

# Lemmatisation of Fixed Expressions: The Case of Proverbs in Northern Sotho\*

I.M. Kosch, *Department of African Languages,  
University of South Africa, Pretoria, South Africa*  
([koschim@unisa.ac.za](mailto:koschim@unisa.ac.za))

---

**Abstract:** The purpose of this article is to make a quantitative and qualitative assessment of the lexicographic treatment and listing of proverbs in the *Wörterbuch der Sotho-Sprache* (Endemann 1911) in comparison to selected Northern Sotho dictionaries. In order to accommodate proverbs, which are fixed multiword expressions, they are customarily entered as sub-lemmas under a particular simple headword, usually one of the key components of a proverb. The selection of a key component relies on the subjective judgement of the lexicographer. This selective approach may result in proverbs falling between the cracks if none of the components strike the compiler as prominent enough to justify the inclusion of a proverb under a particular headword. This seems to have been the case in the dictionary under investigation, given the dearth of proverbs taken up in this work. On the other hand their omission could simply be ascribed to a practical consideration such as limited space in a printed dictionary. A dictionary user might find it challenging to look up a desired proverb, especially if the individual words have a very low general frequency or are even obsolete in modern life. In that case, an electronic format of a dictionary would be most enabling, allowing for an electronic search. Special purpose dictionaries dedicated to culturally-birthed sayings such as proverbs, will go a far way in safeguarding their knowledge for posterity.

**Keywords:** MULTIWORD EXPRESSIONS, PROVERBS, LEXICOGRAPHIC TREATMENT, KEY COMPONENT, HEADWORD, GENERAL DICTIONARY, SPECIAL PURPOSE DICTIONARY

**Opsomming: Lemmatisering van vaste uitdrukkings: Die geval van spreekwoorde in Noord-Sotho.** Die doel van hierdie artikel is om 'n kwantitatiewe en kwalitatiewe bepaling te maak van die leksikografiese hantering en lemmatisering van spreekwoorde in die *Wörterbuch der Sotho-Sprache* (Endemann 1911) vergeleke met geselekteerde Noord-Sotho woordeboeke. Ten einde spreekwoorde, wat vaste meervoudige woorduitdrukkings is, te huis-

---

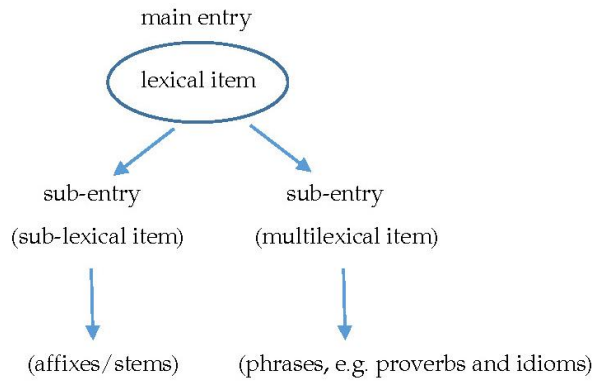
\* Paper presented at ALASA Conference (African Language Association of Southern Africa) at Walter Sisulu University in Umtata, South Africa, 30 June 2014.

ves, word hulle gewoonlik as sub-lemmas onder 'n bepaalde enkelvoudige hoofwoord geakkommodeer, gewoonlik een van die sleutelkomponente van die spreekwoord. Die keuse van 'n sleutelkomponent hang af van die subjektiewe oordeel van die leksikograaf. Hierdie selektiewe benadering kan tot gevolg hê dat spreekwoorde uit die oog verdwyn as geen van hul komponente prominent genoeg deur die samesteller geag word om die insluiting van die spreekwoord onder 'n bepaalde hoofwoord te regverdig nie. Dit blyk die geval te wees in die woordeboek wat hier onder bespreking is, gegewe die skaarsste van spreekwoorde wat daarin opgeneem is. Aan die ander kant kan die weglating van spreekwoorde ook eenvoudig aan 'n praktiese oorweging toegeskryf word, soos beperkte spasie in 'n gedrukte woordeboek. 'n Woordeboekgebruiker mag dit uitdagend vind om 'n verlangde spreekwoord op te soek, veral as die individuele woorde 'n baie lae frekwensie het of selfs in onbruik geraak het in die moderne lewe. In dié geval, sal 'n woordeboek in elektroniese formaat die gebruiker bemagtig deur die gebruik van 'n elektroniese opsoekmetode. Woordeboeke vir spesiale doeleindes wat hul toespits op kultuur-gebaseerde segswyses soos spreekwoorde, speel 'n groot rol om dié kennis vir die nageslag te bewaar.

**Sleutelwoorde:** MEERVOUDIGE WOORDUITDRUKKINGS, SPREEKWOORDE, LEKSIKOGRAFIESE BESKRYWING, SLEUTELKOMPONENT, HOOFWOORD, ALGEMENE WOORDEBOEK, SPESIALE DOEL WOORDEBOEK

## 1. Introduction

A dictionary, if it is to serve as an "authoritative container of knowledge" (Gouws and Prinsloo 2005: 1), needs to be a reflection of the culture of a particular linguistic community. Hence it "compels lexicographers to contextualise the language in terms of the more general world of the relevant speech community" (Gouws and Prinsloo 2005: 2 referring to Zgusta 1971). Proverbs are bearers of culture and hence they also deserve to be lemmatised. Lemmatisation refers to how and where a lexicographer chooses to list a word or phrase in a dictionary. Lexicographers are faced with a challenge when it comes to the lemmatisation of proverbs — a fact which is evident from the haphazard inclusion or exclusion of proverbs in different dictionaries for a language like Northern Sotho, for example. Proverbs are multiword expressions, but the design of the traditional dictionary is not conducive to their lemmatisation. On the contrary, single-word items are the elements which by default constitute the macrostructure of a dictionary. This is corroborated by Gouws and Prinsloo (2005: 86) who state that "dictionaries have often been characterised and dominated by a word-bias". Thus any other items smaller or larger than a word (sub-lexical or multilexical respectively), would not constitute independent look-up items, but would only be found as sub-entries under main entries as illustrated in Figure 1.



**Figure 1:** Lemmatisation of sub-lexical and multilexical items

Gouws and Prinsloo (2005: 86) argue that in a lexical-based approach to the macrostructure of a dictionary *all* types of lexical items should have their own lemmatic status, be they sub-lexical items such as affixes and stems or multilexical items such as proverbs and idioms. They convey meaning like any other lexical unit and therefore deserve comprehensive treatment in a dictionary, just like traditional lexical items. Mulhall (2010: 1355) adds that words and phrases share an equal status as units of meaning in the lexicon. Despite this, lexicographers usually make a distinction in how they record words as opposed to phrases: Words are usually listed as main-entries while phrases are listed as sub-entries of main-entries and do not enjoy their own lemmatic status.

In this article the lemmatisation of proverbs in a number of Northern Sotho dictionaries is compared and particular reference is made to the *Wörterbuch der Sotho-Sprache* (Dictionary of the Sotho languages) by Endemann, published in 1911, to highlight some of the complexities of lemmatising proverbs.

## 2. Identification of proverbs in Endemann's dictionary

Endemann's dictionary is a bilingual, unidirectional Sotho-German dictionary. The term 'Sotho' in the title encompasses all three of the Sotho languages, namely Tswana, Southern Sotho and Northern Sotho. The proverbs were easily identified by the label *Spruchwort* (German for 'proverb'), which appears in the internal article structure of an entry, invariably given either in full or as an abbreviation (*Sprichw.*), with or without parenthesis, before or after a proverb (see labelling in bold in examples in (1) to (3)). For the purpose of this discussion the current orthography is used instead of Endemann's. In the examples below, Endemann's entries are followed first by the researcher's own literal English translation and then by an interpretation extracted from another source which lists the relevant proverb.

(1) Endemann (1911: 242)

kgori, Pl. di-, "wilder Pfau". Kgori e bona mae, moutlo ga e o bone. "der wilde Pfau sieht die Eier, die Falle sieht er nicht" (**Spruchwort**).

Lit: 'The kori bustard sees the eggs, it doesn't see the trap.'

Interpretation: 'Have your mind on what you want to do without thinking of the consequences; if one has set one's heart on something one is inclined to see its bright side only' (Ziervogel and Mokgokong 1975: 681).

(2) Endemann (1911: 692)

boganolemare "Schmeichelei". Vergl. unser "Speichelleckerei". **Spruchwort**: Gano le mare ga di lemogwe "Schmeichelei wird nicht gemerkt".

Lit: 'The mouth cavity and saliva are not recognised (they don't understand each other).'

Interpretation: *Man kann es einem Lügner und Betrüger nicht ansehen, ob er die Wahrheit spricht oder lügt* (Kuhn 1929/1930: 48). 'One cannot perceive whether a liar and deceiver is speaking the truth or is lying' (own translation).

(3) Endemann (1911: 527)

tadi, Pl. di-, "gestreifte Feldmaus". Ngwana-tadi o tsebja ka mereto "das Kind vom *tadi* wird an den Streifen erkannt" (**Spruchw**).

Lit: 'The child of the field mouse is recognised by its stripes.'

Interpretation: 'like father like son; a chip of the old block; a tree is known by its fruit' (Ziervogel and Mokgokong 1975: 890).

In total 24 labelled proverbs were found in the 690 pages of the dictionary. The search was done by manually paging through the dictionary and checking each entry. This figure might be slightly higher, as at least one proverb was encountered which had not been marked with the label 'proverb'. Other unmarked proverbs might have escaped detection.

### 3. Internal rigidity of proverbs

Proverbs, like idioms, are multiword expressions, but they are more rigid in structure than idioms. They have their own fixed microtext, but small changes may nevertheless occur, compare (3) above with the same proverb in Ziervogel and Mokgokong (1975: 890) in (4) below:

(4) Ngwana wa tadi o tsebja ka mebala/merêtô

In (4) the full possessive form (*ngwana wa tadi*) is used instead of the compound

word *ngwana-tadi*. An alternative lexical item *mebala* (colours) is given in addition to *merêtô* (stripes).

Manyawu (2012: 214) describes both proverbs and idioms as "invariable metaphoric utterances that are almost as stable as standard lexical items". As part of the lexicon of the language, even these slightly unstable forms should be incorporated in a dictionary. According to Čermák (2014) proverbs can come in two forms, namely unmodified (reproduction) or based on a partial change (modification). The partial changes may be syntagmatic, paradigmatic or mixed. In his study of English proverbs, Čermák (2014: 25) found that the longer a stable and familiar proverb is, the greater the possibility of variability. It is their very familiarity which gives speakers or writers the creative freedom to quote only the first half of a proverb, if they so wish, because they know their listeners or readers will easily be able to infer the second half from the context, as in *lerumo le tee ...* 'one assegai...' which the addressee will be able to complete with *...ga le bolaye kgomo* '...will not kill a head of cattle', meaning that a person on his/her own cannot succeed in a task. An investigation into the types of variations of proverbs falls outside the scope of this paper, but could form the topic of another study, especially if the proverbs can be culled from a large corpus of written texts and oral passages, based on real usage.

As is well known, the meanings of proverbs can seldom be guessed from the meanings of the individual words which make up the proverbs. As stated by Akande and Mosobalaje (2014: 35) "Proverbs are, often, short value-laden expressions with multi-layered meanings that can be decoded only by those who possess a good mastery of the oral art and culture that produce them". They are figurative and didactic expressions and embrace the inherited wisdom and experience of a people. They are reflective of a nation's values, norms and morality and are used to bring across messages to guide, edify and admonish.

#### 4. The problem of proverb selection and headword selection

When a dictionary design includes the lemmatisation of fixed combinations of words such as proverbs, the lexicographer will need to ask him-/herself the following questions:

- Which type of source is most likely to contain proverbs for inclusion in the dictionary?
- How many and which proverbs should be included?, and thirdly
- How and where is the best place to enter the proverbs in the dictionary?

In a study conducted on the British National Corpus, Čermák (2014: 28) advanced the following regarding the texts most likely to contain proverbs:

It is now generally recognized that proverbs are not limited to a single text domain only, although there might be a tendency for them to occur more often in some. What one may only presume so far, is a different and *perhaps higher use and distribution in the spoken language*. [own emphasis].

The suggestion that proverbs are more prevalent in spoken than in written language makes sense, as proverbs have a communicative function and entail the impartation of wisdom, caution or admonition, normally by one person to another or to an audience in a verbal exchange. Should the lexicographer therefore base his data on an entirely written corpus, he is not likely to encounter many proverbs. His corpus should include examples from written as well as spoken language to increase his chances of coming across proverbs, more particularly well-established proverbs. The written texts should cover a wide spectrum of types, such as periodicals, newspapers, novels (fiction), books (non-fiction) and other miscellaneous works (Čermák 2014: 28). Čermák (2014: 48) concludes that the "nonexistence of proverbs, for example, in chemistry, mathematics or physics, may suggest that proverbs appear in the traditional fields linked to practical human life and its recurrent aspects and repeated patterns shared by the whole community over a period of time. Thus, it seems that proverbs in their coverage and use belong to social sciences (including linguistics), rather than to exact ones ...". In Northern Sotho written texts, proverbs are most likely to occur in works of a didactic nature. They also seem to have a heightened occurrence as titles of books, films or theatres, as opening and concluding lines. Sometimes they may also be used as the heading of a chapter or section, for example *bana ba tau ga re jane, re molokomong* 'children of the lion we do not eat each other, we are family' (*Tšhupa-Mabaka a Kereke* 1959: 36, a Church Bulletin) to aptly convey the gist of a message.

Regarding the second question on the *number and type* of proverbs, the lexicographer has to keep the needs of the target user in mind and make a selection of the most suitable proverbs from the overwhelming number of proverbs that exist in Northern Sotho. It goes without saying that a comprehensive listing of all existing proverbs is not possible within the constraints of a paper-based dictionary and that any selection will of necessity be incomplete and subjective, because criteria such as popularity and frequency of use are relative. No frequency studies have been carried out on proverbs for Northern Sotho to determine the most common ones expressing the most common meanings. One gets the impression that the inclusion of proverbs is not necessarily a deliberate exercise in a general purpose dictionary, but that proverbs are included randomly as example sentences where they can serve as a suitable context in which to illustrate the use of a lemma. The *Oxford Bilingual School Dictionary. Northern Sotho and English* (De Schryver 2007) is an example of a dictionary in which the proverbs which occur as example sentences were not purposefully chosen by the lexicographer, but emanated from the corpus on which the dictionary was based. Users may not be aware that they are being presented with proverbs as example sentences in an article treatment in de Schryver (2007), since the fixed expressions are not marked by a specific label. Had these proverbs been marked by a special label, this would have served a useful additional educational purpose. For example, an unmarked proverb *Se bone thola boreledi, teng ga yona go a baba* is used as an example sentence in de Schryver (2007: 28) under the head word *boreledi* (smoothness) with only its literal

English translation 'don't look at the smoothness of the bitter apple, the inside is bitter'. The figurative meaning (i.e. not everything is as it seems) is not provided. A comparable example in English would be: 'not all that glitters is gold'.

The third question is probably the most challenging for a lexicographer, namely where to *enter* the proverbs in a dictionary. It requires the lexicographer to identify the most suitable allocation lemma from the elements which constitute the multiword expression.

In the conventional word-based approach, according to Mulhall (2007) "the decision to list an idiom, or any phrase, as a sub-entry necessitates the lexicographer to choose an element of the phrase, which they [sic] believe to be the most suitable point of entry as well as being the most identifiable to the dictionary user listing". In this process one lexicographer might choose one lexical item as key component, while another may find another element more suitable as key component of the same proverb. This is responsible for discrepancies in the lemmatisation of proverbs across dictionaries. In Endemann (1911: 527), for example, the proverb in example (3) is listed under *tadi* (field mouse), while it features under the lemma *ngwana* (child) in Ziervogel and Mokgokong (1975: 874) and under both *ngwana* and *mereto* (stripes) in Kriel (1965: 149 and 114 respectively).

In some cases the search for a proverb under any of the assumed key words is unsuccessful. In Endemann, for example, one would have expected with a greater than chance probability that the proverb *ngwana yo a sa llego, o hwela tharing* (the baby that doesn't cry, dies in the carry sling, i.e. if you don't make your needs known, you will not receive assistance) would have featured under the most easily-guessed-at lemma *thari*. Instead, only an explanation of *thari* is given (sling made of skin to carry a baby on its mother's back), while the proverb does not feature in the dictionary at all. The absence of this proverb under the headword *thari* is also observed in other Northern Sotho dictionaries that were consulted (Kriel 1965, 1976a, 1976b, 1977; 1983; Kriel, Van Wyk and Makopo 1989; Prinsloo and Sathekge 1996; Kriel, Prinsloo and Sathekge 1997; Mojela, Mphahlele, Mogodi and Selokela 2006 and De Schryver 2007). One would have thought that a lexical item such as *thari* would have triggered the entry of the proverb in these dictionaries because of the collocational sense relation between this word and the proverb, but this was evidently not the case. In all fairness, the omission of the proverb might have been dictated by the nature of the dictionaries, some of which are concise dictionaries, designed for learners at an elementary level where it is assumed that the priority is finding the meaning of single word items, rather than that of fixed expressions. Only Ziervogel and Mokgokong (1975) list this proverb, and they do so under the lemmas *ngwana* as well as *thari*. As discussed in the next section, it is thus possible for the same proverb to be listed more than once in the same dictionary, that is, under different lemmas, to make it more accessible to users who will be guessing at its location in the dictionary. This reveals that the choice regarding the inclusion of proverbs is a very subjective issue, but partly also dictated by the dictionary plan and purpose.

#### 4.1 Entry points

Given their multilexical composition, there are a number of possible entry points for proverbs, which complicates the look-up process, because users may not be able to predict these entry points.

If a proverb has multiple listings, in other words, if it is listed under more than one of the lexical items contained in the proverb, it inevitably takes up extra space in a traditional paper dictionary. A further disadvantage is that, as sub-entries, proverbs only enjoy a reduced visibility due to the volume of surrounding information. In this case a special symbol or label to mark the proverb would be useful to make the multiword expression stand out among the other information. On the positive side, multiple listings will create more chances for a user to find a proverb while he/she conducts his/her search by trial and error. Cross-references could be used to guide the user to the sought-after proverb. Multiple listing is a compromise on the part of the lexicographer who would otherwise have to rely on his intuition and subjective judgement as to which of the elements deserves to be the most suitable key word under which a proverb should be listed.

The following example illustrates that the same proverb can have one listing at one entry point or multiple listings under various lemmas. In Ziervogel and Mokgokong (1975) the proverb *moeng o naka di maripa* is listed under the lemmas -ENG, -NAKA and RIPA. In Kriel (1977 and 1983) the same proverb is listed at two entry points, namely MOENG and MARIPA. In Endemann, however, only one listing was found, namely under -RIPA:

(5) **Proverb:** *moeng o naka di maripa*

Ziervogel and Mokgokong (1975)	Kriel (1977, 1983)	Endemann (1911)
-ENG, mo-/ba- a visitor should not interfere in matters concerning his host (p. 187)	MOENG 'n gas mag nie sy gasheer kritiseer nie, hy moet stilbly (1977: 147, 1983: 191)	-
-NAKA, le-/ma- a newcomer/a visitor does not meddle in other people's affairs (p. 837)	-	-
RIPA a visitor should always be well-behaved (p. 1118)	MARIPA 'n gas mag nie uitgesproke wees nie (1977: 132, 1983: 172)	-RIPA der Gast ist einer, dem die Hörner verstutzt sind, d.h. ein Gast darf nicht stossen, nicht grob sein (p. 436)



With the problems that have been outlined, it becomes clear that due to their structural nature, proverbs cannot satisfactorily be integrated in a dictionary for general purposes. Their inclusion and placement are not predictable, as theoretically, one lexicographer may decide to enter a proverb under one lemma, while another may decide to include the same proverb under another lemma, which seems more appropriate or salient to him or her. Mpofu (2007: 364) calls such prominent lexical components, 'semantically heavy' words. In the African languages there are many proverbs in which it would be difficult to identify a 'semantically heavy' word, for example, the subject may be unexpressed or the constituents may be of equal lexical and semantic importance. Here it would be difficult to decide under which lemma the proverb should be listed, e.g.

(6) *A di fule, di hlakane* (Erasmus n.d.: 7)

'Let them (domestic animals) graze and mix' (said of herd boys who should reconcile and become friends again after a quarrel — their livestock should graze together again).

The basic form of any of the verb stems in the above example (*-fula* 'graze' and *-hlakana* 'come together, mix') could technically serve as a point of entry. In this regard one could ask whether some proverbs fall between the cracks for the very reason that none of their items are considered prominent enough to trigger their inclusion under a particular lemma.

There is a tendency by users to look up proverbs under the first major component or first key word in an expression. Korhonen (2011) says that for German, the first noun, verb or adjective of a proverb usually determines under which lemma a proverb should be listed. This is perhaps more easily said of European languages than of African languages. In African languages such salient components may only appear towards the end of a proverb and one component may not necessarily be more salient than the other. It is thus difficult to decide which component of a proverb should be chosen as its allocation lemma. Example (6) above illustrates the challenge of deciding on an appropriate key component, as the example does not even contain an overt subject noun, while both verbs carry equal semantic weight. Looking up a proverb under the first key word (e.g. *moeng* in example (5)), would have been rewarded immediately in Ziervogel and Mokgokong (1975) and Kriel (1977, 1983), but not so in the case of Endemann's (1911) entry. Between the three dictionaries in example (5), it would have taken the longest to look it up in Endemann's dictionary, as it only uses the last key word (*-ripa*) as its entry point.

In special purpose dictionaries (see next section), on the other hand, in which proverbs are arranged in alphabetic order according to the first part of the phrase, be it a sub-lexical or full lexical item, no thought needs to be spared for salient lexical items.

## 4.2 Lemmatisation principle in special purpose dictionaries: First-element or index-based

### 4.2.1 Lemmatisation according to the first element

In dictionaries for special purposes such as for multiword expressions, the lemmatisation of proverbs can be done alphabetically according to the first element that they begin with. In the latter case, no decision needs to be taken on which element(s) constitute(s) salient lexical items for lemmatisation purposes, as the proverbs are merely listed according to the first part of the phrase, be it a unit smaller than a word (a morpheme) or a full lexical item. The listing of proverbs in such a publication is deliberate or purposive unlike in dictionaries for general purposes, where proverbs are often included at the lexicographer's discretion.

Rakoma's (1949) impressive collection of idioms and proverbs (*Marema-ka-dika tša Sesotho sa Transvaal*), was inspired by the concern that the older generation would take the knowledge of many fixed expressions to their grave, if not documented. Apart from idioms, the collection contains 1041 proverbs with explanations in Northern Sotho. The entries are arranged alphabetically according to the first element of the proverbs (introducers), but follow no specific alphabetical order in the microstructure under each alphabetical entry. The introducers are of various kinds, from lexical to non-lexical elements which one would not necessarily encounter as lemmatised items in general purpose dictionaries, such as negative morphemes, subject concords, hortative particles, etc.

In his collection of proverbs entitled *Uitgesoekte Noord-Sotho spreekwoorde* (Selected Northern Sotho proverbs) Erasmus (n.d.) also followed an alphabetic listing according to the introducers of the proverbs. Lemmatisation according to initial components is conducive to the user finding the required proverb quickly, even if the same element appears numerous times as the first component under a particular article stretch, e.g. G, as illustrated in the examples in (7) beginning with "Ga go" ('there is not') (Afrikaans explanations by Erasmus rendered in English by the researcher).

(7a) **Ga go** moedi mo-tlhoka-semenya (Erasmus n.d.: 25)

'there is not a valley which lacks a hollow' — nobody is perfect, every household carries a burden

(7b) **Ga go** 'šaka la poo-pedi (Erasmus n.d.: 26)

'there is not a kraal of two bulls' — you cannot serve two masters

(7c) **Ga go** sekiswe khutswane, rakhudu a le gona (Erasmus n.d.: 26)

'the little tortoise is not charged as long as father tortoise is there' — parents are responsible for their children's misdeeds, because they educate them

Finding these proverbs in a general purpose dictionary will take more effort, because the first element (which is often a sublexical item such as a prefix) may not occur in the macrostructure, but rather in the microstructure of another lemma in a word-based dictionary. Even if **ga** should occur as a headword, the users would have to navigate through a long list of sub-entries for **ga** before finding the sought-after proverb. When there are a number of proverbs which start in the same way, the microstructure of the entry becomes congested, particularly if, in addition, the proverbs are quite long. This would complicate or delay successful information retrieval. In a dictionary dedicated specially to proverbs, items can be arranged according to the letters of the alphabet as Rakoma and Erasmus have done and the length of proverbs would not constitute a problem.

A special label to mark an expression as a proverb would facilitate the identification of a proverb under a particular lemma. This practice has not been observed in any of the Northern Sotho dictionaries consulted for this research, except in the dictionary by Endemann. On the other hand, if proverbs are listed alphabetically in a special purpose dictionary, users will find it to be user-friendly, because they would approach such a dictionary with a different expectation, compared to general purpose dictionaries where the proverbs would be concealed under various entries.

#### 4.2.2 Lemmatisation which is index-based

Another collection of proverbs in which a different approach was used compared to Erasmus' (n.d.) approach referred to above, appeared in an article published by Kuhn in the *Zeitschrift für Eingeborenen Sprachen* in 1929/1930. Kuhn listed 702 proverbs, which he remembered from his childhood days, growing up amongst the Bapedi as a missionary child. The proverbs are listed numerically in random alphabetic order, but a list with German key terms/concepts in alphabetic order appears in the appendix, with numerical references to the proverbs in which the key terms occur. For 'mouth/mouth cavity' (*Mund/Mundhöhle*), for example, there are references to 27 proverbs. Of these 27 proverbs, 11 are indexed by a second key term, 6 by a third key term and 7 by a fourth key term. Thus, this is a very useful index, as the same proverb can be located using more than one key term, increasing the chances of the user to find the sought-after proverb.

In 1938 a list of 124 proverbs in Tlokwa, a dialect of Northern Sotho, was published by Krüger, but not in any particular alphabetic or thematic order. Nevertheless, the list constitutes an important part of cultural heritage.

More recently, also driven by the passion to preserve proverbs for future generations, Motana (2004) produced an illustrated booklet of 102 Sepedi (Northern Sotho) proverbs, listed alphabetically under the initial letter of key-words that appear in his English translations; compare the following examples

under the letter B (cf. italicised *baboon*) and V (cf. italicised *vulture*) respectively in (8a) and (8b):

(8a) *Go diega ga tšhwene ke go gadimela morago*

**To look backward delays a *baboon*.**

People who keep on delaying, or do not focus on what they are supposed to do, often fail to accomplish a task. (Motana 2004: 11)

(8b) *Nong ye kgolo ga e rutwe go fofa.*

**An old *vulture* is not taught how to fly.**

You cannot teach an expert the basics. (Motana 2004: 49)

## 5. Dearth of proverbs in general purpose dictionaries

The *Groot Noord-Sotho Woordeboek* by Ziervogel and Mokgokong (1975) by far provides the largest number of proverbs compared to the other dictionaries that were consulted for this study (Endemann 1911, Kriel 1965, 1976a, 1976b, 1977, 1983; Kriel, Van Wyk and Makopo 1989; Prinsloo and Sathekge 1996; Kriel, Prinsloo and Sathekge 1997; Mojela, Mphahlele, Mogodi and Selokela, 2006 and De Schryver 2007). To date Ziervogel and Mokgokong's (1975) publication is still the most comprehensive dictionary that exists for Northern Sotho, but there are nevertheless some proverbs encountered in Endemann which couldn't be found in Ziervogel and Mokgokong, underlining the need for dedicated dictionaries on proverbs. One could surmise that the limited coverage of proverbs in most of these dictionaries can be ascribed to the fact that they are general purpose dictionaries whose priority is to explain the meaning of single-word tokens at word-level, and not to offer information at phrase-level. In dedicated volumes of proverb collections, on the other hand, proverbs do not have to compete for space under lemmas which are judged to be key words in the proverb and the listing is much more comprehensive.

The sporadic occurrences of proverbs in Endemann (1911), speak of an incidental inclusion of proverbs. The proverbs serve as illustrative sentences to demonstrate the application of a particular headword in context. They seem to have been included in a 'by-the-way' fashion, probably as and where a lexical item had a collocational sense relation with a proverb (in other words where there was a mutual expectancy that the word and the fixed expression would co-occur).

## 6. Microstructure of proverbs in Endemann (1911)

Looking at the microstructure of proverbs in Endemann's dictionary, one notices that not all proverbs have been treated equally. In some cases only a literal translation is provided, with no explanation, in others both a literal

translation and an explanation are provided, while in others a literal translation and an equivalent proverb in the target language are provided. Since there are gaps in the provision of meanings or interpretations of the proverbs, it would appear that the lexicographer took a decision on behalf of the user as to which proverbs are transparent and self-explanatory and which ones needed further elucidation.

The different combinations of the treatment of proverbs in Endemann (1911) as opposed to the treatment of the same proverbs in Ziervogel and Mokgokong (1975) are exemplified in examples (9) to (11). The captions of the examples in (9) to (11) reflect the information as encountered in Endemann. (Note that for the purpose of this discussion complete entries are not given, but only the headword and information pertaining to the proverb under that particular headword).

**(9) Proverb, literal translation, no equivalent, no explanation**

Endemann (1911: 72)	Ziervogel and Mokgokong (1975: 215)
Lemma: mo-fatamollo	Lemma: FATA
Proverb: <i>Mofatamollo wa marega o a iphatela</i>	Proverb: <i>sefatamollo se a iphatela</i>
Lit. "der im Winter das Feuer scharrt, der scharrt für sich selbst"	Lit. <i>None</i>
Equivalent: <i>None</i>	Equivalent: reap what one sows
Explanation: <i>None</i>	Explanation: <i>None</i>

**(10) Proverb, literal translation, no equivalent proverb, an explanation**

Endemann (1911: 279)	Ziervogel and Mokgokong (1975: 768)
Lemma: le-loka	Lemma: -LOKA
Proverb: <i>Tlou e hlabja ka lloka, tšhukudu mošimane</i>	Proverb: <i>Tlou e hlabja ka diloka, tšhukudu e be mošimane</i>
Lit. "der Elefant wird vom lloka gestochen, das Nashorn ist ein Junge"	Lit. <i>None</i>
Equivalent: <i>None</i>	Equivalent: a labourer is worthy of his hire
Explanation: Auch ein Grosser kann durch ein kleines Ding überwunden werden.	Explanation: if you work hard you attain your goal.

It is interesting to note that the imperative form of the verb is employed in the above proverb in Kuhn (1929/1930: 124) and Rakoma (1949: 135) (*Hlabang tlou ka dilokwa, tšhukudu mošimane*). The respective explanations given by these authors for this proverb amount to the following: 'through small things one can

achieve something great' and 'through patient perseverance one will reach one's goal'.

(11) **Proverb, literal translation, equivalent proverb, no explanation**

Endemann (1911: 259)	Ziervogel and Mokgokong (1975: 752)
Lemma: le-leka	Lemma: -LEKA
Proverb: <i>Maleka ga se makgone, makgone ke maboeletša</i>	Proverb: <i>Maleka ga se makgone, makgone ke maboeletša</i>
Lit. "Neulinge sind keine Künstler, Künstler sind die (etwas) immer wiederholen"	Lit. <i>None</i>
Equivalent: Übung macht den Meister	Equivalent: Rome was not built in a day
Explanation: <i>None</i>	Explanation: perseverance will be rewarded

7. **Conclusion**

In a general purpose dictionary certain proverbs may fall into oblivion, if their components are not considered prominent enough to trigger their inclusion under a particular headword. Special purpose dictionaries dedicated to the documentation of proverbs, on the other hand, will be able to document such proverbs as custodians of the culture of a people, be they of a bygone era or of modern origin. Proverbs are an integral part of a society, although life has changed to such a degree that many proverbs are no longer used. They can give insight into the way of life of earlier periods in a nation's history or cultural development and play a "social role as repositories of the ancient wisdom of a given speech community" (Manyawu 2012: 214). In the words of Čermák (2014: 47) "there seems to be a consensus that there is no substitute for them in social communication, even in our modern times and society".

Traditionally proverbs occur as sub-lemmas in the microstructure of a lemma instead of the macrostructure of a dictionary. They convey meaning like any other lexical unit and thus deserve comprehensive treatment just like other single lexical items. Otlogetswe (2012: 232) bewails the situation whereby only single words are considered as candidates for dictionary entry because it "impoverishes a dictionary and betrays a rudimentary understanding of what constitutes a word in a language". Mphahlele (2003: 163) argues that "[m]ulti-lingual items are lexical items that consist of more than one word. This combination of words is always a unit and should be treated likewise in a dictionary. Although multiword lexical items consist of more than one word, they should, according to Gouws (1991: 78), be regarded as single lexical items. These items should therefore be included as multilexical lemmata in the macrostructure of dictionaries."

Mpofu (2007: 364) states that "As a result of the changing trends in lexi-

cography, dictionaries are now lemmatising units larger than the word in a bid to meet the needs of different users". A multi-word lexical unit is easier accessed from the macrostructure "than to search for it within another entry in the microstructure" (Mpofu 2007: 361). Even easier would be the location of proverbs in an electronic dictionary (a discussion which falls outside the aim of this article), which allows access at various points and does not have the restrictions of a paper dictionary. It would also counteract the "old, unacceptable trend trying to squeeze the complex meaning of the proverb under a single word label" (Čermák 2014: 145).

A paper-based dictionary for general purposes cannot adequately cater for all the proverbs one could possibly encounter in a language. The choice as to which proverbs should be included and under which main lexical items they should be lemmatised, remains the lexicographer's prerogative and is also dictated by the target user's needs. The inclusion of fixed expressions in Endemann (1911) is clearly not a user-driven one, as entries of proverbs are ad hoc and incidental to the lemmatisation process. Only a handful of proverbs were encountered in his otherwise quite comprehensive dictionary which also lemmatises many sub-lexical items such as subject concords, possessive concords, class prefixes, etc.

In a word-based dictionary, multiple lemmatisation of proverbs under different headwords, increases the user's chances of finding the proverb and it avoids the lexicographer having to subjectively judge which element is more important compared to another.

Mphahlele (2003:166) proposed that lexicographers have a choice of treating fixed expressions such as idioms and proverbs as multilexical lemmas in the macrostructures of general or special dictionaries. General dictionaries receive more attention as "tools for achieving language standardisation, documentation and preservation", but not as repositories for "specialised knowledge domains" (Chabata 2013: 55). Specialised dictionaries, on the other hand, can facilitate access to proverbs, arranged according to topics or themes (e.g. health, God, ancestors, social relations, etc.), thus narrowing down the search field of the user and making information retrieval much more effective.

From a linguistic perspective culturally-birthered sayings and their phraseology are of particular interest to language historians. Special purpose dictionaries, dedicated to the documentation of proverbs, would go a far way in safeguarding their knowledge for posterity. E-dictionaries would particularly ensure quick and easy access to a variety of proverbs and their (minimal) individual variations. An advantage is that the dictionary searching skills of users are not taken for granted. Space restrictions typical of a paper dictionary are not a limiting factor and new proverbs can be added any time. To avoid the situation whereby proverbs are permanently lost to posterity, dedicated dictionaries for proverbs should continually be developed. If a dictionary of the calibre of Rakoma (1949) could be produced as a bilingual resource, for example, it would reach a wider readership, more so, if made available electronically.

## References

- Akande, A.T. and A. Mosobalaje.** 2014. The Use of Proverbs in Hip-hop Music: The Example of Yoruba Proverbs in 9ice's Lyrics. *Proverbium* 31: 35-58.
- Čermák, F.** 2014. *Proverbs: Their Lexical and Semantic Features*. Burlington, Vermont: University of Vermont.
- Chabata, E.** 2013. The Language Factor in the Development of Africa: A Case for the Compilation of Specialised Dictionaries in Indigenous African Languages. *South African Journal of African Languages* 33(1): 51-58.
- De Schryver, G-M. (Editor-in-chief).** 2007. *Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English*. Cape Town: Oxford University Press South Africa.
- Endemann, K.** 1911. *Wörterbuch der Sotho-Sprache*. Volume VII of Abhandlungen des Hamburgischen Kolonialinstituts. Hamburg: L. Friedrichsen & Co.
- Erasmus, J.G.** n.d. *Uitgesoekte Noord-Sotho spreekwoorde*. Johannesburg, Cape Town: APB.
- Gouws, R.H.** 1991. Toward a Lexicon-based Lexicography. *Dictionaries* 13: 75-90.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PRESS.
- Korhonen, J.** 2011. *Phraseologie und Lexikografie: Phraseologismen in ein- und zweisprachigen Wörterbüchern mit Deutsch*. Burlington, Vermont: University of Vermont.
- Kriel, T.J.** 1965. *Pukantšu Noordsottho-Afrikaans, Afrikaans-Noordsottho*. Pretoria: Dibukeng.
- Kriel, T.J.** 1976a. *Popular Northern Sotho Dictionary (N.Sottho-English/English-N.Sottho)*. Second edition. Pretoria: J.L. van Schaik.
- Kriel, T.J.** 1976b. *The New English-Northern Sotho Dictionary (English-Northern Sotho/Northern Sotho-English)*. Fourth revised edition. King William's Town: Educum.
- Kriel, T.J.** 1977. *Pukuntšu Woordeboek. Noord Sotho-Afrikaans, Afrikaans-Noord Sotho*. Second revised and extended edition. Pretoria: J.L van Schaik.
- Kriel, T.J.** 1983. *Pukuntšu Woordeboek. Noord Sotho-Afrikaans, Afrikaans-Noord Sotho*. Third edition. Pretoria: J.L van Schaik.
- Kriel, T.J., D.J. Prinsloo and B.P. Sathekege.** 1997. *Popular Northern Sotho Dictionary, Northern Sotho-English, English-Northern Sotho*. Fourth edition. Cape Town: Pharos.
- Kriel, T.J., E.B. van Wyk and S.A. Makopo.** 1989. *Pukuntšu Woordeboek (Noord Sotho-Afrikaans, Afrikaans-Noord Sotho)*. Fourth revised and extended edition. Pretoria: J.L van Schaik.
- Krüger, F.** 1938. Tlokoa Überlieferungen. *Afrikanische Studien* 41(3): 16-35.
- Kuhn, G.** 1929/1930. Sotho-Sprichwörter. *Zeitschrift für Eingeborenen-Sprachen* 20: 34-73, 120-148, 208-221.
- Manyawu, A.T.** 2012. An Intertextual and Interdiscursive Analysis of Contemporary Thinking about Womanhood in Mokitimi's (1997) *The Voice of the People: Proverbs of the Basotho*. *South African Journal of African Languages* 32(2): 213-222.
- Mojela, M.V., M.C. Mphahlele, M.P. Mogodi and M.R. Selokela.** 2006. *Sesotho sa Leboa-English Pukuntšu ya dipolelo tše pedi le dihlošo/English-Sesotho Bilingual and Explanatory Dictionary*. Cape Town: Phumelela Education.
- Motana, N.** 2004. *Sepedi (Northern Sotho) Proverbs*. Cape Town: Kwela Books.
- Mphahlele, M.C.** 2003. The Lexicographic Treatment of Sublexical and Multilexical Items in a Northern Sotho Monolingual Dictionary: A Challenge for Lexicographers. *Lexikos* 13: 154-167.



- Mpofu, N.** 2007. The User Perspective in Lexicography: The Lemmatisation of Fixed Expressions in *Duramazwi Guru reChiShona*. *Lexikos* 17: 360-371.
- Mulhall, C.** 2007. *Semantics and the Lemmatisation of Idiomatic Expressions: An Analysis of Bilingual Italian–English Dictionaries*. Available at: [https://www.essex.ac.uk/langling/documents/langue/langue\\_2007\\_proceedings.pdf](https://www.essex.ac.uk/langling/documents/langue/langue_2007_proceedings.pdf). [Accessed 29 January 2016.]
- Mulhall, C.** 2010. A Semantic and Lexical-based Approach to the Lemmatisation of Idioms in Bilingual Italian–English Dictionaries. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*: 1355-1371. Ljouwert: Afûk.
- Otlogetswe, T.** 2012. Enriching a Dictionary Database with Multi Word Expressions. Ndinga-Koumba-Binza, H.S. and S.E. Bosch (Eds.). 2012. *Language Science and Language Technology in Africa: A Festschrift for Justus C. Roux*: 215-236. Stellenbosch: SUN PRESS.
- Prinsloo, D.J. and B.P. Sathekge.** 1996. *New Sepedi Dictionary (English–Sepedi (Northern Sotho)/Sepedi (Northern Sotho)–English)*. Pietermaritzburg: Shuter & Shooter.
- Rakoma, J.R.D.** 1949. *Marema-ka-dika tša Sesotho sa Transvaal*. Pretoria: Van Schaik.
- Tšhupa-Mabaka a Kereke.* 1959. Eerste Fabrieke, Pretoria: Northern Sotho Book Depot.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Pukuntšū ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*. Pretoria: J.L. van Schaik.