The Lemmatisation of Nouns in Tshivenḓa Dictionaries

Mashudu Nthambeleni *(mashudu.nthambeleni@univen.ac.za)*
and
Nelson Mbulaheni Musehane *(musehanen@univen.ac.za)*
M.E.R. Mathivha Centre for African Languages, Arts and Culture,
School of Human and Social Sciences, University of Venda,
Thohoyandou, Limpopo, South Africa

Abstract: Tshivenḓa is one of the official languages of South Africa, mostly spoken in Limpopo Province at Vhembe District. From 1994, when the Republic of South Africa became a democratic government, speakers of the Tshivenḓa language spread to all nine provinces of South Africa. Tshivenḓa has different word categories, like nouns, verbs, adjectives and many more. When words are looked up in dictionaries, one should know what types of categories are given, e.g. their spelling, pronunciation and meaning. If this is not clearly represented one would not be able to use such a dictionary. This paper seeks to investigate how nouns are lemmatised in Tshivenḓa dictionaries. Attention is given to word and stem lemmatisation. It also looks at the lemmatisation of singular and plural nouns, and the lemmatisation of deverbative and diminutive nouns. This will be accomplished by analysing published Tshivenḓa dictionaries in circulation.

Keywords: LEMMATISATION, NOUNS, TSHIVENḒA, DEVERBATIVE NOUNS, DIMINUTIVE NOUNS, SINGULAR, PLURAL

Opsomming: Die lemmatisering van substantiewe in Tshivenḓa woordeboeke. Tshivenḓa is een van die amptelike tale van Suid-Afrika wat hoofsaaklik in die Vhembe-distrik van die Limpopoprovinsie gepraat word. Sedert 1994, toe die Republiek van Suid-Afrika 'n demokratiese regering gekry het, het sprekers van Tshivenḓa na al nege provinsies van Suid-Afrika versprei. Tshivenḓa het verskeie woordsoorte, soos substantiewe, werkwoorde, adjektiewe en so meer. Wanneer woorde in woordeboeke nageslaan word, moet 'n mens weet watter datatipes aangebied word, bv. spelling, uitspraak en betekenis. As dit nie duidelik weergegee word nie, sal 'n mens nie in staat wees om so 'n woordeboek te gebruik nie. Hierdie artikel wil ondersoek hoe substantiewe in Tshivenḓa woordeboeke gelemmatiseer word. Daar word aandag geskenk aan woord- en stamlemmatisering. Daar word ook gekyk na die lemmatisering van enkel- en meervoudige substantiewe, asook die lemmatisering van deverbatiewe en diminutiewe substantiewe. Dit word gedoen deur bestaande Tshivenḓa woordeboeke te ondersoek.

Sleutelwoorde: LEMMATISERING, SUBSTANTIEWE, TSHIVENḒA, DEVERBATIEWE SUBSTANTIEWE, DIMINUTIEWE SUBSTANTIEWE, ENKELVOUD, MEERVOUD

1. Introduction

If dictionaries are not compiled according to the required acceptable standard,

users of such dictionaries would find it difficult to use them. In other words such dictionaries will not be user-friendly. There should be proper planning to determine what information about spelling, pronunciation as well as meaning should be included. This process is termed *lemmatisation*. Hartmann and James (1998: 83) define *lemmatisation* as "the reduction of a paradigm of variant word forms to a canonical form; e.g. the inflected forms of English verbs to the infinite." They add that lemmatisation is the process by which the compilers of dictionaries establish the canonical forms of headwords either by removing or retaining inflections.

On the other hand, Plisson et al. (2005: 369) define lemmatisation as "the process of finding the normalized form of words" as their arrangement into alphabetical order whereas Sinclair (1991: 173) defines lemmatisation as "the process of gathering word-forms and arranging them into lemma or lemmata".

From these definitions, it can be deduced that lemmatisation is the selection of words or data to be included in a dictionary. According to Gove (1961: Preface 4a), the lemmatisation of words is determined by the degree to which they are most likely to be looked for.

A noun in Tshivenḓa is *dzina* "name". According to the *Macmillan English Dictionary for Advanced Learners* (2002: 967) a noun is a word or group of words used for referring to a person, thing, place or quantity whereas the Chambers-Macmillan *South African Student's Dictionary* (1996: 639) defines a noun as a word used to refer to a person, thing or quality. By that is meant that it can refer to anything that can occupy a space and have weight as well as some abstract qualities.

Several wordlists and dictionaries have been published in Tshivenḓa, e.g. those by Marole (1955), Marole and De Gama (1958), Joubert and Rapea (1972), Wentzel and Muloiwa (1982), Hartshorne (1984), Van Warmelo (1989) and that by the Tshivenḓa National Lexicography Unit, *Tshivenḓa/English Ṱhalusamaipfi/Dictionary* (henceforth abbreviated as TETD) (2006). This article will analyse the lemmatisation of nouns, particularly in Wentzel and Muloiwa (1982), Van Warmelo (1989) and TETD (2006).

When the current dictionaries of South African languages, including Tshivenḓa, are investigated, it is found that nouns are lemmatised principally in two ways:

(a)   using the whole word (noun), and

(b)   using the noun stems.

2.   Lemmatisation of nouns

Lemmatising the whole word in languages such as Tshivenḓa, Sesotho sa Leboa and Xitsonga has become tradition. Lemmatising according to the noun stem is found in the Nguni languages, where it originated, although it also occurs in Sesotho sa Leboa.

2.1   Lemmatisation of the whole word

From the following examples, it can be seen that words such as *musadzi* in Tshivenḓa, *mosadi* in Sesotho sa Leboa and *wansati* in Xitsonga have been lemmatised as whole words.

(a)   Tshivenḓa: the word *musadzi* 'woman' in Van Warmelo (1989: 235), Wentzel and Muloiwa (1982: 43), and TETD (2006: 48)

(b)   Sesotho sa Leboa: the word *mosadi* 'woman' in Kriel and Van Wyk (1983: 172)

(c)   Xitsonga: the word *wansati* 'woman' in Cuenod (1967: 224)

Every language has its own way of writing words. In other words, Tshivenḓa, Sesotho sa Leboa and Xitsonga have a disjunctive way of writing, as shown in the following sentences, e.g.

Tshivenḓa: *Musadzi u amba na ṅwana* 'The woman talks to the child'

Sesotho sa Leboa: *Mosadi o bolela le ngwana* 'The woman talks to the child'

Xitsonga: *Wansati u vulavula na n'wana* 'The woman talks to the child'

In the Nguni languages, a conjunctive way of writing is traditional, as shown in the following sentences, e.g.

IsiZulu: *Umfazi ukhuluma nomntwana* 'The woman talks to the child'

The preceding examples indicate that in Tshivenḓa, Sesotho sa Leboa and Xitsonga, there are five words in a sentence, including morphemes, whereas in the Nguni languages there are only three words. This is seen in Van Warmelo (1989: 235 and 289), where the verb stem *-amba* 'talks' and nouns *musadzi* 'woman' and *ṅwana* 'child' are lemmatised as individual words. This is also found in Wentzel and Muloiwa (1982: 7, 43, 58). In the preceding isiZulu example, however, the sentence is formed by only three words. All these words have been lemmatised as stems, as can be seen in Doke et al. (1971: 201 and 609) where the nouns *umfazi* and *nomntwana* are lemmatised as *-fazi* and *-ntwana* respectively.

The conjunctive way of writing results in lemmatisation using stems in the Nguni languages, whereas the disjunctive way of writing brings about lemmatisation of nouns using the whole word. The difference between the two ways of lemmatisation lies in the use of a hyphen before the stem, while the word can stand on its own.

Although a dictionary is not a grammar book, it includes the morphology of words because the dictionary cannot be used without some knowledge of their morphology. This is why users should be aware that at some point they would have to be able to find the stems of verbs, nouns and adjectives of the

specific language, so that they can utilise the dictionary. It is also because compilers of dictionaries use morphology when writing their dictionaries, which indicates the correct way of constructing words. This method provides the meaning of such words.

When the three dictionaries by Wentzel and Muloiwa (1982), Van Warmelo (1989), and TETD (2006) were examined, it was found that they have lemmatised the nouns using the complete word and not just the stems.

2.2   Lemmatisation of singular and plural nouns

Nouns can be lemmatised in singular or plural form. In Tshivenḓa, there are different class noun prefixes, which indicate the singular or plural. Such prefixes are affixed to the noun stem, as shown in the following examples:

(a)   *mushumi* 'worker' < *mu-* + *-shumi* in Van Warmelo (1989: 237),

(b)   *musidzana* 'girl' < *mu-* + *-sidzana* in Van Warmelo (1989: 237),

(c)   *vhashumi* 'workers' < *vha-* + *-shumi* is not lemmatised in Van Warmelo (1989),

(d)   *mushumi* 'worker' < *mu* + *-shumi* in TETD (2006: 49),

(e)   *vhashumi* 'workers' < *vha-* + *-shumi* in TETD (2006: 83),

(f)   *musidzana* 'girl' < *vha-* + *-sidzana* in TETD (2006: 49),

(g)   *vhasidzana* 'girls' < *vha-* + *-sidzana* in TETD (2006: 83),

(h)   *mushumi* 'worker' < *mu-* + *-shumi* in Wentzel and Muloiwa (1982: 43), and

(i)   *musidzana* 'girl' mu- + -shumi in Wentzel and Muloiwa (1982: 44).

When the preceding examples are considered, it can be observed that Van Warmelo (1989) and Wentzel and Muloiwa (1982) have lemmatised the noun *mushumi* 'worker' and *musidzana* 'girl' in the singular only; whereas TETD (2006) has lemmatised both *musidzana* and *mushumi* in the singular and *vhasidzana* and *vhashumi* in the plural form. For learners who are not mother-tongue speakers of Tshivenḓa it will be problematic when they come across the nouns *vhashumi* 'workers' and *vhasidzana* 'girls' but cannot find them in the dictionaries.

TETD tried to solve this problem by lemmatising nouns in the singular and plural forms, as shown above. However, they did not lemmatise other Tshivenḓa nouns in their plural forms, e.g. *mulilo* 'fire', is lemmatised in TETD (2006: 48). However, its plural, *mililo* 'fires', is not. Also, *mulambo* 'river', is lemmatised in TETD (2006: 47). However, its plural *milambo* 'rivers', is not.

The problem here is that there is no consistency in the procedure of the compilers of these dictionaries, because some words have been lemmatised in the singular and plural, and others in the singular only.

In Tshivenḓa, there are 21 noun class prefixes in the singular and plural forms. Among these prefixes, some are infinite and some locative class prefixes. The noun class prefixes are as follows:

Class	Singular		Plural
Class 1	*mu-*	Class 2	*vha-*
Class 3	*mu-*	Class 4	*mi-*
Class 5	*ḽi-*	Class 6	*ma-*
Class 7	*tshi-*	Class 8	*zwi-*
Class 9	*n-*	Class 10	*dzi-*
Class 11	*u-*	Classes 10/14	*dzi-/vhu-*
Class 14	*vhu-*	Class 6	*ma-*
Class 15	U	Infinite noun class	
Classes 16/17/18	*fha-/ku-/mu-*	Locative noun classes	
Class 20	*ku-*	Class 8	*zwi-*
Class 21	*ḓi-*		

A noun in Tshivenḓa is formed by two morphemes; namely the noun class prefix plus the noun stem. The singular or plural prefix indicates the singular or plural form. In addition, every noun in Tshivenḓa is categorized under these classes, e.g.

Singular		Plural
musadzi 'woman'	>	*vhasadzi* 'women'
muthu 'person'	>	*vhathu* 'persons'
mulilo 'fire'	>	*mililo* 'fires'

The examples above show that the noun *musadzi* falls under the prefix *mu-* in the singular and under the prefix *vha-* in the plural *vhasadzi.* The noun *mulilo* falls under the singular class prefix *mu-*, whereas it comes under the plural prefix *mi-* in *mililo.* Among these nouns there are instances where the prefix is not visible in the singular form whereas in the plural it is, e.g.

Singular	Plural
mme 'mother' class 1a	*vhomme* 'mothers' class 2b
khotsi 'father' class 1a	*vhokhotsi* 'fathers' class 2b

In the preceding examples, it may appear that there is no class prefix attached to the nouns. What is seen here are the stems *mme* and *khotsi.* To these stems, which are recognized nouns, the plural prefix *vho-* can be affixed to form the plural nouns *vhomme* and *vhokhotsi.* This prefix *vho-* indicates that the noun to which it is affixed is in the plural form. These words should be interpreted as two different words. Lemmatising nouns using a singular or a plural prefix started some time ago; for example, Van Warmelo (1989) and Wentzel and Muloiwa (1982). In Wentzel and Muloiwa (1982: 43) lemmatisation has been

done as follows:

musadzi (vha-)

This shows that the noun *musadzi* is in the singular form, and its plural form is indicated as *vha-* in brackets. Van Warmelo (1989: 235) has not indicated the plural of the noun *musadzi*. He has given it as follows in the singular form only:

musadzi 1 (Class 1)

This merely shows that the noun *musadzi* falls under the first class prefix of nouns, and is in the singular form. Not all nouns in the plural form are lemmatised in this way; others are represented as follows:

Miḓali is the plural form of *muḓali*
mililo is the plural form of *mulilo*

These nouns appear to have been lemmatised in Van Warmelo (1989: 196 and 197) in alphabetical order. However, *muḓali* in the singular and *miḓali* in the plural form have been omitted in TETD (2006). The difference between *mulilo* and *mililo* is brought about by the class prefix *mu-* of the singular, meaning that the word refers to a single entity, while the plural class prefix *mi-* indicates that the word refers to more than one thing. This way of lemmatisation on the basis of the singular and plural forms has advantages and disadvantages.

2.2.1 Advantages of the lemmatisation of singular and plural nouns

It is user-friendly when all nouns will be lemmatised in the singular and the plural. This has been emphasised by Prinsloo and De Schryver (1999: 267) as follows:

> The major advantage of this lemmatization procedure is user-friendliness rendering a practical theory.

With this way of lemmatisation, the dictionary is easy to use, because no intensive knowledge of the language is expected from users. What is required is just knowledge of the alphabet of Tshivenḓa. When we look at the Tshivenḓa noun *mushumi* 'worker/labourer', it is lemmatised in the singular by Van Warmelo (1989: 237). He has also lemmatised the plural form of *mushumi* 'worker/ labourer', which is *vhashumi* 'workers/labourers'. This is a user-friendly format even for a person who is not *Muvenḓa*.

Wentzel and Muloiwa (1982: 43) also lemmatised the noun *mushonga* 'medicine', and the plural for *mushonga* 'medicine' is shown by the prefix *mi-*. By doing this, he is indicating that when one wants to get the plural form of *mushonga* 'medicine' one should just add the prefix *mi-* to the noun stem *-shonga* to make it *mishonga* 'medicines'. This is a user-friendly approach.

2.2.2 Disadvantages of the lemmatisation of singular and plural nouns

There are also problems associated with the lemmatisation of both singular and plural nouns. Firstly, because plurality appears in all Tshivenḓa nouns, compilers of dictionaries will have to pose the question: Which nouns should be selected for lemmatisation? As a result compilers often fail to lemmatise important, frequently occurring nouns, thus rendering the dictionary ineffective to its users.

The following nouns have been lemmatised in the singular form in Van Warmelo (1989: 237 and 235):

> *musidzana* 'girl'
> *musadzi* 'woman'

However, the following plurals of these nouns have not been lemmatised:

> *vhasidzana* 'girls'
> *vhasadz*i 'women'

The dictionary seems to have been written only for users who are conversant in Tshivenḓa; that is users who would know that the plural of the noun *musidzana* is *vhasidzana.* By lemmatising the nouns in the singular only, the plurals of important nouns in Tshivenḓa are left unlemmatised.

In Wentzel and Muloiwa (1982: 42 and 44) the following nouns have been lemmatised in both the singular and plural, and the plural is indicated by *vha-* as shown below:

> *munna (vha-)*
> *musidzana (vha-)*

This means that the noun *musidzana* is the singular form, and to form the plural the prefix *vha-* must be added. This way of lemmatisation is very effective for those who know Tshivenḓa. However, users who do not know Tshivenḓa cannot consult the dictionary to check plurality. When they come across *vhasidzana,* they would not know that it is the plural of *musidzana*. When they come across the plural *vhasidzana*, instead of the singular *musidzana*, the users would be unaware that it is the plural of *musidzana.* It therefore renders the dictionary ineffective to users who are not conversant in Tshivenḓa. In this way, dictionaries omit important words in the plural.

2.3   Lemmatisation of deverbative nouns

Poulos (1990: 14) defines *deverbative nouns* as follows:

> A number of nouns in some of the classes are derived from other parts of speech, for example, from verbal radicals (or verb roots).

In Tshivenḓa, the derivation of nouns from verb stems or adjective stems occur by affixing singular and plural prefixes to these stems and suffixes or endings *-i, -e, -o, -u*.

Nouns	derived from verbs
ndima	*-lima* 'to plough'
murengi	*-renga* 'to purchase'
mulimi	*-lima* 'to plough'
muimbi	*-imba* 'to sing'
tshibiki	*-bika* 'to cook'
vhuhulu	*-hula* 'to be big'
vhusekene	*-sekena* 'to be thin'
vhulapfu	*-lapfa* 'to be long'

(Nthambeleni 2008: 49)

Other deverbatives formed from verbs with suffixes or endings added are the following:

murungo 'seam'
muvhuso 'government'

Because they are problematic, the preceding nouns should have been lemmatised in the dictionary. Compilers should include deverbatives like these in their dictionaries.

In African languages, nominal prefixes are therefore important to such an extent that prefixes of class nouns and suffixes or endings such as *-a, -e, -i, -o,* and *-u* should be considered when deverbatives are discussed.

Nouns formed in this manner are not lemmatised in great numbers in dictionaries, as is the case in Tshivenḓa dictionaries. This results in ineffective dictionaries being produced, which disadvantages dictionary users. Lemmatisation of these words in the dictionary depends on the way speakers use them. When they are used in language, compilers of dictionaries should select them for lemmatisation.

Focusing again on the above-mentioned items, compilers of Tshivenḓa dictionaries have fulfilled an important function in lemmatising deverbatives. However, it is problematic when some deverbative nouns have not been lemmatised as in Van Warmelo (1989: 52). Here the verb stem *-fhaṱa* 'to build' has been given a deverbative noun by affixing the class prefix *mu-* to a verb which is not given. This shortcoming causes problems for non-native users of the dictionary. Such users will find the noun *mufhaṱi* 'builder' but not know its meaning. This is also the case with the noun *murengi* 'buyer'. Although they are used frequently in Tshivenḓa, these nouns have not been lemmatised.

2.4   Lemmatisation of diminutive nouns

Poulos (1990: 81) defines *diminutives* as follows:

> Those expressions in a language that basically express the idea of 'smallness', 'shortness' and where appropriate the 'young of' some or other noun.

As users would look up diminutives in the dictionary, it is important to lemmatise them. In Tshivenḓa, diminutives can be formed by affixing the different class prefixes and suffixes to the nouns, such as the suffixes *-ana, -nyana*; and class prefixes 7 *tshi-*, 11 *lu-*, 20 *ku-* in the singular and class 8 *zwi-* and 14 *vhu-* in the plural, as in the following examples (not from the dictionaries mentioned above).

    Class 7 *tshi- + -budzi > tshibudzi* (small fat goat)
    Class 20 *ku- + -budzi > kubudzi* (small young goat)

Each of these nouns *tshibudzi* and *kubudzi* has a different meaning. They are therefore lemmatised in dictionaries.
    When the suffix *-ana* is added to nouns, it forms diminutive nouns

    mbudzi + -ana > mbudzana

The suffix *-nyana* also forms diminutives and, when added to nouns, it results in the following:

(a)   When it is added to nouns it may refer to a liquid which is little in quantity.

    (i)   *maḓi + -nyana > maḓinyana* (little water)
    (ii)   *mafhi + -nyana > mafhinyana* (little milk)
    (iii)   *malofha + -nyana > malofhanyana* (little blood)*

(b)   When it is used with a class prefix and a verb stem, it may indicate that there is not much of something.

    (i)   *mu + -shum + -o + -nyana > mushumonyana* (little work)

(c)   When it is added to nouns, it may indicate smallness in size.

    (i)   *musidzana + -nyana > musidzanyana* (small girl)

The noun prefix plus suffix function together when added to the noun stem to form a diminutive noun, as in the following examples:

    tshi- + -budz- + -ana > tshibudzana (small goat)
    zwi- + -budz- + -ana > zwibudzana (small goats)
    vhu- + -budz- + -ana > vhubudzana (thin goats)

From the examples shown above, it becomes evident that, just like other nouns, diminutives are important in a language. In addition, in a language like Tshivenḓa, they are used frequently.

When the different dictionaries in Tshivenḓa are examined, one will find that they have lemmatised diminutive nouns. The noun *mbudzi* (goat) has been lemmatised, whereas its diminutive form *tshibudzana* (small goat) has not been listed in Van Warmelo (1989), although it is used frequently. Omitting it from the dictionaries is problematic for the users who do not know Tshivenḓa because they would not be able to find the word *tshibudzana* and its meaning.

When nouns are looked up in Tshivenḓa dictionaries such as Wentzel and Muloiwa (1982), Van Warmelo (1989) and TETD (2006), it becomes clear that several diminutive nouns have not been lemmatised. Compilers should ensure that all such nouns are lemmatised. Only when this is done, will good dictionaries be compiled that will satisfy users' needs.

3.   Conclusion

When the current Tshivenḓa dictionaries are examined, it becomes clear that all nouns should be lemmatised. The main aim of lemmatisation of nouns is to enable compilers of dictionaries to produce useful dictionaries, and not repeat the same mistakes made in previous dictionaries. For the production of useful dictionaries that will benefit users, nouns should be lemmatised in a well-considered manner. This will also support De Schryver and Prinsloo's (2000) idea of 'Simultaneous Feedback', since a dictionary should be written in such a way that users can find information easily. It has been established that Tshivenḓa nouns are lemmatised in both the singular and plural forms; as diminutives and deverbatives. Including nouns in singular and plural form, as well as diminutive and deverbative nouns, should be formulated as a lexicographic principle. Failure to meet this requirement would render the dictionary not user-friendly at all.

References

Chambers-Macmillan. 1996. *South African Student's Dictionary*. Manzini: Macmillan Boleswa.

Cuenod, R. 1967. *Tsonga–English Dictionary*. Braamfontein: Sasavona.

De Schryver, G.-M. and D.J. Prinsloo. 2000. The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.

Doke, C.M. et al. 1971. *English–Zulu Dictionary.* Johannesburg: Witwatersrand University.

Gove, P.B. (Ed.). 1961. *Webster's Third New International Dictionary of the English Language.* Springfield, Mass.: Merriam-Webster.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.

Hartshorne, K.B. 1984. *Dictionary of Basic English: Venḓa*. Assisted by J.H.A. Swart and E. Posselt. Johannesburg: Educum.

Joubert, P.J. and D.N. Rapea. 1972. *Verklarende Woordelys/Mutevhe wa Maipfi o Ṱalutshedzwaho: Afrikaans–English–Venda/Tshivhuru–Tshiisimane–Tshivenḓa. Vol.1 A–C.* SABC.

Kriel, T.J. and E.S. van Wyk. 1983. *Pukuntšu Dictionary.* Pretoria: J.L. van Schaik.

Macmillan English Dictionary for Advanced Learners. (Edited by M. Rundell and G. Fox.) 2002. London: Macmillan.

Marole, L.T. 1955. *Afrikaans–Venḓa: Vocabulary and Phrase Book.* Sibasa: Marole Book Depot.

Marole, L.T. and F.J. de Gama. 1954. *English–Venda Vocabulary.* Third edition. Sibasa: Marole Book Depot.

Nthambeleni, M. 2008. *Thaidzo musi hu tshi Lemathaizwa Madzina na Maiti kha Ṱhalusamaipfi dza Tshivenḓa.* Unpublished M.A. Dissertation. Thohoyandou: University of Venda.

Plisson, J., D. Mladenić, N. Lavrač and T. Erjavec. 2005. A Lemmatization Web Service Based on Machine Learning Techniques**.** Vetulani, Z. (Ed.). 2005. *Proceedings of the 2nd Language & Technology Conference, April 21–23, 2005, Poznań, Poland*: 369-372. Poznań: Wydawnictwo Poznańskie.

Poulos, G. 1990. *A Linguistic Analysis of Venda.* Pretoria: Via Afrika.

Prinsloo, D.J. and G.-M. de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *Southern African Journal of African Languages* 19(4): 258-275.

Sinclair, J.M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Tshivenḓa National Lexicography Unit. 2006. *Tshivenḓa/English Ṱhalusamaipfi/Dictionary.* Cape Town: Phumelela Publishers.

Van Warmelo, N.J. 1989. *Venda Dictionary: Tshivenḓa–English.* Pretoria: J.L. van Schaik.

Wentzel, P.J. and T.W. Muloiwa. 1982. *Ṱhalusamaipfi ya nyambotharu yo khwiniswaho: Luvenda–Luvhuru–Luisimane/Verbeterde drietalige woordeboek: Venḓa–Afrikaans–Engels/Improved Trilingual Dictionary: Venḓa–Afrikaans–English.* Pretoria: University of South Africa.