

A Critical Evaluation of the Paradigm Approach in Sepedi Lemmatisation — The *Groot Noord-Sotho Woordeboek* as a Case in Point*

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

Abstract: This article gives a critical evaluation of the paradigm approach of the *Groot Noord-Sotho Woordeboek* to the lemmatisation of verbs and nouns derived from verbs. The verb stem *-roba* 'break' with its complicated system of derivations will be taken as a case in point. The paradigm presented for *-roba* will be evaluated in terms of structure, occurrence in Sepedi corpora and dictionaries, actual use by mother-tongue speakers, user-friendliness, contextualisation versus decontextualisation in relation to the cross-referencing system and space utilisation. Bringing together, and lexicographically treating all these forms for a single verb surely is a lexicographic achievement. The question, however, is to what extent such an approach is useful in respect of forms likely to be looked up by dictionary users, whether all of these forms actually exist, how user-friendly the approach and presentation is, if comment on semantics is sufficient and consistent and whether such a lumping approach actually saves space in contrast to entering derivations as main lemmas in a splitting approach.

Keywords: LEXICOGRAPHY, SEPEDI, LEXICOGRAPHIC TRADITIONS, LEMMATISATION STRATEGIES, THE VERB STEM *-ROBA*

Opsomming: 'n Kritiese evaluering van die paradigmabenedering tot Sepedi-lemmatisering — Die *Groot Noord-Sotho Woordeboek* as voorbeeld.

Hierdie artikel gee 'n kritiese evaluering van die paradigmabenedering tot die *Groot Noord-Sotho Woordeboek* tot die lemmatisering van werkwoorde en naamwoorde wat van werkwoorde afgelei is. Die werkwoordstam *-roba* 'breek' met sy komplekse sisteem van afleidings word as voorbeeld geneem. Die paradigma wat vir *-roba* aangebied word, sal in terme van struktuur, werklike gebruik deur moedertaalsprekers, voorkoms in Sepedikorpusse, gebruikersvriendelikheid, kontekstualisering versus dekontekstualisering ten opsigte van die kruisverwysingstelsel en ruimtebenutting geëvalueer word. Die byeenbring, en leksikografiese bewerking van al hierdie vorme vir 'n enkele

* This article was the basis of a shortened version presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014.

werkwoord is sonder twyfel 'n leksikografiese prestasie. Die vraag is egter tot watter mate dit nuttig is ten opsigte van vorme wat waarskynlik deur woordeboekgebruikers opgesoek sal word, of al hierdie vorme werklik bestaan, hoe gebruikersvriendelik die benadering en aanbieding is, of semantiese kommentaar voldoende en konsekwent is, en of so 'n saamgevoegde benadering werklik ruimte bespaar in teenstelling met die aanbieding van afleidings as afsonderlike hooflemmas in 'n opgedeelde benadering.

Sleutelwoorde: LEKSIKOGRAFIE, SEPEDI, LEKSIKOGRAFIESE TRADISIES, LEMMATISERINGSTRATEGIEË, DIE WERKWOORDSTAM -ROBA

1. Introduction

The aim of this article is to give a critical evaluation of the paradigm approach to the lemmatisation of verbs, and nouns derived from verbs (deverbatives) in the *Groot Noord-Sotho Woordeboek/Comprehensive Northern Sotho Dictionary/Pukuntšú ya Sesotho sa Leboa* (Ziervogel and Mokgokong 1975). The complicated verb stem *-roba* 'break' will be taken as a case in point. The paradigm presented for *-roba* will be evaluated in terms of (a) structure, (b) real life use as reflected by occurrence in Sepedi (also referred to as Northern Sotho or Sesotho sa Leboa) corpora and dictionaries, as well as actual use by mother-tongue speakers, (c) user-friendliness of the paradigm approach in respect of lumping versus splitting, (d) contextualisation versus decontextualisation in relation to the cross-referencing system and (e) space utilisation. As a prerequisite, a brief theoretical background on lemmatisation approaches, traditions and strategies will be presented with special emphasis on the paradigm approach.

2. A brief theoretical background on lemmatisation approaches, traditions and strategies

Prinsloo (2009) distinguishes five aspects of importance for lemmatisation in African languages given in table 1. These aspects are discussed in more detail for nouns in Prinsloo and De Schryver (1999) and for verbs in Prinsloo (1994).

The Sepedi lexicographer has to deal with all of the aspects and subcategories in A to E in table 1. As far as A is concerned the *traditional* way to compile dictionaries, especially in the pre-corpus era was for the lexicographer to select lemmas on intuition/introspection. The advent of corpora enabled lexicographers to use *frequency* counts of words in a corpus as a major criterion for the inclusion or omission of lemmas. The *paradigm* approach could be described as an attempt to physically include all derivations of especially verbs in the dictionary. This is the approach in the *Groot Noord-Sotho Woordeboek* (GNSW) which will be outlined and evaluated in detail in this article. The *rule-orientated* approach stands in contrast to the paradigm approach in the sense that the urge is to reduce the number of lemmas presented for a specific paradigm to the absolute minimum. So, for example, only singular forms of nouns are lem-

matised and only stem forms (i.e. root + *-a*), without any extensions, of verbs are included as lemmas.

Table 2: Lemmatisation approaches, lexicographic traditions and lemmatisation strategies in Bantu languages

	A Lemmatisation approaches	B Orthography of the language	C Lexicographic traditions	D Lemmatisation strategies : verbs	E Lemmatisation strategies : nouns
1	Traditional	Disjunctive	Stem tradition	Strict stem	Strict stem
2	Rule-orientated	Conjunctive	Word tradition	Left-expanded stem	Left-expanded stem
3	Paradigm				Singular only
4	Frequency				Singular and plural
5					First and 3rd letter

Prinsloo (2009: 153)

The *orthography* of the language (B) plays an important role in the choice of the lexicographic tradition (C). A *disjunctively* written language such as Sepedi will e.g. write the phrase "I love you" as four orthographic words, i.e. *ke a go rata* and a conjunctively written language such as isiZulu as a single orthographic word, i.e. *ngiyakuthanda*. Both have exactly the same structure, i.e. subject concord + present tense marker + object concord + verb stem.

Disjunctively written languages such as Sepedi favour the *word tradition*, i.e. lemmatising nouns with their prefixes while the *stem* tradition is mostly chosen by lexicographers for conjunctively written languages. So, for example, *monna* 'man' will be lemmatised according to the word tradition on its full form under M while on its stem *-nna* under N in a stem dictionary such as GNSW.

3. The paradigm approach

In lexicography lumping versus splitting are mostly used in the literature in relation to the presentation of different senses of a word. In this article lumping versus splitting will be used in a grammatical sense i.e. grouping different derivations of a specific verbal stem under a single lemma or presenting each of the derivations as main lemmas. This brings the so-called paradigm approach following stem lemmatisation of GNSW in contrast with the traditional word lemmatisation approach of the *Pukuntšu* dictionaries (PUKU1 and PUKU2). GNSW lemmatises a verb under its stem form and all derivations of the verb including deverbatives will then be lumped together. The different forms will also be lemmatised *separately* as untreated lemmas with an implicit cross-refer-

ence to the main verb stem.

In the paradigm approach in GNSW the basic micro-architecture of an article is designed in terms of a modular layout aimed at bringing together all derivations of e.g. a verb stem. So, for example, the article of the lemma ROBA in GNSW consists of 32 modules distinguished on the basis of derived forms by suffixes and combinations of suffixes.

1. ROBA (-rôba, -rôbilê, -rôbja, -rôbilwê)
2. ROBAGANA (-rôbagana, -rôbagane, -rôbaganwa, -rôbaganwe)
3. ROBAGANELA (-rôbagenêla, -rôbaganêtše, -rôbaganêlwa, -rôbaganêtšwe)
4. ROBAGANTŠHA (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê)
5. ROBAGANTŠHETŠA (-rôbagantšhêtsa, -rôbagantšhêditšê, -rôbagantšhêtswa, -rôbagantšhêditšwê)
6. ROBAGANTŠHETŠANA (-rôbagantšhêtsana, -rôbagantšhêtsane, -rôbagantšhêtsanwa, -rôbagantšhêtsanwe)
7. ROBAGANYA (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê)
8. ROBAGANYETŠA (-rôbaganyêtsa, -rôbaganyêditšê, -rôbaganyêtswa, -rôbaganyêditšwê)
9. ROBAGANYETŠANA (-rôbaganyêtsana, -rôbaganyêtsane, -rôbaganyêtsanwa, -rôbaganyêtsanwe)
10. ROBAKA (-rôbaka, -rôbakilê, -rôbakwa, -rôbakilwê)
11. ROBAKANA (-rôbakana, -rôbakane, -rôbakanwa, -rôbakanwe)
12. ROBAKANTŠHA (-rôbakantšha, -rôbakantšhitšê, -rôbakantšhwa, -rôbakantšhitšwê)
13. ROBAKANTŠHETŠA (-rôbakantšhêtsa, -rôbakantšhêditšê, -rôbakantšhêtswa, -rôbakantšhêditšwê)
14. ROBAKANTŠHETŠANA (-rôbakantšhêtsana, -rôbakantšhêtsane, -rôbakantšhêtsanwa, -rôbakantšhêtsanwe)
15. ROBAKANYA (-rôbakanya, -rôbakantšê, -rôbakanywa, -rôbakantšwê) ¹
16. ROBAKANYETŠA (-rôbakanyêtsa, -rôbakanyêditšê, -rôbakanyêtswa, -rôbakanyêditšwê)
17. ROBAKANYETŠANA (-rôbakanyêtsana, -rôbakanyêtsane, -rôbakanyêtsanwa, -rôbakanyêtsanwe)
18. ROBEGA (-robêga, -robêgilê)
19. ROBEGETŠA (-rôbêgêtsa, -rôbêgêditšê, -rôbêgêtswa, -rôbêgêditšwê)
20. ROBELA (-rôbêla, -rôbêtše, -rôbêlwa, -rôbêtšwe)
21. ROBELANA (-rôbêlana, -rôbêlane, -rôbêlanwa, -rôbêlanwe)
22. ROBESÊTŠA (-rôbêsêtsa, -rôbêsêditšê, -rôbêsêtswa, -rôbêsêditšwê)
23. ROBIŠA (-rôbiša, -rôbišitšê, -rôbišwa, -rôbišitšwê)
24. ROBIŠANA (-rôbišana, -rôbišane, -rôbišanwa, -rôbišanwe)
25. ROBOKA (-rôbôka, -rôbôkilê, -rôbôkwa, -rôbôkilwê)
26. ROBOKANA (-rôbôkana, -rôbôkane, -rôbôkanwa, -rôbôkanwe)
27. ROBOKANTŠHA (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê)
28. ROBOKANTŠHETŠA (-rôbôkantšhêtsa, -rôbôkantšhêditšê, -rôbôkantšhêtswa, -rôbôkantšhêditšwê)
29. ROBOKANTŠHETŠANA (-rôbôkantšhêtsana, -rôbôkantšhêtsane, -rôbôkantšhêtsanwa, -rôbôkantšhêtsanwe)
30. ROBOKANYA (-rôbôkanya, -rôbôkantšê, -rôbôkanywa, -rôbôkantšwê)
31. ROBOKANYETŠA (-rôbôkanyêtsa, -rôbôkanyêditšê, -rôbôkanyêtswa, -rôbôkanyêditšwê)
32. ROBOKANYETŠANA (-rôbôkanyêtsana, -rôbôkanyêtsane, -rôbôkanyêtsanwa, -rôbôkanyêtsanwe)

The full article of ROBA in GNSW is given in the appendix.

In module 1 the lemma is the basic stem (root *-rob-* plus the terminative *-a*) without any suffixes. The stem is repeated followed by the perfect, passive and perfect plus passive forms. Prinsloo and De Schryver (1999) refer to the latter three as "standard modifications". Modules 2-32 give the root plus a suffix or combination of suffixes with the standard modifications. For example, ROBELANA in Module 21 consists of the root plus applicative suffix (*-el-*) plus reciprocal

(-an-) plus the verbal ending followed by the perfect form *-rôbêlane*, passive *-rôbêlanwa*, and perfect plus passive *-rôbêlanwe*.

The module layout includes comments on form and on semantics, mainly giving translation equivalents in Afrikaans and English as well as examples of usage and deverbatives.

Consider the treatment of ROBAGANA in Module 2:

2. ROBAGANA (-rôbagana, -rôbagane, -rôbaganwa, -rôbaganwe) ass.; in stukke gebreek w. // b. broken into pieces; morobagani, ba- (morôbagani) pers. dev.; morobagano, me- (morôbaganô) man. dev.

Treatment in this module includes four nouns which are derived from *-robagana* i.e. *morobagani*, *barobagani*, *morobagano* and *merobagano*.

The entire article of ROBA consists of 265 nominal and verbal forms of *roba*:

barobagani, barobagantšhetšani, barobagantšhetši, barobagantšhetšo, barobagantšhi, barobaganyetšani, barobaganyetši, barobaganyi, barobakantšhetšani, barobakantšhetši, barobakantšhi, barobakanyetšani, barobakanyetši, barobakanyi, barobaki, barobedi, barobegetši, barobegi, barobelani, barobeseditši, barobi, barobišani, barobiši, barobja, barobokantšhetšani, barobokantšhetši, barobokantšhi, barobokanyetšani, barobokanyetši, barobokanyi, baroboki, barokantšhetšani, borobaganelo, diroba, dirobaganedi, dirobakani, dirobakantšhwa, dirobakanywa, dirobja, dirobokani, dirobokantšhwa, dirobokantšwa, dirobokanywa, dithobego, ithoba, ithobela, merobaganelo, merobagano, merobagantšhetšano, merobagantšho, merobaganyetšano, merobaganyetšo, merobaganyo, merobakano, merobakantšhetšano, merobakantšhetšo, merobakantšho, merobakanyetšano, merobakanyetšo, merobakanyo, merobako, merobegetšo, merobego, merobelano, merobelo, merobeseditšo, merobišano, merobišo, merobokano, merobokantšhetšano, merobokantšhetšo, merobokantšho, merobokanyetšano, merobokanyetšo, merobokanyo, meroboko, morobaganelo, morobagani, morobagano, morobagantšhetšano, morobagantšhetši, morobagantšhetšo, morobagantšhi, morobagantšho, morobaganyetšano, morobaganyetši, morobaganyetšo, morobaganyi, morobaganyo, morobakano, morobakantšhetšano, morobakantšhetši, morobakantšhetšo, morobakantšhi, morobakantšho, morobakanyetšano, morobakanyetši, morobakanyetšo, morobakanyi, morobakanyo, morobaki, morobako, morobedi, morobegetši, morobegetšo, morobegi, morobego, merobelano, merobelo, morobesetši, morobesetšo, morobi, morobišano, morobiši, morobišo, morobja, morobokano, morobokantšhetšano, morobokantšhetši, morobokantšhetšo, morobokantšhi, morobokantšho, morobokanyetšano, morobokanyetši, morobokanyetšo, morobokanyi, morobokanyo, moroboki, moroboko, roba, robagana, robagane, robaganela, robaganelwa, robaganetše, robaganetšwe, robagantše, robagantšha, robagantšheditše, robagantšheditšwe, robagantšhetša, robagantšhetšana, robagantšhetšane, robagantšhetšanwa, robagantšhetšanwe, robagantšhetšwa, robagantšhitšwe, robagantšhitše, robagantšhwa, robagantšwe, robaganwa, robaganwe, robaganya, robaganyeditše, robaganyeditšwe, robaganyetša, robaganyetšana, robaganyetšanwa, robaganyetšanwe, robaganyetšwa, robaganywa, robagenela, robaka, robakana, robakane, robakantše, robakantšha, robakantšheditše, robakantšheditšwe, robakantšhetša, robakantšhetšana, robakantšhetšane, robakantšhetšanwa, robakantšhetšanwe, robakantšhetšwa, robakantšhitše, robakantšhitšwe, robakantšhwa, robakantšwe, robakanwa, robakanwe, robakanya, robakanyeditše, robakanyedi-

tšwe, robakanyetša, robakanyetšana, robakanyetšane, robakanyetšanwa, robakanyetšanwe, robakanyetšwa, robakanywa, robakile, robakilwe, robakwa, robega, robegeditše, robegeditšwe, robegetša, robegetšwa, robegile, robela, robelana, robelane, robelanwa, robelanwe, robelwa, robeseditše, robeseditšwe, robesetša, robesetšwa, robetše, robetšwe, robile, robilwe, robiša, robišana, robišane, robišanwa, robišanwe, robišitše, robišitšwe, robišwa, robja, roboka, robokana, robokane, robokantše, robokantšha, robokantšheditše, robokantšheditšwe, robokantšhetša, robokantšhetšana, robokantšhetšane, robokantšhetšanwa, robokantšhetšanwe, robokantšhetšwa, robokantšhitše, robokantšhitšwe, robokantšhwa, robokantšwe, robokanwa, robokanwe, robokanya, robokanyeditše, robokanyeditšwe, robokanyetša, robokanyetšana, robokanyetšane, robokanyetšanwa, robokanyetšanwe, robokanyetšwa, robokanywa, robokile, robokilwe, robokwa, seroba, serobagamedi, serobakantšhwa, serobakanywa, serobja, serobokantšhwa, serobokantšwa, serobokanywa, thobego

Certain omissions in the paradigm completion were however noticed:

1. Module 6: *morobagantšhetšani* (only the plural *barobagantšhetšani* is given)
2. Module 9: *morobaganyetšani* (only the plural *barobaganyetšani* is given)
3. Module 11: *serobakani* to be added
4. Module 14: *morobakantšhetšani* to be added
5. Module 17: *morobakanyetšani* to be added
6. Module 21: *morobelani* to be added
7. Module 24: *morobišani* to be added
8. Module 26: *serobokani* to be added
9. Module 29: *morobokantšhetšani* to be added
10. Module 31: *morobokanyetšani* to be added

Accounting for, and to lexicographically treat all these forms for a single verb surely is a lexicographic achievement and succeeds in bringing together all derived forms of *-roba* in a single article. The question, however, is to what extent is it useful to the dictionary user in respect of (a) how many of these forms are likely to be looked-up by the users, (b) do all of these forms actually exist, (c) how user friendly is the approach/presentation, (d) is comment on semantics sufficient and consistent, (e) does such a lumping approach really save space in contrast to giving derivations as main lemmas? These issues will be briefly addressed in the following paragraphs.

4. Usage versus presumed usage and existence of words in the language

Ziervogel (1965) says that the basic meaning of a word lies in its root, e.g. for *-roba* in *-ROB-* and by adding a series of pre- and suffixes the root can obtain a variety of senses/meanings which have to do with the basic meaning.

Hoewel die wortel selde 'n selfstandige gebruik in die taal het, dui hy wel altyd 'n begrip aan ... Deur die toevoeging van 'n reeks voor- en/of agtervoegsels kan die wortel 'n verskeidenheid begrippe kry wat met die basiese betekenis te doen het. Die voor- en agtervoegsel het wel 'n betekenisinhoud maar nie noodwendig

'n ekwivalente betekenis in Afrikaans nie. (Ziervogel 1965: 47)

Thus for *-roba* it means that *-ROB-* is the ideal point of departure for building a paradigm of derivations by means of affixes in order to reflect the variety of different meanings. In Ziervogel's view this also means that stem lemmatisation is the ideal lemmatisation strategy, e.g. for a systematised representation of word formation.

Ziervogel (1965: 45) claims that: Entries must be arranged under their stems with cross-references where necessary. This method is scientifically sound. A systematised survey of word formation in the languages is given; it shows word and lexical relationship and prevents repetition. (Van Wyk 1995: 85)

Van Wyk's severe and detailed criticism of the GNSW's approach is focused on the deficiencies of employing a stem lemmatisation strategy instead of a word lemmatisation strategy for a disjunctively written language such as Sepedi. He rejects Ziervogel's claims that stem lemmatisation is scientifically more sound than word lemmatisation, or that it prevents repetition. Of special importance to this article is Van Wyk's statement that it is the task of a grammar book and not a dictionary to give a systematic survey of word formation.

In this article the focus is on presumed aspects of user-unfriendliness in relation to problematic aspects of the presentation and especially the selection of lemmas.

As for the selection of lemmas Ziervogel acknowledges inclusion versus omission of lemmas as being important and problematic and suggests that the written language should be the point of departure for an effort to include all written forms.

Die probleem van wat in 'n woordeboek opgeneem moet word, is nogal van belang ... Ek glo 'n mens moet begin by die geskrewe taal en alle geskrewe woorde opneem. (Ziervogel 1965: 50)

He continues that for a comprehensive dictionary it is important to document the derivations but that the question is to what extent reduplications (repetition of a word with added affixes) and reflexives (i.e. do something to oneself) should be included.

Vir 'n redelik volledige woordeboek is dit tog van belang om aan te teken watter afleidings gemaak kan word. Die vraag is natuurlik vir hoever afleidings soos reduplikasie en refleksiewe opgeneem moet word. (Ziervogel 1965: 52)

This brings us to the core of the issue, i.e. what is the duty of the lexicographer in terms of what to include and what to omit from the dictionary.

Gouws and Prinsloo (2005: 86) state that in a general dictionary (with a text reception function) "the user should be able to find the words encountered in the day to day general language usage ...". The lexicographer should include a selection from the lexical stock of the language. It should not be limited to

words found in written texts but also include words from the spoken language. The question, however is whether the lexicographer should invent words in the sense of e.g. derivations that are theoretically possible, i.e., that could possibly be derived in terms of the grammatical rules of the language? In the case of *-roba* one would have to ask whether all of the derivations given in GNSW are really in use in the language and what the likelihood would be for them to actually be looked up by the target users of the dictionary.

In order to determine the likelihood of the different derivations of *-roba* to be looked up as well as their actual use in the language, (a) their actual occurrence in the Pretoria Sepedi Corpus (PSC) was determined, (b) the treatment of *-roba* in Sepedi dictionaries was studied and (c) two mother tongue speakers of Sepedi were requested to indicate which of these forms they know.

As for the PSC, only 35 (13%) of the 265 words occurred in the corpus, frequencies given in brackets.

robetše (642)*, roba (229), robega (119), robile (81), robegile (63), robja (60), seroba (40), diroba (28), robetšwe (25)*, robaganya (24), robela (14), thobego (13), robilwe (13), ithoba (6), roboka (5), robagana (3), robagantšwe (3), robesetša (3), robaka (2), robelwa (2), robokanya (2), robokile (2), robiša (2), robaganywa (2), robaganyeditše (1), robeseditšwe (1), robagantšha (1), robaganyetša (1), robagantše (1), robagane (1), ithobela (1), ithobela (1), robakanya (1), robagantšha (1), robokwa (1)

*Frequency count includes homonyms: *thoko* 'side', *robetše* 'sleep, slept', *robetšwe* 'was slept'

Only ten derivations of *-roba* are lemmatised and treated in seven randomly selected Sepedi dictionaries as given in table 2.

Table 2: Derivations of *-roba* in Sepedi dictionaries

POP=*Popular Northern Sotho Dictionary*

NEND=*New English–Northern Sotho Dictionary*

PUKU1=*Pukuntšu Dictionary* (1983)

SLEPD=*Sesotho sa Leboa/English Pukuntšu Dictionary*

ONSD=*Oxford Bilingual School Dictionary: Northern Sotho and English*

PUKU2=*Pukuntšu Dictionary* (1989)

PTLH=*Pukuntšutlhaloši ya Sesotho sa Leboa*

	POP	NEND	PUKU1	SLEPD	ONSD	PUKU2	PTLH
roba	√	√	√	√	√	√	√
robega	√	√	√	√	√	√	√
robaganya	√	√	√			√	
robja	√	√	√	√			
robile	√			√	√		
robakanya			√			√	
robaka		√					
robela		√					
robegile	√						
robagantše	√						

From the 265 forms of *-roba* the mother tongue speakers only identified 54 (20%) and 77 (29%) respectively as known to them. Forms which they both know are indicated in bold.

barobagani	merobego	robaganela	robelana
barobagantšhetšani	merobelano	robaganelwa	robelane
barobagantšhi	merobelo	robaganetše	robelanwe
barobaganyi	merobišano	robagantše	robelwa
barobakanyi	merobišo	robagantšha	robesešša
barobaki	merobokantšho	robagantšheditše	robetše
barobegi	merobokanyo	robagantšheditšwe	robetšwe
barobelani	meroboko	robagantšhetša	robile
barobi	morobagano	robagantšhetšana	robilwe
barobišani	morobagantšhi	robagantšhetšane	robišana
barobiši	morobaganyi	robagantšhitše	robišana
barobja	morobakantšhi	robagantšwe	robišane
barobokantšhetšani	morobakantšho	robaganya	robišitše
barobokantšhi	morobakanyi	robaka	robišitšwe
barobokanyi	morobegetši	robakana	robišwa
borobaganelo	morobegi	robakane	robja
diroba	morobego	robakantše	robokantše
dirobakantšhwa	morobelano	robakantšha	robokantšha
dirobja	merobelo	robakantšhitše	robokanya
dirobokanywa	morobeseššo	robakantšhitšwe	seroba
dithobego	morobi	robakantšwe	serobagamedi
ithoba	morobišano	robakanya	serobakantšhwa
ithobela	morobiši	robakile	serobakanywa
merobagantšho	morobišo	robakwa	serobja
merobaganyo	roba	robega	serobokanywa
merobakano	robagana	robegile	thobego
merobegetšo	robagane	robela	

The very limited number of derivations of *-roba* that (a) occur in the Sepedi Corpus, (b) are lemmatised and treated in Sepedi dictionaries and (c) were recognised by the mother-tongue speakers, suggest that the actual extent of use of derivations of *-roba* is much smaller than suggested by GNSW.

The question is thus whether most of these words given by GNSW for *-roba* actually *exist* in the language or whether the compilers mainly focused on categorically completing morphological/grammatical paradigms? Does the task of the lexicographer go beyond the lemmatisation and treatment of words actually used in a language to those that can *potentially exist* because they are e.g. possible through morphological reduplication rules or might come into use as new inventions? As for the latter it is true that one could never claim that a specific reduplication/combination will never come into use. For example, the author once objected to the inclusion of *deurgans* as a noun 'door goose' in a spelling checker lexicon for Afrikaans only to find that goose door stops exist in English, cf. images at http://vintagepatterns.wikia.com/wiki/Patch_Press_379.

It is therefore possible that *deurgans* can become a common word in Afrikaans. As a second example, it would have been unlikely a few years ago to include a nominal *like(s)* as a lemma in a dictionary but today it is commonly used on websites, e.g. 34 likes.

The English language is notoriously fast in adapting to the changing world. New words enter English from every area of life where they represent and describe the changes and developments that take place from day to day. Here are some words and expressions that have been coined in recent years. Some can be found in official dictionaries; others may never make their way there, but new words will continue to appear as the English language adapts to innovations and trends:

<http://www.learn-english-today.com/new-words/new-words-in-english.html>

Consider the following examples stated: *breadcrumbing* (a navigation technique which helps users by displaying a list of links to the pages they have visited when exploring a website), *copyleft* (opposite of copyright ... allows freedom of use for all), *crowdfunding* (raising money for a project by getting a large number of people to make a small financial contribution), *cyberbully* (a person who uses the Internet to harm another person), *textspeak* (language used in text messages), etc. <http://www.learn-english-today.com/new-words/new-words-in-english.html>

Be that as it may, it is not the task of the lexicographer to provide for possible future use/existence of words in a dictionary. Lemma selection should not be influenced by words that the lexicographer would like to see as part or to become part of the language. In terms of Wells (1973), Hartmann (1983) and Gove (1961) the duty of the lexicographer is to record language and to include words which are actually used by the speakers of the language in the dictionary.

The responsibility of a dictionary is to record the language, not set its style ... The only area in which the truth may be found is actual usage. In fine, the function of a dictionary is to reflect the facts of usage as they exist. A dictionary neither permits nor prevents. (Wells 1973: 84)

Lexemes become entries in a dictionary only when they are socialised, that is when they are used by a sufficient number of speakers. (Hartmann 1983: 71)

The basic aim is nothing less than coverage of the current vocabulary of standard spoken and written English. (Gove 1961: 4a)

The lexicographer's attention should be limited to the treatment of existing words in the lexicon especially given the fact that it is hardly possible to cover the existing words, even in a comprehensive multivolume dictionary. Currently available corpora which reflect actual use of words and indicating their frequency of occurrence are the ideal sources to guide the lexicographer in the selection of lemmas.

5. User-friendliness of the paradigm approach

Gouws and Prinsloo (2005: 39) emphasize the importance of the user-perspective:

The user-perspective, so prevalent in modern-day metalexigraphy, 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. Good dictionaries are 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.

Bothma and Prinsloo (2013) emphasize that the user may not want to read or browse through a long article with much irrelevant information in terms of his/her specific information need at a given time. In most cases (s)he only requires the information needed to solve the current information need. The lexicographer should therefore guard against excessive offering of information and rather guide the user more directly to the required information. Haas' remark of five decades ago still holds true:

A good dictionary is one in which you can find the information you are looking for — preferably in the very first place you look. (Haas 1962: 48)

Consider in this regard Prinsloo et al. (2011) where users are guided through decision trees directly to the required information.

An approach to lump information together in long dictionary articles as for ROBA in GNSW runs against the desire to quickly and directly find the information that the user is looking for at a given time.

5.1 Evaluation of GNSW in terms of user-friendliness

The GNSW is generally regarded as user-unfriendly. Prinsloo and De Schryver (1999: 258) state that the user-perspective was not seriously considered in the compilation. Their main criticism in this regard is against the use of phonemic sorting on lemmas, and stem lemmatisation. As for the sorting order the compilers of GNSW deviate from an ordinary alphabetical sorting of the entries and utilize a phonemic one, namely: A, B, BJ, D, E, F, FS, FŠ, G, H, HL, I, J, K, KG, KH, L, M, N, NG, NX, NY, O, P, PH, etc., because this is in their opinion 'more scientific'. To the user it is nothing more than sheer frustration to eventually find, for example, a word commencing on *bj* alphabetically after *bu* in the dictionary, (Prinsloo and De Schryver 1999: 261).

The layout of the complex article of *-roba* as given in the appendix is user-unfriendly in many ways. First it is very long. Secondly, although the derivations are alphabetically ordered as sublemmas, they are presented in a run-on layout which makes it difficult to detect them as the starting point for most of the 32 modules. Thirdly the use of capital letters to mark them is compromised by the use of the same convention to indicate the derivation from which a specific sublemma was derived. Consider Modules 3-9 in this regard.

3. ROBAGANELA (-rôbagenêla, -rôbaganêtše, -rôbaganêlwa, -rôbaganêtšwe)
4. ROBAGANTŠHA (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê)
5. ROBAGANTŠHETŠA (-rôbagantšhêtsa, -rôbagantšhêditšê, -rôbagantšhêtswa, -rôbagantšhêditšwê)
6. ROBAGANTŠHETŠANA (-rôbagantšhêtsana, -rôbagantšhêtsane, -rôbagantšhêtsanwa, -rôbagantšhêtsanwe)
7. ROBAGANYA (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê)
8. ROBAGANYETŠA (-rôbaganyêtsa, -rôbaganyêditšê, -rôbaganyêtswa, -rôbaganyêditšwê)
9. ROBAGANYETŠANA (-rôbaganyêtsana, -rôbaganyêtsane, -rôbaganyêtsanwa, -rôbaganyêtsanwe)

For each of the starting points for Modules 3-9, it is indicated by means of the symbol "<" that they are derived from *robagana*, i.e. < ROBAGANYA:

(morôbaganô) man. dev.; **ROBAGANELA** (-rôbagenêla, -rôbaganêtše, -rôbaganêlwa, -rôbaganêtšwe) appl. < **ROBAGANA**; **borobaganelo** (borôbaganêlô) lo. dev.; **morobaganelo**, **me-** (morôbaganêlô) man. dev.; **serobaganeli**, **di-** (serôbaganêdi) impers. dev.; **ROBAGANTŠHA** (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê) caus. < **ROBAGANA**; in stukke breek // break into pieces; **morobagantšhi**, **ba-** (morôbaganantšhi) pers. dev.; **morobagantšho**, **me-** (morôbaganantšhò) man. dev.; **ROBAGANTŠHETŠA** (-rôbagantšhêtsa, -rôbagantšhêditšê, -rôbagantšhêtswa, -rôbagantšhêditšwê) caus. appl. < **ROBAGANA**; **morobagantšhetši**, **ba-** (morôbaganantšhetši) pers. dev.; **morobagantšhetšo**, **me-** (morôbaganantšhetšò) man. dev.; **ROBAGANTŠHETŠANA** (-rôbagantšhêtsana, -rôbagantšhêtsane, -rôbagantšhêtsanwa, -rôbagantšhêtsanwe) caus. appl. rec. < **ROBAGANA**; **barobagantšhetšani** (barôbaganantšhetšani) pers. dev.; **morobagantšhetšano**, **me-** (morôbaganantšhetšanò) man. dev.; **ROBAGANYA** (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê) caus. < **ROBAGANA**; in stukke breek // break into pieces; **morobaganyi**, **ba-** (morôbagananyi) pers. dev.; **morobaganyo**, **me-** (morôbagananyò) man. dev.; **ROBAGANYETŠA** (-rôbaganyêtsa, -rôbaganyêditšê, -rôbaganyêtswa, -rôbaganyêditšwê) caus. appl. < **ROBAGANA**; **morobaganyetši**, **ba-** (morôbagananyetši) pers. dev.; **morobaganyetšo**, **me-** (morôbagananyetšò) man. dev.; **ROBAGANYETŠANA** (-rôbaganyêtsana, -rôbaganyêtsane, -rôbaganyêtsanwa, -rôbaganyêtsanwe) caus. appl. rec. < **ROBAGANA**;

This way of indicating the source of derivation in a run-on layout thus obscures the capitalised starting point of the modules making it more difficult for the user to find the sublemma. Starting each of the 23 modules on a new line would have substantially increased user-friendliness of the layout.

5.2 Inadequate comment on semantics

The predicament of the user however does not end with the difficulty of locating the specific derivation for which (s)he wants to find the meaning. In most cases (s)he will find the specific sublemma with its presumed standard modifications neatly spelled out but without any comments on semantics. The use of actual comment on semantics in the article of ROBA is very limited, especially in relation to the length of the article. So, for example, no comment on semantics is given for the entire stretch of modules 24-32, i.e. ROBIŠANA to ROBOKANYE-TŠANA:

me- (morôbišô) man. dev.; *ROBIŠANA* (-rôbišana, -rôbišane, -rôbišanwa, -rôbišanwe) caus. rec.; *barobišani* (barôbišani) pers. dev.; *morobišano*, me- (morôbišanô) man. dev.; *ROBOKA* (-rôbôka, -rôbôkilê, -rôbôkwa, -rôbôkilwê) it., v. *ROBAKA*; *moroboki*, ba- (morôbôki) pers. dev.; *moroboko*, me- (morôbôkô) man. dev.; *ROBOKANA* (-rôbôkana, -rôbôkane, -rôbôkanwa, -rôbôkanwe) rec. < *ROBOKA*; v. *ROBAKANA*; *dirobokani* (dirôbôkani) pers. dev.; *morobokano*, me- (morôbôkanô) man. dev.; *ROBOKANTŠHA* (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < *ROBOKANA*; (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < *ROBOKANA*; v. *ROBAKANTŠHA*; *morobokantšhi*, ba- (morôbôkantšhi) pers. dev.; *morobokantšho*, me- (morôbôkantšhó) man. dev.; *serobokantšhwá*, di- (serôbôkantšhwa) impers. pass. dev.; *ROBOKANTŠHETŠA* (-rôbôkantšhêtša, -rôbôkantšhêditšê, -rôbôkantšhêtšwa, -rôbôkantšhêditšwê) caus. appl. < *ROBOKANA*; *morobokantšhetši*, ba- (morôbôkantšhêtši) pers. dev.; *morobokantšhetšo*, me- (morôbôkantšhêtšó) man. dev.; *ROBOKANTŠHETŠANA* (-rôbôkantšhêtšana, -rôbôkantšhêtšane, -rôbôkantšhêtšanwa, -rôbôkantšhêtšanwe) appl. rec. < *ROBOKANTŠHA*; *barobokantšhetšani* (barobokantšhêtšani) pers. dev.; *morobokantšhetšano*, me- (morôbôkantšhêtšanô) man. dev.; *ROBOKA-*

NYA (-rôbôkanya, -rôbôkantšê, -rôbôkanywa, -rôbôkantšwê) caus. < **ROBOKANA**; v. **ROBAKANYA**; morobokanyi, ba- (morôbôkanyi) pers. dev.; morobokanyo, me- (morôbôkanyô) man. dev.; serobokanywá, di- (serôbôkanywa) impers. pass. dev.; **ROBOKANYETŠA** (-rôbôkanyêtša, -rôbôkanyêditšê, -rôbôkanyêtšwa, -rôbôkanyêditšwê) appl. < **ROBAKANYA**; morobokanyetši, ba- (morôbôkanyêtši) pers. dev.; morobokanyetšo, me- (morôbôkanyêtšô) man. dev.; **ROBOKANYETŠANA** (-rôbôkanyêtšana, -rôbôkanyêtšane, -rôbôkanyêtšanwa, -rôbôkanyêtšanwe) appl. rec. < **ROBOKANYA**; barobokanyetšani (barôbôkanyêtšani) pers. dev.; morobokanyetšano, me- (morôbôkanyêtšanô) man. dev.

This reflects a serious imbalance between comment on form versus comment on semantics which is detrimental to the main reason for looking up words in a dictionary, i.e. to find its meaning. Gouws and Prinsloo (2005: 48) refer to the "main assignment" of linguistic dictionaries "i.e. to give an explanation of the meaning of the lemma in monolingual dictionaries and to provide target language translation equivalents for a source language lemma in bilingual and multilingual dictionaries". It could be argued that the compilers of GNSW were so obsessed to include all possible derived forms that comment on semantics was neglected. Prinsloo and De Schryver (1999: 261) call it an 'enter-them-all-syndrome'.

In the article of *phêfa* the compilers apparently concentrated so hard on completing the modular paradigms that they 'forgot' to give any translation equivalents in Afrikaans and English for the entire article.

PHÉFA (-phêfa, -phêfilê, -phêfša, -phêfilwê) (Pb.) v. **PHEMA**; mophêfi, ba- (mophêfi) pers. dev.; phêfo, (n-)/di- (phêfô) man. dev.; v. **phemo**; **PHÉFANA** (-phêfana, -phêfane, -phêfanwa, -phêfanwe) rec.; v. **phemana**; baphêfani (baphêfani) pers. dev.; phêfáno, (n-)/di- (phêfanô) man. dev.; **PHÉFÉGA** (-phêfêga, -phêfêgilê) neutr.; v. **phemega**; **PHÉFÉLA** (-phêfêla, -phêfêtše, -phêfêlwa, -phêfêtšwe) appl.; mophêfédi, ba- (mophêfêdi) pers. dev.; phêfélo, (n-)/di- (phêfêlô) man. dev.; **PHÉFÉLANA** (-phêfêlana, -phêfêlane, -phêfêlanwa, -phêfêlanwe) appl. rec.; baphêfélani (baphêfélani) pers. dev.; phêfélano, (n-)/di- (phêfêlanô) man. dev.; **PHÉFÍŠA** (-phêfiša, -phêfišitšê, -phêfišwa, -phêfišitšwê) caus.; mophêfiši, ba- (mophêfiši) pers. dev.; phêfišo, (n-)/di- (phêfišô) man.

dev.; PHÉFÍŠANA (-phêfišana, -phêfišane,
-phêfišanwa, -phêfišanwe) caus. rec.; ba-
phêfišani (baphêfišani) pers. dev.; phêfišano,
(n-)/di- (phêfišanô) man. dev.

5.3 Efficiency of the medio-structure

The lumping approach in GNSW also reduces the effectiveness of the medio-structure (system of cross-referencing) which is crucial in a lumping approach i.e. to guide the users from a reference position outside the article where the derivation was lemmatised in the alphabetical stretch, to the reference address inside the main article where the derivation in question is treated. Gouws and Prinsloo (2005: 181) state that one of the important functions of the medio-structure of a dictionary is to combat the decontextualisation brought about by alphabetical ordering. In a simplified way one could say that alphabetical ordering of lemmas in a dictionary has the detrimental effect of decontextualizing words that belong together. By way of comparison, words indicating *fruit* such as *apple*, *pear*, *banana* and *orange* belong together but are scattered over the dictionary as they belong to different alphabetical stretches. Dictionaries consequently attempt to combat such decontextualisation e.g. by means of a colour plate for fruit given in the back matter or another reference address in the dictionary. In principle the same holds true for what could be termed as *grammatical decontextualisation* in the sense of different derivations of e.g. *-roba* such as *ithoba*, *seroba* and *diroba* that will alphabetically be scattered over the dictionary. For *-roba* this would mean lemmatisation of all derivations in their appropriate alphabetical positions (reference positions) thus decontextualised, to be contextualised by cross-reference to the main lemma *-roba* and its treatment.

The lumping approach surely brings all these derivations together so that they can be treated together and studied as a grammatical set. Contextualisation is further supported by GNSW lemmatising derived forms separately with implicit reference to ROBA. The article of ROBA in GNSW is followed by no less than 51 derivations entered as untreated lemmas cross-referenced to *-roba*:

robagana v. ROBA
robaganedi, se-/di- v. ROBA
robaganelo bo- v. ROBA
(See the appendix for the complete list)

The value of such cross-references for derived forms where all the affixes are suffixes in a stem dictionary is questionable because they all end up alphabetically directly following the article of the lemmatised and treated stem, i.e. ROBA in this case. It does not help the user much if he/she looks up, for example, *robagana* only to be referred to the article of ROBA directly above where (s)he has to work down through the entire user-unfriendly article layout anyway. A more precise reference address within the article of ROBA, i.e. *robagana*

would have been more helpful.

In terms of space utilisation and especially Ziervogel's (1965: 45) claim that the paradigm lumping approach prevents repetition, Van Wyk (1995: 88, 91-2) has shown in a critical review of this dictionary that in following this approach the compilers did *not* manage to avoid repetition. In his view they *introduced* redundancy by having to resort to unnecessary cross-referencing.

This brings no gain in economy compared with word dictionaries. The number of entries is the same for both types, the only difference being the structure and the alphabetic classification of the entries. (Van Wyk 1995: 88)

It also results in overuse of the medio structure.

Should the lexicographer really wish to include entire paradigms of verbal derivations, a splitting approach would be more user friendly: modules 1-32 would be given as main lemmas, each with treatment and will naturally alphabetically be grouped together anyway. Thus there will not be loss in economy and because they will alphabetically be in close proximity, morphological relations would to a large extent be visible and cross-referencing will be limited.

6. Conclusion

The GNSW is the most comprehensive dictionary ever compiled for Sepedi and as such remains an invaluable reference source even after four decades — it is a monument for the language. The GNSW scores high marks as a grammar reference source.

Viewed from many other angles however, GNSW is less effective as a dictionary, especially on different aspects pertaining to lemma selection, user-friendliness and comment on semantics. Initial criticism by sources such as van Wyk (1995) and Prinsloo and De Schryver (1999) were aimed at detrimental aspects of alphabetical ordering and the lemmatisation approach. They concluded among others that stem lemmatisation is the wrong option for a disjunctively written language and that a phonemic ordering is highly problematic from a user perspective.

In this article the selection and presentation of the lemmas were critically evaluated. It is highly unlikely that most of the lemmas will be looked for by target users. The lexicographer should not be creative in the sense of inventing words. He remains a recorder of the language and in the words of Phillip Gove (1961) should not attempt to set its style. (S)he should reflect what is real, the real language as used in print and speech, not that which is possible. Precious dictionary space should rather be used to include more words from the living language than artificially created possible reduplications.

The compilers focused on the completion of grammatical modular paradigms to the extent that the actual existence of most lemmas are questionable as supported by a limited user study, corpus evidence and treatment in other Sepedi dictionaries. Comment on semantics, the most important information

type that users are looking for was grossly neglected. Finally it was argued that the medio structure is largely ineffective.

Note

1. In GNSW *-rôbakantše* and *-rôbakantšwe* incorrectly appear with a circumflex on the *-e*.

Acknowledgement

This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85763).

The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.

References

Dictionaries

- GNSW = Ziervogel, D. and P.C. Mokgokong. 1975. *Pukuntšu 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.
- NEND = Kriel, Theunis J. 1976. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English*. Johannesburg: Educum Publishers.
- ONSD = De Schryver, G.-M. et al. (Eds.). 2008. *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.
- POP = Kriel, T.J., D.J. Prinsloo and B.P. Sathekge. 1997. *Popular Northern Sotho Dictionary, Northern Sotho–English, English–Northern Sotho*. Cape Town: Pharos.
- PTLH = Mojela, M.V. (Ed.). 2007. *Pukuntšuthaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.
- PUKU 1 = Kriel, T.J. 1983. *Pukuntšu woordeboek*. Pretoria: J.L. van Schaik.
- PUKU 2 = Kriel, T.J. and E.B. van Wyk. 1989. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.
- SLEPD = Mojela, M.V., M.C. Mphahlele, M.P. Mogodi en M.R. Selokela. 2006. *Sesotho sa Leboa/English Pukuntšu Dictionary*. Cape Town: Phumelela.

Other sources

- Bothma, T.J.D. and D.J. Prinsloo. 2013. Automated Dictionary Consultation for Text Reception: A

- Critical Evaluation of Lexicographic Guidance in Linked Kindle-dictionaries. *Lexicographica* 29(1): 165-198. Berlin: De Gruyter.
- Gouws, Rufus H. and Danie J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gove, Philip B. (Ed.)**. 1961. *Webster's Third New International Dictionary of the English Language*. Springfield, Mass.: Merriam-Webster.
- Haas, M.R.** 1962. What Belongs in the Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1962. *Problems in Lexicography*: 45-50. Bloomington: Indiana University/The Hague: Mouton.
- Hartmann, Reinhard R.K. (Ed.)**. 1983. *Lexicography: Principles and Practice*. London/New York: Academic Press.
- New Words in English*. <http://www.learn-english-today.com/new-words/new-words-in-english.html>. Accessed 12 June 2014.
- Prinsloo, D.J.** 1994. Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 14(2): 93-102.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the Oxford Northern Sotho Dictionary. *International Journal of Lexicography* 22(2): 151-178.
- 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, Danie J., Ulrich Heid, Theo J.D. Bothma and Gertrud Faaß.** 2011. Interactive, Dynamic Electronic Dictionaries for Text Production. Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011*: 215-220. Bled: Trojina.
- PSC = Pretoria Sepedi Corpus. University of Pretoria.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.
- Vintage patterns*. http://vintagepatterns.wikia.com/wiki/Patch_Press_379. Accessed 12 June 2014.
- Wells, Ronald A.** 1973. *Dictionaries and the Authoritarian Tradition. A Study in English Usage and Lexicography*. Janua Linguarum. Series Practica 196. Berlin/The Hague: Mouton de Gruyter.
- Ziervogel, D.** 1965. Probleme van leksikografie in die Suid-Afrikaanse Bantoetale. *Taalfasette* 1: 45-53.
- 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.

Appendix: The verb stem *ROBA* in GNSW

-ritša, bo-	1122	ROBA
<p>waar/waarheen stadig geloop word // place where/to which one walks slowly; moritšetši, ba- (moritšētši) pers. dev.; moritšetšo, me- (moritšētšo) man. dev.; RITŠETŠANA (-ritšētšana, -ritšētšane, -ritšētšanwa, -ritšētšanwe) appl. rec.; baritšetšani (baritšētšani) pers. dev.; moritšetšano, me- (moritšētšanō) man. dev.; RITŠISA (-ritšiša, -ritšišitšē, -ritšišwa, -ritšišitšwē) caus.; moritšiši, ba- pers. dev.; moritšišo, me- (moritšišō) man. dev.; RITŠISANA (-ritšišana, -ritšišane, -ritšišanwa, -ritšišanwe) caus. rec.; baritšišani pers. dev.; moritšišano, me- (moritšišanō) man. dev.</p>		<p>argument, subdue, be indifferent to what is said, give a hiding; ~ <i>pele</i> ontmoedig, iemand se hart breek // discourage, break one's heart; ~ <i>thokō</i> dom wees // be stupid; <i>thutëlō-bogolo e a</i> ~ dis nie maklik om te leer as jy oud is nie // it is not easy to learn when one has grown old; ~ <i>dikēnō</i> belofte verbreek // break one's promise; ~ <i>molaō</i> wet oortree // break/transgress the law; <i>ithōba</i> (-ithōba) refl.; ~ <i>dingalō</i> die vrugte van jou arbeid pluk // reap the fruits of one's labour; morōbi, ba- (morōbi) pers. dev.; morōbjá, ba- (morōbja) pers. pass. dev.; serobá, di- (serōba) stock, kerie // stick; serobjá, di- (serōbja) impers. pass. dev.; ROBAGANA (-rōbagana, -rōbagane, -rōbaganwa, -rōbaganwe) ass.; in stukke gebreek w. // b. broken into pieces; morobagani, ba- (morōbagani) pers. dev.; morobagano, me- (morōbaganō) man. dev.; ROBAGANELA (-rōbagenēla, -rōbaganētšē, -rōbaganēlwa, -rōbaganētšwē) appl. < ROBAGANA; morobaganelo (borōbaganēlō) lo. dev.; morobaganelo, me- (morōbaganēlō) man. dev.; serobagane, di- (serōbaganēdi) impers. dev.; ROBAGANTŠHA (-rōbagantšha, -rōbagantšhitšē, -rōbagantšhwa, -rōbagantšhitšwē) caus. < ROBAGANA; in stukke breek // break into pieces; morobagantšhi, ba- (morōbagantšhi) pers. dev.; morobagantšho, me- (morōbagantšhō) man. dev.; ROBAGANTŠHETŠA (-rōbagantšhētšā, -rōbagantšhēditšē, -rōbagantšhētšwa, -rōbagantšhēditšwē) caus. appl. < ROBAGANA; morobagantšhetšā, ba- (morōbagantšhētšā) pers. dev.; morobagantšhetšo, me- (morōbagantšhētšō) man. dev.; ROBAGANTŠHETŠANA (-rōbagantšhētšana, -rōbagantšhētšane, -rōbagantšhētšanwa, -rōbagantšhētšanwe) caus. appl. rec. < ROBAGANA; barobagantšhetšani (barōbagantšhētšani) pers. dev.; morobagantšhetšano, me- (morōbagantšhētšanō) man. dev.; ROBAGANYA (-rōbaganya, -rōbagantšē, -rōbaganywa, -rōbagantšwē) caus. < ROBAGANA; in stukke breek // break into pieces; morobaganyi, ba- (morōbaganyi) pers. dev.; morobaganyo, me- (morōbaganyō) man. dev.; ROBAGANYETŠA (-rōbaganyētšā, -rōbaganyēditšē, -rōbaganyētšwa, -rōbaganyēditšwē) caus. appl. < ROBAGANA; morobaganyetšā, ba- (morōbaganyētšā) pers. dev.; morobaganyetšo, me- (morōbaganyētšō) man. dev.; ROBAGANYETŠANA (-rōbaganyētšana, -rōbaganyētšane, -rōbaganyētšanwa, -rōbaganyētšanwe) caus. appl. rec. < ROBAGANA; barobaganyetšani (barōbaganyētšani) pers. dev.; morobaganyetšano, me- (morōbaganyētšanō) man. dev.; ROBAKA (-rōbaka, -rōbakilē, -rōbakwa, -rōbakilwē) it.; <i>ronden-bont</i> breek // break indiscriminately; <i>bošego bo rōbakilwē</i> dit word lig // it is dawning; morobaki, ba- (morōbaki) pers. dev.; morobako, me- (morōbakō) man. dev.;</p>
<p>-ritša, bo- v. RITŠA</p>		
<p>-ritša, se-/di- v. RITŠA</p>		
<p>ritšetša v. RITŠA</p>		
<p>ritšetšana v. RITŠA</p>		
<p>-ritšetšani, ba- v. RITŠA</p>		
<p>-ritšetšano, mo-/me- v. RITŠA</p>		
<p>-ritšetši, mo-/ba- v. RITŠA</p>		
<p>-ritšetšo, bo-/ma- v. RITŠA</p>		
<p>-ritšetšo, mo-/me- v. RITŠA</p>		
<p>-ritši, se-/di- v. RITŠA</p>		
<p>ritšiša v. RITŠA</p>		
<p>ritšišana v. RITŠA</p>		
<p>-ritšišani, ba- v. RITŠA</p>		
<p>-ritšišano, mo-/me- v. RITŠA</p>		
<p>-ritšiši, mo-/ba- v. RITŠA</p>		
<p>-ritšišo, mo-/me- v. RITŠA</p>		
<p>-ritšo, mo-/me- v. RITŠA</p>		
<p>-ritwá, se-/di- v. RITA</p>		
<p>-RŌ, bo-/ma-, boROFA, klou (hond, leeu) // paw (dog, lion)</p>		
<p>-RO, mo-/me- (morō) sop, sous // soup, gravy; ~ <i>wa pitsi</i>, <i>se mphišē</i>, <i>ke rilē mōla ba go lahla</i>, <i>ka go ja mphagō ka go nwa</i> stank vir dank is die wêreld se loon // get more kicks than halfpence; ~, <i>motsefēla-bašebi</i>, <i>bašomi dipelo tša baba</i> 'n mens moet vir klein dingetjies ook dankbaar wees // one must be thankful for small mercies</p>		
<p>-RŌ, se-/di- ['n soort eetbare bol] // a kind of edible bulb</p>		
<p>-ROBA, le-/ma-, cf. leŠOBA, gat, (pl.) steekwonde // hole, (pl.) stab wounds</p>		
<p>-ROBA, mo-/me- (morōba) gemak, ledigheid // comfort, idleness; <i>-bolawa ke</i> ~ op jou dooie gemak werk // work at leisure; <i>-ba morōbēng</i> op jou gemak wees // be at one's leisure</p>		
<p>-ROBA, se-/di- ingang aan agterkant van die lapa // back entrance to lapa</p>		
<p>-robá, se-/di- v. ROBA</p>		
<p>ROBA (-rōba, -rōbilē, -rōbja, -rōbilwē) breek, afbreek, swaar w. // break, break off, b. heavy; ~ <i>monō o tee nege</i> // nine; ~ <i>menō ē mebēdi</i> agt // eight; ~ <i>diēma</i> onkontroleerbaar word, jou ongewoon gedra // become uncontrollable, behave unusually; ~ <i>kgarebē leoto</i> 'n meisie swanger maak // impregnate a girl; ~ <i>molala</i> 'n argument wen, betoel, tem, jou nie steur aan wat gesê word nie, 'n pak slae gee // win an</p>		

ROBA	1123	ROBA
<p>ROBAKANA (-rôbakana, -rôbakane, -rôbakanwa, -rôbakanwe) rond-en-bont gebreek w. // be broken indiscriminately; <i>kgôrông go rôbakanwe</i> die raad het uiteengegaan // the meeting broke up; dirobakani (dirôbakani) impers. dev.; morobakano, me- (morôbakano) man. dev.; ROBAKANTŠHA (-rôbakantšha, -rôbakantšhitšê, -rôbakantšhwa, -rôbakantšhitšwê) caus. < ROBAKANA; rond-en-bont breek // break indiscriminately; morobakantšhi, ba- (morôbakantšhi) pers. dev.; morobakantšho, me- (morôbakantšhó) man. dev.; serobakantšhwá, di- (serôbakantšhwa) impers. pass. dev.; ROBAKANTŠHETSÁ (-rôbakantšhêtsá, -rôbakantšhêditšê, -rôbakantšhêtswana, -rôbakantšhêtsane) caus. appl. < ROBAKANA; morobakantšhetšani (barôbakantšhêtšani) pers. dev.; morobakantšhetšano, me- (morôbakantšhêtšanó) man. dev.; ROBAKANYA (-rôbakanya, -rôbakantšê, -robakanywa, -robakantšwê) caus. < ROBAKANA; rond-en-bont breek // break indiscriminately; morobakanyi, ba- (morôbakanyi) pers. dev.; morobakanyo, me- (morôbakanyó) man. dev.; serobakanywá, di- (serôbakanywa) impers. pass. dev.; ROBAKANYETSÁ (-rôbakanyêtsá, -rôbakanyêditšê, -rôbakanyêtswana, -rôbakanyêditšwê) caus. appl. < ROBAKANA; morobakanyetši, ba- (morôbakanyetši) pers. dev.; morobakanyetšo, me- (morôbakanyêtšó) man. dev.; ROBAKANYETSANA (-rôbakanyêtsana, -rôbakanyêtsane, -rôbakanyêtsanwa, -rôbakanyêtsanwe) caus. appl. rec. < ROBAKANA; barobakanyetšani (barôbakanyêtšani) pers. dev.; morobakanyetšano, me- (morôbakanyêtšanó) man. dev.; ROBEGA (-robêga, -robêgilê) neutr.; breekbaar w. // b. breakable, b. fragile; -<i>otlolla mohlašana e sa le ô monana, wa gola o tlô</i> ~ jy moet die boompie buig solank hy nog jonk is // as the twig is bent so the tree grows; <i>pelo e robêgilê</i> gekwets raak, seerkry (fig.) // get hurt (fig.); morobegi, ba- (morôbêgi) pers. dev.; 'n persoon van wie 'n liggaamsdeel gebreek is // a person whose limb is broken; morobego, me- (morôbêgô) man. dev.; thobego, (n-)/di- (thôbêgo) man. dev.; breekbaarheid // breakability; †beenbreuk // †fracture; ROBEGETSÁ (-rôbêgêtsá, -rôbêgêditšê, -rôbêgêtswana, -rôbêgêditšwê) v. ROBESETSÁ; ~ <i>lehlôkwana tsêbêng</i> raad gee // advise; ~ <i>lerapô têng</i> iemand swaar straf // punish someone severely; morobegetši, ba- (morôbêgêtsi) pers. dev.; morobegetšo, me- (morôbêgêtsó) man. dev.; ROBELA (-rôbêla, -rôbêtše, -rôbêlwa, -rôbêtšwe) appl.;</p>	<p>~ <i>lerapô têng</i> iemand hartseer maak // break someone's heart; ithôbela (-ithôbêla) refl.; ~ <i>legong hlôgông</i> jouself in die moelikheid laat beland // bring trouble upon oneself; morobedi, ba- (morôbêdi) pers. dev.; morobelo, me- (morôbêlô) man. dev.; ROBELANA (-rôbêlana, -rôbêlane, -rôbêlanwa, -rôbêlanwe) appl. rec.; barobelani (barôbêlani) pers. dev.; morobelano, me- (morôbêlanó) man. dev.; ROBESETSÁ (-rôbêsêtsá, -rôbêsêditšê, -rôbêsêtswana, -rôbêsêditšwê), cf. ROBEGETSÁ, binne-in afbreek, afbreek sodat 'n stuk agterbly // break off inside, break off in such a way that a piece remains behind; morobesetši, ba- (morôbêsêtsi) pers. dev.; morobesetšo, me- (morôbêsêtsó) man. dev.; ROBIŠA (-rôbiša, -rôbišitšê, -rôbišwa, -rôbišitšwê) caus.; laat/help breek, ens. // help/cause to break, etc.; morobiši, ba- (morôbiši) pers. dev.; morobišo, me- (morôbišó) man. dev.; ROBIŠANA (-rôbišana, -rôbišane, -rôbišanwa, -rôbišanwe) caus. rec.; barobišani (barôbišani) pers. dev.; morobišano, me- (morôbišanó) man. dev.; ROBOKA (-rôbôka, -rôbôkilê, -rôbôkwa, -rôbôkilwê) it., v. ROBAKA; moroboki, ba- (morôbôki) pers. dev.; moroboko, me- (morôbôkô) man. dev.; ROBOKANA (-rôbôkana, -rôbôkane, -rôbôkanwa, -rôbôkanwe) rec. < ROBOKA; v. ROBAKANA; dirobokani (dirôbôkani) pers. dev.; morobokano, me- (morôbôkanó) man. dev.; ROBOKANTŠHA (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < ROBOKANA; (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < ROBOKANA; v. ROBAKANTŠHA; morobokantšhi, ba- (morôbôkantšhi) pers. dev.; morobokantšho, me- (morôbôkantšhó) man. dev.; serobokantšhwá, di- (serôbôkantšhwa) impers. pass. dev.; ROBOKANTŠHETSÁ (-rôbôkantšhêtsá, -rôbôkantšhêditšê, -rôbôkantšhêtswana, -rôbôkantšhêditšwê) caus. appl. < ROBOKANA; morobokantšhetši, ba- (morôbôkantšhêtšhi) pers. dev.; morobokantšhetšo, me- (morôbôkantšhêtšó) man. dev.; ROBOKANTŠHETSANA (-rôbôkantšhêtšana, -rôbôkantšhêtšane, -rôbôkantšhêtšanwa, -rôbôkantšhêtšanwe) appl. rec. < ROBOKANTŠHA; barobokantšhetšani (barobokantšhêtšani) pers. dev.; morobokantšhetšano, me- (morôbôkantšhêtšanó) man. dev.; ROBOKANYA (-rôbôbakanya, -rôbôbakantšê, -rôbôbakanywa, -rôbôbakantšwê) caus. < ROBOKANA; v. ROBAKANYA; morobokanyi, ba- (morôbôbakanyi) pers. dev.; morobokanyo, me- (morôbôbakanyó) man. dev.; serobokanywá, di- (serôbôbakanywa) impers. pass. dev.; ROBOKANYETSÁ (-rôbôbakanyêtsá, -rôbôbakanyêditšê, -rôbôbakanyêtswana, -rôbôbakanyêditšwê) appl. < ROBAKANYA; morobokanyetši, ba- (morôbôbakanyetši) pers. dev.; morobokanyetšo, me- (morôbôbakanyêtšó) man. dev.</p>	

-robádi, bo-

1124

ROBALA

- dev.; *ROBOKANYETŠANA* (-róbókanyētšana, -róbókanyētšane, -róbókanyētšanwa, -róbókanyētšanwe) appl. rec. < *ROBOKANYA*; barobokanyetšani (baróbókanyētšani) pers. dev.; morobokanyetšano, me- (moróbókanyētšanò) man. dev.
- robádi, bo- v. ROBALA
- robadi, mo-/ba- v. ROBALA
- robadi, se-/di- v. ROBALA
- ROBADILÉPE, mo-/me- (moróbadilepê) doppruimboom // Papea fulva
- robadiša v. ROBALA
- robadišana v. ROBALA
- robadišani, ba- v. ROBALA
- robadišano, mo-/me- v. ROBALA
- robadiši, mo-/ba- v. ROBALA
- robadišo, mo-/me- v. ROBALA
- robadišwá, mo-/ba- v. ROBALA
- robagana v. ROBA
- robagamedi, se-/di- v. ROBA
- robaganela v. ROBA
- robaganelo, bo- v. ROBA
- robaganelo, mo-/me- v. ROBA
- robagani, mo-/ba- v. ROBA
- robagano, mo-/me- v. ROBA
- robagantšha v. ROBA
- robagantšhetša v. ROBA
- robagantšhetšana v. ROBA
- robagantšhetšani, ba- v. ROBA
- robagantšhetšano, mo-/me- v. ROBA
- robagantšhetši, mo-/ba- v. ROBA
- robagantšhetšo, mo-/me- v. ROBA
- robagantšhi, mo-/ba- v. ROBA
- robagantšho, mo-/me- v. ROBA
- robaganya v. ROBA
- robaganyetša v. ROBA
- robaganyetšana v. ROBA
- robaganyetšani, ba- v. ROBA
- robaganyetšano, mo-/me- v. ROBA
- robaganyetši, mo-/ba- v. ROBA
- robaganyetšo, mo-/me- v. ROBA
- robaganyi, mo-/ba- v. ROBA
- robakanyo, mo-/me- v. ROBA
- robaka v. ROBA
- robakana v. ROBA
- robakani, ba- v. ROBA
- robakano, mo-/me- v. ROBA
- robakantšha v. ROBA
- robakantšhetša v. ROBA
- robakantšhetšana v. ROBA
- robakantšhetšani, ba- v. ROBA
- robakantšhetšano, mo-/me- v. ROBA
- robakantšhetši, mo-/ba- v. ROBA
- robakantšhetšo, mo-/me- v. ROBA
- robakantšhi, mo-/ba- v. ROBA
- robakantšho, mo-/me- v. ROBA
- robakantšhwá, se-/di- v. ROBA
- robakanya v. ROBA
- robakanyetša v. ROBA
- robakanyetšana v. ROBA
- robakanyetšani, ba- v. ROBA
- robakanyetšano, mo-/me- v. ROBA
- robakanyetši, mo-/ba- v. ROBA
- robakanyetšo, mo-/me- v. ROBA
- robakanyi, mo-/ba- v. ROBA
- robakanyo, mo-/me- v. ROBA
- robakanywá, se-/di- v. ROBA
- robaki, mo-/ba- v. ROBA
- robako, mo-/me- v. ROBA
- ROBALA (-róbala, -róbêše, -róbálwa, -róbêšwe) slaap, lê, geslagtelik verkeer // sleep, lie down, have sexual intercourse; *o robêše* hy is aan die slaap // he is asleep; *go mo róbala* geslagtelik met haar verkeer // have sexual intercourse with her; ~ *borókò* nie uitgeslaape genoeg wees nie, gemaklik lewe // not to be shrewd enough, live comfortably; ~ *botala* kaal slaap, geen komberse om onder te slaap hê nie // sleep naked, have no blankets to sleep under; ~ *ka leihlò le tee* voorgee dat jy slaap, voorgee dat jy dom is // pretend to go and sleep, pretend to be stupid; ~ *le dikgogo* saam met die hoenders gaan slaap // go to bed early; ~ *le dipudi* baie onnet wees // be very untidy; ~ *le letšatši* v. ~ *le dikgogo*; ~ *le matólò* sonder 'n vroumens slaap // sleep without a woman; ~ *le taba* 'n geheim bewaar, eerste kennis dra van // keep a secret, first to know; ~ *phokéng* swaarkry // suffer; ~ *sepatikane* 'n dutjie neem voordat jy gaan slaap // take a nap before going to bed; ~ *sebutla/semmutla* met oop oë slaap // sleep with eyes open; ~ *seróbalela* v. *serobalela*; ~ *sethaka* saam op een plek slaap soos seuns van dieselfde ouderdom // sleep together in one place as boys of the same age-group do; *naga e sa robêše* die mense is nog onderontwikkeld, vredenstyd // people are still under-developed, peace time; *ke ilò* ~ *le wéna kobò e tee* ek sal jou agtervolg // I shall follow you; *boróbadi* (boróbadi) slaperigheid, slaapsiekte // sleepiness, sleeping sickness; *borobalo*, ma- (boróbalò) lo. dev.; slaapplek, akkommodasie // sleeping place, accommodation; -*kgopéla maróbalò* slaapplek vra // apply for accommodation; *morobadi*, ba- (moróbadi) pers. dev.; slaper // sleeper; *morobalo*, me- (moróbalò) man. dev.; *morobalwá*, ba- (moróbalwa) pers. pass. dev.; 'n vrou met wie geslagtelik verkeer w. // a woman with whom one has sexual intercourse; *serobadi*, di- (seróbadi) impers. dev.; *serobalo*, di- (seróbalò) kombers // blanket; *thobalo*, (n-)/di- (thóbalò) man. dev.; slaap, geslaap // sleep, sleeping; *ROBADIŠA* (-róbadiša, -róbadišitšê, -róbadišwa, -róbadišitšwê) caus.; cf. *ROBATŠA*, aan die slaap maak // cause to sleep; ~ *kgogo borókò* kul // deceive; ~ *mmutla* kul // deceive; ~ *nóga ya mpa* 'n bietjie eet // eat a little; *morobadiši*, ba- (moróbadiši) pers. dev.; *morobadišo*, me- (moróbadišò) man. dev.; *morobadišwá*, ba- (moróbadišwa) pers. pass. dev.; persoon wat aan die slaap gemaak word // a person who is caused/helped to sleep; *thobadišo*, (n-)/di- (thó-badišò) man. dev.; *ROBADIŠANA* (-rò-