

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 Tshivenda 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'
vhasadzi 'women'

The dictionary seems to have been written only for users who are conversant in Tshivenda; 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 Tshivenda 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 Tshivenda. However, users who do not know Tshivenda 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 Tshivenda. 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 Tshivenda, 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
<i>ndima</i>	<i>-lima</i> 'to plough'
<i>murengi</i>	<i>-renga</i> 'to purchase'
<i>mulimi</i>	<i>-lima</i> 'to plough'
<i>muimbi</i>	<i>-imba</i> 'to sing'
<i>tshibiki</i>	<i>-bika</i> 'to cook'
<i>vhuhulu</i>	<i>-hula</i> 'to be big'
<i>vhusekene</i>	<i>-sekene</i> 'to be thin'
<i>vhulapfu</i>	<i>-lapfa</i> 'to be long'

(Nthambeleni 2008: 49)

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

<i>murungo</i>	'seam'
<i>muvhuso</i>	'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, -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 Tshivenda 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 Tshivenda 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 *-fhaṭa* '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 *muḥḥaṭi* 'builder' but not know its meaning. This is also the case with the noun *murengi* 'buyer'. Although they are used frequently in Tshivenda, 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 Tshivenḡa, 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) *maḡi* + *-nyana* > *maḡinyana* (little water)
 - (ii) *maḡhi* + *-nyana* > *maḡhinyana* (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 Tshivenḡa, they are used frequently.

When the different dictionaries in Tshivenda 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 Tshivenda because they would not be able to find the word *tshibudzana* and its meaning.

When nouns are looked up in Tshivenda 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 Tshivenda 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 Tshivenda 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: Venda*. 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 Țalutshedzwaho: Afrikaans-English-Venda/Tshivhuru-Tshiisimane-Tshivoenda. 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–Venda: 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 Tshivenda*. 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.

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

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

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