The Compilation of Bilingual Dictionaries between African Languages in South Africa: The Case of Northern Sotho and Tshivenda*

Kwena J. Mashamaite, *Department of Northern Sotho, University of the North, Pietersburg, Republic of South Africa* *(mashamaitek@unin.unorth.ac.za)*

Abstract: Bilingual dictionaries between South African languages have existed for more than three centuries (Smit 1996: 232). These dictionaries have English or Afrikaans as the source language and an African language as the target language. There has never been a case of the opposite except for bi-directional bilingual dictionaries. Still more, there seems to be no record of an attempt ever made to date to compile bilingual dictionaries between African languages.

This scenario illustrates the trend and history of lexicographic practice in South Africa. It is evident from most lexicographic research and products such as the existing kinds of dictionaries that Africans were never considered to be the first target group of users or prospective beneficiaries of such lexicographic products. The absence of bilingual dictionaries between African languages again provides evidence of who the lexicographers were and which population groups they represented.

To fill this void, a model called the hub-and-spoke is proposed in this paper for the compilation of such kind of dictionaries. The model has been chosen for its purported economy of use. Northern Sotho and Tshivenda are the African languages that will be used as examples in the application of the model.

A purposive sampling technique will be used to select lexemes that will constitute the entry words which form the central list of the proposed dictionary.

Keywords: BI-DIRECTIONAL, CENTRAL LIST, CONCEPTUAL EQUIVALENCE, ENTRY WORDS, HUB-AND-SPOKE MODEL, LEXEMES, LEXICAL ITEM, LEXICAL UNIT, LEXICALISATION STATUS, PRAGMATIC CONTRAST, PURPOSIVE SAMPLING TECHNIQUE, SOURCE LANGUAGE, TARGET LANGUAGE, VARIANT STATUS

Opsomming: Die samestelling van tweetalige woordeboeke tussen Afrikatale in Suid-Afrika: Die geval van Noord-Sotho en Tshivenda. Tweetalige woordeboeke tussen Suid-Afrikaanse tale bestaan vir meer as drie eeue (Smit 1996: 232). Hierdie woordeboeke het Engels of Afrikaans as brontaal en 'n Afrikataal as doeltaal. Daar was nooit 'n

---

* This paper was presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

geval van die teenoorgestelde nie buiten vir tweerigting tweetalige woordeboeke. Nog meer, daar skyn geen rekord te wees van 'n poging tot op hede om tweetalige woordeboeke tussen Afrikatale saam te stel nie.

Hierdie scenario illustreer die verloop en geskiedenis van leksikografiese praktyk in Suid-Afrika. Dit is duidelik uit die meeste leksikografiese navorsing en produkte soos die bestaande woordeboeksoorte dat Afrikane nooit beskou is as die eerste doelgroep gebruikers of waarskynlike voordeeltrekkers uit sulke leksikografiese produkte nie. Die afwesigheid van tweetalige woorde-boeke tussen Afrikatale gee ook 'n aanduiding van wie die leksikograwe was en aan watter bevolkingsgroepe hulle behoort het.

Om hierdie leemte te vul, word 'n naaf-en-speek-model vir die samestelling van hierdie soort woordeboeke in dié artikel voorgestel. Die model is gekies vir sy beweerde ekonomiese aard. Noord-Sotho en Tshivenda is die Afrikatale wat as voorbeelde gebruik sal word in die toepassing van hierdie model.

'n Doelgerigte toetsingstegniek sal gebruik word om lekseme uit te soek wat die inskrywings in die sentrale lys van die voorgestelde woordeboek sal vorm.

Sleutelwoorde: TWEERIGTING SENTRALE LYS, KONSEPTUELE EKWIVALENSIE, IN-SKRYWINGS, NAAF-EN-SPEEK-MODEL, LEKSEME, LEKSIKALE ITEM, LEKSIKALE EEN-HEID, LEKSIKALISASIESTATUS, PRAGMATIESE KONTRAS, DOELGERIGTE TOETSINGS-TEGNIEK, BRONTAAL, DOELTAAL, VARIANTSTATUS

1.   Introduction

In South Africa there are eleven official languages and other non-official lan-guages such as Khoi and San which in future are also likely, if developed, to acquire official status. South Africa as a multilingual country needs bilingual dictionaries, especially bi-directional learner's dictionaries so that the speakers of these languages can learn each other's language.

The absence of bilingual dictionaries between the African languages can be ascribed to the shortage of trained African lexicographers. Fortunately the Pan South African Language Board, in collaboration with the African Associa-tion for Lexicography, has pledged to address this backlog by providing the necessary training.

One of the major tasks bilingual dictionaries in South Africa has to per-form, is to assist speakers of various languages to learn to speak each other's language in order to promote multilingualism.

Bilingual dictionaries may be described as dictionaries in which lexical items of the native language are given translation equivalents in a foreign lan-guage. The native language may be the source language whereas the foreign language is the target language or vice versa.

Bilingual dictionaries may serve different purposes depending more on the communicative needs of the dictionary users than on the amount of infor-mation supplied by the compiler. Mdee (1997: 95) argues that a bilingual dic-

tionary has two functions, the first being that of comprehension, that is, reading and listening, and the second that of production, which means writing and speaking. One may add a third function which is that of translation.

For a bilingual dictionary to be useful to speakers of both languages it has to be bi-directional. *Bi-directional* refers to a condition whereby a bilingual dictionary is structured in such a manner that the speakers of both languages may use it for either encoding, decoding or any other purpose, including translation. This condition circumvents a situation in which four mono-functional mono-directional dictionaries per language of a language pair would be required for the same functions. In a South African context in which there are eleven official languages, one is speaking of 55 different language pairs times four dictionaries per language pair, yielding a total of 220 different kinds of dictionaries. Surely a developing country like South Africa cannot afford to produce such numbers of dictionaries.

It is against this background, as well as because of the need to have such dictionaries, that a more economic solution is sought to achieve the same goal. Consequently, the hub-and-spoke model proposed in this paper seems to be just the kind of solution needed.

According to Mashamaite (1995) and Martin (1995) the main contention of the hub-and-spoke model is that several bilingual dictionaries can be derived by linking the lexical items of the spoke languages, that is, the source languages in a bilingual dictionary, to those of a common hub language which is the target language in a language pair. The condition for the application of the model is that one must have at least two bilingual dictionaries which share a common target language or hub. For instance, in the example used in this paper, Northern Sotho and Tshivenda are the spoke or source languages sharing a common hub or target language which is English.

The application of the model does not require prior collection of data or the building of corpora. It therefore reduces costs normally incurred in the process of hiring fieldworkers and computer experts needed for the building of corpora and lexical databases as well as time spent on this procedure. This is why the model is thought to be cost-effective. It only uses already existing data in the form of dictionaries.

A purposive sampling technique will be used in this study to select lexemes which will form entry words that constitute the central list of the proposed dictionary. Although a non-probability technique, it is, according to Neuman (1997), an acceptable sampling technique used for special situations. One of these special situations is when a researcher wants to identify particular types of cases for in-depth investigation. It has been chosen for this study in particular because it allows one to choose lexemes that might pose an obstacle to the application of the model and consequently affords one the opportunity to suggest a probable solution, something that no other kind of sampling is able to do.

In the application of the model special attention will be paid to the explanation and use of the parameters of the hub-and-spoke model, namely concep-

tual equivalence, pragmatic contrast, variant status, and lexicalisation status.

2.   The hub-and-spoke model

The hub-and-spoke model has been explained in the introduction above. This section will look into the parameters of the hub-and-spoke model and the role they play in the application of the model. The parameters have been borrowed from Omkeerbare Bilinguale Lexicale Databanken (OMBI), translated as Reversible Bilingual Lexical Databases, which is a language editor developed by the Dutch software house Software Engineering Research Centre (SERC) in Utrecht in the Netherlands (Martin 1995).

The purpose of the parameters is to ensure reversibility of translation equivalence. Translation equivalence according to Svensén (1993: 140) is achieved when a bilingual dictionary is able to "provide words and expressions in the source language with counterparts in the target language which are as near as possible, semantically and as regards style level (register)". It is important to note that in providing translation equivalents one is dealing with meaning equivalence rather than with word equivalence. To support this assertion Martin (1995) argues that in translation relations "it is not words that are translated into other words, but rather words in a specific meaning". He illustrates his argument by the use of the terms *form unit* (FU) to refer to a word form, and *lexical unit* (LU) to refer to a word-specific meaning. One cannot give a precise translation equivalent of an FU because an FU can have more than one meaning. Only meanings or lexical units (LUs) can be given translation equivalents which are also LUs and not word forms in a target language. For example, the Northern Sotho FU *panka* has the following LU: 1. bench 2. desk 3. bank. These three meanings (LUs) have the following translation equivalents which are also the LUs in the target language namely *panka/senno*, *teseke/panka* and *panka* respectively in Northern Sotho.

On the other hand, the English FU *bank* has the following meanings or LUs: 1. panka 2. leriba 3. lebopo 4. khwiti (Kriel 1988). It is apparent from the preceding examples that the English FU *bank* and the Northern Sotho FU *panka* have only one common LU which is *bank* (in the sense of money). It is therefore incorrect and misleading too to assume that the FU *bank* has only one translation equivalent in Northern Sotho whereas there are four meanings as indicated. To capture all four meanings one should focus on the translation equivalents of the lexical units attached to the form unit and not the form unit itself. This confirms the assertion that only lexical units and not form units can have translation equivalents in a target language.

3.   Parameters of the hub-and-spoke model

To ensure that only meanings and not words are translated, the following conditions in the form of parameters should be met.

3.1   Conceptual equivalence

The term treats of a situation in which the target language translation equivalent refers exactly to the same concept as that referred to by the source language lexical item. It is not always easy to find this kind of translation equivalence. The difficulty lies in the fact that every language has a set of concepts which are perceived differently from or are even nonexistent in another language as a result of historical, geographical, cultural, social and economic differences between countries where these languages are spoken (Svensén 1993).

In a situation where the target translation equivalent refers to the same concept as that in the source language, one would speak of complete equivalence. For example, Northern Sotho *noka* refers to English *river.* It happens sometimes that the conceptual relationship between the target and the source language lexical items is not complete or exactly the same. One would refer to such a situation as partial equivalence. This may occur in the following three ways:

(a)   The source language lexical item may have two or more target language translation equivalents, a state referred to as divergence. It is a situation in which the meaning relation between the source language lexical item is that of a hyponym of the lexical item of the target language. For example, English *tick* refers to Northern Sotho *kgofa; patšane*; English *uncle* refers to Northern Sotho *malome; ramogolo; rangwane* (Kriel 1988).

(b)   Several lexical items in the source language may have one translation equivalent in the target language. This situation is known as convergence. The target language translation equivalent becomes the super-ordinate. For example, Northern Sotho *moeti, mosepedi* refers to English *tourist* (Kriel 1988). When the translation equivalents cannot be conceptually regarded as either hyponyms or hyperonyms but show various degrees of overlap between each other, the condition is known as related equivalence. For example, English *shine* refers to Northern Sotho *kganya, nyedima, phadima* (Kriel 1988).

(c)   The lexical item of the source language may not have any appropriate translation equivalent in the target language. This situation is known as non-equivalence. To overcome this problem the source language lexical item is either substituted by an explanatory equivalent or a phrase, or by a borrowed word, or by another word with a close conceptual meaning, that is, a near-equivalent. Non-equivalents occur very often with the translation of terms relating to culture-specific concepts. For example, Northern Sotho *hlatswadirope* is translated in English as *ancillary wife* (Ziervogel and Mokgokong 1975). The translation equivalent is deficient in the sense that it does not reveal the cultural significance underlying the purpose of having an ancillary wife. An ancillary wife is not any second wife but the principal wife's sister who is married with the specific purpose of bearing children for her barren sister.

3.2   Pragmatic contrast

The term *pragmatic* refers to the contextual adequacy of a word. A translation equivalent should have the same contextual meaning as its source language counterpart. The pragmatic component consists of the subject field, style, connotation, chronology (time), frequency and geography (place). For the translation equivalent to be exact it should have no contrast with the source language counterpart on the basis of the stated pragmatic components. This condition will be specified as **nil** which means that pragmatic components are the same for both words. Sometimes the translation equivalent falls short of one or two of the pragmatic components and, depending on which component is missing, could be described as **obsolete**, **formal**, **neutral**, or **informal**. For example, the translation of the English *window* in Northern Sotho is *letsikangope* which can be labelled as obsolete because of chronology and frequency.

3.3   Variant status

This term refers to a situation where the lexical item of the source language has more than one translation equivalent in the target language. These translation equivalents may be regarded as synonyms although they might have functional restrictions depending on the context in which they are used. The status of the variant may be specified as main native synonym, loan word, hybrid term, and loan translation. Native synonyms are target language lexical items derived from the native language lexicon. Loan words are the adopted lexical items. Hybrid terms are words one part of which is derived from one language and the other part from another language. For example, *condensed milk* is translated into Tshivenda as *mafhi a khondasi* (Wentzel and Muloiwa 1982). *Mafhi* is a Tshivenda native word for *milk* and *khondasi* is a loan word from English *condensed*. A loan translation is a compound word whose constituents are literal translations of the constituents of a corresponding compound in another language (Martin et al. 1992: 43). For example, English *medical examination* is translated into Northern Sotho as *tlhahlobo ya kalafo* (Kriel 1988).

3.4   Lexicalisation status

The concept refers to a situation in which the lexical item of the source language may or may not have a translation equivalent which is a lexical item in the target language. In a situation where the translation equivalent is a lexical item, it will be specified as **lexicalised.** Some lexical items of a given language express culture-bound concepts which may not form part of the conceptual world of the target language. In this case, a translation equivalent may not be a lexicalised word but a description or phrase which explains the conceptual meaning of the source language concept. For example, Northern Sotho *thiswana* means in English *little basin, small vessel* (Kriel 1988). The two English transla-

tion equivalents do not appear in the central list as English entry words. This is an indication that they are not lexicalised simply because they have no cultural referent in English. Such non-lexicalised words are specified as either **semi-lexicalised** when they can be inserted in some contexts as equivalents or **non-lexicalised** when they are not insertible in the place of the source language lexical item but only serve to explain or describe its meaning (Mashamaite 1995).

4.   Application of the hub-and-spoke model

The process of providing source language lexical items with appropriate target language translation equivalents applying the parameters of the hub-and-spoke model will look as follows:

(a)   (i)   Source language: Northern Sotho — **thelebišene**
            Target language: English — **television**
            Lexicalisation status: lexicalised
            Variant status: main
            Pragmatic contrast: nil
            Conceptual equivalence: complete

       (ii)   Source language: Tshivenda — **thelevishini**
            Target language: English — **television**
            Lexicalisation status: lexicalised
            Variant status: main
            Pragmatic contrast: nil
            Conceptual equivalence: complete

       (iii) Source language: Northern Sotho — **thelebišene**
            Target language: Tshivenda — **thelevishini**
            Lexicalisation status: lexicalised
            Variant status :   main, loan word
            Pragmatic contrast: nil
            Conceptual equivalence: complete

       (iv)   **thelebišene,** thelevishini

(b)   (i)   Source language: Northern Sotho — **setšidifatši**
            Target language: English — **refrigerator**
            Lexicalisation status: lexicalised
            Variant status: main
            Pragmatic contrast: nil
            Conceptual equivalence: same

(ii) Source language: Tshivenda — **tshirothodzi**
Target language: English — **refrigerator**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same

(iii) Source language: Northern Sotho — **setšidifatši**
Target language: Tshivenda — **tshirothodzi**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same

(iv) **setšidifatši,** tshirotodzi

(c) (i) Source language: Northern Sotho — **phasa**
Target language: English — **bring sacrifice to ancestral spirits**
Lexicalisation status: non-lexicalised
Variant status: substitution by explanation
Pragmatic contrast: nil
Conceptual equivalence: same

(ii) Source language: Tshivenda — **phasa**
Target language: English — **bring sacrifice to ancestral spirits**
Lexicalisation status: non-lexicalised
Variant status: substitution by explanation
Pragmatic contrast: nil
Conceptual equivalence: same

(iii) Source language: Northern Sotho — **phasa**
Target language: Tshivenda — **phasa**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same

(iv) **phasa,** phasa

In the above examples (a) and (b) have the same parameters and it is therefore easy to derive the translation equivalents of Tshivenda and Northern Sotho as illustrated by (iv) in both cases.

The difference in one parameter in (c) (ii) indicating a non-lexicalised English translation equivalent does not pose a serious obstacle to the process since the substitution by explanation does not contrast with the pragmatic and conceptual framework of the source language.

What makes the situation even simpler is the fact that the Tshivenda translation equivalent matches all the parameters of the source language lexical item in Northern Sotho thereby becoming a complete equivalent.

The above exercise can be summarised by the following diagrammatic representation.


```
———————    Direct link from spoke to hub
—————→     Direct link from spoke to spoke
```
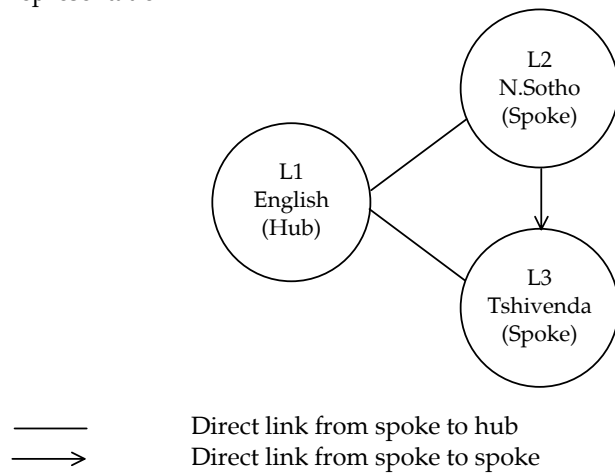
5.   Conclusion

This article, which acknowledges the multilingual nature of the South African population, has expressed a concern about the lack of bilingual dictionaries between African languages. The primary purpose of bilingual dictionaries is to assist the speakers of the various languages spoken in South Africa to learn each others' languages in order to promote multilingualism. One of the reasons mentioned for the lack of such dictionaries is the shortage of trained African lexicographers. Fortunately, the Pan South African Language Board, in collaboration with the African Association for Lexicography, has pledged to address this backlog by providing the necessary training.

Three broad functions of bilingual dictionaries have been identified, namely

(a)   that of comprehension, that is, reading and listening,
(b)   that of production, that is, writing and speaking, and
(c)   that of translation.

The article has again acknowledged the fact that the process of making dictionaries is an expensive one both in terms of personnel, time and funding. This

reason alone has served as motivation to look for a less expensive alternative which is the hub-and-spoke model proposed here.

An attempt has been made to explain and illustrate how the model can be applied. My proposal is that it should be tried.

References

Kriel, T.J. 1988. *Popular Northern Sotho Dictionary.* Pretoria: J.L. van Schaik.

Martin, W. 1995. *Lexicographical Resources in a Multilingual Environment: An Orientation.* Unpublished paper. Amsterdam: Vrije Universiteit.

Martin, W. et al. 1992. Standards for the Terminological Description of Lexical Items. Multilex report. Amsterdam: Vrije Universiteit.

Mashamaite, K.J. 1995. *The Hub-and-Spoke Model: A Recipe for Making Bilingual Dictionaries between African Languages in South Africa.* Unpublished doctoral paper. Amsterdam: Vrije Universiteit.

Mdee, J.S. 1997. Language Learners' Use of a Bilingual Dictionary: A Comparative Study of Dictionary Use and Needs. *Lexikos* 7: 94-106.

Neuman, W.L. 1997. *Social Research Methods.* Boston: Allyn and Bacon.

Smit, M. 1996. *Wiegand's Metalexicography as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa.* Unpublished D.Litt. thesis. Stellenbosch: University of Stellenbosch.

Svensén, Bo. 1993. *Practical Lexicography.* Oxford: Oxford University Press.

Wentzel, P.J. and T.W. Muloiwa. 1982. *Ṱhalusamaipfi ya Nyambotharu yo Khwiniswaho. Ḽuvenda. Luvhuru. Luisimane.* Pretoria: University of South Africa.

Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho Woordeboek.* Pretoria: J.L. van Schaik.