lekwa or Vaal River into the modern Free State.

Mogopa and the remaining Kwena groups, which included the Modimesana clans and those that later formed the Botswana group of the Bakwena migrated to Mabjanamatshwana along the Odi River to its confluence with Madikwe and there built a settlement named Rathatheng. After a period of very strenuous or difficult existence owing to scarcity of food and water, Mogopa migrated back to Mabjanamatshwana in the modern Brits district of the former Transvaal.

Thus, partly as a result of the droughts and famines that occurred during the generation c.1625–c.1655, two Kwena kingdoms in the former western Transvaal emerged. These were the Bakwena-Mogopa based at Mabjanamatshwana and the Bakwena-Kgabo at Rathatheng.

Splits caused by droughts and famines also applied to the Bahurutshe state which gave birth to the Manyana and Gopane chiefdoms, the Bakaa and the Phuduhutswana-Tlhaping hived off from the Barolong-Tshidi, and the Bakwena-Modimosana split up and formed the four chiefdoms known as Ramanela, Maaka, Mmatau and Matlhaku. It is believed that it was at Rathatheng that Kgabo II was succeeded by his son Motshodi although according to some traditions, he (Kgabo II) led the migration of his followers across the Madikwe into present-day Botswana.

Towards the end of the long reign of Motshodi, the Kwena-Kgabo kingdom broke up. From this fission two new independent states came into being; the Ngwato and Ngwaketse kingdoms. Parsons (1973: 84) estimates the Ngwato secession to have occurred around 1790. Towards the end of the 18th century a group of Bangwato led by Tawana seceded to form an independent tribe on the shores of Lake Nghabe.

The above discussion gives a broad picture of some of the Batswana splits and fissions which subsequently gave rise to distinct dialectal features some of which are discussed in the development of this paper.

Some phonological differences between Setswana dialects

Because of the various separations amongst the Batswana, there are a number of linguistic variations that developed. Many of these are lexical while others are phonological with lexical implications. We therefore consider a few phonological differences between the various Batswana groups. By phonology we refer to the characteristic pronunciation patterns of a speech community. We only discuss those which have had impact on written Setswana. Phonological differences are important markers of linguistic variation and have been previously used to differentiate dialects (Batibo 1999; Cole 1955). Phonological features discussed here are only those that are reflected in certain spelling peculiarities.

a. The [tl/t] and [t^h/t^h] distinction

One of the obvious differences between the northern Setswana dialects (Sengwato and Setawana found around the areas of Serowe and Maun) and central dialects (Sengwaketse, Serolong and Sehurutshe found around Kanye and Mahikeng areas) is the distinction between [t^h/t] and [t^h/t^h] as in *batla/bata* (want) and *kgotlha/khotha* (poke). Northern Setswana uses [t^h/t] where all other Setswana dialects use [t^h/t^h] as well as where central dialects use [tl/t] (Batibo 1999). The northern Setswana dialect is sometimes mistakenly labelled by some South Africans Setswana speakers as *Setswana sa Botswana* (Setswana spoken in Botswana). This is because the northern Setswana dialect speakers, Bangwato and Batawana, are found only in Botswana and not in South Africa compared to Bakgatla ba ga Kgafela, Barolong and Bahurutshe who are found both in Botswana and South Africa. Table 1 presents comparative data of how Northern Setswana words are pronounced compared to central Setswana dialect speech.

Table 1: A comparison [tl] and [t] in Sengwato and other dialects

Sengwato dialect	Southern dialects	English
<i>thaga</i>	<i>tllhaga</i>	grass
<i>othaya</i>	<i>otlhaya</i>	discipline
<i>setha</i>	<i>setllha</i>	bladder
<i>thaba</i>	<i>tllhaba</i>	pierce; kill
<i>ta</i>	<i>tla</i>	come
<i>bata</i>	<i>batla</i>	search for
<i>tota</i>	<i>tlotla</i>	respect

Phonologically, this manner of pronunciation has led to lexical ambiguity between certain words which does not exist elsewhere in other Tswana dialects. Table 2 demonstrates the challenges that may result as a consequence of pronouncing [tl] as [t] and [tʰ] as [tʰ].

Table 2: Ambiguity as a result of the conflation of [tl] and [t]

Southern dialect	Sengwato	Ambiguous with
<i>batla</i> "want"	<i>bata</i>	<i>bata</i> (bath)
<i>tladi</i> "thunderbolt"	<i>tadi</i>	<i>tadi</i> (striped cat)
<i>tlala</i> (fill)	<i>tala</i>	<i>tala</i> (green)
<i>tlhaba</i> (pierce)	<i>thaba</i>	<i>thaba</i> (hill)
<i>tlhaka</i> (letter)	<i>thaka</i>	<i>thaka</i> (pupil of eye)
<i>tlhapa</i> (bath/wash)	<i>thapa</i>	<i>thapa</i> (employ)

Setswana orthographies over the years have standardized the Setswana writing system so that though northern Setswana speakers can speak using [tʰ/t] instead of [tʰ/tl], in writing they are taught to write [tʰ/tl] instead of [tʰ/t] (Ministry of Education 1981; Chebanne 2008 and Chebanne et al. 2008). Although this is the case, it is common that many of the writings of northern Setswana speakers reflect their speech peculiarities resulting in lexical ambiguity. The problem outlined in this section is much more common in Botswana than in South Africa since the northern Setswana dialect is spoken only in Botswana.

The challenge for a Setswana lexicographer in this instance is not severe since the problem is limited to a northern Setswana dialect and it is rare in written text though it exists phonologically and in some students' essays and in the social media such as Facebook text.

b. The difference between [lo-] and [le-]

One conspicuous dialectal difference between northern and central Setswana varieties is the distinction between the noun class singular prefix [le-] of class 5 nouns and [lo-] singular prefix of class 11. In the central dialects such as Sengwaketse and Serolong the distinctions between the two are still maintained while in the northern dialects such as Sengwato and Setawana the distinctions have been collapsed into singular prefix [le-] of class 5. This difference has been identified by Cole (1955). All the words which in other dialects are class 11 nouns are class 5 nouns in northern dialects. This is illustrated in Table 3 below.

Table 3: The [lo-] and [le-] in Setswana dialects

Southern dialects	Northern dialects	Gloss
<i>logong</i>	<i>legong</i>	wood
<i>lokotswana</i>	<i>lekotswana</i>	wall
<i>lobadi</i>	<i>lebadi</i>	door
<i>lofeelo</i>	<i>lefeelo</i>	broom

According to general linguists Setswana books (e.g. Mogapi 1984; Cole 1955) and Setswana orthography, the determination of whether a word takes a [le-] or [lo-] prefix is based on the plural prefix that the word takes (Chebanne et al. 2008). Class 11 nouns when pluralised take a class 8 prefix [di-] while class 5 nouns when pluralised take a class 6 prefix [ma-] (Cole 1955; Chebanne et al. 2008). This can be summarised as follows:

Singular noun class prefix	→	Plural noun class prefix
noun class 5 [le-]	→	noun class 6 [ma-]
noun class 11 [lo-]	→	noun class 8 [di-]

The rule above can be applied to the following linguistic data as follows:

Table 4: Formation of class 5 and 11 plurals

Noun	English meaning	Plural form	NOT
<i>logong</i>	wood	<i>dikgong</i>	<i>legong > magong</i>
<i>lobone</i>	light/bulb	<i>dipone</i>	<i>lebone > mabone</i>
<i>lobota</i>	wall	<i>dipota</i>	<i>lebota > mabota</i>
<i>lebante</i>	belt	<i>mabante</i>	<i>lobante > dipante</i>
<i>lerato</i>	love	<i>marato</i>	<i>lorato > dithato</i>
<i>leeto</i>	travel/visit	<i>maeto</i>	<i>loeto > dieto</i>

Although these linguistic rules exist, they have been applied inconsistently by Setswana writers. A lexicographer is therefore bound to find in a Setswana corpus spelling variation as a consequence of the inconsistent use of [le-] and [lo-] in the language. For instance in a twenty million-token Setswana corpus (Otlogetswe 2008), *logong* occurs 511 times while *legong* occurs 180 times. *Loeto* occurs 925 times while *leeto* occurs 558 times. To resolve these spelling variations, a lexicographer has two possible approaches to adopt. He can enter both variants in a dictionary and at the *wrong* spelling offer a cross-reference to the properly spelt word and indicate that the offered spelling is unacceptable. Alternatively, a lexicographer could only lemmatise those headwords which are consistent with Setswana spelling and treat grammatical information, including spelling rules, in the front matter of the dictionary.

c. Borrowings and variants

Setswana has borrowed extensively from both English and Afrikaans. Many borrowed words result with spelling variation since words are usually borrowed as they are heard. Since the Setswana language, especially in Botswana, Zimbabwe and Namibia lacks a language board which fixes spellings of borrowed terms, the variants have remained in Setswana texts. A good example is the word *computer* which has been borrowed into Setswana using any of the following spelling variants: *khompiutara*, *khomphiutara*, *khomputara*, *komputara* or *khompiuta*. The 1988 *Terminology and Orthography No.4* (Setshedi et al. 1988: 129) gives *khomphutara* as the Setswana equivalent for *computer*. The word *bus* has been borrowed into Setswana as *base*, whose pronunciation is closer to the English pronunciation of *bus*, and *bese*, whose pronunciation is significantly distinct from English. Table 5 shows borrowings from English and their variants in Setswana. The data is derived from a twenty million-word corpus.

Table 5: Variants of borrowed terms

Borrowed variants	Original borrowed term
<i>Baebele, Baebela, Bible</i>	Bible
<i>boronse, boronso</i>	bronze
<i>diabolo, diabolose</i>	devil
<i>enfolopo, enfelopo, enfolopo</i>	envelope
<i>Keresemese, Keresemose, Khirisimore</i>	Christmas
<i>khompiutara, khomphiutara, khomputara, komputara</i> or <i>khompiuta</i>	computer

The challenge that confronts a lexicographer is how a dictionary should capture the spelling variations that are found in a corpus. For a general dictionary variants are interesting and important to include in a dictionary so that users may find what they are looking for. This is particularly important if the variation occurs at the beginning of a word and therefore necessitates distinct lemmatization of variants in different parts of a dictionary. Variants are, however, a challenge since they take space in a dictionary and contribute no distinct meaning to their variant form. In some dictionaries (e.g. Otlogetswe 2012) spelling variants are treated as synonyms. However, treating them as synonyms implies that they constitute separate lexical items (lexemes) when in fact all variants constitute a single lexeme only having alternate forms. Such forms are therefore better marked as *variants*, that is, identical terms that only happen to have an idiosyncratic spelling. Writing about the English language Svensén observes that:

Many words have spelling variants. Since we are dealing here with well-established languages with a long lexicographic tradition, it is seldom difficult to ascertain which spelling is to be preferred and, consequently, is to appear as a

lemma. Frequent spelling variants should be included, although as a rule only in dictionaries entirely or partly intended for reception; in production dictionaries they are mostly unnecessary (Svensén 2009: 110).

Svensén's observation holds true for English as well as for Setswana. As has been demonstrated before, there are various reasons which may lead to variation in Setswana and as we will show in the development of this paper, sometimes it is not clear which variant form should be considered the preferred lemma to which other variants should be cross-referenced.

The data presented in the tables that follow demonstrates how widespread the variant phenomenon is across different Setswana word classes. In our study we have identified one thousand variants in *Tlhalosi ya Medi ya Setswana* dictionary (Otlogetswe 2012) which constitutes 6.5% of all of the dictionary head-words. We start with evidence of spelling variation amongst nouns.

Table 6: Variants amongst nouns

Variants	Word equivalent
<i>bokoso, lebokisi, lebokose, lebokoso, bokose, lepokisi</i>	box
<i>bolakaboroto, boroto</i>	blackboard, board
<i>kgonagalo, kgonafalo</i>	a possibility, a chance, a likelihood
<i>poresidente, moporesitente, moporesidente, poresitente</i>	president
<i>boperesiti, boperesita</i>	priesthood
<i>dithutopuisano, dithutapuisano, dithutopuisanyo</i>	workshops
<i>dithutopuisano, dithutapuisano, dithutopuisanyo</i>	blouse
<i>lekwaladikgwang, lekwalodikgang</i>	a newspaper
<i>letlebekwane, letemekwane, letemelekwane</i>	a dumbling
<i>bopelokhutshwane, bobelokhutshwane</i>	impatience
<i>sekarete, sekarese, sekerete, sakerete</i>	cigarette
<i>fatuku, faatuku, fatukwi</i>	a table cloth
<i>base, bese</i>	bus
<i>sepalamo, sepagamo, sepalangwa</i>	a transport
<i>moitseanape, moitsaanape, maitseanape</i>	an expert
<i>ntšhe, ntšhwe, mmantšhe, mmantšhwe,</i>	ostrich
<i>khompiutara, khomphiutara, khomputara, komputara</i>	computer
<i>leje, lejwe</i>	a stone
<i>lerapo, lesapo</i>	a bone
<i>Bopelotshetlha, bobelotshetlha</i>	greed
<i>Tlhotlhwa, tlhwatlhwa</i>	a price

It is not only nouns that are characterized by variants. Verbs also have numerous variants as shown in Table 7.

Table 7: Variants amongst verbs

Verb spelling variants	Meaning/equivalent
<i>atamêlana, atamalana, atumêlana</i>	come closer to each other
<i>abêlana, abalana</i>	share
<i>atolosa, katolosa</i>	widden
<i>Tšhemola, tšhamola, tšhwemola, tšhwamola</i>	slip away
<i>batola, bitola</i>	beat severely
<i>Fologa, gologa</i>	step down
<i>Gagoga, kgagoga</i>	tear apart
<i>Katogana, katologana</i>	spread out
<i>Kgorometsa, kgarametsa</i>	push
<i>Palama, pagama</i>	climb onto
<i>Pepetla, papetla</i>	quash
<i>Reetsa, theetsa</i>	listen
<i>Rutlolola, rutlomolola</i>	unroof
<i>Thela, tshela</i>	pour
<i>Thologa, tshologa</i>	pour out
<i>Tsisa, tlisa</i>	bring
<i>Thathologa, thanthologa</i>	unwind

The spelling variations are also found not just in common nouns but also in proper nouns as shown in Table 8.

Table 8: Variants of Names of persons and places

Setswana variants	English
<i>Aforika, Aferika</i>	Africa
<i>Baebele, Babele, Beibele</i>	Bible
<i>Ennyelane, Engelane, Enngelane, Enyelane</i>	England
<i>Faro, Farwe</i>	Pharaoh
<i>Balete, Bamalete</i>	Balete
<i>Baphoting, Baphuting</i>	Baphoting
<i>Borithani, Borithane</i>	Britain
<i>Diphalane, Phalane</i>	September
<i>Goleate, Goliata, Goliata, Goliathe</i>	Goliath
<i>Gouteng Gauteng</i>	The city of gold (Johannesburg)
<i>Hirikgong, Herikgong, Ferikgong, Firikgong</i>	January
<i>Keresemose, Keresemese, Keresemose, Khirisemose</i>	Christmas
<i>Lenyeseleman, lenyeseleman, leesemane, moesemane</i>	An Englishman man
<i>Mmantaga, Mantaga</i>	Monday
<i>Ngwanaitseele, Ngwanatsele, Ngwanaatsele</i>	November
<i>Jorotane, Joretane</i>	Jordan

Spelling variants are also to be found amongst interjectives in Setswana. We present these in Table 9.

Table 9: Variants on interjectives

Interjective variants	Interjective meanings
<i>uši, utšhi, kuši, kutšhi, iši, itšhi,</i>	an expression of pain
<i>haa! haaa! haahaaa!</i>	used to represent laughing
<i>haleluja, haleluya</i>	an expression of praise to god
<i>ijaa! ija!</i>	an expression of shock
<i>še! šeng!</i>	an expression of shock or surprise
<i>thakaa! thakanaa!</i>	an expression of shock or surprise
<i>tshi! tshikhi!</i>	an expression used to capture a cold environment
<i>tshuu! tshutshu! tšhuutšhuu!</i>	an expression used to show that one is burning
<i>tswée, tswéetswée, tswédintswerere</i>	a plea

There are also lexical variations which are as a consequence of the variation of [f] and [h] in Setswana words

Table 10: Variants caused by f/h alternation

Words	English
<i>fora, hora</i>	deceive
<i>funolola, hunolola</i>	untie
<i>futswela, hutswela</i>	add milk to porridge
<i>fuduga, huduga</i>	move to another place
<i>mohaladi, mofaladi</i>	a stranger from elsewhere
<i>sehatlha, sefatlha</i>	a naughty person
<i>lefuto, lehuto</i>	a knot
<i>bofalele, bohahalele</i>	carelessness
<i>mahura, mafura</i>	fat/oil
<i>firisa, hirisa</i>	to rent

The tables above have demonstrated how widespread variants are in Setswana across different word classes. There are variants of different types. Some variants are caused by a difference in a single vowel or consonant, for instance *fata/hata* (dig), *Gouteng/Gauteng* (Johannesburg), *Aforika/Aferika* (Africa), *reetsa/theetsa* (listen), *Jorotane/Joretane* (Jordan) etc. Other variants are caused by the presence of an additional syllable(s) in one of the variants, for instance, *Phalane/Diphilane* (October), *Ngwanatsele/Ngwanaitseele* (November), *Mmantaga/Mantaga* (Monday), *katogana/katologana* (part/move away from each other), and *bokoso/lebokoso* (a box).

Treatment of variation in Setswana monolingual dictionaries

In this section we test how three Setswana monolingual dictionaries have lemmatized a randomly selected group of words with spelling variants. We consider three dictionaries of the past twenty years: *Thanodi ya Setswana* (Kgasa and Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). Kgasa and Tsonope's (1995) dictionary was for a long time the definitive monolingual Setswana dictionary which was widely used in Botswana and South Africa. For years it was the prescribed Setswana dictionary in Botswana schools. Mareme (2007) is the largest Setswana dictionary to come out of South Africa. It has much in common with Kgasa and Tsonope (1995). It was produced by the Setswana Lexicographic Unit based at the North West University (the Mahikeng campus). Otlogetswe (2012) is the most recent of the three dictionaries compiled with a great dependency on a large Setswana corpus. The three dictionaries are therefore contemporary monolingual dictionaries which are in general circulation and use.

Table 11: Measure of variation in Setswana dictionaries

Word	Kgasa and Tsonope (1995)	Mareme (2007)	Otlogetswe (2012)
<i>lebele</i> (breast)	✓	✓	✓
<i>letsela</i> (breast)	✓	✓	✓
<i>iši</i> (ouch!)	✓	x	✓
<i>uši</i> (ouch!)	x	x	✓
<i>itšhi</i> (ouch!)	x	x	✓
<i>utšhi</i> (ouch!)	✓	x	✓
<i>gofele</i> (the youngest child)	x	x	✓
<i>gofejane</i> (the youngest child)	✓	✓	✓
<i>kgonotšwe</i> (thumb)	x	x	✓
<i>kgonotswe</i> (thumb)	x	✓	x
<i>kgonope</i> (thumb)	✓	✓	✓
<i>kgonojwe</i> (thumb)	x	✓	✓
<i>khomphiutara</i> (computer)	✓	x	✓
<i>khompiutara</i> (computer)	x	x	✓
<i>khomputara</i> (computer)	x	✓	✓
<i>khomphutara</i> (computer)	x	✓	x
<i>khomphuta</i> (computer)	x	x	x
<i>jala</i> (plant)	x	✓	✓
<i>jwala</i> (plant)	✓	x	✓
TOTAL	8	9	16

The results as shown in Table 11 indicate that different Setswana dictionaries adopt different strategies in their inclusion of variants. Of the twenty variants tested in Table 11 Kgasa and Tsonope (1995) have eight variants. Mareme

(2007) has nine while Otlogetswe (2012) has sixteen. The dominant approach with Kgasa and Tsonope (1995) seems to be a choice of one of the variants and including it as a headword while other variants are excluded though this approach is abandoned in the treatment of other variants such as *lebele* and *letsele*.

It isn't clear on what basis one of the variants was chosen amongst the many to refer to the other variants.

With Mareme (2007), there isn't much of a discernible pattern in the treatment of spelling variants since in some cases the variants are lemmatized as in the case of *letsele/lebele* and *kgonope/kgonojwe*, while in other cases one of the variants is lemmatized to the exclusion of one of the variants. For instance *jala* is lemmatized while *jwala* is not. *Khomputara* and *khomphutara* are lemmatized while *khomphiutara*, and *khompiutara* are not lemmatized. Otlogetswe (2012) tends to lemmatize many of the variants with the exception of only a few. The Setswana monolingual dictionaries therefore adopt different strategies in their lemmatization of spelling variants in the dictionary.

Below we give examples of how the variants have been treated in the three dictionaries. We only consider *lebele/letsele* (breast) from Kgasa and Tsonope (1995), *kgonojwe/kgonotswe* (thumb) from Mareme (2007) and *khomphiutara*, *khompiutara* and *khomputara* (computer) in Otlogetswe (2012) for brevity.

lê·bêlê TGG *ln/5*. ma-. nama e e mo sehubeng sa motho, mo basading e ipotokile, mo diphologolong e fa gare ga maoto kana mo mpeng; e kgona go ntsha maši mo ditsheding tse di namagadi

lê·tsêlê TGG *ln/5*. ma-. nama e e mo sehubeng e e amusang mo diphologolong tse di namagadi
(Kgasa and Tsonope 1995: 115, 134)

Kgasa and Tsonope (1995) do not link the two variants through cross-referencing; instead they offer the full treatment of tone, part of speech and definition to both *lebele* and *letsele*, though *lebele*'s definition is more elaborate than that of *letsele*. We now consider *kgonotswe/kgonojwe* variants in Mareme (2007).

kgonojwe (*ln*) (9/10) di monwana o motona wa leoto kgotsa seatla
kgonotswe (*ln*) (9/10) di- BONA **kgonojwe**
(Mareme 2007: 230)

Mareme (2007) links *kgonotswe* to *kgonojwe* through cross-referencing and does not offer any definition at *kgonotswe* only offering it at *kgonojwe*. He however does not link *kgonojwe* to *kgonotswe*. This means that when users consult *kgonojwe* they are unlikely to know that *kgonojwe* has a variant *kgonotswe*. We now move to consider the *khompiutara* variants.

khomphiutara /k^hõmp^híútàrá/ •*ln*. 9. n-, *10. *din*-• = KHOMPIUTARA
⇔ SeE: computer

khompiutara /k^h̀̀m̀p̀íútárá/ •*ln.* 9. *n-*, *10. *din-•* motšhine o o dirisiwang go boloka ditlhaka, ditshwantsho, le tse dingwe, o dirisiwa go kwala, dipalo le dipuisano tsa imeile jalojalo ⇐ SeE: computer
khomputara /k^h̀̀m̀p̀útárá/ •*ln.* 9. *n-*, *10. *din-•* = KHOMPIUTARA ⇐ SeE: computer
 (Otlogetswe 2012: 210)

Otlogetswe (2012) only offers a definition at *khompiutara* and then links *khomphiutara* and *khomputara* to *khompiutara* through cross-referencing. However at *khompiutara* he does not show the other variants to which *khompiutara* is linked. This means that when one consults a dictionary for *khompiutara* one is unlikely to know its variants.

It is however not clear if any of the Setswana dictionaries have handled the matter of cross-referencing of variants appropriately.

It is the argument of this paper that the decision of what to lemmatize and what constitutes a primary form to which other variants refer, should be based on corpus frequency. We also argue that to preserve space only one of the variants should have full lexicographic treatment and carry a definition or definitions in the case of polysemous entries. Such a variant, we call a primary form. Frequency list information can assist, not just in determining which variants to include, but also in deciding which of the variant forms is the primary form on which a full lexical treatment can be made. Below we test the treatment of *khompiutara*, *lebele* and *kgonotswe* variants against corpus evidence to determine the appropriate primary form of each. The table below presents the corpus frequencies of the variants

Table 12: Frequency of selected variants in a corpus

Variant	Frequency
<i>khomputara</i>	181
<i>khompiutara</i>	38
<i>khomphiutara</i>	6
<i>khompiotara</i>	6
<i>khomphuitara</i>	4
<i>khomphutara</i>	2
<i>lebele</i>	412
<i>letsele</i>	136
<i>kgonotswe</i>	60
<i>kgonope</i>	29
<i>kgonojwe</i>	17
<i>kgonotšwe</i>	11

The corpus data demonstrates that amongst the *khomputara*, *khompiutara*, *khomphiutara*, *khompiotara*, *khomphuitara* and *khomphutara*, variants, *khomputara* is

the most frequent variant with a frequency of 181 that qualifies to be entered in the dictionary as a primary variant form. It should therefore have the full meaning treatment with all the other variant forms linked to it through cross-referencing. At the *khomputara* entry all the other variants should be listed at the end of the entry as variant forms. This is so that a user who consults the dictionary meaning at *khomputara*, may also be aware of other variant forms.

Lebele is the most frequent variant with a frequency of 412 in the corpus. It would have the full meaning treatment in the dictionary as a primary variant with *letsele* linked to it through cross-referencing. The same approach will be applied to *kgonotswe* and its variants as above. *Kgonotswe* would receive a full lexicographic treatment with *kgonope*, *kgonojwe* and *kgonotšwe* cross-referenced to it since *kgonotswe* has the highest frequency comparatively. The treatment of *lebele/letsele* variants by Kgasa and Tsonope in which both variants are defined is therefore not recommended since it fails to show the link between the two variants. A more preferred approach is the one similar to that of Otlogetswe (2012) in the treatment of *khomputara* as the primary form and having all the other variants cross-referenced to it. Corpus evidence however shows that instead of *khompiutara* receiving full lexicographic treatment, as Otlogetswe (2012) has done, it is *khomputara* that must receive full treatment and have all the other variants cross-referenced to it since it is the most frequent of all the variants. The inclusion of variants cross-referenced to a primary form is attractive since it is user friendly and assists the learner (Svensén 2009). Moon prefers this approach also arguing that:

Some, particularly learners' dictionaries, help users by routinely setting cross-references to the place of treatment. This solves the problem of handling variations, and at least means that users are supported during their search for information. (Moon 1994: 342)

Linking variants by cross-referencing brings together related dictionary material which may be spread throughout the dictionary. It saves space by giving one of the variants full lexicographic treatment and linking related headwords.

Multi-word expression variants

Variants do not only exist at the spelling level of the word, but also occur at the level of multi word expressions. By multi word expression is meant:

... any word combination (adjacent or otherwise) that has some feature (syntactic, semantic or purely statistical) that cannot be predicted on the basis of its component words and/or the combinatorial processes of the language. Such units need to be included in any language description that hopes to account for actual usage (Bannard 2007: 1).

MWEs therefore include idioms, phrasal verbs, proverbs, compound words, etc. (Otlogetswe 2009). English examples include *ad hoc*, *by and large*, *kick the*

bucket, in step, take up, take off, shake up, telephone booth, pull strings, fresh air, fish and chips, salt and pepper etc. Setswana examples include *solegela molemo* (benefit), *kukega maikutlo* (be upset), *iphaga dikoro* (involve yourself in other people's business), *tsholetsa maoto/dinao* (walk faster), *opisa tlhogo* (cause trouble), *tsaya karolo* (participate), *tsaya tsia* (pay attention), *nna le seabe* (take part), and *ja monate* (enjoy).

Moon has observed that:

idioms are always difficult to treat lexicographically. This is not just because of the problems of variation and lexical form. There are other problems presented by idioms: how to convey the meaning and usages of what are essentially context-bound items, with vague or plastic meanings and heavy connotations (Moon 1994).

MWEs just like single word forms also display variation. For instance English has the following variants:

Get a raw deal	Hit the roof
Have a raw deal	Hit the ceiling
Have cold feet	At least
Develop cold feet	At the least
Have one's feet on the ground	At the very least
Keep one's feet on the ground	Shake in one's shoes
With one's feet on the ground	Shake in one's boots
With one's nose in the air	Quake in one's shoes
Stick one's nose in the air	Quake in one's boots
Drag one's feet	Quiver in one's boots
Drag one's heels	Quake in one's Doc Marten's

(Moon 1994: 90-100)

Just like English, Setswana also has MWEs that display variation. Examples include the following:

Table 13: Variation amongst MWEs

MWE VARIANTS	MEANING
a. <i>bolwetse jwa pelo</i> (Noun Phrase) b. <i>bolwetsi jwa pelo</i>	Heart disease
a. <i>koma ka tlhogo</i> (Idiom) b. <i>koma tlhogo</i>	To nod one's head
a. <i>kgomo e e maši ga e itsale</i> (Proverb) b. <i>e e maši ga e itsale</i>	Just because a parent is good, it doesn't mean that they will have good children

a. <i>kgwele ya dinao</i> (Noun Phrase) b. <i>kgwele ya maoto</i>	Football
a. <i>matlho a ngwana a roga mogolo ga a gake</i> (Proverb) b. <i>matlho a roga mogolo ga a gake</i>	A person with a secret is revealed by how they behave
a. <i>bana ba motho ba kgaogana tlhogo ya ntsi</i> (Proverb) b. <i>bana ba motho ba kgaogana tlhogo ya tsie</i>	People who associate in some form share the little that they have
a. <i>mogama kgomo tsa mafisa o di gama a lebile tsela</i> (Proverb) b. <i>mogama kgomo tsa mafisa o gama a gadimile tsela</i> c. <i>mogama kgomo ya mafisa o e gama a gadimile kwa morago</i>	A person who is using something that is borrowed, uses it knowing that it may be wanted back any time
a. <i>monate o tswa ka ditsebe</i> (Idiom) b. <i>monate o tswa ka phogwana</i>	This idiom is used to mean something (e.g. food) is very nice
a. <i>monna o bolawa ke se a se jeleng</i> (Proverb) b. <i>monna o jewa ke se a se jeleng</i>	It is the actions of an individual which bring them harm
a. <i>moswela Tebele ga a jewe o tshwana le moswela gagabo</i> (Proverb) b. <i>moswela gae ga a jewe o tshwana fela le moswela nageng</i>	People can live where they can best make a living even though it is not where they were born
a. <i>bodiba bo jeleng ngwana'a mmago sika ka bone o bo kakologe</i> (Proverb) b. <i>bodiba ba go ja ngwana wa ga mmago, e re o feta ka jone o bo dikologe</i> c. <i>bodiba ba go ja ngwana wa ga mmago, e re o feta ka jone o bo tlarologe</i>	An individual must learn from other people's mistakes and avoid danger

Treatment of MWEs in Setswana dictionaries

Table 13 has demonstrated the extent as well as types of variation in Setswana MWEs. As in variation amongst individual words, the MWEs such as idioms and proverbs have variations which pose unique challenges to their lemmatisation in a monolingual general dictionary.

Svensén observes that:

There are two options as regards the positioning of an idiom in dictionaries. Either it can be entered as an indication in the entry for one of the component words, or it can be entered as a lemma, and in the latter either as an independent lemma or sublemma. (Svensén 2009: 194)

Kgasa and Tsonope (1995) do not lemmatize the MWEs as headwords, they include them as subentries or in Svensén terminology, as sublemmata. For instance the idiom *go bona ka bobi jwa segokgo* (to see a little bit) is entered as a subentry of *bona* (see):

bôna GT tpt. fa o lebile sengwe ka matlho o a se bôna *go bona ka bobi
jwa segokgo = go bona go le go nnye
(Kgasa and Tsonope 1995: 25)

Mareme (2007) doesn't deviate from Kgasa and Tsonope's (1995) approach. For instance:

kakabolola (*ldr*) 1 go utlwise botlhoko ka go itaya kgotsa go tsietsa thata
Go kakabolola ditsebe: go betsa thata 2 go thibolola 3 go bulela
(Mareme 2007: 189)

In this example *go kakabolola ditsebe* (to beat severely) is a subentry of the headword *kakabolola* (unblock).

However Svensén argues that:

Presenting an idiom as an indication under one of its components, which is the traditional and probably still prevailing method, is actually contrary to the idiom's nature of an independent lexical item. (Svensén 2009: 194)

This is because the idiom or proverb is semantically non-compositional and usually semantically unrelated to the headword under which it is subsumed. For instance *kabolola ditsebe* is not semantically related to *kabolola* or *ditsebe*. It is therefore flawed to associate it with any of the words that constitute it. Otlogetswe (2012) adopts a different approach from Kgasa and Tsonope (1995) and Mareme (2007). He enters MWEs such as idioms and proverbs as independent lemmas in the dictionary without subsuming them under a headword. For instance:

botlhale jwa phala bo tswa phalaneng, phala e se nang phalana lesilo
•*seane*• bagolo ba ka thusiwa ke bana ka megopolo
(Otlogetswe 2012: 63)

This treatment of the MWE as a separate lemma is the one we support since an idiom is "an independent lexical item having an opaque meaning" (Svensén 2009: 194).

Treatment of MWE variants in Setswana dictionaries

Returning to the matter of entering variants in dictionaries, an examination of Kgasa and Tsonope (1995) and Mareme (2007) dictionaries reveal that they both exclude variants whilst Otlogetswe (2012) includes them. For instance the

case of *go itaya kgomo lenaka*:

itaya TTT | itaa tpt. > itaya -ile. utlwisa botlhoko ka go betsa ka letswele kana go ngata. * *go itaya kgomo lenaka* = *go bua selo fela jaaka se ntse kana nnete*
(Kgasa and Tsonope 1995: 72)

The idiom *go itaya kgomo lenaka* (to get something right) has multiple variants that are not captured in either Kgasa and Tsonope (1995) or Mareme (2007). These are *go opa kgomo lonaka* and *go otla kgomo lonaka* which are included in Otlogetswe (2012) though they are included as synonyms:

itaya kgomo lonaka •leele• go bua nnete jaaka e ntse = ÔPA MAGÔGWÊ MO NKONG, ÔPA KGOMO LONAKA.
(Otlogetswe 2012: 148)

ôtle kgomo lonaka •leele• = ÔPA MAGÔGWÊ MO NKONG, ÔPA KGOMO LONAKA.
(Otlogetswe 2012: 452)

ôpa kgomo lonaka •leele• go bua se e leng buammaaruri *Le fa o ne o bua fela o se na bosupi, o ne wa opa kgomo lonaka ka go diragetse fela jaaka o buile*
(Otlogetswe 2012: 450)

The evidence therefore demonstrates that the three Setswana dictionaries adopt different strategies to deal with multi-word expression variants. Kgasa and Tsonope (1995) and Mareme (2007) do not include variants while Otlogetswe (2012) does, although some of them are given a full treatment with meaning. In the following section we propose three different ways of handling multi-word expressions in Setswana.

Three strategies of handling MWE variants in a Setswana dictionary

We propose that multi-word expressions be handled in any one of the three different ways in a general dictionary.

1. Where a difference between the two or more multi-word expressions is just because of a single word in the MWE in the same part of a phrase then a forward slash [/] may be used in the entry of a MWE or fixed expression to save space to show alternates. This approach is the one used by the *Macmillan English Dictionary for Advanced Learners* (Rundell 2007) in the treatment of multi-word expressions, for instance, in the treatment of the word *framework*.

framework [...] 1. ...

- ◆ **provide/establish/create a framework** *Einstein's research provided much of the theoretical framework for particle physics* (Rundell 2007: 595)

For Setswana, this means that the following multi-word expressions will be treated in the following manner:

The variants *monate o tswa ka ditsebe* and *monate o tswa ka phogwana* only differ in terms of the use of the words *ditsebe* and *phogwana* in the same spot which engender variation between the two structures. We therefore propose that these be treated in the following manner:

monate o tswa ka ditsêbê/phogwana •*leele*• monate o o kwa godimo *Re ne re itumetse kwa moketeng, monate o tswa ka ditsebe phogwana*

This approach therefore avoids the approach that follows which uses much space in the dictionary.

monate o tswa ka ditsêbê •*leele*• monate o o kwa godimo *Re ne re itumetse kwa moketeng, monate o tswa ka ditsebe*

monate o tswa ka phogwana •*leele*• monate o mogolo = MONATE O TSWA KA DITSÊBÊ (Otlogetswe 2012: 376)

Additionally *monna o bolawa ke se a se jeleng* and *monna o jewa ke se a se jeleng* only differ in the use of the words *bolawa* and *jewa*. The proverb would therefore be represented as follows in the dictionary:

monna o bolawa/jewa ke se a se jeleng •*seane*• ditlamorago tsa sengwe di bakwa ke se motho a se dirileng *O ne a tshwanetse go solofela sone se ka gore monna o bolawa ke se a se jeleng*

The attractive matter about this approach is that it saves space since it doesn't unnecessarily repeat words.

2. The second solution to dealing with MWEs is that in instances where variation is caused by the presence or absence of a word in a MWE, then brackets may be used around a word that may be left out. For instance the variants *kgomo e e maši ga e itsale* and *e e maši ga e itsale* may be treated as

(kgomo) e e maši ga e itsale •*seane*• motsadi ga se gantsi a tshola bana ba ba ditiro dintle jaaka ene = E E MAŠI GA E ITSALÉ

This approach differs from the Otlogetswe (2012) approach below that consumes space.

kgomo e e maši ga e itsale •seane• motsadi ga se gantsi a tshola bana ba ba ditiro dintle jaaka ene = E E MAŠI GA E ITSALÉ (Otlogetswe 2012: 197)

and

e e maši ga e itsale •seane• ga se gore motho yo o ditiro dintle o tlaa nna le bana ba ba ditiro dintle jaaka ene *Ngwana yo ga a na botho mo go tlhabisang ditlhong, mme mmaagwe ke motho yo o maitseo tota; e le ruri e e maši ga e itsale* (Otlogetswe 2012: 91).

3. The third solution applies to cases where the variants differ in too many places such that it would be much more elegant to treat them as separate entries. For instance:

- moswela *Tebele* ga a jewe o tshwana le moswela *gagabo*
- moswela *gae* ga a jewe o tshwana *fela* le moswela *nageng*

The two variants differ in *Tebele/gae* and *gagabo/nageng* and one of the variants has *fela* that is absent in the other. The argument here is that it is inelegant to combine solution one and two above to come up with a single entry as below:

moswela Tebele/gae ga a jewe o tshwana (fela) le moswela gagabo/nageng •seane• motho o gololesegile go nna gongwe le gongwe kwa a ka tshelang sentle teng mo lefatsheng, le fa e se gagabo ka motsing a swang ga a ka ke a itemogela pharologano epe ka a tlaabo a setse a sule

The above entry is confusing since it ends up generating variants that do not exist in Setswana. *Moswela Tebele/gae ga a jewe o tshwana (fela) le moswela gagabo/nageng* generates the following variants:

1. moswela Tebele ga a jewe o tshwana le moswela nageng
2. moswela Tebele ga a jewe o tshwana fela le moswela nageng
3. moswela Tebele ga a jewe o tshwana le moswela gagabo
4. moswela Tebele ga a jewe o tshwana fela le moswela gagabo
5. moswela gae ga a jewe o tshwana le moswela nageng
6. moswela gae ga a jewe o tshwana fela le moswela nageng
7. moswela gae ga a jewe o tshwana le moswela gagabo
8. moswela gae ga a jewe o tshwana fela le moswela gagabo

In this instance, the use of brackets and forward slashes complicates matters and results with unintended variants that do not exist in the language. This paper therefore proposes that in this instance both variants

should be entered as separate headwords and one of the variant forms be cross referenced to the primary form as demonstrated below:

moswêla gae ga jewe, o tshwana fêla le moswêla nageng •seane•
motho o gololesegile go nna gongwe le gongwe kwa a ka tshelang
sentle teng mo lefatsheng, le fa e se gagabo ka motsing a swang ga a
ka ke a itemogela pharologano epe ka a tlaabo a setse a sule

and

moswêla Tebele ga a jewe o tshwana le moswêla gagabô •seane• =
MOSWÊLA GAE GA JEWE, O TSHWANA FÊLA LE MOSWÊLA
NAGENG

Conclusion

Variants in the Setswana language are important since many words and expressions are written in a variety of ways. This may be because of dialectal variation or as a consequence of pronunciation pattern amongst speakers. Some variation may be because of the way words are borrowed into the language. The variants occur even though Setswana orthographies have been developed over the years both in Botswana (Chebanne et al. 2008) and South Africa (PanSALB 2008) as well as by independent research centres such as Centre for Advanced Studies of African Society (Chebanne et al. 2003). This paper has shown that variants are treated differently in Setswana dictionaries – mostly in a non-consistent manner. It has argued that a well-designed Setswana corpus is an important source in exposing variants that could be lemmatised in a dictionary and it is useful in showing the variant to which a cross-reference should be made. This paper has also proposed a variety of ways of treating MWEs in a monolingual dictionary.

References

- Bannard, C.** 2007. A Measure of Syntactic Flexibility for Automatically Identifying Multi-word Expressions in Corpora. *Proceedings of the ACL Workshop on a Broader Perspective on Multi-word Expressions, Prague, Czech Republic, June 2007*: 1-8.
- Batibo, H.M.** 1999. A Lexicostatistical Survey of the Setswana Dialects Spoken in Botswana. *South African Journal of African Languages* 19(1): 2-11.
- Chebanne, A.** 2008. A Sociolinguistic Perspective of the Indigenous Communities of Botswana. *African Study Monographs*. 29(3): 93-118.
- Chebanne, A., M. Mokitimi, L. Matlosa, R. Nakin, M.W. Nkolola, S. Mokgoatsana and M. Machobane.** 2003. The Standard Unified Orthography for Sotho/Tswana Languages: Botswana, Lesotho, Namibia, South Africa, Zambia and Zimbabwe. Cape Town: CASAS.
- Chebanne A., T.J. Otlogetswe, P. Seloma, A. Molosiwa, B. Moatlhodi, N. Kgolo and M. David.** 2008. *Mokwalo o o lolameng wa Setswana*. Gaborone: Pentagon Publishers.

- Cole, D.T.** 1955. *An Introduction to Tswana Grammar*. Cape Town: Longman.
- Crystal, D. and D. Davy.** 1969. *Investigating English Style*. London, Longman.
- Honey, J.** 1997. *Language is Power: The Story of Standard English and its Enemies*. London: Faber and Faber.
- Kgasa, M.L.A. and J. Tsonope.** 1995. *Thanodi ya Setswana*. Gaborone: Longman Botswana.
- Mareme, G.B.** 2008. *Thanodi ya Setswana*. Pietermaritzburg: Nutrend Publishers.
- Ministry of Education.** 1981. *Setswana Standard Orthography of 1981*. Gaborone: Ministry of Education.
- Mogapi, K.** 1984. *Thutapuo ya Setswana*. Gaborone: Longman.
- Moon, R.** 1994. *Fixed Expressions and Text: A Study of the Distribution and Textual Behaviour of Fixed Expressions in English*. Unpublished Ph.D. Thesis, University of Birmingham.
- Ngcongco, L.D.** 1979. Origins of the Tswana. *Pula: Botswana Journal of African Studies* 1(2): 21-46.
- Otlogetswe, T.J.** 2008. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Otlogetswe, T.J.** 2009. Populating Sub-entries in Dictionaries with Multi-word Units from Concordance Lines. *Lexikos* 19: 446-457.
- Otlogetswe, T.J.** 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.
- Otlogetswe, T.J.** 2014. Dialectical Variations among Tswana Peoples of South Africa and Botswana: An Historical Linguistic Perspective. *Boleswa Journal of Theology, Religion and Philosophy* 4(2): 326-346.
- PanSALB.** 2008. *Melawana ya mokwalo le mopeleto Setswana*. Pretoria: PanSALB.
- Parsons, Q.N.** 1973. On the Origins of the bamaNgwato. *Botswana Notes and Records* 5: 82-103.
- Rundell, M. (Ed.).** 2007. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan.
- Schapera, I.** 1963. Kinship and Politics in Tswana History. *Journal of the Royal Anthropological Institute* 93(2): 159-173.
- Setshedi, J.E., J.H.K. Malao, O.S. Mathule, C.D. Ditsele, P.M. Rakgokong, W.M. Legotlo, G.R. Letsholo, J.W.P. Mashike, R.O. Komane, A.M.I. Rabaji, T.L. Tsambo and K.D. Rahlaga.** 1988. *Setswana Terminology and Orthography No 4*. Pretoria: Department of Education and Training and Department of Education.
- Southerland, R.H. and F. Katamba.** 1996. Language in Social Contexts. O'Grady, W., M. Dobrovolsky and F. Katamba (Eds.). 1996. *Contemporary Linguistics: An Introduction*: 540-590. Third edition. London/New York: Longman.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Tlou, T. and A. Campbell.** 1997. *History of Botswana*. Gaborone: Macmillan Botswana Publishing.
- Wells, J.C.** 2000. *Longman Pronunciation Dictionary*. Second Edition. Harlow: Pearson Education.