

**THE SEMANTICS OF NORTHERN SOTHO VERBS
AS TRANSLATED FROM ENGLISH:
COMMENTS ON THE AFRICAN LANGUAGES WORDNET PROJECT**

Mampaka Lydia Mojapelo

University of South Africa (UNISA), Pretoria, South Africa

***Abstract:** African Languages WordNet is an ongoing project which is based on the English WordNet. WordNet is an electronic lexical database that groups words in synonym sets (synsets). In this project, words are translated from English to African Languages. As such, this paper comments on the semantic aspects of verbs in one of the African Languages, namely Northern Sotho, that are translated from English. Verbs are generally understood as expressions of action or state of being. These two languages are typologically dissimilar, with different cultural-historical backgrounds. Northern Sotho is a Bantu¹ language, agglutinating with extensive and productive use of affixes while English is not. In the first place, structural differences between these two languages pose equivalence challenges, both linguistically and computationally. Secondly, verb equivalents may not be affected by the same collocational restrictions in the source and target languages. Another issue is that a verb in one language may invoke certain connotations, which may not apply to its equivalent in another language. Finally, some concepts may be foreign and others culture-specific to one language and not the other, thus resulting in omission of some target language concepts. Attention to these equivalence challenges may enhance technological development of the target language lexicon.*

1. Introduction

This paper is based on an ongoing African Languages WordNet project, which was started in South Africa in 2007. It is a collaborative project between Unisa's Department of African Languages and Northwest University's Centre for Text Technology. The project is based on existing global networks such as the English WordNet, Princeton (cf. <http://wordnet.princeton.edu>; Fellbaum, 1998). In this project, the synonym sets (henceforth referred to as synsets) are translated from English into African Languages. The envisaged result of the project is the WordNets of official South African languages – the lexical database that will support automatic text analyses and other applications. The languages currently involved in the project are Northern Sotho, Setswana, Zulu and Xhosa. The synsets translated and quality-controlled to date include nouns and verbs. The discussion in this paper, however, is limited to verbs, that is, the semantic aspects of Northern Sotho verbs within the context of WordNet. In other words, although WordNet is multidisciplinary in nature, the paper focuses specifically on the linguistic aspects of the project.

A verb is simply defined as a word that expresses action or state of being. Yule (2003: 88), for example, defines this part of speech as “words used to refer to various kinds of actions and states involving the ‘things’ in events”. Semantically, verbs may be classified under those expressing change, change of possession, communication, existence, experience, contact, motion, creation, weather, bodily care and function, perception and social interaction (Du Plessis, 1999). Under each of these classes, there are narrower classifications according to semantic contents.

Verbs can further be classified into main and deficient verbs. Lexical items that are used as deficient verbs require main verbs to form a complete predicate. Ziervogel and Mokgokong (1969) affirm that often a deficient verb is derived from an independent verb and it assumes a different meaning and a different role of modifying the main verb. Deficient verbs provide additional information to the basic

¹ Bantu is used in this paper as a linguistic classification term. It bears no reference to the South African socio-political connotation.

meaning of the main verb which is similar to what the adverb does. For instance, the verb stem **-šala** (remain, past tense: **-šetše**) can be used as a main verb or as a deficient verb, as illustrated below:

- 1) (a) Main verb:
 - i. Baratho ba Modupi [ba [šetše].
(Modupi's younger siblings remained)
 - ii. Baratho ba Modupi [ba [šetše] thabeng.
(Modupi's younger siblings remained on the mountain.)
- (b) Deficient verb:

Baratho ba Modupi [ba [šetše] [ba [boile].
(Modupi's younger siblings are already back.)

In example 1) (a) the main verb expresses “remained” and the main verb in 1) (b) expresses “are back”. **-šetše** in 1) (b) is a deficient verb stem, providing more information about the main verb.

Furthermore, the verb in Bantu languages has a complex structure because of the languages' agglutinating nature. Besides the basic meaning of the verb, morphological units that may be affixed to the verbal root or stem carry with them some additional semantic content. For instance, the causative verbal extension **-iš-** has two basic semantic interpretations: *cause to* and *assist to*, such as in: **-loga** (knit; crotchet; plait) > **-log-iš-a** > **-logiša** (cause to/assist to knit; crotchet; plait). The causation interpretation excludes the *causer* from the action while assistance implies the involvement of the *assistor*. However, the two semantic interpretations do not apply in every context, as illustrated by the following underlined translations: **-rekiša** < **-rek-iš-a** (sell), **-sepediša** < **-sepel-iš-a** (cause to walk; assist to walk - includes accompanying - ; walk hurriedly) and **-botšiša** < **-botš-iš-a** (ask). With regard to **-bolediša** < **-bolel-iš-a** (cause to talk; talk to), the causation interpretation is indirectly implied in *talk to*, in the sense that when you talk to someone you cause them to talk back.

Verbs also assign thematic roles to their arguments. Jackendoff (1992:18) cites selectional restrictions as an example of semantic well-formedness conditions, and adds that knowledge of the language and/or of the real world are the bases on which violation or compliance to these restrictions occur. A verb's subcategorisation conditions are therefore sensitive to semantic alignment with their arguments. For instance, if the verb stem **-ya** (go) acquires an internal argument, by virtue of its semantic content, its argument has to have a thematic role of *location*.

The main aim of translation is to provide the semantic equivalent of the linguistic item in the source language, in the target language. However, according to Baker (1992), translators are often faced with challenges of non-equivalence. Given the above scenarios regarding the Northern Sotho verb and the verb in general, it becomes inevitable that challenges relating to equivalence will be encountered during a translation process. While differences in morphology and syntax are bound to impact on semantic equivalence, drawing from morphological and syntactic relevance of a particular language may inform decisions regarding equivalence. Owing to these and many other factors, the general-specific conceptual representation may not be linguistically in unison in the source and target languages. Languages have different grammatical structures, the speakers of languages have different historical and cultural experiences and all of these reflect on their languages.

2. Material and methods

The source language in the African Languages WordNet is English and, in this case, Northern Sotho is the target language. The source language provides lexical entries, definitions and examples of usage. The domain of use and suggested upper merged ontology (SUMO) tags help to clarify the concepts. The examples used in this paper are sampled from 750 (three batches of 250 synsets each) verb entries. The envisaged end product will be equivalent entries in the target language, which are semantically related. The synsets will be edited, made available for inputs and, finally, the database will be open source; aimed at language development.

Apart from definitions and usages provided, available resources include the project manual detailing procedures for the translation of synsets and documented lexical support in the form of dictionaries. The constraints include the lexical limitation of dictionaries or of the target language, cultural-linguistic and structural dissimilarity as well as specific-general conceptual dissimilarity between the source and target languages. Although the ideal for good translation is that a translator should be both bilingual and bicultural, it is not always easy to find that kind of a match. However, in the case of this paper, both translation and quality control are done by mother tongue speakers with reasonable knowledge of the target language and culture.

3. Translation and equivalence

Etymologically speaking, translation (from Latin *translatum*) is to carry over, to bring over, to remove from one place to another, to turn from one language to another (Online Etymology Dictionary). According to Nida and Taber (2003: 19), the primary aim of translation must be to reproduce the message. The process must reproduce in the target language the closest natural equivalent of the source language message in terms of meaning and style (Nida & Taber, 2003). They further differentiate between *equivalence* and *identity* by explaining that the process of obtaining *equivalence* carries over the message while obtaining *identity* preserves the form. This conforms to Nida's two types of equivalence, namely: *formal equivalence* and *dynamic equivalence*. While the former preserves the form thereby bringing across the text literally, the latter carries over the central meaning expressed in the source language. The process of providing the semantic equivalent of the source text/lexical unit while preserving the natural features of the source language requires a lot of grammatical and lexical adjustment (Nida & Taber, 2003). Although Nida's main translation involvement was with the bible, and therefore text translation, his theory is very relevant to lexical translation such as in the case of WordNet. Morphological and syntactical differences between the source and the target languages need to be observed so that, in conveying the semantic equivalent, the target language is not short-changed in terms of grammatical structure.

Equivalence is known to be the main challenge for translation. Al-Kasimi (1983:63) maintains that "absolute equivalents which have exactly the same semantics and grammatical functions in both languages are rare". On equivalence, Newmark (1995: 90) asserts that it is "an unfortunately named term implying approximate equivalence". According to Culler, the problem with equivalence is that "each language articulates or organises the world differently" (Baker, 1992: 10). Over the years proponents and opponents of the Sapir-Whorf hypothesis debated the issue of the connection between language and cognition. Jackendoff (1992), for instance, points out that the semantic representations are tightly integrated into the cognitive system of the human mind. The way in which different societies experience the world affect how their language is structured and presented, and vice versa. Equivalence challenges can be attributed to a number of factors. Among those, this paper comments and provides illustrations on factors relating to collocations, connotation, specificity, grammatical structure, foreign concepts and culture-specific concepts.

Among sense relations that exist between linguistic items are broad categories of paradigmatic and syntagmatic relations. Paradigmatic relations, on the one hand, exist between linguistic items which are substitutable by one another in the same syntactic environment. Syntagmatic relations, on the other hand, exist between linguistic items which co-occur in the same syntactic relationship. These sense relations form the basis for semantic aspects that are under discussion in this paper. For instance, collocational restrictions apply at syntagmatic level while hyponymy is a paradigmatic relation of inclusion (cf. Cruse, 2000: 150). With regard to verbs, the term "troponym" was introduced and is used instead of "hyponym" (cf. Fellbaum, 1990; Fellbaum & Miller, 1990). Fellbaum (*op. cit.*: 285) explains the troponymy relation between verbs with the formula: "To V_1 is to V_2 in some particular manner", whereby V_1 is a troponym of the superordinate V_2 .

The complexity of language is such that, as far as semantic content is concerned, one lexical item may be affected by more than one of the aspects mentioned above. Regarding language structure and expression, one language may need an adverb to express that which is inherently contained in the semantics of a verb stem in another language. Therefore, while meaning itself is a complex concept, not every verb stem will necessarily be translated as just a verb stem. Some verbs' semantic contents inherently express manner and direction. By way of illustration, the following set of verb stems have the basic meaning of **-sepela** (walk) but each one also inherently carries additional information as to how or in which direction the act is being carried out – there is no need for an adverb to express that. **-sepela** is the superordinate or hypernym and the rest of the verbs in 2) (a) – (l) are troponyms:

- 2) (a) -nanya (walk slowly)
- (b) -phakiša (walk hurriedly)
- (c) -sepediša (walk hurriedly)
- (d) -gwataka (walk fast defiantly or arrogantly)
- (e) -nanabela (walk with the aim of catching or reaching for someone or something you stalk)
- (f) -babaela (walk or tread like someone with sore feet; walk like a cat on hot bricks)
- (g) -khukhuna (walk about or roam around as if a predator; crouch; prowl; slink)
- (h) -nyonyoba (walk stealthily; walk smartly; stalk)
- (i) -katakata (walk backwards)
- (j) -tsatsanka (walk boastfully; swagger; strut)
- (k) -sesella (walk about slowly and aimlessly as if ashamed of something; loiter)
- (l) -tsherema (walk slowly as if making a great effort against psychological reluctance or physical incapacitation, without lifting feet)

The superordinate **-sepela** entails troponyms (a) to (l), but is not entailed by them. Fellbaum (1990: 45) maintains that “lexical entailment is a unilateral relation: if a verb V_1 entails another verb V_2 , then it cannot be that case that V_2 entails V_1 ”. She further explains that if the verbs are mutually entailing, then they are in a synonymous relationship and not in a superordinate-troponym relationship. In support of this view Croft and Cruse (2004: 142) suggest the following formulation for lexical entailment: “X is a hyponym of Y iff F(X) entails, but is not entailed by F(Y)”. The meaning of a troponym is therefore more specific and has additional meaning. While all of the verb stems have the semantic content expressed by *walk*, each of them also has an additional meaning that is not contained in *walk* and that they do not necessarily share. For the verb stems listed in 2) (a) – (l) above, some languages may need an adverb to capture some of added specific semantic contents inherently contained for Northern Sotho. Northern Sotho may also use a word group to express the meaning of a single lexical item from another language. With reference to the examples above, there are potentially many different manners of walking, but each language will encode some and not the others, such that no two languages encode exactly the same. As evident from 2) (j) above **-tsatsanka** corresponds to English “*swagger/strut*” while there may be no direct lexical equivalent of **-babaela**. Some of the semantic aspects that have a bearing on equivalence challenges, as were mentioned earlier in this section, will be illustrated in the following paragraphs.

3.1 Collocations

Collocations are defined by Croft and Cruse (2004: 249) as “combinations of words that are preferred over other combinations that otherwise appear to be semantically equivalent.” Leech (1981: 17) explains collocational meaning as “what is communicated through association with words which tend to occur in the environment of another word.” Collocation restriction therefore refers to a situation where one lexical item may be barred from co-occurring with another because of the existence of an-

other one with the same conceptual meaning that is semantically suited to appear in that environment.

In this project, lexical equivalents inserted are verb stems. In the grammar of the target language, verb stems are preceded by a hyphen (as reflected in the verb stems in 2) (a) – (l) above) to show that the prefixal morphemes that complete the predicate have been left out. This is so because the morphemes will be extensively varied, consistent with the nominal class system. For computational convenience in the project, the hyphens are left out of Northern Sotho verb stems. Therefore, from this point on, the verb stems will appear without hyphens. The examples in tables 1-5 relate to collocational restrictions.

Table 1 shows two separate entries of the source language, *dress* and *wear*, defined as *put on clothes* and *put clothing on one's body*, respectively.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	dress, get dressed	apara, tšwara, tšwala, rwala	put on clothes	Can the child dress by herself?	Fashion	Putting
v	wear, put on, get into, assume	apara, tšwara, tšwala, rwala	put clothing on one's body	He got into his jeans.	Fashion	Putting

Table (1)

Both entries have the target language equivalents: **apara**, **tšwara**, **tšwala** and **rwala**. The equivalents represent the same conceptual meaning as the source language synsets, as their definitions seem to require. These Northern Sotho verb stems are, however, subject to collocational restrictions. For instance, **rwala** cannot substitute **tšwara** in a syntactic environment. The semantic content of the verb may select a specific noun as internal argument and other nouns may be prohibited from appearing in the syntactic environment. **Tšwara** and **tšwala** are dialectal variants. While **apara** is a superordinate, the other three (or two) are collocationally determined troponyms – their usage will depend on the type of garment and, more specifically, the part of body being covered or clothed. **Tšwara/tšwala** collocate with pants and underwear; **rwala** with hat, earrings and shoes. For the upper body, the general word is used. Figure 1 illustrates the semantic relation of the three lexical entries. English may not encode the concepts in a similar way.

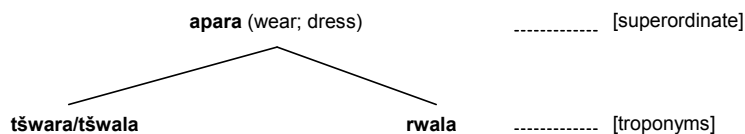


Fig. (1): Superordinate-troponym relation

Table 2 illustrates other two separate entries.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	reproduce; procreate; multiply	tswala, belega, hlatša, phaphaša	have offspring or young	The Catholic Church tells people to procreate, no matter what their economic situation may be.	biology	Sexual Reproduction

POS	Eng_Synset	Tgt_Language Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	give birth; deliver; bear; birth; have	tswala, belega, hlatša	give birth (to a newborn)	My wife had twins yesterday.	zoology	Sexual Reproduction

Table (2)

Northern Sotho equivalents **tswala**; **belega**; **hlatša** and **phaphaša** are conceptual equivalents; but they differ in terms of collocations. **Phaphaša** is excluded from the second synset because it means *hatch*. It applies to fowls and egg-laying reptiles. **Tswala** applies to cows, goats and other such animals, **belega** to human beings and **hlatša** to dogs and cats – they *litter*. Therefore, all the Northern Sotho equivalents appear for the first synset because the concept is general.

Other source language entries whose equivalents are subject to collocational restrictions are in table 3.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	have	na le; ba le; tšwa	suffer from; be ill with	She has arthritis.	factotum	ProbabilityRelation

Table (3)

The target language's first two equivalents, **na le** and **ba le**, are subject to grammatical structure, more specifically the verbal form or mood. Example 3 illustrates the use of the two verbal forms:

- 3) (a) Ke na le maphone. (I have blisters.) – [Indicative mood]
 (b) Ke na le bolwetši bja pelo. (I have heart disease.) – [Indicative mood]
 (c) Go ba le maphone go palediša motho go šoma. (To have blisters makes one unable to work.) --- [Infinitive mood]
 (d) Go ba le bolwetši bja pelo go a tšhoša. (To have heart disease is frightening.) – [Infinitive mood]

The equivalents in table 3 translate back as *have*, but it is not the only way of expressing having an illness. Other ways of expressing the same concept, subject to collocational selections, are the following: **bolawa ke/bolaya ke**², **swarwa ke/swara ke**^{*}, **opša ke/opa ke**^{*}, **rengwa ke/rema ke**^{*}, **longwa ke/loma ke**^{*} and **tswenywa ke/tshwenya ke**^{*}. They are linguistically represented in the passive construction, with the body part or type of illness serving as “agent”. For example,

- 4) (e) Ke swerwe ke mokgohlane.
 (I have flu – literally “I am caught by flu”; not “I caught flu”)

The choice of any of these verb stems will be determined by the affected body part or the type of illness. **Bolawa ke/bolaya ke**^{*} and **swarwa ke/swara ke**^{*} are hypernyms with troponyms **opša ke/opa ke**, which will select head (headache) or tooth (toothache), **rengwa ke/rema ke**^{*} will collocate only with head (headache), **longwa ke/loma ke**^{*} will select tummy (tummy ache). **Tšwa** (literally *come out*) is not used with the name of the body part affected, it is rather used with the name of the disease which can be visually attested. The past tense of **tšwa** is **tšwile**, also written as **tšwele**. The past tense is the most commonly used in this case. It expresses the state of the patient. For example,

² The asterisk sign is used to mark the forms prevalent in spoken language – cf. Kosch (2003).

- 5) Kedibone o tšwile/tšwele mauwe.
 [Kedibone (she) came out mumps*], expressing:
 [On Kedibone, mumps came out*], simply meaning:
 (Kedibone has mumps).

The following illnesses/conditions can replace **mauwe** in example (5) above: **digaroga** (rash), **sekaku/lentshwe/lentsho/sekutu** (boil/abscess), **sepšhatlapšhatlane** (chicken pox), **mooko** (measles), and other visual types of illnesses and conditions.

Very closely related to the expressions in the previous paragraph are the following, which in a very narrow sense express the type or nature of pain or discomfort, issues which will inevitably link to the body part likely to be so affected:

- 6) (a) thunya ke; thunya ke* (as of tooth ache or painful bone)
 (b) tšhatšhamelwa (ke); tšhatšhamela (ke)* (burning pain on the skin)
 (c) hlohlonwa (ke); hlohlonwa (ke)* (itch)
 (d) hlabja ke; hlabja ke* (literally *stabbed* by, which is figurative in nature but never substituted by other forms. It selects *madi* (literally *blood*) and *mašimatho* as “agents”).

The following examples, with transliterations, illustrate syntactic representations of (6) (a) – (d) above:

- (aa) Ke thunya ke leino [I am pained by tooth].
 (bb) Ke tšhatšhamelwa ke menyabidi [I am burnt by allergic reaction].
 (cc) Ke hlohlonwa ke seatla [I am itched by hand].
 (dd) Ke hlabja ke madi [I am stabbed by blood].
 (to experience internal muscular or organic pain in the torso).
 (ee) Ke hlabja ke mašimatho [I am stabbed by ‘stitch’].
 (to experience athletics-related muscular pain on the sides of one’s abdominal wall – ‘stitch’).

It becomes clear from the examples above that, in such contexts, Northern Sotho does not lexicalise possession of the body part involved (for body parts syntax, cf. Mojapelo, 2007: 123-124).

Table 4 contains other Northern Sotho equivalents, **goela; feka; kopa; robala (le); robalana; nyoba; nyobana**, that are subject to collocational restrictions.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	copulate; mate; pair; couple	goela, feka, kopa, robala, robala le, roba- lana, nyoba, nyobana	make love	Birds mate in Spring.	sexuality	Sexual Reproduc- tion

Table (4)

The first three verbs would select examples such as **kgomo** (cattle), **mpša** (dog), **kgogo** (chicken), respectively. The last four (which are actually two basic verbs) apply to human beings. Syntactically, all of these verbs can only select a [+male] external argument while the internal argument is normally [-male]. As far as human beings are concerned, the verbs can also express reciprocation by adding extensional suffix **-an-**.

Table 5 is the last example illustrating collocational selections/restrictions.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	chew, masticate, masticate, jaw	sohla, sotla, hlahuna, phura, phuphura, ota, sesena	chew (food)	The cows were masticating the grass.	gastronomy	Eating

Table (5)

The concept concerns a step in the digestive process. Northern Sotho **sohla**, **sotla**, **hlahuna**, **phura**, **phuphura** and **ota**'s collocational selections involve also the texture of the object and the stage of the digestive process. **sohla/sotla/hlahuna** generally collocate with soft objects. **Sesena** is when the front teeth are used to chew without putting the whole object in the mouth. **Phura** and **phuphura** collocate only with hard objects like bones, hard candies and pits of fruit and vegetable. Only when the object has been gnawed and crunched can it collocate with **sohla/sotla/hlahuna**. **Ota** means to chew the cud.

The situation, therefore, is that one or more English synsets will have a number of same Northern Sotho words as equivalents, but which will not necessarily substitute one another in all contexts. They are collocationally restricted to certain arguments. Another factor, with the exception of what Baker (1992: 13) terms propositional meaning, is "what other semantic implications do speakers of a language as a group or as individuals attach to a particular linguistic item?". The next section will address those kinds of associations.

3.2 Connotation

A word's basic meaning is termed its conceptual, connotative, denotative or logical meaning (cf. Leech, 1981: 9, 23), similar to Baker's (1992) propositional meaning. According to this kind of meaning assignment, a word is defined by semantic features that make its referent differentiable from others. That is, the meaning can be conceptualised by componential analysis. It covers the "basic, essential components of meaning which are conveyed by the literal use of a word" (Yule, 2003: 114). Over and above the conceptual meaning of a word, there are other implications or associations present. "Connotations are relatively unstable ... they vary according to culture, historical period and the experience of an individual" (Leech, 1981: 13).

Examples in table 6 illustrate this type of meaning.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	reduce; melt off; lose weight; slim; slenderize; thin; slim down	phophothega, fokotšega, fela, ota, gwamelwa	take off weight	[empty]	body_care	Removing
v	gain; put on	oketšega, akola, šušumoga, nona,	increase (one's body weight)	She gained 20 pounds when she stopped exercising.	body_care	Increasing

Table (6): Connotation

Equivalents of the first synset **phophothega; fokotšega; fela; ota; gwamelwa** are subject to connotations, and in table 6 they are arranged from the neutral or the least negative down to the most in-

clination towards negative connotation. **Fela** is subject to how the speaker uses it; it can carry either a positive or a negative connotation. To induce an unambiguously non-negative connotation, often loan words are used. Compare the following sentences of example 6 with the same conceptual meaning, namely “**Mosima has reduced weight**”:

- 7) (a) Mosima [o [slimile]. (From English *slim*)
 (b) Mosima [o [gwametšwe].

In contrast to 7) (a), (b) paints an unsightly picture of Mosima’s weight reduction. 7) (a) depicts Mosima’s weight reduction as pretty, unless the remark is meant to be sarcastic. The antonym synset *gain; put on* defined as *increase (one’s body weight)* has in the target language **oketšega, šušumoga, akola** and **nona**.

In the first place it depends on the cultural-historical orientation or the mental orientation of the speaker, the hearer and the society; whether having a big body is viewed in a positive light as compared to having a small body. Whether a big body is a reflection of good life, healthy life, neglect or unhealthy life will call for selection of a word which will either be neutral or reflect envy, a compliment, disgust, pity or sarcasm. The situation triggers a mental continuum ranging from obese to skinny, with either the neutral, positive or negative connotation attached. The connotation also reflects the speaker’s feelings or attitude either towards the person talked about or towards the particular body size or condition. In giving an equivalent, all of them are relevant but the selection for use will be determined by connotations.

3.3 General-specific conceptual (mis)alignment

A concept may be general in one language while it is specific in another language and vice versa. Another scenario is that a general or specific concept may possibly not be lexicalised or may be represented as a word group in another language. Tables 7-9 illustrate this point. In table 7 the source language synset has target language equivalents **imiša, senya** and **roba leoto**. While **imiša** is general, **senya**’s meaning is specific.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	impregnate; knock up; bang up; prang up	imiša, senya; roba leoto	make pregnant	He impregnated his wife again.	biology	Sexual Reproduction

Table (7): generality or specificity in the source and target languages may not necessarily correspond

The semantic content includes the relationship between the parties involved as well as the state of the female prior to the condition. **Senya** happens only to a first time unmarried mother-to-be; **roba leoto** is its figurative counterpart.

Table 8 illustrates a case where the target language lacks lexical expression of the general concept expressed in the source language.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	administer, dispense	fa, nweša, noša, tlotša, tsikitla, tšhela	give or apply (medications)	[empty]	medicine	TherapeuticProcess

Table (8): The target language may lack a lexical expression of the general concept expressed in the source language

Equivalents **fa**, **nweša/noša**, **tlotša**, **tsikitla** and **tšhela** can be back-translated as *give*; *make to drink* (including pills); *apply* for example, cream or ointment, *rub* and *pour*, respectively. Northern Sotho does not lexicalise the general concept *administer*. **Fa** (give) will be inappropriate since it will exclude *apply* and *pour*.

In table 9 the target language lacks lexical expression of the specific concept expressed in the source language.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	widow	bolaela	Cause to be without spouse	The war widowed many women in the former Yugoslavia.		Death

Table (9): The target language may lack a lexical expression of the specific concept expressed in the source language

In this case, Northern Sotho does not lexicalise the specific concept. Lexicalisation is general, and the internal argument will shed light on who is it that has been killed. Northern Sotho word for *kill* is **bolaya**. The equivalent will require affixation of applicative **-el-** > **bolaela** followed by the word for husband, which will perfectly be represented by a word group rather than a single lexical item. Applicative **-el-** in this case means “*to the detriment of*”. Alternatively, a phrase which will be back-translated as “*change/turn her into a widow*”, **fetola/fetoša mohlologadi**, will convey the same meaning.

Another scenario is that the target language verb may require assistance of an adverb to express the concept from the source language, as illustrated in table 10.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	revolutionize; revolutionise; overturn	fetoša kudu, fetoša ka maatla	change radically	Email revolutionised communication.	factotum	Process

Table (10): The target language may need an adverb to express the concept only by a verb in the source language

In this case, Northern Sotho requires the assistance of an adverb to modify a word for *change*, **fetoša**, in order to capture the scale of the act or process, with phrases [**fetoša [kudu]**] and [**fetoša [ka maatla]**]. The fact that an adverb would be needed to express the scale of the act or that a concept may be expressed differently is an indication that languages are structured differently. The next section concerns grammatical structure, which also reflects on equivalence challenges.

3.4 Grammatical structure

While Northern Sotho is a Bantu language of the Niger-Congo family, English is an Indo-European language. The grammatical structures of the two languages differ on many levels, including the agglutinating nature of Northern Sotho. The verbal affixes in Northern Sotho are extensively productive, including the shades of meaning that they add to the basic meaning of verbs. From the English definition and usage in table 11, the Northern Sotho equivalent **tsentšha/ tsenya** will apply only to the context of the example of usage provided.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	get; let; have	[empty]	cause to move; cause to be in a certain position	He got a girl into trouble.	factotum	Process

Table (11): Lack of one-to-one corresponding lexical expression due to difference in grammatical structure

On the other hand, the causative morpheme **-iš-** affixed to whatever verb that expresses the act that the subject is being caused to do seems to include what is represented in the definition. The challenge in entering affixes as equivalents is that the equivalent is expected to be a verb stem, which should not have hyphens around it. Secondly, while leaving out the hyphens may misrepresent the entry as a verb stem, retaining the hyphens may interfere with the computational part of the project (the programming script that converts the Excel file into XML must be able to read the data properly). To avoid misinterpretation, the target language synset slot is left open.

3.5 Foreign concepts

Another challenge in finding an equivalent is when the entry in the source language represents a concept that is foreign to the target language. An example of this is in table 12.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	tease; fluff	[empty]	ruffle (one's hair) by combing to- wards the ends towards the scalp, for a full effect.	[empty]	body_care	ShapeChange

Table (12): Lack of lexical expression due to the concept being foreign to the target language

Hair can be ruffled to make it untidy with equivalent: **hlahlamola**, but it does not seem to be what this entry is about. This seems to imply some relative length between the scalp and the end of the hair. One wonders what the relevance of the concept is for Northern Sotho: what full effect? Language reflects the experience of the speakers, be it anatomic, cultural or historical. The concept seems foreign to the target language.

3.6 Culture-specific concept

According to Al-Kasimi (1983: 62), there is a close relationship between language and culture. Baker (1992: 21) identifies culture-specific concepts as one of the challenges for finding an equivalent in translation, but her point of departure in this respect is the source language. She explains culture-specific challenges as situations where the linguistic item in the source language is culture-specific, and is not culturally relevant to the target language. In this sub-section, though, culture-specific concepts are being viewed from the target language perspective. Firstly, a concept may be familiar, but due to culture-specific factors can be expressed only idiomatically as far as humans are concerned, or else it can be perceived as an insult or abhorrent to say. Secondly, as translation progresses, some gaps emerge regarding some senses that are not necessarily peripheral in the target language, but have not been catered for, probably because they are not central in the source language.

The case of 'miscarry' in table 13 illustrates this point.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	miscarry	folotša , goma tseleng, boa tseleng	suffer miscarriage	[empty]	biology	Death

Fig. (13): Non-lexicalisation or idomatisation due to cultural observance

Northern Sotho equivalents are **folotša**, **boa tseleng** and **goma tseleng**. **Folotša** is the only non-idiomatic one, but it is never used for persons. Only animals **folotša**. The idiomatic ones used for persons literally mean “turn back while on the way” and this implies that the subject, namely the pregnant woman, has not been able to reach her destination. In terms of pregnancy, **boa/goma tseleng** (turn back while on the way) means not running full term.

Another factor is that the translation process may reveal a gap that may never be filled in the target language because the concept is culture-specific. Multiple appearance of a specific verb as the target language equivalent of various source language entries brings several concepts involving the verb to mind. Most of these concepts may find the way in, but some may never appear. While **swara** appears several times as the target language equivalent, as illustrated in table 14, it becomes clear that some senses linked to the verb may never be included in the project.

POS	Eng_Synset	Tgt_Language_Synset	Eng_Definition	Eng_Usage	Domain	SUMO
v	catch	khetšha; swara	be the catcher	Who is catching?	baseball	Sport
v	catch; grab; take hold of	swara	take hold of so as to seize or restrain or stop the motion of	Grab the elevator door!	factotum	Touching
v	handle; palm	swara	touch, lift, or hold with the hands	Don't handle the merchandise.	factotum	Touching
v	seize; pre-hand; clutch	swara	take hold of; grab	Birds of prey often seize small mammals.	factotum	Grasps
v	seize	thopa; swara	take or capture by force	The rebels threaten to seize civilian hostages.	factotum	Unilateral Getting
v	collar; nail; apprehend; arrest; pick up; nab; cop	swara; golega	take into custody	The police nabbed the suspected criminals.	law	Imprisoning

Table (14): The same verb as an equivalent of various source language entries

For example, among the meanings of **-swarega** < **-swar-eg-a** there are cases wherein the conventional meaning of neuter **-eg-** “potential” does not apply. Affixed to the verb stem **-swara**, the neuter extension brings out different meanings such as “to be held up and be unable to perform a certain duty or to go somewhere, to be very ill, to have a need to respond to the call of nature, and to be caught in the act and be unable to move”. The last semantic interpretation is also associated with witchcraft – likened to being caught with a hand in the cookie jar and freezing when caught. Whether or not the situation is possible is beside the point; the point is that it is part and parcel of the idiom of the language.

A similar situation involves the verb stem **-bona** (see) and its extensional affixes. While the verb stem and its extensions would serve as equivalent to senses such as *see*, *visible*, *understand* and *real-*

ise, there may not be a chance to capture, for instance, all senses of **-bonela** < **-bon-el-a**. Applicative affix **-el-** has semantic interpretations: *on behalf of, for the benefit of, to the detriment of, onto and towards*. One of the senses, which is culturally significant, may not be lexicalised in the source language as it is in the target language. It means “*seeing private parts or undergarments of*”. Traditionally, a child would know that he/she must look away when, for instance, he/she sees an older person standing next to a footpath relieving him/herself so that he/she does not **-bonela** them – and the myth attached to the taboo is that should anyone not abide by that they will turn blind. Young girls are also taught correct postures for sitting down, bending and standing up from the sitting position so that no one **-bonela** them. This interpretation is culture-specific, being part of traditional informal curriculum for raising children with respect for themselves and others in a society that upholds a certain level of moral standards. One may argue that such concepts are outdated and that societies have moved on with times, but leaving out cultural fundamentals is tantamount to destroying a language. Taking note of such gaps may ensure that in one way or another some target language gaps get filled. Culture-specific concepts or words like the two discussed above do not hamper the translation process, but they are conspicuously missing.

4. Conclusion

With the challenges of equivalence discussed and illustrated in this paper, it will be interesting to see how the pre-editing stage is going to unfold; and how the resultant trial ontologies will prove and provide. It remains to be seen whether semantic aspects such as collocational restrictions, connotations, general-specific, structural differences, foreign concepts and culture-specific and other (non)-equivalence issues will enrich or impoverish the languages the project seeks to advance. As translation progresses, some lexical items and/or concepts are clearly missing, a fact which points to the two languages’ different cultural-historical backgrounds, which in turn affects their linguistic make-up. The spin-off of this is that it may provide an opportunity to have original words of the target language also documented and preserved.

It may also be interesting to experiment with a situation where the target and source languages swap places. While Northern Sotho keeps technologically abreast with other languages, it is also important that the language’s idiom does not fall through the cracks. The project should ideally be a useful tool to also document, preserve and contextualise words and concepts that are original to the target languages.

Acknowledgements

I acknowledge with gratitude the helpful comments from of the following scholars, who read earlier drafts of this paper:

1. Prof James Mafela (Unisa): first draft
2. Prof Christiane Fellbaum (Princeton): pre-final draft.

References

- Al-Kasimi, A.M. (1983), *Linguistics and bilingual dictionaries*. Leiden: EJ. Brill.
- Baker, M. (1992), *In other words: a coursebook on translation*. London & New York: Routledge.
- Chow, I.C. & Webster, J.J. (2008), “Supervised clustering of the WordNet VerbHierarchy for systemic functional process type identification”. *ICGL 2008 International Conference on Global Interoperability for Language Resources*, 51 – 58.
- Croft, W & Cruse, D.A. (2004), *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Cruse, D A. (2000), *Meaning in language*. Oxford: Oxford University Press.
- Du Plessis, J.A. (1999), *Lexical semantics and the African languages*. Stellenbosch: Stellenbosch University.
- Fellbaum, C. (Ed.) (1998), *WordNet: An electronic lexical database*. Cambridge, Mass: MIT Press.
- Fellbaum, C. (1990), “English verbs as a semantic net”. *International journal of Lexicography* 3(4):278-301.
- Fellbaum, C & Miller, G.A. (1990), „Folk psychology or semantic entailment? A reply to Rips and Conrad (1989)”. *Psychological review* 97:565-570.

- Jackendoff, R. (1972), *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Kosch, I.M. (2003), "Sandhi under the spotlight". *South African Journal of African Languages* 23(3) 144-153.
- Leech, G. (1981), *Semantics*. Middlesex: Penguin Books.
- Mojapelo, M.L. (2007), *Definiteness in Northern Sotho*. Unpublished D.Litt. dissertation. Stellenbosch: University of Stellenbosch.
- Newmark, P. (1995), *A textbook of translation*. New York: Phoenix.
- Nida, E.A & Taber, C.R. (2003), *The theory and practice of translation*. Leiden: Koninklijke Brill NV.
- Poulos, G & Louwrens, L.J. (1994), *A linguistic analysis of Northern Sotho*. Pretoria: Via Afrika.
- Yule, G. (2003), *The study of language*. Cambridge: Cambridge University Press.
- Ziervogel, D, Lombard, D.P & Mokgokong, P.C. (1969), *A handbook of the Northern Sotho language*. Pretoria: Van Schaik.
- Ziervogel, D & Mokgokong, P.C. (1975), *Pukuntšu ye kgolo ya Sesotho sa Leboa*. Pretoria: Van Schaik.

Web references

- Online Etymology Dictionary* Available at <http://www.etymonline.com/> accessed on 2010-05-28.
- http://en.wikipedia.org/wiki/Linguistic_relativity accessed on 2010-08-31.
- <http://wordnet.princeton.edu> accessed on 2010-04-02