

A large, stylized blue graphic element that resembles a calligraphic flourish or a stylized letter 'G' with a pen nib at the top right. It frames the main title text.

GWNC 2014 Tartu, Estonia

Proceedings of the Seventh Global Wordnet Conference

Tartu, Estonia, January 25-29, 2014

Editors: Heili Orav, Christiane Fellbaum, Piek Vossen

Volume editors

Heili Orav

University of Tartu

e-mail: heili.orav@ut.ee

Christiane Fellbaum

Princeton University

e-mail: fellbaum@princeton.edu

Piek Vossen

VU University Amsterdam

e-mail: piek.vossen@vu.nl

ISBN 978-9949-32-492-7

Morphosyntactic discrepancies in representing the adjective equivalent in African WordNet with reference to Northern Sotho

Mampaka Lydia Mojapelo

University of South Africa

Department of African Languages

mojapml@unisa.ac.za

Abstract

This paper aims to highlight morphosyntactic discrepancies encountered in representing the adjective equivalent in African WordNet, with reference to Northern Sotho. Northern Sotho is an agglutinating language with rich and productive morphology. The language also features a disjunctive orthographic system. The orthography determines the attachment selection of morphemes. The immediate issue, in this paper, is the absence of a one-to-one correspondence between the adjective in English and that in Northern Sotho. The meaning equivalent of the English adjective covers more than one morphosyntactic category in Northern Sotho. In addition, the categories' structural diversity has a bearing on representation considerations. In some of these categories the stem suffices to represent the specific category unambiguously while in others there is a need to incorporate affixes with the stem. The challenge is to categorize semantic equivalents of the English adjective as such, while retaining their separate morphosyntactic tags in Northern Sotho, in harmony with the typology of the language. The present paper proposes morphologically feasible ways of representing this varied equivalent of the English adjective in Northern Sotho.

1 Introduction

African WordNet¹ seeks to build WordNets for all indigenous official languages of South Africa, which will be linked to one another. Northern

Sotho² is one of the languages in African WordNet. So far in the project the work covers the verbs, nouns, and few adjectives. This presentation is based on the experiences with the adjective in the project so far. Like many languages African WordNet is expanded from the Princeton WordNet³. Being cognisant of dissimilar typologies of the source and target languages, as well as language-specific cultural and historical orientations, African WordNet is geared towards customisation to the African context.

The aim of this paper is to highlight morphosyntactic discrepancies encountered with the Northern Sotho equivalent of the adjective in African WordNet. It also proposes morphologically feasible ways in which the equivalent can be represented. Synsets are linked to one another through conceptual-semantic and lexical relations. WordNet therefore links together not only lexical items but, more significantly, the senses that the lexical items represent. It may be possible for a sense to be lexicalised in both the source and the target language without necessarily carrying the same morphosyntactic tag. This presentation will not go into the broad theoretical issues attending adjectives; rather the focus will be on the meaning equivalent of the English adjective in Northern Sotho, which is the target language, given typological differences between the two languages and differences in morphological structures of the equivalents in the target

¹ <http://www.globalwordnet.org>

² Northern Sotho (Sesotho sa Leboa) also known as Sepedi, one of the dialects, is a Niger-Congo Bantu language (Guthrie's zone S30)

³ <http://wordnet.princeton.edu>

language. Each morphosyntactic category in Northern Sotho will be discussed separately and will conclude with a proposed representation in the database.

2 Semantic function of the adjective

English will be used as springboard here because it is the source language for the expand approach adopted for African WordNet. The semantic function of the English adjective, be it attributive or relational (Miller, 1978), is universal, namely to modify the noun. Morphologically, apart from the core adjectives which may also be morphologically affected through inflection, English adjectives include denominals and deverbals (Peters and Peters, 2000). Furthermore, there are also other different morphosyntactic constructions that are used in modifying the noun, such as the genitive and relative clause. For the purpose of this presentation and in context with African WordNet, the discussion will be confined to the English lexical entry with POS tag adjective, such as *purple*, *murdered*, *cute* and *little* and how they are rendered in the African WordNet.

The immediate issue, first of all, is the absence of a one-to-one correspondence between the adjective in English and that in Northern Sotho (Poulos and Louwrens, 1994). Northern Sotho has a limited number of adjectival stems, which is by no means a reflection of the language's capacity to produce qualifications for the noun. It is not always possible to use an adjective to convey a concept in Northern Sotho that is expressed by an English adjective. Traditional Northern Sotho grammars identify four morphosyntactic categories (the adjective, descriptive possessive [genitive], relative and enumerative) to perform this semantic function (Ziervogel et al., 1969; Poulos and Louwrens 1994). Moreover, each of these equivalents of the English adjective assumes a different prefix depending on the class of the noun it modifies. Some of the stems are unambiguous without affixes and some need affixes to make sense or to identify them with the relevant functional category. The issue is that a lexicalised equivalent of the sense expressed by an English adjective cannot be ignored on the grounds that it is not an adjective, nor can it be categorized as an adjective while it is not. It remains a challenge, specifically in this word category, that

the source and target language differ on structural level. The next sections explore the ways in which each of the morphosyntactic categories can be represented, given their dynamic structures.

3 Northern Sotho equivalents of the English adjective

The English adjective can be rendered by an adjective, possessive, relative or enumerative in Northern Sotho. The next sections discuss three of these morphosyntactic categories, illustrating and substantiating proposed representation strategies.

3.1 The adjective

Some English concepts expressed by adjectives are also expressed by adjectives in Northern Sotho. The structure of a Northern Sotho adjective is sketched as follows:

(Head noun)	Adjective	
	adjectival agreement	adjectival stem
	Demonstrative	Adjectival prefix

Figure 1: The structure of a Northern Sotho adjective

The following example has a class 1 noun as head:

Monna [*yo motelele*]

CL1-man CL1-Dem CL1-Pref-tall

/man that is tall/

'A tall man'

The head noun is given consideration in the structure because it influences the morphological structure of the adjective as a whole. For example, both parts of the adjectival agreement (in bold italics) agree with the head noun and will therefore change every time a

and Mandes, 2006) – and similarly, the verbal relative to the verb.

4 Lexical semantic and morphosyntactic challenges to sort out along the way

Concepts such as worse (232954-a) and worst (2309979-a) are not easy to represent without including that which is ‘worse or worst’, or an adverb. Selection restrictions also have a bearing on this point as ‘their meaning is determined ... by the headnoun that they modify’ (Fellbaum, 1998).

Other strategies used in the language to extend or refine a qualifying concept include the diminutive affix and reduplication. For example, yo motelele*nyana*/yo motelele*šana* (diminutive) and yo motelele*telele* (reduplication), which normally serve for gradability of various adjectival concepts as is the case with English degrees of comparison. While it is generally not necessary to include degrees of comparison in the database, some English concepts are perceived as being at various points on a continuum, where reduplication and adverbs are employed to differentiate them from others. Other challenges attending these forms include the frequent case that the diminutive involves phonological processes; whereas in reduplication there is no limit to the number of times the adjectival stem can be repeated, and for reduplication involving monosyllabic stems the adjectival prefix has to interfere repeatedly, for example:

Adjectival stem *-so* (black; dark):

borokgo bjo boso (A pair of black trousers); *borokgo bjo bosobosoboso* (for intensity)

5 Concluding remarks

The lack of one-to-one correspondence between the adjective in English and in Northern Sotho results in the English adjective equivalent being represented by various morphosyntactic categories in Northern Sotho. Given their structural differences, these Northern Sotho equivalents require distinctive consideration in representing them in a manner that will be consistent with the language system. The proposal is that while it is understandable that only stems be considered, invariant parts that are

separate from the stem but that will help to disambiguate it be retained (for example, *a* in the descriptive possessive construction). The suffix *go* or *ng* of the verbal relative also marks it as different from the verb. The challenge with the representation in African WordNet is that while they are all meaning equivalents of the same English word category, they straddle a number of morphosyntactic categories in Northern Sotho, which nevertheless share a semantic function.

While the nominal relative base is a noun, it selects nouns from classes 11 and 14 and is unlikely to be problematic. The enumerative has been left out of the discussion because their occurrence is not as wide as that of the categories discussed.

References

- Denis Creissels. 2011. The ‘new adjectives’ of Tswana. Paper presented at the WC2010 conference, Rome, 24-26 March 2010, revised April 2011. Viewed from http://reissels-new_adjectives.pdf
- Christiane Fellbaum. 1998. A semantic network of English: the mother of all wordnets. In Piek Vossen (ed.) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht, pages 137-148.
- Palmira Marrafa and Sara Mendes. 2006. Modeling adjectives in computational relational lexica. *Proceedings of the COLING/ALC Main Conference Poster Session*, Sydney, July 2006, pages 555-562.
- Katherine J. Miller. 1998. Modifiers in WordNet. In Christiane Fellbaum (ed.) *WordNet: an electronic lexical database*, The MIT Press, Cambridge, MA, pages 47-68.
- Ivonne Peters and Wim Peters. 2000. The Treatment of Adjectives in SIMPLE: Theoretical Observations, *Proceedings of LREC 2000*.
- George Poulos and Louis J. Louwrens. 1994. *A linguistic analysis of Northern Sotho*. Via Afrika, Pretoria.
- Dirk Ziervogel, Daniel P. Lombard and Pothinus C. Mokgokong. 1969. *A Handbook of the Northern Sotho Language*, Van Schaik, Pretoria.

noun from a different class is being modified. For this reason only the basic adjectival stem is captured as equivalent of the English adjective. The following examples illustrate the point made:

Monna [*yo motelele*]

‘A tall man’

Monna [*yo mošweu*]

‘A light-complexioned man’

Banna [*ba bantšij*]

‘Many men’

The adjectival stem *-telele* (tall/long), for example, can be used to qualify nouns from various classes, as illustrated below:

Class 1: *Monna* [*yo motelele*]

‘A tall man’

Class 3: *Mohlare* [*wo motelele*]

‘A tall tree’

Class 5: *Lephodisa* [*le letelele*]

‘A tall policeman’

Class 6: *Maphodisa* [*a matelele*]

‘Tall policemen’

Class 7: *Setimela* [*se setelele*]

‘A long train’

Class 9: *Kota* [*ye telele*]

‘A tall/long pole’

For the reasons mentioned and illustrated above the Northern Sotho adjective stem *-botse* (beautiful/cute/precious/dinky/pretty) appears in African WordNet as illustrated in Figure 2 to Figure 4:

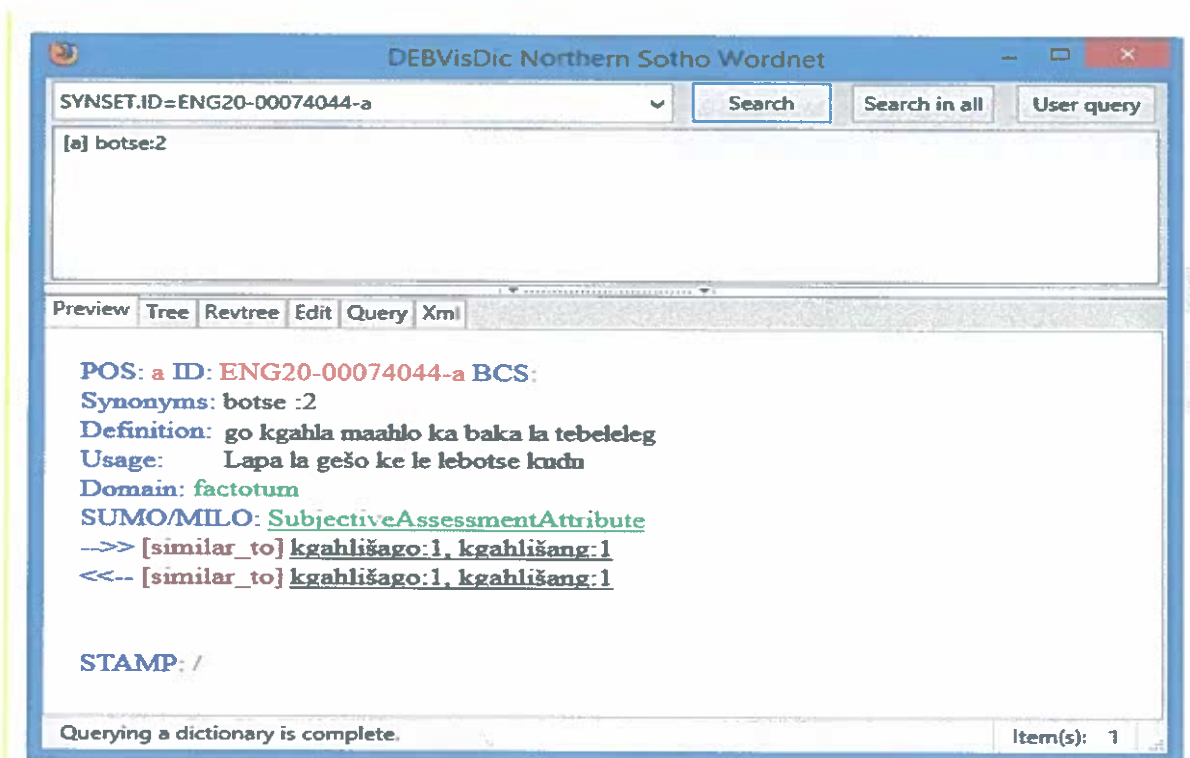


Figure 2: Adjective *beautiful:2*: aesthetically pleasing

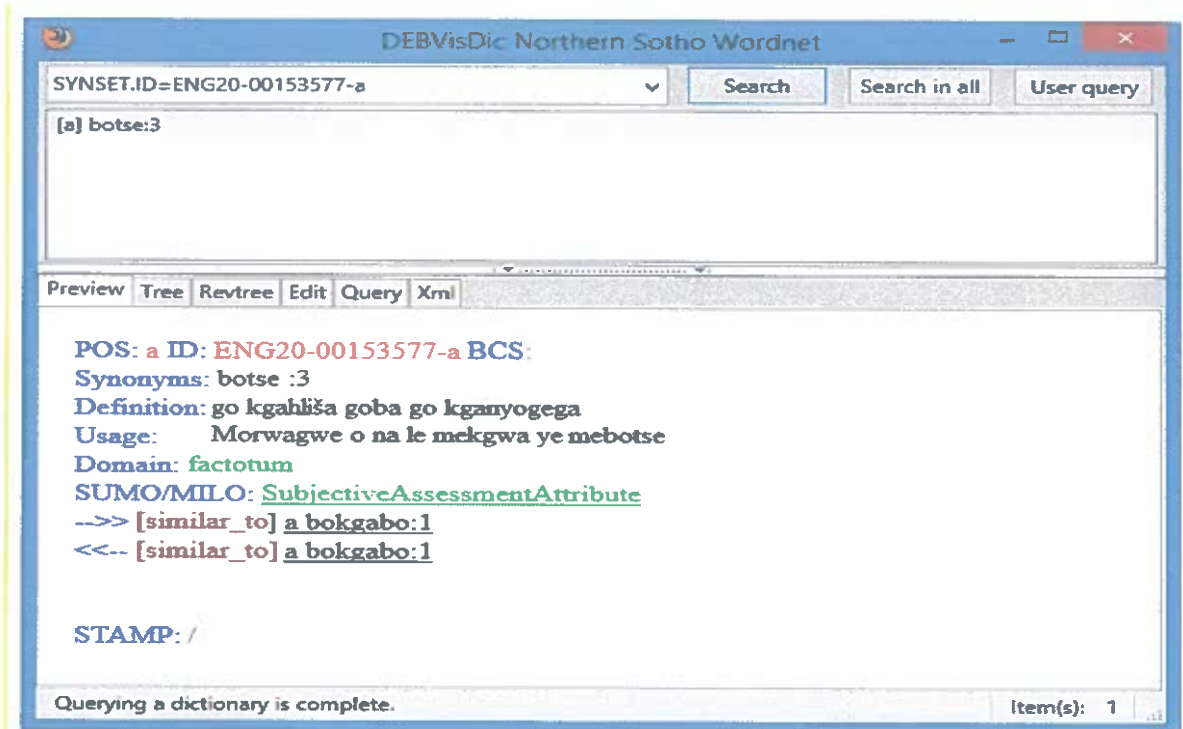


Figure 3: Adjective *cute:2,precious:3*: obviously contrived to charm

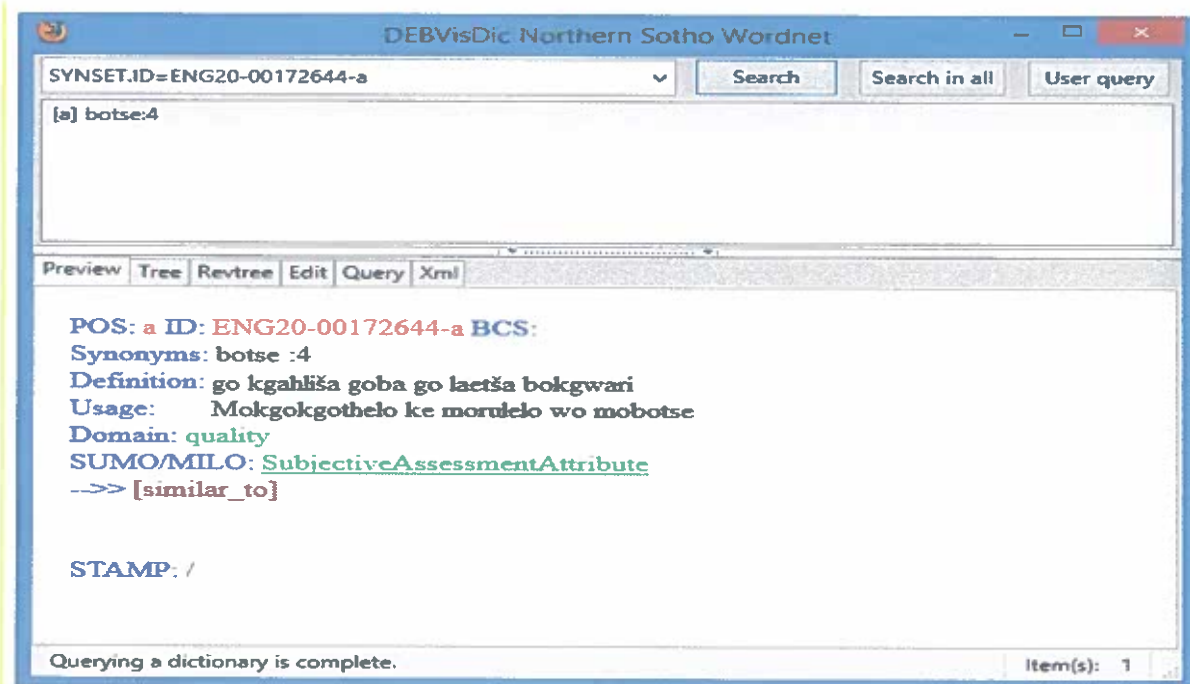


Figure 4: Adjective *dinky:2* (British informal) pretty and neat

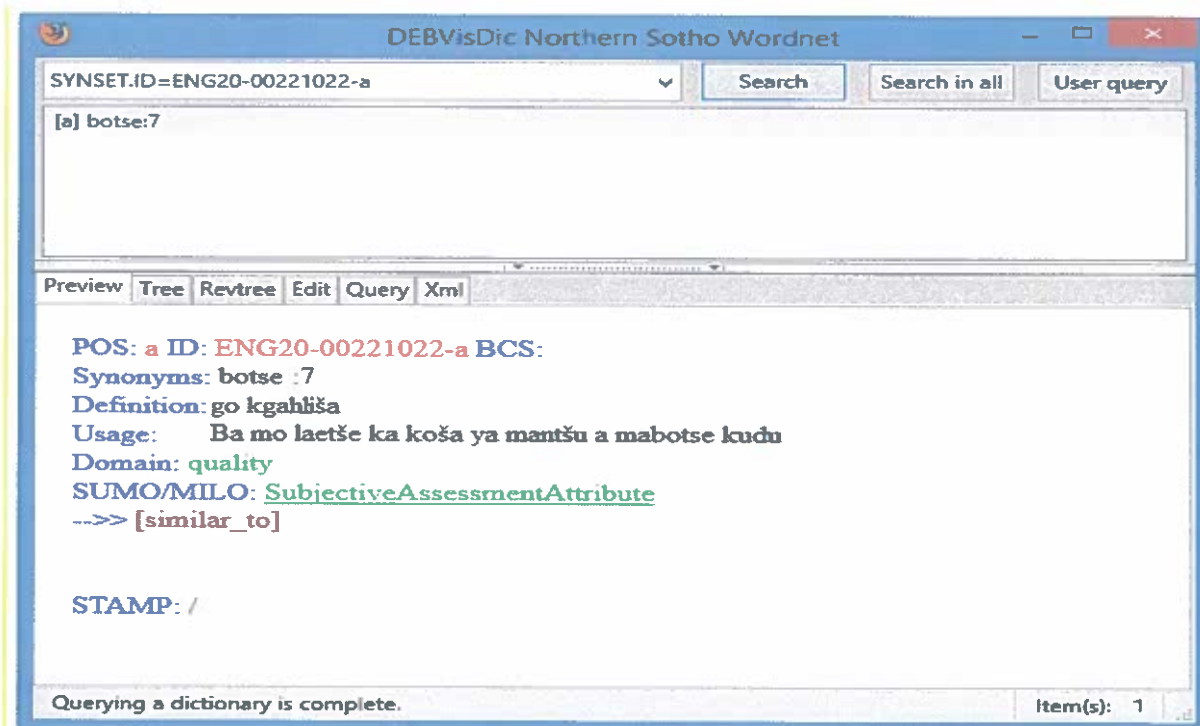


Figure 5: Adjective *pretty:1* pleasing by delicacy or grace; not imposing

Some English adjectives do not have adjective equivalents in Northern Sotho, but the senses are represented by different morphosyntactic categories.

3.2 Descriptive possessive/genitive

The genitive or possessive construction in general serves two semantic functions. It can be used for direct possession or ownership, and for describing the feature or quality of the noun (Poulos and Louwrens, 1994). It is the latter that is under discussion here. The descriptive possessive or genitive construction may serve as the cognitive-semantic equivalent of the English adjective. The general genitive/possessive structure is as follows:

(Head noun)	Genitive	
	genitive agreement	Noun
	subject agreement	genitive a

Figure 6: The genitive/possessive construction

The following example of a possessive construction has a class 1 noun as head:

Monna [*wa senatla*]

CL1-man CL1-Dem CL7-strong individual

/man of strong individual/

‘A strong man’

The first issue with the genitive is that the agreement comprises two components which behave differently. The subject agreement component is dependent on the head noun while genitive *a* is invariant. Secondly, the complement is a noun phrase, which is just another noun without the genitive agreement. To encode it unambiguously we need to include the invariant part of the genitive agreement with the complement, which is the descriptive part serving as equivalent to the English adjective. First, the invariant part of the genitive agreement is applicable to every head noun and, secondly, it makes the complement noun phrase duly interpreted as a descriptive. The first part of the genitive agreement will thus be unreliable as illustrated below (in italics):

Class 1: *Monna* [*wa senatla*]

‘A strong man’

Class 5: *Leho* [la go tia]

‘A strong wooden spoon’

Class 9: *Kala* [ya boleta]

‘A soft branch’

Class 10: *Dinku* [tša bohlokwa]

‘Important sheep’

The Northern Sotho descriptive possessive/ genitive as equivalent of the English adjective appears in African WordNet as follows:

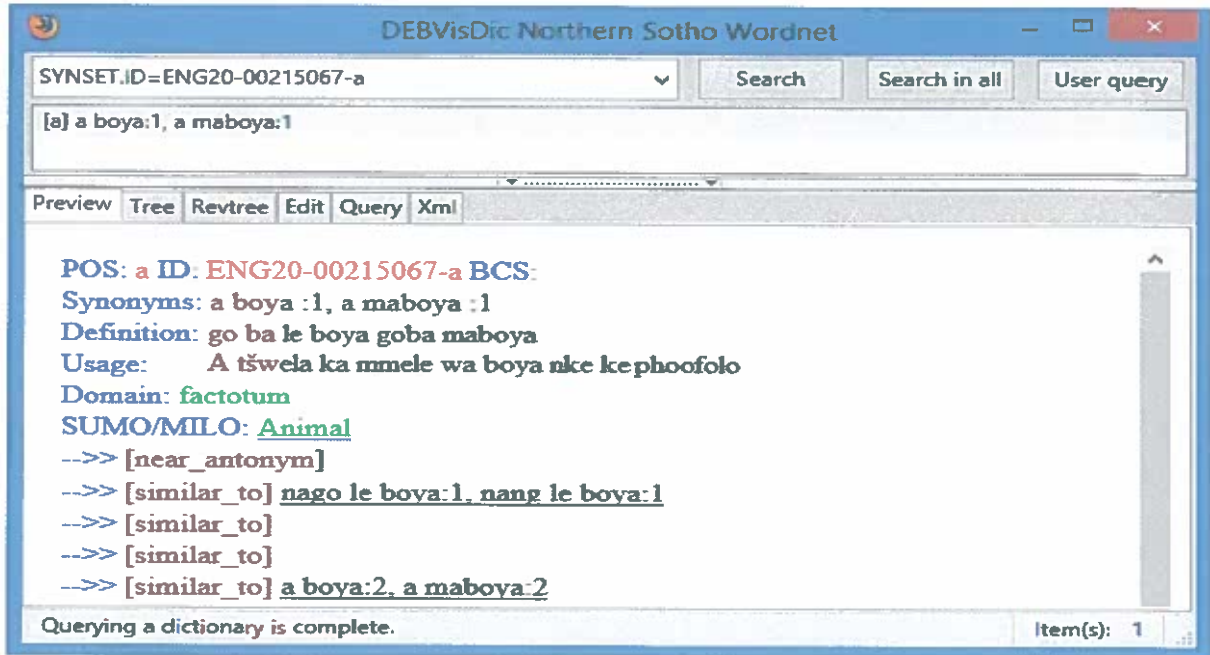


Figure 7: Adjective *hairy:1* having or covered with hair

3.3 The relative

Traditional Northern Sotho grammars (and those of other Sotho languages) distinguish between the verbal and nominal relative (Poulos and Louwrens, 1994; Ziervogel, Lombard and Mokgokong, 1969). What is traditionally known as a nominal relative is called a ‘new attributive adjective’ by Creissels (2011) based on differentiation between word level and phrase level. The reason for this difference or overlap is that the Northern Sotho relative (both verbal and nominal) can also be a conceptual-semantic equivalent of the English adjective.

The verbal relative is further divided into the direct and indirect forms.

Verbal relative

A class 5 noun serves as head in the following examples:

Direct: *Lephodisa* [le le thuntšhago]

CL5-policeman CL5-Dem CL5-SM shoot-SUFF-go

/Policeman that shoots/

Indirect: *Lephodisa* [le ba le thuntšhago]

CL5-policeman CL5-Dem CL1-SM CL5-OC shoot-SUFF-go

/Policeman that they shoot/

‘Policeman that is being shot’

For illustration we shall use only the direct relative clause, given that the same principles apply to the indirect relative. Figure 8 illustrates the structure of the direct verbal relative in Northern Sotho, as equivalent of the English adjective:

(Head noun)	verbal relative		
	Relative agreement	Verb stem	Suffix <i>go/ng</i>
	Dem	subject agreement	

Figure 8: The structure of the direct relative

Both parts of the relative agreement, namely the demonstrative (Dem) and the subject

agreement depend on the head noun. Northern Sotho has two variant suffixes for the verbal relative, namely *-go* and *-ng*. The affixes *-go* and *-ng* on the verb stem indicate that its function is not to be a verb, but to qualify the noun. Both suffixes are equally recognised in Northern Sotho. Exclusion of variant parts of the verbal relative is not problematic because they are written disjunctively from the stem. Therefore only the verb stem, with the attached suffix, is recorded.

The Northern Sotho verbal relative as equivalent of the English adjective appears in African WordNet as follows:

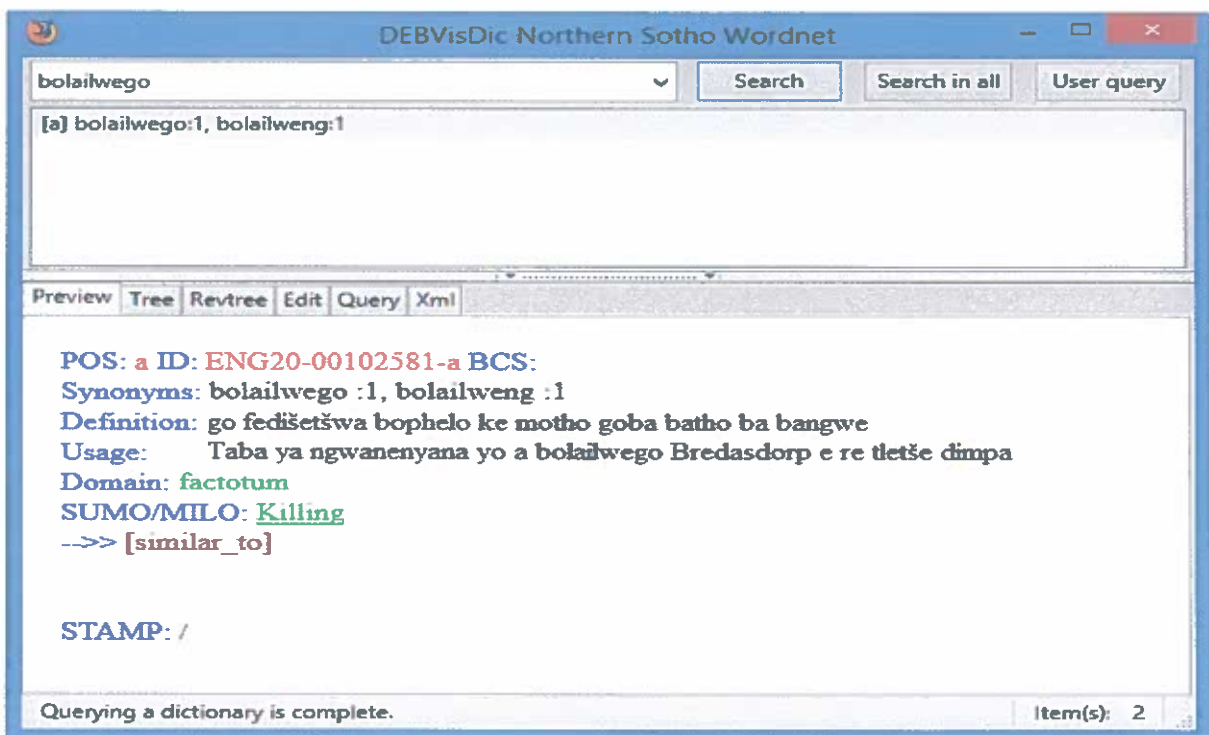


Figure 9: Adjective *murdered:1* killed unlawfully

Nominal relative

The structure of the nominal relative is as follows:

(Head noun)	Nominal relative	
	Nominal relative agreement (resembles Dem)	Noun

Figure 10: The structure of the nominal relative

A class 7 noun serves as head in the following example:

Segotlane [*se bohale*]
 CL7-toddler CL7-Dem CL14-cleverness
 /toddler that is clever/
 'clever toddler'

Semantically the nominal relative can link to the noun through cross-POS relations (Marrafa

and Mandes, 2006) – and similarly, the verbal relative to the verb.

4 Lexical semantic and morphosyntactic challenges to sort out along the way

Concepts such as worse (232954-a) and worst (2309979-a) are not easy to represent without including that which is ‘worse or worst’, or an adverb. Selection restrictions also have a bearing on this point as ‘their meaning is determined ... by the headnoun that they modify’ (Fellbaum, 1998).

Other strategies used in the language to extend or refine a qualifying concept include the diminutive affix and reduplication. For example, yo motelelenyana/yo moteleletšana (diminutive) and yo moteleletelele (reduplication), which normally serve for gradability of various adjectival concepts as is the case with English degrees of comparison. While it is generally not necessary to include degrees of comparison in the database, some English concepts are perceived as being at various points on a continuum, where reduplication and adverbs are employed to differentiate them from others. Other challenges attending these forms include the frequent case that the diminutive involves phonological processes; whereas in reduplication there is no limit to the number of times the adjectival stem can be repeated, and for reduplication involving monosyllabic stems the adjectival prefix has to interfere repeatedly, for example:

Adjectival stem *-so* (black; dark):

borokgo bjo boso (A pair of black trousers): *borokgo bjo bosobosoboso* (for intensity)

5 Concluding remarks

The lack of one-to-one correspondence between the adjective in English and in Northern Sotho results in the English adjective equivalent being represented by various morphosyntactic categories in Northern Sotho. Given their structural differences, these Northern Sotho equivalents require distinctive consideration in representing them in a manner that will be consistent with the language system. The proposal is that while it is understandable that only stems be considered, invariant parts that are

separate from the stem but that will help to disambiguate it be retained (for example, *a* in the descriptive possessive construction). The suffix *go* or *ng* of the verbal relative also marks it as different from the verb. The challenge with the representation in African WordNet is that while they are all meaning equivalents of the same English word category, they straddle a number of morphosyntactic categories in Northern Sotho, which nevertheless share a semantic function.

While the nominal relative base is a noun, it selects nouns from classes 11 and 14 and is unlikely to be problematic. The enumerative has been left out of the discussion because their occurrence is not as wide as that of the categories discussed.

References

- Denis Creissels. 2011. The ‘new adjectives’ of Tswana. Paper presented at the WC2010 conference, Rome, 24-26 March 2010, revised April 2011. Viewed from http://reissels-new_adjectives.pdf
- Christiane Fellbaum. 1998. A semantic network of English: the mother of all wordnets. In Piek Vossen (ed.) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht, pages 137-148.
- Palmira Marrafa and Sara Mendes. 2006. Modeling adjectives in computational relational lexica. *Proceedings of the COLING/ALC Main Conference Poster Session*, Sydney, July 2006, pages 555-562.
- Katherine J. Miller. 1998. Modifiers in WordNet. In Christiane Fellbaum (ed.) *WordNet: an electronic lexical database*, The MIT Press, Cambridge, MA, pages 47-68.
- Ivonne Peters and Wim Peters. 2000. The Treatment of Adjectives in SIMPLE: Theoretical Observations, *Proceedings of LREC 2000*.
- George Poulos and Louis J. Louwrens. 1994. *A linguistic analysis of Northern Sotho*. Via Afrika, Pretoria.
- Dirk Ziervogel, Daniel P. Lombard and Pothinus C. Mokgokong. 1969. *A Handbook of the Northern Sotho Language*, Van Schaik, Pretoria.