The Lemmatization of Copulatives in Northern Sotho[*]

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (prinsloo@postino.up.ac.za)*

Abstract: For learners of Northern Sotho as a second or even foreign language, the copulative system is probably the most complicated grammatical system to master. The encoding needs of such learners, i.e. to find enough information in dictionaries in order to actively use copulatives in speech and writing, are poorly served in currently available dictionaries. The aim of this article is to offer solutions to the lemmatization problems regarding copulatives in Northern Sotho and to propose guiding entries for paper and electronic dictionaries which could serve as models for future dictionaries. It will be illustrated that the maximum utilisation of macrostructural and microstructural strategies as well as the mediostructure is called for in order to reach this objective. Prerequisites will be to reconstruct the entire copulative system in a user-friendly way, to abstract the rules governing the use of copulatives and to isolate the appropriate lemmas. The treatment of copulatives in Northern Sotho dictionaries will also be critically evaluated, especially in terms of frequency of use and target users' needs.

Keywords: LEXICOGRAPHY, LEMMATIZATION, COPULATIVES, INFORMATION RETRIEVAL, ACCESS STRUCTURE, ELECTRONIC DICTIONARY, MACROSTRUCTURE, MICROSTRUCTURE, CROSS-REFERENCING, MEDIOSTRUCTURE, DICTIONARY, AFRICAN LANGUAGES

Opsomming: Die lemmatisering van kopulatiewe in Noord-Sotho. Vir aanleerders van Noord-Sotho as tweede of vreemde taal is die kopulatief waarskynlik die mees komplekse grammatiese sisteem om te bemeester. Die enkoderende behoeftes van sulke aanleerders, dit is om genoegsame inligting in woordeboeke te verkry ten einde kopulatiewe in spraak en skrif aktief te kan gebruik, word nie bevredig in beskikbare woordeboeke nie. Die doel van hierdie artikel is om oplossings aan die hand te doen vir die lemmatiseringsprobleme ten opsigte van kopulatiewe in Noord-Sotho en om gidsinskrywings voor te hou wat as modelle kan dien vir toekomstige papier en elektroniese woordeboeke. Daar sal aangetoon word dat maksimale benutting van makrostrukturele en mikrostrukturele strategieë asook die mediostruktuur 'n vereiste is vir bereiking van hierdie doelwit. 'n Voorvereiste is dat die struktuur van die totale kopulatiewe sisteem op gebruikersvriendelike wyse voorgestel moet word. Die reëls wat die gebruik van die kopulatief bepaal kan dan geabstraheer word. Daarna moet die toepaslike lemmas geïdentifiseer word. Die bewerking van kopulatiewe in Noord-Sothowoordeboeke sal ook krities geëvalueer

---

word, veral in terme van gebruiksfrekwensie en gebruikersbehoeftes.

Sleutelwoorde: LEKSIKOGRAFIE, LEMMATISERING, KOPULATIEWE, INLIGTINGS-ONTSLUITING, TOEGANGSTRUKTUUR, ELEKTRONIESE WOORDEBOEK, MAKROSTRUK-TUUR, MIKROSTRUKTUUR, KRUISVERWYSING, MEDIOSTRUKTUUR, WOORDEBOEK, AFRIKATALE

Introduction

From a learner's point of view, the copulative in Northern Sotho is probably the most complicated grammatical system to master. Northern Sotho grammars and dictionaries often fail in many respects to treat copulatives in a satisfactory way. Grammar books generally do not give learners a *comprehensive overview* of the copulative *in a logical way*.[1] Copulatives are either described in bits and pieces or treated in such an elaborate and *ad hoc* manner that the learner fails to get an overall picture that is essential for comprehension of such a complicated system. Currently available dictionaries for Northern Sotho also do not treat copulatives in a systematic way. Copulatives in these dictionaries are lemmatized and treated at random and in an *ad hoc* manner as they come under the compiler's attention.

For the lexicographer it is imperative to depart from giving an overall account or the full scope of copulatives when planning the macrostructure and microstructure in respect of their treatment. An attempt will firstly be made to construct such a comprehensive account in a user-friendly way by means of an *example-driven* approach which as such is suitable to form the back matter entry for copulatives. From this detailed (re)construction it will be attempted to abstract the rules underlying or governing the use of copulatives in Northern Sotho. Once these rules have been deducted/defined, candidates for lemmatization will be isolated and the compilation of model entries for both paper and electronic dictionaries will be attempted. It will be clearly illustrated that maximum utilization of access structures in terms of the macrostructure, microstructure and mediostructure is essential for the successful treatment of copulatives. The construction of back matter entries in particular, and the required mediostructural strategies required to successfully link central text entries with the back matter is crucial.

A brief theoretical perspective

Theoretical aspects pertaining to the lemmatization of copulatives in Northern Sotho in respect of macrostructure, microstructure, mediostructure and access structure will firstly be outlined briefly. Definitions of the two key concepts that constitute the title of this article, namely *copulatives* and *lemmatization* will be given first.

Louwrens (1994: 40) defines the term *copulative* as follows:

copulative (leba, kopulatief)
>   A term used to refer to structures which incorporate a variety of *prefixes* (called
>   particles in some grammatical descriptions of Northern Sotho) of which some
>   resemble the *subject concord* (e.g. *o* **bohlale,** *le* **bohlale,** *e* **bohlale**, etc. 'he/it is
>   clever'), whereas others have a unique form like **ke** (e.g. *ke* **monna** 'it is a man')
>   and **ga se** (e.g. *ga se* **monna** 'it is not a man'), as well as a series of verb forms
>   with the *copulative verb stems* **-bê** (e.g. **...** *gore a bê* **bohlale** 'so that he may become
>   clever'), **-ba** (e.g. **...** *a ba* **bohlale** 'and then he became clever'), **-le** (e.g.
>   **...** gê *a le* **bohlale** 'if he is clever'), **-se** (e.g. **...** gê *a se* **bohlale** 'if he is not clever')
>   and **-na** (e.g. *o na* **le lehufa** 'she is jealous'). The copulative expresses the English
>   verb *to be*, and is therefore translated with 'is', 'was', 'will be', 'has been', etc.

Taljard (1999: 3) justifiably remarks that linguists disagree on terminology in respect of copulatives:[2]

>   Wat die terminologie rakende die kopulatief en sy samestellende dele betref, is
>   dit by nadere ondersoek duidelik dat die linguis hom hier in 'n taalkundige
>   mynveld bevind. Aan die een kant word verskillende terme deur grammatici
>   gebruik om na dieselfde begrippe te verwys. Aan die ander kant word dieselfde
>   terme gebruik, maar verskillend geïnterpreteer deur verskillende grammatici.

Louwrens (1991: 63) states that clear terminological distinctions should be drawn between the terms 'copulative', 'subject', 'copula' and 'complement'. He cites the example: *Madika ke morutiši* 'Madika is a teacher' where 'copulative' is the umbrella term for the entire construction, *Madika* = subject, *ke* = copula and *morutiši* = complement. In Poulos and Louwrens (1994: 290) they, however, decide not to draw a distinction between the terms 'copula' and 'copulative'. From a lexicographic point of view, the value of such a distinction is question-able anyhow and therefore 'copulative' will be used throughout in this article.

Hartmann and James (1998) defines *lemmatization* as 'the reduction of a paradigm of variant word forms to a canonical form e.g. the inflected forms (-s, -ed, -ing etc.) of English verbs to the infinitive'. They define *canonical form* as 'the base form under which several variants of a word or phrase can be cited as a headword for a dictionary entry'. Thus, in terms of this definition *walk*, *walks*, *walked*, *walking*, etc. can be lemmatized as *walk*. For more advanced users of a Northern Sotho dictionary the words *moreki* 'buyer', *morekiši* 'seller', *rekela* 'buy for', *theko* 'price', etc. can all be lemmatized as *reka*.

Their use of the term *headword* focuses the attention on a 'word' being the key element as lemma in the macrostructure of a dictionary. Gouws (1989: 84), however, rightly emphasizes that such a traditional focus on the word as repre-sentative of the lexicon should be shifted to lemmas representing the lexical items of the particular language.

>   Die leksikon [bevat] veel meer as net woorde … Leksikale items … sluit ook ele-
>   mente wat groter as woorde en elemente wat kleiner as woorde is, in. Die

> struktuur van die leksikon moet dus gesien word as gebaseer op leksikale items. Hierdie leksikaalgebaseerde struktuur van die leksikon staan in opposisie tot die tradisionele woordgebaseerde struktuur van woordeboeke. … Die leksikograaf [moet] in sy keuse van makrostruktuurelemente die fokus verskuif vanaf trefwoorde wat die woorde van die taal verteenwoordig, na lemmas wat die leksikale items van daardie taal verteenwoordig. … Die lemmas van 'n woordeboek [moet] die hele reeks leksikale items verteenwoordig … woorde asook elemente kleiner en groter as woorde [moet] lemmastatus kry.

For the lemmatization of copulatives in Northern Sotho the final remark, namely that lexical items often consist of elements bigger or smaller than orthographic words, is of particular importance. It will be argued that the copulatives *ga se, e le, e se, e ba*, etc. are all excellent examples of multiword lemmas which have not been entered as such in Northern Sotho dictionaries.[3]

The lemmatization of copulatives in modern dictionaries cannot be described without focusing on the so-called *user-perspective* which has emerged as an all important criterion in the selection and lexicographical treatment of lexical items. Lexicographers such as Gouws and Prinsloo (1998), Hartmann and James (1998), Prinsloo and De Schryver (1999), Gouws (2000), etc. emphasize the importance of the user-perspective in modern dictionary compilation. A dictionary should not primarily reflect the attitude of the lexicographer but it should rather be aimed at specific needs of a well-defined target user. Dictionary makers are constantly striving to improve the quality of their reference works, especially in respect of enhanced information retrieval strategies. The aim should be to present the dictionary text as user-friendly as possible. Gouws and Prinsloo (1998: 18) write:

> The user-perspective, so prevalent in modern-day metalexicography, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality. Good dictionaries do not only display a linguistically sound treatment of a specific selection of lexical items. They are also products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are.

Lemmatizing the complex copulative system in a user-friendly way poses a major challenge to the lexicographic skills of a dictionary compiler and can be regarded as a text-book example of the lexicographer's duty as a *mediator* between a complicated linguistic issue on the one hand and the dictionary user on the other. His/her task as mediator is to find ways to lexicographically treat copulatives in such a way that it is user-friendly. According to Prinsloo and Gouws (1996: 103), lexicographers compiling Northern Sotho dictionaries have to rely on insights from the field of metalexicography:

> As the lexicographer is the mediator between theoretical linguistics and the everyday language user, it is important that a dictionary should convey enough information to satisfy the needs of the user and it should present the information in such a way that the relevant linguistic patterns and peculiarities of the target language can be emphasized. Within the boundaries of metalexicographical criteria a lexicographer should also be innovative enough to apply and adapt the theoretical insights, which are not language specific, to a specific language.

In addition to satisfying the *decoding* needs of the target user, namely to correctly interpret copulatives, particular attention should also be given to the *encoding* needs of the user, i.e. providing sufficient assistance to the user to produce correct copulative structures in speech or writing. Fox (1987: 138-139) states in this regard:

> It is important that we realize that learners' needs for encoding are at least as urgent as their needs for decoding. More so, perhaps, in these days when such strong emphasis is placed on communication. … We must ensure that the information we give students will genuinely help them to produce language of their own which is as near as possible to that produced by native speakers. In the past, dictionaries have been seen more as aids to decoding than aids of encoding. There is no reason why they should not succeed in doing both. … Dictionaries have traditionally concentrated more on the grammatical, partly perhaps because they have been seen as tools for interpreting language rather than for creating it.

In order to achieve these objectives the lexicographer should optimise the three major structural components of dictionaries, namely the *macrostructure*, *microstructure* and *access structure* (cf. Gouws 1996: 19). On the *macrostructural* level the lexicographer must ensure that all copulative forms, be they particles, concords or verb forms, are lemmatized. On the *microstructural* level the treatment of each copulative lemma should be done with the decoding and encoding needs of the target user in mind. In terms of Prinsloo and Gouws (1996: 104), the linguistic and communicative needs of the target user should determine both the macrostructural and microstructural selection.

The dictionary should be a practical linguistic instrument. Hartmann and James (1998: 2) define *access* as 'the relative ease with which information can be located in a reference work' and state that 'in a general dictionary this may be the alphabetical order in which headwords appear'. Compare also Gouws and Prinsloo (1998: 18):

> The quality of dictionary use, that is the degree of success a user experiences when consulting a dictionary and employing the retrieved information, is determined by a variety of features but one of the most important characteristics of a good dictionary is its accessibility. Accessibility leads to an unambiguous retrieval of the information presented on both the macro- and microstructural levels.

The *outer access structure* (cf. Gouws 1996: 19 and Louw 1999: 110) should successfully guide the user to the lemmatized canonical form or multiword lemma of each and every copulative likely to be looked for by the target user. The *inner access structure* then guides the user to the specific information categories within the article of each lemmatized copulative. As will be illustrated below, a dedicated, preferably corpus-based, approach towards the compilation of the latter microstructural categories is required for copulatives. However, it will be emphasized that, even rich and well-constructed articles for copulatives are at best insufficient for the encoding needs of the target user and should be *strongly* supported by the *mediostructure* of the dictionary. The mediostructure or *system of cross-referencing* can be described in terms of Wiegand (1996) as referring the dictionary user from a reference position to a reference address where additional information can be obtained. This is done by a variety of reference markers such as *see*, *compare*, →, etc. (See Gouws and Prinsloo 1998: 19 for a detailed discussion.) Of particular importance for copulatives are the so-called *external reference addresses* where the cross-reference exceeds the boundaries of the article, referring the user from the central text to the *back matter* or even to an external source such as a grammar book. For copulatives, it will be argued, the entire route of information retrieval, from the outer access structure right up to one (or more) external reference address(es), should entail a comprehensive and logical *build-up of information* for the user. Here cross-reference to well-compiled tables and additional information reflecting the *total scope* of the copulative is essential. It is here that the user must be presented with the 'overall picture' of the copulative system, or more formally, that the mediostructure should be employed to counter the decontextualisation which is brought about by the alphabetical ordering of copulative lemmas in the central text.[4]

Complexity of the copulative

It has been claimed above that the copulative system in Northern Sotho is complicated. Poulos and Louwrens (1994) devote 37 pages to the discussion of copulatives and summarises it as follows (Poulos and Louwrens 1994: 320-326):

CONSOLIDATION OF COPULATIVE PATTERNS[5]

The various patterns for the different basic tense forms of the copulative are listed below, for ease of reference. It must be remembered that **CB**, i.e. the *complement* of the copula represents different entities or word categories, depending on the type of copulative that is used. The following abbreviations are used: **SC** *subject concord,* **CP** *copulative prefix,* **CB** *copulative base* which is also termed the *complement.*

The identifying copulative
The indicative series *The present tense Principal Identifying* pos. lst and 2nd persons: **SC - CB** Classes: **CP - CB** neg. 1st and 2nd persons: **ga - SC - CB** Classes: **ga - se - CB** *Participial* pos. 1st and 2nd person: **SC - le - CB** Classes: **CP - le - CB** neg. lst and 2nd person: **SC - se - CB** Classes: **CP - se - CB**

The future tense Principal pos. 1st and 2nd person: *SC - tlô/tla - ba + CB* Classes: *CP - tlô/tla - ba + CB* neg. 1st and 2nd person: *SC - ka - se - bê + CB SC* Classes: *CP - ka - se - bê + CB Participial* pos. 1st and 2nd person: *SC - tlô/tla - ba + CB* Classes: *CP - tlô/tla - ba + CB* neg. 1st and 2nd person: *SC - ka - se - bê + CB* Classes: *CP - ka - se - bê + CB The past tense Principal* pos. 1st and 2nd person: *SC - bilê + CB* Classes: *CP - bilê + CB* neg. 1st and 2nd person: *ga - se - SC - be + CB ga - se - SC2 - a - ba + CB ga - SC2 - a - ba + CB* Classes: *ga - se - CP - bê + CB ga - se - SC2 - a - ba + CB1 ga - SC2 - a - ba - CB Participial* pos. lst and 2nd person: *SC - bilê + CB* Classes: *CP - bilê + CB* neg. lst and 2nd person: *SC - sa - ba + CB* Classes: *CP - sa - ba + CB*

The potential   Principal and participial lst and 2nd person: pos. *SC - ka - ba + C* neg. *SC - ka - se - bê + CB* Classes: pos. *CP - ka - ba + CB* neg. *CP - ka - sê - bê + CB*

The subjunctive 1st and 2nd person: pos. *SC - bê + CB* neg. *SC - se - bê + CB* Classes: pos. *CP - bê + CB* neg. *CP - se - bê + CB* Note also the compound negative *SC/CP - se - kê + SC2 - a - ba + CB*

The consecutive lst and 2nd person: pos. *SC2 - a - ba + CB* neg. *SC2 - a - se - bê + CB* Classes: pos. *SC2 - a - ba + CB* neg. *SC2 - a - se - bê + CB* Note also the compound negative *SC2 - a - se - ke + SC2 - a - ba + CB*

The habitual 1st and 2nd person: pos. *SC - be + CB* neg. *SC - se - be + CB - be + CB* Classes pos. *CP - be + CB* neg. *CP - se - be + CB*

The infinitive pos. *go - ba + CB* neg. *go - se - bê + CB*

The imperative pos. *e - ba - ng + CB* or *ba - a - ng + CB* neg. *se - bê - ng + CB*

The descriptive and locational copulative

The indicative series   Principal Persons and classes: pos. *SC + CB* neg. *ga - SC + CB Participial* Persons and classes: pos. *SC - le + CB* neg. *SC - se + CB The future tense Principal and participial* Persons and classes pos. *SC - tlô/tla - ba + CB* neg. *SC - ka - se - bê + CB The past tense Principal* Persons and classes: pos. *SC - bilê + C* neg. *ga - se - SC - ba + CB ga - se - SC - be + CB ga - SC2 - a - ba + CB Participial* Persons and classes: pos. *SC - bilê + CB* neg. *SC - sa - ba + CB*

The potential   Principal and participial Persons and classes: pos. *SC - ka - ba + CB* neg. *SC - ka - se - bê + CB*

The subjunctive Persons and classes: pos. *SC - bê + CB* neg. *SC - se - bê + CB* Note the compound negative *SC - se - kê + SC2 - a - ba + CB*

The consecutive Persons and classes: pos. *SC2 - a - ba + CB* neg. *SC2 - a - se - bê + CB* Note the compound negative *SC2 - a - se - kê + SC2 - a - ba + CB*

The habitual Persons and classes: pos. *SC - be + CB* neg. *SC - se - be + CB*

The infinitive Persons and classes: pos. *go - ba + CB* neg. *go - se - bê + CB*

The imperative Persons and classes: pos. *e - ba - ng + CB* or *ba - a - ng + CB* neg. *se - bê - ng + CB*

The associative copulative

The indicative series   Principal Persons and classes: pos. *SC - na - le + CB* neg. *ga - SC + na - (le) + CB Participial* Persons and classes: pos. *SC - na - le + CB* neg. *SC - se - na - (le) + CB* The stem remains *-na* in both the positive and the negative. *The future tense Principal and participial* Persons and classes: pos. *SC - tlo/tla - ba - le + CB* neg. *SC - ka - se - bê - le + CB The past tense Principal* Persons and classes pos. *SC - bilê - le + CB* neg. *ga - se - SC2 - a - ba - le + CB ga - se - SC - be - le + CB ga - SC2 - a - ba - le + CB Participial* Persons and classes pos. *SC - bilê - le + CB* neg. *SC - sa - ba - le + CB*

The potential   Principal and participial Persons and classes: pos. *SC - ka - ba - le + CB* neg. *SC - ka - se - bê - le + CB*

The subjunctive Persons and classes: pos. *SC - bê - le + CB* neg. *SC - se - bê - le + CB* Note the compound negative S*C - se - kê + SC2 - a - ba - le + CB*

The consecutive Persons and classes: pos. *SC2 - a - ba - le + CB* neg. *SC2 - a - se - bê - le + CB* Note the compound negative *SC2 - a - se - kê + SC2 - a - ba - le + CB*
The habitual Persons and classes: pos. *SC - be - le + CB* neg. S*C - se - be - le + CB*
The infinitive Persons and classes: pos. *go - ba - le + CB* neg. *go - se - bê - le + CB*
The Imperative Persons and classes: pos. *e - ba - ng - le + CB* or *ba - a - ng - le + CB* neg. *se - bê - ng - le + CB*

This effort to summarise or consolidate copulative patterns firstly proves that copulatives in Northern Sotho is a *complicated issue*. Secondly, it is clear that such a summary is, unfortunately, user-unfriendly and not useful in such a format for, e.g., inclusion in the back matter of a dictionary. The lexicographer in his/her role as mediator between a complex grammatical pattern on the one hand and the inexperienced target user on the other, must find ways to present these copulatives in a logical and user-friendly way in dictionaries. Before lexicographers can consider appropriate lemmatization strategies for copulatives in Northern Sotho, they themselves must have absolute clarity on the nature and full scope of the occurrences of copulatives. An example-driven, oversimplified approach will now be taken in order to compile a comprehensive overview of copulatives in Northern Sotho. This will form the basis of, or input to, suggested lemmatization strategies for these copulatives in subsequent paragraphs. Prospective compilers of dictionaries for African languages should take note of the fact that in many cases it will be necessary to do work and even research on a linguistic level before lemmatization strategies can be considered for a specific phenomenon. Compare in this regard the lemmatization of reflexives (Prinsloo 1992) and the lemmatization of adjectives (Gouws and Prinsloo 1997).

An example-driven reconstruction of copulatives in Northern Sotho

In order to master the copulative it is crucial for learners to understand that:

— 	two types of copulatives can be distinguished, namely *static* (in a state of rest) and *dynamic* (in motion or changing);

— 	copulatives express three different semantic *relations* between a subject and a complement, namely *identification/equality*, *descriptive* or *associative;*

— 	copulatives can consist of a particle e.g. *ké*, or a copulative subject concord, *ke, re, o, le, ba*, etc. or a copulative verb, *e le, e se, o ba, o na*, etc.[6]

Consider the following examples of a copulative particle, a copulative subject concord and copulative verbs of *static* and *dynamic* copulatives in *identifying*, *descriptive* and *associative* relations. These relations are expressed between *monna* 'man' and *morutiši* 'teacher' or *bohlale* 'wisdom' or *mpša* 'dog' in Table 1, A–C respectively.

Table 1

A Identifying copulative: The relation is one of *identification/equality*, i.e. subject **=** complement			
Static:	**monna**	**ke**	**morutiši**
	the man	is	a teacher
(**ke** = copulative particle)			
Dynamic:	**monna**	**e ba**	**morutiši**
	the man	becomes	a teacher
(**e ba** = copulative verb)			
B Descriptive copulative: The relation is one of *description*, i.e. complement describes subject			
Static:	**monna**	**o**	**bohlale**
	the man	is	clever
(**o** = copulative subject concord)			
Dynamic:	**monna**	**o ba**	**bohlale**
	the man	becomes	clever
(**o ba** = copulative verb)			
C Associative copulative: The relation is one of *association*, i.e. subject is associated with complement			
Static:	**monna**	**o na**	**le mpša**
	the man	is	with a dog (has a dog)
(**o na** = copulative verb)			
Dynamic:	**monna**	**o** tlo **ba**	**le mpša**
	the man	will be	with a dog (will have a dog)
(**o** tlo **ba** = copulative verb containing the future prefix)			

Dynamic copulatives occur in all moods.[7] It is advisable for the lexicographer (and the compiler of a basic Northern Sotho grammar) to use the user's pre-sumed basic knowledge of the noun class system and the moods, tenses and aspects of common verbs as point of departure. Learners normally master the nominal and verbal systems first when studying an African language.

Consider firstly copulatives such as *o* in Table 1, B, which Poulos and Louwrens (1994: 307) describe as 'a form resembling the subject concord'. The lexicographer should immediately link such copulatives to the target user's presumed knowledge of subject concords as summarized in an example-driven layout in Table 2.

Table 2: Subject Concords

Person or noun class	**Example**	**Sc.**
1st Person singular	nna 'I'	**ke**
1st Person plural	rena 'we'	**re**
2nd Person singular	wena 'you' (singular)	**o**
2nd Person plural	lena 'you' (plural)	**le**
Class 1	monna 'man'	**o/a**
Class 2	banna 'men'	**ba**

Class 3	monwana 'finger'	**o**
Class 4	menwana 'fingers'	**e**
Class 5	lesogana 'young man'	**le**
Class 6	masogana 'young men'	**a**
Class 7	selepe 'axe'	**se**
Class 8	dilepe 'axes'	**di**
Class 9	nku 'sheep (singular)'	**e**
Class 10	dinku 'sheep (plural)'	**di**
Class 14	bogobe 'porridge'	**bo**
Class 15	go sepela 'to walk'	**go**
Class 16	fase 'below'	**go**
Class 17	godimo 'above'	**go**
Class 18	morago 'behind'	**go**
invariable[9]	ge e le ... 'if he/she/it ... is ...'	**e**
indefinite[10]	go tseba mang? 'who knows?'	**go**

Secondly, the lexicographer should link the dynamic copulative verb *ba (be, bile)* to the user's knowledge of the moods, tenses, etc. of ordinary verbs. This is exactly what is suggested for columns 5, 6 and 7 in comparison with columns 1-4 in Table 3. Note the degree of similarity between common verbs and the copulative verb *ba*. In an attempt to save space, the complete initial part, *mosadi* or *ge mosadi*, or *mosadi yo*, etc. is not repeated for the copulative in columns 5, 6 and 7.

In Table 3 no fewer than 34 copulative forms for 3 different copulative relations were given, covering only class 1. Multiplied by the roughly 20 different sets of concords for persons and classes in Table 1, this means about 34 x 3 x 20 = 2,040 possible candidates for lemmatization of the dynamic copulative. It stands to reason that a *drastic reduction strategy* is required for dynamic copulatives in order to stay within the physical limitations of a paper dictionary. The same is applicable to static copulatives for which Table 4 is a condensed, example-based layout.

Thus, for instance, the final example in the last column of Table 4, the copulative *se ... nago* can be described as a static, associative, copulative verb, in the relative mood in the present tense, negative.

Static copulatives occur in the indicative, situative and relative moods only. Once again, it has advantages for the user to link it to his/her presumed knowledge of common verbs,[11] and now also building on his/her knowledge of the copulative subject concords and dynamic copulative verbs.

Table 4 reflects, in a simplistic way, the use of first, second and third persons (the latter is divided into noun classes in Northern Sotho), for the positive and negative forms of the three different relations for the indicative, situative and relative moods, including the progressive (+prog.).

Table 3: Dynamic Copulatives

Column 1: MD. = MOOD, IND. = INDICATIVE, SIT. = SITUATIVE, REL. = RELATIVE, SUB. = SUBJUNCTIVE, CON. = CONSECUTIVE, INF. = INFINITIVE, IMP. = IMPERATIVE, HAB. = HABITUAL

Column 2: PRES. = PRESENT, FUT. = FUTURE, PAS. = PAST +Pot. = containing the Potential

Column 3: ACT. = ACTUALITY (p. = positive, n. = negative)

MD.	TENSE	ACT.	Common verb	Identifying	Descriptive	Associative
IND.	PRES.	p.	mosadi o reka dipuku	e ba morutiši	o ba bohlale	o ba le mpša
		n.	mosadi ga a reka dipuku	ga e be morutiši	ga a be bohlale	ga a be le mpša
	+Pot.	p.	mosadi a ka reka dipuku	e ka ba morutiši	a ka ba bohlale	a ka ba le mpša
		n.	mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	FUT.	p.	mosadi o tlo/tla reka dipuku	e tlo/tla ba morutiši	o tlo/tla ba bohlale	o tlo/tla ba le mpša
		n.	mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	PAS.	p.	mosadi o rekile dipuku	e bile morutiši	o bile bohlale	o bile le mpša
		n.	mosadi ga se a reka dipuku	ga se ya ba morutiši	ga se a ba bohlale	ga se a ba le mpša
SIT.	PRES.	p.	ge mosadi a reka dipuku	e eba morutiši	a eba bohlale	a eba le mpša
		n.	ge mosadi a sa reke dipuku	e sa be morutiši	a sa be bohlale	a sa be le mpša
	+Pot.	p.	ge mosadi a ka reka dipuku	e ka ba morutiši	a ka ba bohlale	a ka ba le mpša
		n.	ge mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	FUT.	p.	ge mosadi a tlo/tla reka dipuku	e tlo/tla ba morutiši	a tlo/tla ba bohlale	a tlo/tla ba le mpša
		n.	ge mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	PAS.	p.	ge mosadi a rekile dipuku	e bile morutiši	a bile bohlale	a bile le mpša
		n.	ge mosadi a sa reka dipuku	e sa ba morutiši	a sa ba bohlale	a sa ba le mpša
REL.	PRES.	p.	mosadi yo a rekago dipuku	e bago morutiši	a bago bohlale	a bago le mpša
		n.	mosadi yo a sa rekego dipuku	e sa bego morutiši	a sa bego bohlale	a sa bego le mpša
	+Pot.	p.	mosadi yo a ka rekago dipuku	e ka bago morutiši	a ka bago bohlale	a ka bago le mpša
		n.	mosadi yo a ka se rekego dipuku	e ka se bego morutiši	a ka se bego bohlale	a ka se bego le mpša
	FUT.	p.	mosadi yo a tlo/tla rekago dipuku	e tlo/tla bago morutiši	a tlo/tla bago bohlale	a tlo/tla bago le mpša
		n.	mosadi yo a ka se rekego dipuku	e ka se bego morutiši	a ka se bego bohlale	a ka se bego le mpša
	PAS.	p.	mosadi yo a rekilego dipuku	e bilego morutiši	a bilego bohlale	a bilego le mpša
		n.	mosadi yo a sa rekago dipuku	e sa bago morutiši	a sa bago bohlale	a sa bago le mpša
SUB.		p.	(gore) mosadi a reke dipuku	e be morutiši	a be bohlale	a be le mpša
		n.	(gore) mosadi a se reke dipuku	e se be morutiši	a se be bohlale	a se be le mpša
CON.		p.	mosadi a reka dipuku	ya ba morutiši	a ba bohlale	a ba le mpša
		n.	mosadi a se reke dipuku	ya se be morutiši	a se be bohlale	a se be le mpša
INF.		p.	go reka dipuku	go ba morutiši	go ba bohlale	go ba le mpša
		n.	go se reke dipuku	go se be morutiši	go se be bohlale	go se be le mpša
IMP.		p.	reka dipuku!	eba morutiši!	eba bohlale!	eba le mpša!
		n.	se reke dipuku!	se be morutiši!	se be bohlale!	se be le mpša!
HAB.		p.	mosadi a reke dipuku	e be morutiši	a be bohlale	a be le mpša
		n.	mosadi a se reke dipuku	e se be morutiši	a se be bohlale	a se be le mpša

32   D.J. Prinsloo

Table 4: Static Copulatives

	Positive (Identifying)	Negative	Positive (Descriptive)	Negative	Positive (Associative)	Negative
Indicative:						
1ps	(Nna) ke morutiši	ga ke morutiši	ke bohlale	ga ke bohlale	ke na le mpša	ga ke na mpša
+prog.	(Nna) ke sa le morutiši	ga ke sa le morutiši	ke sa le bohlale	ga ke sa le bohlale	ke sa na le mpša	ga ke sa na mpša
1pp-2pp -----	-----	-----	-----	-----	-----	-----
1	Monna ke morutiši	ga se morutiši	o bohlale	ga a bohlale	o na le mpša	ga a na mpša
+prog.	Monna e sa le morutiši	ga e sa le morutiši	o sa le bohlale	ga a sa le bohlale	o sa na le mpša	ga a sa na mpša
2-8 -----	-----	-----	-----	-----	-----	-----
9	Kgoši ke morutiši	ga se morutiši	e bohlale	ga e bohlale	e na le mpša	ga e na mpša
+prog.	Kgoši e sa le morutiši	ga e sa le morutiši	e sa le bohlale	ga e sa le bohlale	e sa na le mpša	ga e sa na mpša
10-18 -----	-----	-----	-----	-----	-----	-----
Situative:						
1ps	(Nna) ge ke le morutiši	ke se morutiši	ke le bohlale	ke se bohlale	ke na le mpša	ke se na mpša
+prog.	(Nna) ge ke sa le morutiši	ke se sa le morutiši	ke sa le bohlale	ke se sa le bohlale	ke sa na le mpša	ke se sa na mpša
1pp-2pp -----	-----	-----	-----	-----	-----	-----
1	Ge monna e le morutiši	e se morutiši	a le bohlale	a se bohlale	a na le mpša	a se na mpša
+prog.	Ge monna e sa le morutiši	e se sa le morutiši	a sa le bohlale	a se sa le bohlale	a sa na le mpša	a se sa na mpša
2-4 -----	-----	-----	-----	-----	-----	-----
5	Ge lesogana e le morutiši	e se morutiši	le le bohlale	le se bohlale	le na le mpša	le se na mpša
+prog.	Ge lesogana e sa le morutiši	e se sa le morutiši	le sa le bohlale	le se sa le bohlale	le sa na le mpša	le se sa na mpša
6-18 -----	-----	-----	-----	-----	-----	-----
Relative:						
1ps	(Nna) yo ke lego morutiši	ke sego morutiši	ke lego bohlale	ke sego bohlale	ke nago le mpša	ke se nago le mpša
+prog.	(Nna) yo ke sa lego morutiši	ke se sa lego morutiši	ke sa lego bohlale	ke se sa lego bohlale	ke sa nago le mpša	ke se sa nago mpša
1pp-2pp -----	-----	-----	-----	-----	-----	-----
1	Monna yo e lego morutiši	e sego morutiši	a lego bohlale	a sego bohlale	a nago le mpša	a se nago le mpša
+prog.	Monna yo e sa lego morutiši	e se sa lego morutiši	a sa lego bohlale	a se sa lego bohlale	a sa nago le mpša	a se sa nago mpša
2-6 -----	-----	-----	-----	-----	-----	-----
7	Sehlalefi se e lego morutiši	e sego morutiši	se lego bohlale	se sego bohlale	se nago le mpša	se se nago le mpša
+prog.	Sehlalefi se se sa lego morutiši	e se sa lego morutiši	se sa lego bohlale	se se sa lego bohlale	se sa nago le mpša	se se sa nago mpša
8-18 -----	-----	-----	-----	-----	-----	-----

Reduction of copulative subject concords, static and dynamic copulatives

The next step in isolating lemmas for copulatives should be directed at reducing the number of possible candidates for lemmatization in a logical, user-friendly way. The reduction of *subject concords* can be done purely logically by eliminating duplicate forms i.e. lemmatising the subject concord for classes 8 and 10 as a single lemma *di*, then by treating it for all of its copulative functions as well as its normal use as subject concord of classes 8 and 10 in the microstructure. Table 5 suggests such a reduction from 22 (grammatical) candidates to 11 logical lemmas.

Table 5

1st Person singular	**ke**	Class 6, 1	**a**
1st Person plural	**re**	Class 7	**se**
2nd Person singular, Cl. 1, 3	**o**	Class 8, 10	**di**
2nd Person plural	**le**	Class 14	**bo**
Class 2	**ba**	Class 15-18, Indefinite Sc.	**go**
Class 4, 9, Invariable Sc.	**e**		

Thus the subject concords *ke*, *re*, *o*, *le*, *ba*, *e*, *a*, *se*, *di*, *bo* and *go* are to be entered as macrostructural elements, firstly for their ability to express a copulative meaning (cf. Table 4 above) and for their occurrence in dynamic and static copulative verbs with *ba, le*, etc.

Based on the assumption that users have mastered the basic verbal system, the entire Table 3 for *dynamic copulatives* can be collapsed into the following structures where *Sc.* represents the entire series of subject concords in Table 5.

(1)   Structure of Dynamic Copulatives
    (a)   (**Sc.**) + (tlo/tla/ka) + (e)ba(go) + (le)
    (b)   ga se + **Sc.** + ba + (le)
    (c)   **Sc.** + be
    (d)   ga + **Sc.** + be
    (e)   **Sc.** + ka se + be(go)
    (f)   **Sc.** + bile(go)

Finally the whole of the complex Table 4 for *static copulatives* can be reduced to the rules formulated in (2).

(2)   Structure of Static Copulatives
    (a)   ké
    (b)   ga se
    (c)   **Sc.**
    (d)   ga + **Sc.**

(e)   **Sc.** + (sa) + le(go)
(f)   **Sc.** + se(go)
(g)   **Sc.** + (sa) + na(go)
(h)   ga + **Sc.** + (sa) + na
(i)   **Sc.** + se + (sa) + na(go)

Selecting the lemmas

The list of subject concords (Table 5) as well as the rules given in (1) and (2) can be very useful as input strings to morphological analysers and machine translation programmes with the aim of generating — or breaking down — copulative constructions in Northern Sotho. However, for the lexicographer a further step is required in the lemmatization process, namely to isolate the final macrostructural elements, i.e. copulative lemmas, from these rules.

For the dynamic copulative *ba(go), be(go)* and *bile(go)* should be lemmatized. Here it is assumed that *subject concords*, the future tense prefix *tlo/tla*, the potential *ka* and the conjunctive *le* have all been lemmatized in their own right with the necessary cross-references to the copulative tables in the back matter.

For the static copulative, *ké, ga se*, the reduced list of *subject concords*, this time for their *copulative use*, as well as *le(go), se(go),* and *na(go)* should be lemmatized. The progressive *sa* is assumed to be lemmatized with a cross-reference to the copulative tables in the back matter.

Consider the number of occurrences of each of these copulatives in a 6-million-word corpus of Northern Sotho:

Table 6

ké	±8,000	**ga se**	±4,000		
ba(go)	±1,200	**be(go)**	±6,000	**bile(go)**	30
na(go)	±12,000	**le(go)**	±30,000	**se(go)**	±2,000

These enormous overall counts clearly indicate not only that they should be included as lemmas but also that *exhaustive treatment* is required/justified especially for the encoding needs of inexperienced target users. The success or failure of a selected number of Northern Sotho dictionaries, *Pukuntšu* (Kriel 1983), *Pukuntšu* (Kriel and Van Wyk 1989), *New English–Northern Sotho Dictionary* (Kriel 1967) and *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975), to satisfactorily treat these copulatives will now be evaluated. These dictionaries are the most comprehensive commercially published dictionaries for Northern Sotho.

Critical evaluation of the treatment of copulatives in Northern Sotho dictionaries

Let us start with what these dictionaries have to offer the learner in terms of

English or Afrikaans dictionaries with Northern Sotho as the target language in selecting the word *is*. For users departing from English as source language, no guidance is available since *is* is not lemmatized in *New English–Northern Sotho Dictionary* (Kriel 1967) and cannot be looked up in *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975) since it is a mono-directional Northern Sotho → English/Afrikaans dictionary. For Afrikaans, it is lemmatized in both versions of *Pukuntšu*.

(3)   *Pukuntšu* (Kriel 1983)
        is, ke; go; *daar* –, go na le; *dit* –, ke; *dit* – *so*, go bjalo; *daar* –, go na le; *dit* – *nie so nie*, ga go bjalo; *daar* – *nie*, ga go, ga dio; ga di gona.

(4)   *Pukuntšu* (Kriel and Van Wyk 1989)
        is, *(dit)* –, ke; *(dit)* – *nie*, ga se; *daar* –, go na le; *dit* – *so*, go bjalo.

At least *is* is included as a lemma in these dictionaries. Treatment does not, however, go beyond giving basic decoding information. It is furthermore limited to the third person only and, in terms of the basic relations in Table 1 above, covers only the first part of A. Note also that in (3), the information '*daar* –, go na le' is given twice.
        A suggested improvement could be (5):

(5)   **is** ke [id. cop.], **ke lengwalo, ga se sephuthana** it is a letter, it is not a parcel; **bohodu ke sebe**. theft is a sin; o/e/le … [des. cop.], **aowa, mosadi yo o bohlale, ga a bogale!** No, this woman is clever, she's not cruel!; o/e/le ... na le [ass. cop.], **Satsope o na le Sara** Satsope is with Sara. ► BM 1.1-1.3

In this relatively short entry, basic distinctive grammatical information in A–C of Table 1 is given. Suitable examples and translations are included which are designed in such a way that the positive and negative forms are contrasted. Contrasting *ke* with *ga se* in the first example is valuable information for learners. Appropriate cross-reference to the back matter is also given where more information can be found. Note also the employment of a new convention *o/e/ le ...* indicating to the user that a whole range of concords can be used. Similar entries should be compiled for English lemmas *am, are, be, become* or Afrikaans *word, wees*, etc.
        When we turn to the treatment of copulatives in Northern Sotho with English or Afrikaans as the target languages, *ke* will be taken as a first example:

(6)   *Pukuntšu* (Kriel 1983)
        ke, *ke,* dit is; *se ke,* moenie; – *go re*, dis om te sê, dit beteken.
        ke, – *mo*, dit is dan wanneer. Ke nna yo ke lego Yena, Ek is wat Ek is.

(7)   *Pukuntšu* (Kriel and Van Wyk 1989)
        ke[1], identifiserende kopula, ind. imp. H: (dit) is, deur
        ke[2], osk. 1 p.e. L: ek.

(8)   *New English–Northern Sotho Dictionary* (Kriel 1967)
        ke, 1st person singular, I
        ké, pron. copulative, it is

(9)   *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
        KE (ke-) ... [1.p.s. cop. formative] I am; *ke mogwêra wa gagwê* ... I am your friend
        KE (ké-) ... [cop. formative] it is, he is, she is, they are; *ké lehôdu* ... he/she/it is a
          thief; *ké mahôdu* ... they are thieves; *ké moka* ... that is all

These entries are technically correct but only offer limited decoding informa-
tion and very little encoding information. The amount of dictionary space
could have been utilized much better. For example, the value of the entry **Ke
nna yo ke lego Yena** is highly questionable. Firstly, it does not occur even once
in the corpus. Then, the user should be informed that the first *ke* is the copula-
tive particle 'it is' and that the second *ke* means 'I' occurring as part of a copula-
tive verb *ke lego* 'I who am'. The inexperienced user is also misguided by the
occurrence of *yena* 'he/him' which is translated in this context as 'I'. All infor-
mation given in Kriel's first entry deals with the copulative except *se ke* 'must'
(in which *ke* is used as an auxiliary verb stem). It could at least have been sepa-
rated from the copulatives. It is also unclear why three separate entries were
made. Kriel (1983), Kriel and Van Wyk (1989) and Kriel (1967) do not treat the
copulative subject concord of the first person singular. Ziervogel and Mokgo-
kong (1975) do much better in this regard in at least including both the copula-
tive particle *ke* and the copulative subject concord of the first person singular as
lemmas with typical examples of usage. They also indicate by means of an
example that, in contrast to English, a single strategy exists in Northern Sotho
expressing the notions 'he is', 'she is', and 'it is'. This is a sign of good lexicography
since learners of Northern Sotho tend to have difficulty with this concept. Note,
however, that they incorrectly translated *gagwe* 'his/her friend' as 'your friend'.
The most serious shortcoming in the treatment of the copulative particle *ke* is
that *none* of these dictionaries lemmatize the negative copulative particle *ga se*,
of which the total occurrence in the corpus is estimated at more than 4,000.
Consider (10) for *ga se* and (11) as a suggested improved entry for *ke*.

(10)   **ga se** [cop. part. neg.] it is not, **ga se phošo ya gago** it is not your fault; he/she/it
          is not, **Satsope ga se morutiši, ke mongwaledi** Satsope is not a teacher, she is
          a secretary; they are not, **dingaka ga se mahodu** doctors are not thieves ► **ke**
          ► BM 1.1; ■■■■■
        se, ga ~ ► **ga se**

(11)   **ke¹** (ké) [id. cop. part.], it is, **ke lengwalo, ga se sephuthana** it is a letter, it is not
          a parcel, **bohodu ke sebe** theft is a sin; he/she is, **Madika ke morutiši**
          Madika is a teacher; they are **dikgoši ke baetapele ba ditšhaba** kings are
          leaders of the nations ► BM 1.1; ►**ke²** ■■■■■

Consider as a second example the treatment of *le* in these dictionaries.

(12)   *Pukuntšu* (Kriel 1983)
　　　le, en, plus, ook, mede –; *ge go* –, as daar is; *mogolegwa* –, medegevangene; ...

(13)   *Pukuntšu* (Kriel and Van Wyk 1989)
　　　le², osk. kl 5, H: dit.
　　　le⁴, osk. 2 p.m. L: julle, u.
　　　le⁷, sit./rel. kop. ww. H: synde.

(14)   *New English–Northern Sotho Dictionary* (Kriel 1967)
　　　le, copulative, le-class.

(15)   *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
　　　LE (le-) ... [cc. 2 p.p.] ... you; ... [cc. cl. le-] ... he, she, it
　　　LE ... [cop. d.v. of part.] ... being; *e sa le ka bošego* ... while it is still night/early; *e bê*
　　　　e le la mang? ... whose was it?; *e bê e le mang?* ... who was it?; *e bê e le*
　　　　nna/Matome ... it was I/Matome; *o bê a le gôna* ... he was present

The treatment in *Pukuntšu* (1989) and *New English–Northern Sotho Dictionary* is
once again technically correct, but not exhaustive enough for a copulative
occurring more than 30,000 times in the corpus. Ziervogel and Mokgokong do
much better in at least including typical examples of usage but could have
varied the choice of subject concord more instead of using *e* five times. The
copulative verb *e le* with an estimated overall count of 20,000 or an average of 2
occurrences per page in a Northern Sotho literary work should seriously be
considered for inclusion as a multiword lemma with treatment analogous to
(11) above.

　　　The same line of argumentation could be followed for the treatment of *se*,
na, etc. In the case of *se* only one dictionary treats it.

(16)   *Pukuntšu* (Kriel and Van Wyk 1989)
　　　se⁵, kop. ww. H; *gê a se gôna*, as hy nie aanwesig is nie.

(17)   *Pukuntšu* (Kriel 1983)
　　　na, met, saam met, het, ...; *se/ga* - nie hê nie, het nie; *... ke na le tšhêlêtê*, ek het geld
　　　　(lett., ek is saam met geld); ...

(18)   *Pukuntšu* (Kriel and Van Wyk 1989)
　　　na³, kop. ww. H: wees (met); *ke na le tšhêlêtê* ek is met/het geld,
　　　na ⁴, part. H: met; *nabô*, met hulle; *nago*, met jou.

(19)   *New English–Northern Sotho Dictionary* (Kriel 1967)
　　　na, aux., verb, to be (with), to have; *o na le*, (lit., he is with), he has.

(20)   *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
　　　NA (-na) (part. **ena**) ... [assosiative copulative verb] ... have, be; *a na naso* ... he
　　　　having it; *o na le bana* ... she has children; *yô o nago bana* ... she who has chil-
　　　　dren; *gê a na nabô* ... if she is with them; *ga re na selô* ... we have nothing; *go bê*
　　　　go se na motho ... nobody was there; *gê a ena le tsebô* ... if he has knowledge; …

From these examples it is clear that the treatment offered in the majority of cases is barely sufficient to fulfil the decoding needs of advanced users. The learner gets no guidance for encoding purposes from any Northern Sotho dictionary.

Lacking in all these dictionaries is treatment of copulatives in the back matter of the dictionary and appropriate use of the mediostructure. Ziervogel and Mokgokong, for example, include a detailed mini-grammar in the front matter of the dictionary, but hardly refer to copulatives, let alone giving an exhaustive discussion.

As far as the dynamic copulatives are concerned, fairly detailed entries are offered for the copulative verb stem *ba*. The need for more encoding information, however, remains. Note also that Kriel (1983) confuses the auxiliary verb stem *ba* with the copulative one but that this has been rectified in Kriel and Van Wyk (1989).

(21)   *Pukuntšu* (Kriel 1983)
        ba, ... hulpww., sal wees; selfs, ook, eindelik, totdat; *go – le*, het, besit; lett., om te wees met; *– gona*, daar wees, *bile*, was gewees; *eba*, sal wees, *go –*, om te wees.

(22)   *Pukuntšu* (Kriel and Van Wyk 1989)
        ba[6], kop. ww. L: wees, word; *go ba le*, om te wees/word met, om te hê/kry; *go ba gôna*, om aanwesig/teenwoordig/wel te wees; *go ba le dinala tšê ditêlêlê*, om lui te wees, om graag te slaan.

(23)   *New English–Northern Sotho Dictionary* (Kriel 1967)
        ba v.n., to be, to become.

(24)   *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
        BA (-ba, -bile) ... be, become; *go ba monna* ... to be/become a man; *lesogana, gê le nyala le ba monna* ... when a young man marries he becomes a man; *go ba setekeng* ... be pursued doggedly.

Be and *bile* occur frequently in the corpus, but are not substantially treated in any of the four dictionaries. The following examples show the complete treatment in all four dictionaries combined, clearly reflecting insufficient treatment.

(25)   *New English–Northern Sotho Dictionary* (Kriel 1967)
        bê, subj., of **go ba**, to be; **e se bê mohlomong**, lest.

(26)   *Pukuntšu* (Kriel 1983)
        bile, *bi.le*, *(ba)*, was (gewees), reeds, selfs; geword het; *e – kaone*, dit was beter.

(27)   *Pukuntšu* (Kriel and Van Wyk 1989)
        bilê[1], kop. ww. HH (perf. van **ba**): was gewees, het geword.

(28)   *New English–Northern Sotho Dictionary* (Kriel 1967)
        bile, aux. v., even, pft., has been.

Constructing the back matter

It has been argued so far that the data types on microstructural level should be designed in such a way that they (a) fulfil the needs of the decoding user, and (b) go some way in assisting the encoding user in his/her information retrieval process. Formulated differently, the article should be constructed in such a way that it contains sufficient decoding information for mother-tongue speakers and advanced users. Thus, in most cases, it will not be necessary for them to follow up on the article-external cross references to the back matter or to an outside source like a grammar book. The learner, on the other hand, will in most cases benefit from following up on all references to the back matter and dictionary external sources.

The back matter should be constructed in such a way as to act as a bridge between the article and the outside source. This means firstly that the relevant section should not be as comprehensive as the discussion in a grammar book. Compare in this regard Ziervogel and Mokgokong (1975) who dedicated 39 pages in small print in the front matter to an all-in-one mini-grammar and user's guide where no exhaustive discussion of copulatives is given anyhow. Secondly, the discussion of copulatives in the back matter should fulfil the basic purpose of cross-reference, namely to be the reference address where the user would indeed find more information on copulatives, structured in such a way that it extends the information the user has obtained in consulting the article of the copulative in the central text. For the encoding user it should thus be 'the next logical step' in explaining the correct use of the copulative. Likewise, the back matter should also be the logical step/link to the outside source — thus *a comprehensive process from dictionary article to back matter to outside source*. In paper dictionaries this does not narrow the gap between dictionary and grammar but at least offers logical steps to the user in the information retrieval process, namely dictionary article to back matter to grammar book. In an electronic dictionary, this information retrieval process is simplified as will be briefly outlined below.

Excellent candidates for inclusion in the back matter are Tables 1–5 above. For the learner who has mastered basic skills regarding nouns and verbs in Northern Sotho, these tables will immediately put the copulative into perspective. Table 1 gives the user a full perspective on the relations, types and variant forms of copulatives. Table 2 brings together all the possible subject concords that can occur in copulative constructions. Table 3 gives the overall picture for the dynamic copulative and at the same time links it directly and visually to the use of ordinary verbs. Table 4, which should be expanded to include examples from all classes, will probably be the most useful table, reflecting the much-needed overall picture for the static copulatives. Translations should be added where applicable. The use of the invariable subject concord *e-* and the so-called indefinite subject concord *go* could also be explained in the back matter or, alternatively, catered for with a dictionary-external cross-reference to one or more grammar books. These tables could also form the basis for what is offered to the encoding user in electronic dictionaries.

Electronic dictionary entries

In contrast to the paper dictionary, an electronic dictionary can offer the user an exciting new range of data-access routes to the contents. (See Prinsloo 2001 for a detailed discussion.) The encoding needs of users who look up copulatives in electronic dictionaries for Northern Sotho can, for example, be satisfied by means of pop-up screens. Compare the following example for (10) above, now presented for an electronic dictionary as (29):

Indicative:	Identifying	
1ps	(Nna) *ke* morutiši	*ga ke* morutiši
+prog.	(Nna) *ke sa le* morutiši	*ga ke sa le* morutiši
1pp-2pp -----		-----
1	Monna *ke* morutiši	*ga se* morutiši
+prog.	Monna *e sa le* morutiši	*ga e sa le* morutiši
2-18 -----		-----
Click here for *Complete Table*		

ga se phošo ya gago *it is not your fault*; he/she/it is not, **Satsope ga se morutiši, ke mongwaledi** *Satsope is not a teacher, she is a secretary*; they are not, **dingaka ga se mahodu** *doctors are not thieves*

(29)   **ga se** … [cop. part. Neg] it is not, ⌂ *structure; examples; pronunciation; combination; frequency; concords; idioms; expressions; picture; copulative relations;* ■■■■□

A Identifying copulative: The relation is one of *identification/equality*, i.e. subject = complement Click here for *Complete Table*	
B Descriptive copulative: The relation is one of *description*, i.e. complement describes subject Click here for *Complete Table*	
C Associative copulative: The relation is one of *association*, i.e. subject is associated with complement Click here for *Complete Table*	

The only effort required of the user is to rest the cursor momentarily on the lemma. A click of the mouse takes him/her down to a next level of information. Such dictionary articles can be short but multifunctional in that they simultaneously serve the decoding needs of the more experienced user, and the encoding needs of the less experienced one. For decoding purposes the skeleton entry might suffice, therefore information boxes only appear if the user wants to see them. For the inexperienced user, the text boxes in this example offer encoding information which could at best be offered in a paper dictionary as a reference address in the back matter. One of the major advantages of such entries in electronic dictionaries is that the gap between dictionary and grammar, which is generally believed to be 'unbridgeable', is starting to close, as suggested by Geeraerts (2000: 77):

In this way, the *differences between dictionary and grammar begin to diminish*: the

dictionary entries are linked to a grammatical description of the language that offers more detail than the grammatical compendium that is sometimes included with paper dictionaries.

Conclusion

Compiling user-friendly dictionaries of a high lexicographic standard for African languages poses a great challenge to prospective lexicographers. They are the mediators between complicated grammatical structures and the decoding and encoding needs of their target users. Complicated structures such as nouns, verbs, copulatives, etc. should not be tackled haphazardly as they come under the attention of the compiler. They should be carefully studied and even researched to obtain a comprehensive overview of the relevant structures. Only then can the lexicographer proceed to plan the macrostructure and microstructure for the lemmatization of a specific construction. On the macrostructural level, candidates for inclusion (or omission) should carefully be considered, preferably based on corpus data. On the microstructural level, data should be presented in such a way that it satisfies both the needs of encoding and decoding users. The mediostructure should be employed in a sensible way to refer the user to reference addresses where more information can be found. Special attention should be given to references to a well-compiled back matter where cohesion of decontextualised items is restored, thus rendering the 'full picture' to the user. In terms of Laufer (1992: 71), this means to really 'know' the copulative:

> When a person "knows" a word, he/she knows the following: the word's pronunciation, its spelling, its morphological components, if any, the words that are morphologically related to it, the word's syntactic behaviour in a sentence, the full range of the word's meaning, the appropriate situations for using the word, its collocational restrictions, its distribution and the relation between the word and other words within a lexical set.

Endnotes

1. An exception is the clear, systematic way in which copulatives are discussed in Louwrens (1991).
2. See Taljard (1999) for a detailed overview of the different standpoints.
3. The term *multiword* in this context refers to more than one *orthographic* word.
4. See Gouws and Prinsloo (1998: 24) for a detailed discussion of this problem and the overuse of mediostructure in this regard to maintain structural links in the African languages.
5. The original layout has been changed to save space.
6. Since copulative verb stems are orthographically written as separate words, they will be indicated as such in this article, namely *le*, *ba*, *be*, *bile* and not as verb stems *-le*, *-ba*, *-be*, *-bile*. Note also that Kriel (1983), and Kriel and Van Wyk (1989) opted for lemmatizing these stems

without hyphens.

7.   The Van Wyk approach (Van Wyk et al. 1992) will be favoured in this article. The debate in respect of moods and alternative classifications fall outside the scope of this article. For the latter, the reader is referred to Poulos and Louwrens (1994).

8.   In this article the term 'copulative subject concords' will be used.

9.   "It is said to be invariable since it's form is not influenced by the particular noun class to which the subject noun belongs." (Louwrens 1994: 91)

10.   "A term used to refer to the concord *go-* which occurs in verbs when the subject of the sentence is *indefinite*." (Louwrens 1994: 84)

11.   In some instances 'linguistic impurity' should be tolerated in favour of user-friendliness. Strictly speaking, the copulative *particle ke* cannot occur in a mood but in this case fruitfully fills an empty slot and should be tolerated as natural scientists tolerate the scientific incorrect statement: 'the sun rises in the east and sets in the west'.

References

Fox, Gwyneth. 1987. The Case for Examples. Sinclair, John M. (Ed.). *Looking Up, An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-49. London: Collins ELT.

Geeraerts, D. 2000. Adding Electronic Value: The Electronic Version of the Grote Van Dale. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress on Lexicography, Stuttgart, Germany, August 8th–12th, 2000*: 75-84. Stuttgart: Stuttgart University.

Gouws, R.H. 1989. *Leksikografie*. Cape Town: Academica.

Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.

Gouws, R.H. 2000. Toward the Formulation of a Metalexicographic Founded Model for National Lexicography Units in South Africa. Wiegand, Herbert E. (Ed.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 109-33. Tübingen: Max Niemeyer.

Gouws, R.H. and D.J. Prinsloo. 1997. Lemmatisation of Adjectives in Sepedi. *Lexikos* 7: 45-57.

Gouws, R.H. and D.J. Prinsloo. 1998. Cross-referencing as a Lexicographic Device. *Lexikos* 8: 17-36.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London: Routledge.

Kriel, T.J. 1967. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English*. Johannesburg: Educum.

Kriel, T.J. 1983. *Pukuntšu*. Pretoria: J.L. van Schaik.

Kriel, T.J. and E.B. Van Wyk. 1989[4]. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.

Laufer, Batia. 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, Harnu et al. (Eds.). 1992. *Euralex '92 Proceedings*: 71-76. Tampere: University of Tampere.

Louw, P.A. 1999. Access Structures in a Standard Translation Dictionary. *Lexikos* 9: 108-118.

Louwrens, L.J. 1991. *Aspects of Northern Sotho Grammar*. Pretoria: Via Afrika

Louwrens, L.J. 1994. *Dictionary of Northern Sotho Grammatical Terms*. Pretoria: Via Afrika

Poulos, G. and L.J. Louwrens. 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.

Prinsloo, D.J. 1992. Lemmatization of Reflexives in Northern Sotho. *Lexikos* 2: 178-191.

Prinsloo, D.J. 2001. The Compilation of Electronic Dictionaries for the African Languages. *Lexikos* 11: 139-159.

Prinsloo, D.J. and G.-M. de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.

Prinsloo, D.J. and R.H. Gouws. 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.

Taljard, Elizabeth. 1999. *Die Kopulatief van Noord-Sotho: 'n Nuwe Perspektief*. Unpublished D.Litt. Thesis. Pretoria: University of Pretoria.

Van Wyk, E.B. et al. 1992. *Northern Sotho for First-Years*. Pretoria: J.L. van Schaik.

Wiegand, H.E. 1996. Über die Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, A. and V.H. Pedersen (Eds.). 1996. *Symposium on Lexicography VII*: 11-43. Tübingen: Max Niemeyer.

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

MD.	TENSE	ACT.	Common verb	Identifying	Descriptive	Associative
IND.	PRES.	p.	mosadi o reka dipuku	e ba morutiši	o ba bohlale	o ba le mpša
		n.	mosadi ga a reke dipuku	ga e be morutiši	ga a be bohlale	ga a be le mpša
	+Pot.	p.	mosadi a ka reka dipuku	e ka ba morutiši	a ka ba bohlale	a ka ba le mpša
		n.	mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	FUT.	p.	mosadi o tlo/tla reka dipuku	e tlo/tla ba morutiši	o tlo/tla ba bohlale	o tlo/tla ba le mpša
		n.	mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	PAS.	p.	mosadi o rekile dipuku	e bile morutiši	o bile bohlale	o bile le mpša
		n.	mosadi ga se a reka dipuku	ga se ya ba morutiši	ga se a ba bohlale	ga se a ba le mpša
SIT.	PRES.	p.	ge mosadi a reka dipuku	e eba morutiši	a eba bohlale	a eba le mpša
		n.	ge mosadi a sa reke dipuku	e sa be morutiši	a sa be bohlale	a sa be le mpša
	+Pot.	p.	ge mosadi a ka reka dipuku	e ka ba morutiši	a ka ba bohlale	a ka ba le mpša
		n.	ge mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	FUT.	p.	ge mosadi a tlo/tla reka dipuku	e tlo/tla ba morutiši	a tlo/tla ba bohlale	a tlo/tla ba le mpša
		n.	ge mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlale	a ka se be le mpša
	PAS.	p.	ge mosadi a rekile dipuku	e bile morutiši	a bile bohlale	a bile le mpša
		n.	ge mosadi a sa reka dipuku	e sa ba morutiši	a sa ba bohlale	a sa ba le mpša
REL.	PRES.	p.	mosadi yo a rekago dipuku	e bago morutiši	a bago bohlale	a bago le mpša
		n.	mosadi yo a sa rekego dipuku	e sa bego morutiši	a sa bego bohlale	a sa bego le mpša
	+Pot.	p.	mosadi yo a ka rekago dipuku	e ka bago morutiši	a ka bago bohlale	a ka bago le mpša
		n.	mosadi yo a ka se rekego dipuku	e ka se bego morutiši	a ka se bego bohlale	a ka se bego le mpša
	FUT.	p.	mosadi yo a tlo/tla rekago dipuku	e tlo/tla bago morutiši	a tlo/tla bago bohlale	a tlo/tla bago le mpša
		n.	mosadi yo a ka se rekego dipuku	e ka se bego morutiši	a ka se bego bohlale	a ka se bego le mpša
	PAS.	p.	mosadi yo a rekilego dipuku	e bilego morutiši	a bilego bohlale	a bilego le mpša
		n.	mosadi yo a sa rekago dipuku	e sa bago morutiši	a sa bago bohlale	a sa bago le mpša
SUB.		p.	(gore) mosadi a reke dipuku	e be morutiši	a be bohlale	a be le mpša
		n.	(gore) mosadi a se reke dipuku	e se be morutiši	a se be bohlale	a se be le mpša
CON.		p.	mosadi a reka dipuku	ya ba morutiši	a ba bohlale	a ba le mpša
		n.	mosadi a se reke dipuku	ya se be morutiši	a se be bohlale	a se be le mpša
INF.		p.	go reka dipuku	go ba morutiši	go ba bohlale	go ba le mpša
		n.	go se reke dipuku	go se be morutiši	go se be bohlale	go se be le mpša
IMP.		p.	reka dipuku!	eba morutiši!	eba bohlale!	eba le mpša!
		n.	se reke dipuku!	se be morutiši!	se be bohlale!	se be le mpša!
HAB.		p.	mosadi a reke dipuku	e be morutiši	a be bohlale	a be le mpša
		n.	mosadi a se reke dipuku	e se be morutiši	a se be bohlale	a se be le mpša

	Positive	Negative	Positive	Negative	Positive	Negative
Indicative:	**Identifying**		**Descriptive**		**Associative**	
1ps	(Nna) ke morutiši	ga ke morutiši	ke bohlale	ga ke bohlale	ke na le mpša	ga ke na mpša
+prog.	(Nna) ke sa le morutiši	ga ke sa le morutiši	ke sa le bohlale	ga ke sa le bohlale	ke sa na le mpša	ga ke sa na mpša
1pp-2pp -----		-----	-----	-----	-----	-----
1	Monna ke morutiši	ga se morutiši	o bohlale	ga a bohlale	o na le mpša	ga a na mpša
+prog.	Monna e sa le morutiši	ga e sa le morutiši	o sa le bohlale	ga a sa le bohlale	o sa na le mpša	ga a sa na mpša
2-8 -----		-----	-----	-----	-----	-----
9	Kgoši ke morutiši	ga se morutiši	e bohlale	ga e bohlale	e na le mpša	ga e na mpša
+prog.	Kgoši e sa le morutiši	ga e sa le morutiši	e sa le bohlale	ga e sa le bohlale	e sa na le mpša	ga e sa na mpša
10-18 -----		-----	-----	-----	-----	-----
Situative:	**Identifying**		**Descriptive**		**Associative**	
1ps	(Nna) ge ke le morutiši	ke se morutiši	ke le bohlale	ke se bohlale	ke na le mpša	ke se na mpša
+prog.	(Nna) ge ke sa le morutiši	ke se sa le morutiši	ke sa le bohlale	ke se sa le bohlale	ke sa na le mpša	ke se sa na mpša
1pp-2pp -----		-----	-----	-----	-----	-----
1	Ge monna e le morutiši	e se morutiši	a le bohlale	a se bohlale	a na le mpša	a se na mpša
+prog.	Ge monna e sa le morutiši	e se sa le morutiši	a sa le bohlale	a se sa le bohlale	a sa na le mpša	a se sa na mpša
2-4 -----		-----	-----	-----	-----	-----
5	Ge lesogana e le morutiši	e se morutiši	le le bohlale	le se bohlale	le na le mpša	le se na mpša
+prog.	Ge lesogana e sa le morutiši	e se sa le morutiši	le sa le bohlale	le se sa le bohlale	le sa na le mpša	le se sa na mpša
6-18 -----		-----	-----	-----	-----	-----
Relative:	**Identifying**		**Descriptive**		**Associative**	
1ps	(Nna) yo ke lego morutiši	ke sego morutiši	ke lego bohlale	ke sego bohlale	ke nago le mpša	ke se nago le mpša
+prog.	(Nna) yo ke sa lego morutiši	ke se sa lego morutiši	ke sa lego bohlale	ke se sa lego bohlale	ke sa nago le mpša	ke se sa nago mpša
1pp-2pp -----		-----	-----	-----	-----	-----
1	Monna yo e lego morutiši	e sego morutiši	a lego bohlale	a sego bohlale	a nago le mpša	a se nago le mpša
+prog.	Monna yo e sa lego morutiši	e se sa lego morutiši	a sa lego bohlale	a se sa lego bohlale	a sa nago le mpša	a se sa nago mpša
2-6 -----		-----	-----	-----	-----	-----
7	Sehlalefi se e lego morutiši	e sego morutiši	se lego bohlale	se sego bohlale	se nago le mpša	se se nago le mpša
+prog.	Sehlalefi se se sa lego morutiši	e se sa lego morutiši	se sa lego bohlale	se se sa lego bohlale	se sa nago le mpša	se se sa nago mpša
8-18 -----		-----	-----	-----	-----	-----