# Lemmatization of Reflexives in Northern Sotho

D.J. Prinsloo, *Department of African Languages, University of Pretoria*

**Abstract:** The aim of this article is to evaluate current strategies in the lemmatization of reflexives in Northern Sotho. In particular the so-called "traditional" approach according to which reflexives are lemmatized randomly, as well as the more 'rule orientated' alternative, will be critically evaluated mainly against the background of principles such as user friendliness, avoidance / tolerance of redundancy, constant application of rules versus ad hoc decisions, and practical versus linguistic / scientific considerations. The scope is furthermore narrowed down to learners' dictionaries with the target user defined as a mother tongue speaker of English or Afrikaans who studies Northern Sotho. Special attention is given to those cases where *sound changes* and or *semantic shift* occur in the formation of the reflexive. The importance or relevance of the category "reflexives", the scope, nature and amount of sound changes, and the viability of an in-depth frequency study on reflexives will be determined from the output of a recently conducted frequency study on 15 randomly selected Northern Sotho books and magazines. It will be concluded that due to serious shortcomings in both the traditional and rule-orientated approaches, reflexives should be lemmatized on the basis of *frequency of use,* which in turn will require extended studies on considerably enlarged data corpora.

**Keywords:** LEMMATIZATION, REFLEXIVES, LEXICOGRAPHY, NORTHERN SOTHO, WORD FREQUENCY STUDY, USER FRIENDLY, PRACTICAL DICTIONARIES, LEARNERS' DICTIONARIES, RULE-ORIENTATED APPROACH, AD HOC DECISIONS

**Opsomming: Lemmatisering van refleksiewe in Noord-Sotho.** Die doel van hierdie artikel is om bestaande strategieë waarvolgens refleksiewe in Noord-Sotho gelemmatiseer word, te evalueer. Die sogenaamde "tradisionele" benadering waarvolgens refleksiewe willekeurig in woordeboeke opgeneem word, asook die meer reëlgerigte benadering word in besonder krities geëvalueer, hoofsaaklik teen die agtergrond van beginsels soos gebruikersvriendelikheid, vermyding / verduring van oortolligheid, konstante toepassing van reëls versus ad hoc-beslissings en praktiese versus linguistiese / wetenskaplike oorwegings. Die bestek word verder beperk tot aanleerderwoordeboeke en die teikengebruiker word gedefinieer as 'n moedertaalspreker van Afrikaans of Engels wat Noord-Sotho aanleer. Gevalle waar *klankverandering* en of *betekenisverskuiwing* plaasvind as gevolg van refleksiefvorming, geniet besondere aandag. Die belangrikheid ofte wel relevantheid van die kategorie "refleksiewe", die omvang en aard van klankveranderinge en die lewensvatbaarheid van 'n dieptestudie van refleksiewe sal bepaal word aan die hand van die resultate behaal in 'n onlangse frekwensiestudie van 15 willekeurig geselekteerde Noord-Sothoboeke en -tydskrifte. Daar sal gekonkludeer word dat as gevolg van ernstige tekortkominge in beide die tradisionele en die reëlgerigte benaderings, refleksiewe gelemmatiseer moet word op grond van *gebruiksfrekwensie,* wat dan verdere studie van aansienlik vergrote datakorpusse sal vereis.

**Sleutelwoorde:** LEMMATISERING, REFLEKSIEWE, LEKSIKOGRAFIE, NOORD-SOTHO, WOORDFREKWENSIESTUDIE, GEBRUIKERSVRIENDELIKHEID, PRAKTIESE WOORDEBOEK, AANLEERDERWOORDEBOEK, REËLGERIGTE BENADERING, AD HOC-BESLISSINGS

## Introduction

Although dictionaries may be readily available for a language like Northern Sotho, Gouws (1990: 53, 55), is correct when he says:

> "An analysis of the dictionary types available in the various [African] languages indicates a complete *lack of lexicographical planning* ... The majority of dictionaries for African languages are the *products of limited efforts not reflecting a high standard of lexicographical achievement* ... with a few exceptions these dictionaries offer only restricted translating equivalents ..."[1]

Pioneering efforts towards real lexicographical study for the African languages have been made by Van Wyk and Prinsloo in the fields of practical versus linguistic dictionaries, word frequency studies, prescriptiveness versus descriptiveness, etc. However, much work still has to be done to put generally accepted lexicographical principles such as these into practice for Northern Sotho.

As far as the lemmatization of nouns and verbs are concerned, no serious effort has been made to formulate a lexicographical policy or even to lemmatize nouns and verbs in a consistent manner, not to mention the arrangement required to make it computer compatible.

Since it is impossible even to scratch the surface of the problems regarding the lemmatization of nouns and verbs in the scope of this article, the discussion will be focused on only one problematic aspect concerning the lemmatization of *verbs*, namely lemmatizing reflexives.

## Morphological and semantic realities facing the lexicographer

The reflexive in Northern Sotho is formed by *prefixing* the morpheme *i-* to the verb. This fact immediately brings the lexicographer to the root of the problem: From a redundancy point of view, it is virtually *impossible* to enter all verbs *twice*, i.e. once under the basic stem, e.g. *thuša* 'help' and again as a separate entry alphabetically under *i, ithuša* 'help oneself'. It stands to reason that unless some sound lexicographical principle could be employed to limit these entries, the redundancy factor would get totally out of hand. The lexicographer has to take a basic stand right from the start.

He firstly must make a decision on the principle of lemmatizing reflexives for say a pocket size or medium size dictionary. He might even opt for one of

the *extremes* namely to re-enter *most of the verbs*, with the prefix *i-*; or the other extreme namely *not to enter reflexives at all*; or *compromise*, in settling for something in between.

"One of the basic problems of lexicography is to decide *what to put in the dictionary and what to exclude*." (Hartmann 1983: 51)

It will subsequently be illustrated that the serious consideration of *frequency of use* can considerably aid the lexicographer towards the selection of reflexives to be entered into a dictionary:

"At the level of the 'macrostructure' ... it may be possible to verify the *frequency of a lexeme, in order to decide for or against its inclusion*." (Hartmann 1983: 70)

"... more and more our dictionary-makers see the language as conditioned by its social functions, and so the choice of words to go into them is determined by sophisticated assessments of *frequency* and of users' needs ..." (Hartmann 1983: 21)

For the purpose of this article reflexives in fifteen randomly selected Northern Sotho literary works and magazines, totalling 220 000 words, likely to be read by the target user, have been studied and counted in order to determine

(i)     the importance or relevance of the category *reflexives* as a whole in the living language, for example in comparison to nouns, ordinary verbs, pronouns, etc., i.e. whether it is worthwhile for the lexicographer to take pains in the lemmatization of reflexives;

(ii)    the scope, nature and amount of sound changes and semantic shift resulting from the prefixing of *i-*; and

(iii)   whether a clear distinction between highly used reflexives and sel-, domly used ones exists on a broad base; in simple terms whether it is possible to establish if a specific reflexive, like *ithuša* 'help yourself' is productively used or not.[2]

Hartmann (1983: 188) states that a word must occur evenly over a broad spectrum of miscellaneous data corpora:

"... a word must occur evenly in a large number of the stratified sub-samples rather than excessively often in a small number of them, given that these two very different cases could show identical 'total-corpus' frequencies."

Table 1

|  | Total no. of occurrences | Text categories (maximum 15) | Distribution in text samples (maximum 500) |
|---|---|---|---|
| only | 1 815 | 15 | 458 |
| time | 1 658 | 15 | 449 |
| Kennan | 29 | 1 | 1 |
| two-day | 4 | 1 | 4 |

(Johansson and Hofland (1989: 43))

In Table 1 it is clear that 'only' and 'time' are highly used words over a broad spectrum in contrast to 'Kennan' and 'two-day'. (See Prinsloo (1991) and Johansson and Hofland (1989) for detailed discussions on the Brown- and Lob Corpora as well as word frequency studies for Northern Sotho.)

Thus a specific reflexive will only be regarded as relatively highly used if it occurs frequently in (a) the corpus as a whole and (b) every book or magazine (and not for example with a high frequency in one book, but not at all in the next five).

As far as (iii) is concerned, it soon became clear that if a reflexive is highly used in any single book or magazine, it is also highly used, or equally spread in all of the others. In fact it proved to be *fairly easy* to establish which reflexives are productively used in Northern Sotho. Reflexives like those given in Table 2, not only occur frequently in the data corpus as a whole, they also occur frequently in virtually every book and magazine.

Table 2

| Book/Magazine no. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|---|---|---|---|---|---|---|---|---|
| ikemišeditše(go) 'intended' | 11 |  | 7 | 1 | 1 | 7 | 1 | 14 | 42 |
| ikhwetša(go) 'find oneself' | 1 |  | 2 |  | 2 | 4 | 9 | 7 | 25 |
| ikwa(go) 'feel/hear oneself' | 1 |  |  | 2 | 1 | 4 | 4 | 21 | 33 |
| ipha(go) 'give to oneself' | 2 | 4 | 3 | 6 | 1 | 2 | 9 | 5 | 32 |
| iphile 'gave to oneself' | 6 | 1 | 2 | 2 | 2 |  | 10 | 1 | 24 |
| ipona(go) 'see oneself' | 1 | 1 | 10 | 8 | 1 | 10 | 3 | 7 | 41 |
| ipotšiša 'ask oneself' | 1 |  |  | 3 | 3 | 5 | 1 | 21 | 34 |
| ithuta(go) 'teach oneself' | 14 | 4 | 1 | 25 |  | 11 | 3 | 11 | 69 |
| itokišetša 'prepare oneself' | 1 | 10 | 3 | 3 |  | 7 |  | 17 | 41 |
| itshola 'blame oneself' |  | 1 | 4 | 4 |  | 13 | 7 | 4 | 33 |

(In order to compare *equal quantities* of words with one another, i.e. by not comparing a magazine consisting of 28 pages with a book consisting of 150 pages, the 15 randomly selected books and magazines were sub-grouped into eight categories as follows:

1. *Tšwelopele*:  Mosegamanye (1979, no. 7), Dibokwana (1979, no. 2), Setemere (1985, no. 9), Lewedi (1979 no. 9), Moranang (1979, no. 4), Hlakola (1979, no. 3), Maatla Mpulele, A re Bale std. 1; 2.  Lukas Motšheletšhele; 3.  Noto ya masogana; 4.  Ditaodišo; 5.  Ke namane ya morago; 6.  Tshehlana ya ka; 7. Moremogolo; 8.  Nonyana ya Tokologo).  (Full particulars are given in the list of references below).

From a lexicographical point of view words such as these which reflect a high overall count as well as an equal distribution across the data base, are beyond any doubt, in Gove's words, those reflexives *most likely to be looked up by the target user.*

> "Selection is guided by usefulness, and usefulness is determined by the degree to which terms most likely to be looked for are included." (Gove 1961: preface 4a)

Wells (1973: 95) in reference to Gove remarks:

> 'The word *usage* today generally refers to 'the way in which words and phrases are actually used', as opposed to some externally derived set of arbitrary rules."

The results in Table 2 are significant and admittedly impressive. However, one must not lose perspective in making too strong claims on a limited data base. For example, it will be risky at this stage to *reject* a specific reflexive that has been entered into a dictionary, simply because it does not occur in this data corpus.  On the other hand a much stronger claim is appropriate for the reflexives in Table 2, due to their *total performance on the text as a whole* and being *evenly spread over a broad base* of miscellaneous data corpora.  Formulated differently, it means that even at this stage, it is clear that the reflexives in Table 2 are highly used in Northern Sotho but that more evidence is needed before reflexives, which did not occur in this data corpus, could be labelled as 'seldomly used'.

Some consolation for not having taken oral communication as yet into consideration for this study is found in Fries' remark:

> "In spite of the modern insistence that the materials of speech constitute the 'language', most practical dictionaries have selected their materials from writing and literature." (Fries (Al-Kasimi 1977: 24))

As far as (i) is concerned the necessity to include reflexives into even the small-est dictionary was underlined by the fact that two to three reflexives occur on every page totalling well over 2 000 for the data corpus as a whole. In regard to (ii) concerning the issue of sound- and semantic changes resulting from the prefixing of *i-*, it became evident that from a lexicographic point of view reflexives should be sub-grouped into four categories namely:

(a)    no sound changes nor any semantic shift,
       go *ithuša* 'to help oneself'
       (*i-* = 'self', *thuša* = 'help')
(b)    no sound changes but semantic shift,
       go *ingwadiša* 'to register'
       (*i-* = 'self', *ngwadiša* = 'cause to write')
(c)    sound changes but not semantic shift,
       go *ithata* 'to love oneself'
       (*i-* = 'self', *rata* = 'love')
(d)    sound changes as well as semantic shift,
       go *ikemela* 'to stand alone', 'to be independent'
       (*i-* = 'self', *emela* = 'stand (for) by', 'wait for')

**The traditional way of lemmatizing reflexives at random**

In this paragraph the attention will mainly be focused on the handling of reflexives by the dictionary pioneer T.J. Kriel in editions of the *Pukuntšu* up to 1983, the *Popular Northern Sotho Dictionary* up to 1988 as well as Ziervogel and Mokgokong's approach in the *Klein Noord-Sotho Woordeboek* up to 1979. In the *Klein Noord-Sotho Woordeboek* (1979), Ziervogel and Mokgokong, apart from formulating a few rules (which will be discussed in the next paragraph), lem-matized only ten reflexives. Such an ad hoc decision is totally acceptable if it is done to reflect extremely high usage, say for example ten of the reflexives listed in Table 2 above. However, with the exception of *ithuta* 'teach oneself' and *itshwara* 'behave oneself' the likeliness of these words to be looked up by the target user is *highly questionable*. Compare their occurrences or even total absence in the 15 selected books and magazines utilized for this study: *ikgata* (2) 'tread on oneself', *ikola* (0) (no clear translation given), *ikwela* (1) 'fall (for) oneself', *ithuta* (63) 'teach oneself', *itiša* (2) 'take care of oneself', *itshelela* (1) 'seek food for oneself', *itshwara* (19) 'behave oneself', *itsomarela* (0) '(no clear transla-tion given), *itswalanya* (4) 'associate oneself with', *ipea* (11) 'place oneself'. For reflexives such as *ipea* and *ikgata*, translations are given which have nothing to do with the reflexive character of these words. In cases such as *itsomarela*, the user is referred to the basic verb stem, e.g. *-tsomarela* in this case, which has not even been entered into the dictionary!

Consider Kriel's *Popular Northern Sotho Dictionary* (1988) as a second example in this regard. Compared to the *Klein Noord-Sotho Woordeboek*, which

is more or less of the same size, Kriel did well in including more than 200 reflexives. However, once again, from a lexicographical point of view, one cannot but come to the conclusion that no lexicographical planning has been done and that the compiler added reflexives to the dictionary 'as he came across them'. No serious consideration has been given to frequency of use, which, needless to say, should be the most important consideration in the compilation of a pocket size dictionary like this one. As far as lexicographical planning or editorial policy is concerned, it is not clear why *itokišetša* (41) 'prepare oneself', (counts given in brackets), and its perfect form *itokišeditše* (11) 'prepared oneself' is given while only *ikemišetša* (17) 'intend' has been entered but not its perfect form *ikemišeditše* (42) 'intended'. This omission surely could not have been done on the ground of frequency of use because *ikemišeditše* is clearly a highly used reflexive in Northern Sotho. The same arguments hold true for the lemmatization of *itaeditše* (4) 'has revealed oneself' but not *itaetša* (12) 'reveal oneself', *ithwalela* (1) 'carry for oneself' but not *ithwala* (4) 'carry oneself', etc.

Furthermore, if there is room in this dictionary for words like *ikubuela* (0) 'scatter over oneself', *ikgara* (0) 'coil oneself up', *ikgolola* (0) 'free oneself', *ikhunela* (0) 'draw oneself up', etc., which lack a single occurrence in these 15 selected books and magazines, why not for highly used words such as *ikemišeditše(go)* (42) 'intended', *ipotšiša* (34) 'ask oneself', *ikhwetša* (25) 'find one-self', *inyakela* (17) 'want something for yourself', *iponatša* (13) 'let oneself see', *itheta* (12) 'praise oneself ', and many more.

In some cases highly used reflexives have been entered but are found 'hidden' in articles of head words where they are not likely to be looked for. In *Pukuntšu* (1983 and previous editions) the entry *ikemišeditše* is given as such, which is of course user friendly. However, it is hidden in the article of the head word *ikemela* which is from a lexicographical point of view incorrect. The chances of the user having missed *ikemišeditše* in the article of *ikemela* is 90%. There should at least have been an entry *ikemiša* or even better *ikemišetša* as is the case in the *New English Northern Sotho Dictionary* (1985) and even in the small pocket size *Popular Northern Sotho Dictionary* (1988).

In the same way it could be illustrated that in dictionaries like the *New English Northern Sotho Dictionary* (1985) precious space has been allocated to reflexives which are unlikely to be looked up by the target user whilst highly used reflexives were omitted. This renders the dictionary impractical.

> "As in its very beginnings twelve centuries ago, the new shape which our dictionaries are taking today is still determined by *practical needs*." (Hartmann 1983: 21) "... dictionaries came into being in response to very practical needs ..." (Osselton (El-Badry 1990: 36))

Unfortunately these findings painfully echo Rufus Gouws' remarks quoted above i.e. *"products of limited efforts not reflecting a high standard of lexicographical*

*achievement ... complete lack of lexicographical planning".*

## The rule-orientated approach

From the outset it can be argued that this approach has an 'underlying' disadvantage:

> "many introductory pages usually allocated to grammatical sketches of the language concerned without the knowledge of which it is deemed hazardous to use the dictionary successfully. We believe, however, that these sections and introductory explanations *are not sufficient provisions for a user friendly product. Dictionary users* are known to *allocate little time to the study of these prefactory matters"* (Busane 1990: 28)

Compare in this regard Ziervogel and Mokgokong's 40 page introduction / mini-grammar in the *Groot Noord-Sotho Woordeboek* (1975). Secondly it stands to reason that the success of any attempt in regulating the lemmatization of reflexives in order to control redundancy is seriously challenged by sound changes and semantic shift as categorized above. Recall instances where

(a)  no sound changes nor any semantic shift,
(b)  no sound changes but semantic shift,
(c)  sound changes but not semantic shift, and
(d)  sound changes as well as semantic shift occur.

Category (a) poses no problem and the success in reducing redundancy is extensive: The rule to follow is simply to remove the *i-* and look the word up under the following letter, for example *ithuša* 'help oneself'; in this case under *thuša* 'help'. Categories (b) and (d) in which semantic shift (with or without sound changes) occurs, cannot be reduced by regulation and have to be entered as separate lemmas anyway. Van Wyk, in Kriel and Van Wyk (1989), admits this fact and lemmatizes such reflexives, which he terms "afwykend", under their reflexive prefix *i-*. Unfortunately (b) and (d) pose a serious setback for this approach because it means that redundancy is not reduced, and secondly, the question arises on what grounds the lexicographer is going to select reflexives from categories (b) and (d) to be entered into the dictionary. Therefore, as far as the lemmatization of reflexives which have undergone semantic changes is concerned, the lexicographer is back to square one.

In the *Klein Noord-Sotho Woordeboek* (1979) Ziervogel and Mokgokong made no provision for categories (b) and (d). In fact, in ignoring the phenomenon of semantic shift, they help their users from the frying pan into the fire by creating the impression that a reflexive, semantically always consists of 'self' + the meaning of the verb, when they guide the user as follows:

"N.B. verbs commencing with the reflexive prefix should be looked up under the first sound of the basic verb; remember the sound changes effected [sic] by the *i-*; cf. the Handboek" (Ziervogel and Mokgokong 1979: 62)

Numerous examples could be listed to illustrate the misinterpretation of meaning which results from the above statement, but the mistake is so blatant that it does not deserve further elaboration. Hartmann underlines the potential danger:

"... foreign language learners are not sophisticated enough to realize that they are poorly served and take the information supplied at face value, the results — their essays, compositions and translations — *reveal the deficiencies of such dictionaries in a most dramatic way.*" (Hartmann 1983: 47)

"'n aanleerderwoordeboek, die sogenaamde 'learners' dictionary', is gerig op die gebruiker wat 'n vreemde taal aanleer. Die inligting moet op so 'n manier aangebied word dat daar *geensins 'n beroep op die taalintuïsie* van die gebruiker gedoen word nie." (Gouws 1989: 71)

The next challenge is to regulate category (c) in such a way that redundancy is reduced without rendering the dictionary user unfriendly, impractical or even irrelevant for looking up reflexives. Ziervogel and Mokgokong's ill fated 'guidance' quoted above, desperately referring the user to the "Handboek" results in the same problems as will now be discussed in terms of Van Wyk for regulating this category.

A fresh approach was taken by Van Wyk in the compilation of the *Pukuntšu* (1989) which is the updated and revised edition of Kriel's *Pukuntšu* (1983). He certainly deserves much credit for doing some serious lexicographical planning, in defining his target user, carefully compiling the mini-grammar / guide towards using the dictionary, and for his serious attempts to reduce redundancy, especially in the lemmatization of nouns and verbs.

As far as category (c) is concerned, he discontinued Kriel's tradition of lemmatizing reflexives as separate entries and included a list of the sound changes which occur when *i-* is prefixed to most of the verb stems in Northern Sotho. Table 3 is an *extract* from this list.

In *ideal circumstances* it simply means that the user has to 'undo' the sound changes or use this guide to establish the basic form of the verb stem and to look it up. Take for example the word *ithata* 'love oneself'. The user simply has to find the appropriate rule in Table 3 and to look up the meaning of *rata* 'to love'.

Table 3

| ip-/mp-: | b-, | ex. ipona/mpona | under bôna |
|---|---|---|---|
|  | p-, | ex. ipatolla/mpatolla | under patolla |
| iph-/mph-: | f-, | ex. iphahla/mphahla | under fahla |
|  | ph-, | ex. iphemela/ | under phema |
|  |  | mphemela |  |
| it-/nt | l-, | ex. itebala/ntebala | under lebala |
|  | d-, | ex. itira/ntira | under dira |
|  | t-, | ex. itanya/ntanya | under tanya |
| ith-/nth-: | r-, | ex. ithata/nthata | under rata |
|  | th-, | ex. ithabiša/nthabiša | under thaba |
| itš-/ntš | j-, | ex. itšela/ntšela | under ja |
|  | tš-, | ex. itšeela/ntšeela | under tšea |
| itsh-/ntsh-: | s-, | ex. itshwara/ntshwara | under swara |
|  | tsh-, | ex. itshepa/ntshepa | under tshêpa |
| itšh-/ntšh-: | š-, | ex. itšhia/ntšhia | under šia |
|  | tšh-, | ex. itšhaša/ntšhaša | under tšhaša |

(Kriel and Van Wyk 1989: Preface)

As for category (a), the success in reducing redundancy in category (c) is *impressive*, but in the following paragraphs an attempt will be made to evaluate the *real* success factor, the impact on the knowledge presupposed of the target user, and to count the cost in terms of user friendliness. Take for example the highly used reflexive *ikemišeditše* (42). This word was included in the 1983 edition of the *Pukuntšu* but omitted in *Pukuntšu* (1989) following the editorial policy according to which the learner is expected to have a knowledge of the affixal system of the language as well as to utilize Table 3. This word consists of a reflexive *i-*, a consonant *-k-* (the result of a sound change when *i-* is used before a vowel), the verbal root *-em-* (stem = *-ema-*), the causative suffix *-iš-*, the applicative suffix *-el-* and the perfect suffix *-il-*. Firstly, it is highly unlikely that the target user of this dictionary (scholars and students) will be able to analyze this word by reversing the sound changes that occur when the different suffixes are combined, as has been done above. This process is of course unavoidable to eventually obtain the precise meaning rendered by the combination of the semantic values of the different suffixes. Secondly, even if he is able to do this he then attempts to establish the meaning of the word and concludes that *i-* 'self'+ *-em-* 'stand'+ *-iš-* 'cause'+ *-el-* 'for'+ *-il-* 'past' means 'caused yourself to stand for' compared to the actual meaning of 'intended'.

In contrast to these examples for which it could be argued that an unfair quantity of knowledge is expected from the target user, words like *ithuta* 'learn oneself / study', and *itapološa* 'refresh oneself' are given as separate entries, i.e. as if belonging to categories (b) and (d) in terms of the editorial policy. How-

ever, the meaning of *lapološa* 'refresh' + *i*- 'self' equals 'refresh oneself' and could have been treated as a category (a) case. The same holds true for *ithuta* 'teach oneself'.

Another problem here is that the user can easily become *strangled* by the very rules designed to help him, for example in looking up reflexives in which the *i*- is followed by *-p*-. In Table 3 and on page 58 of *Pukuntšu* (1989), if the combination is *ip*-, he is directed to *b*- and *p*-; if *-p*- is followed by *-h*-, then he is directed to *f*- and *ph*-. If, however, *-p*- is followed by *-š*-, he should try *bj*- or *pš*-. If *pš*- is followed by *-h*- then the search should be continued under *fš*- or *pšh*-. The user will probably give up before he finds the appropriate verb stem or he will eventually be in doubt whether he has arrived at the correct solution after all.

It is thus questionable whether the dictionary, in terms of Barnhart, answers those questions asked by the user:

"It is the function of the popular dictionary to *answer the questions that the user of the dictionary asks*, and the dictionaries on the commercial market will be successful in proportion to the extent to which they answer these questions of the buyer. This is the basis on which the editor must determine the type of information to include." (Barnhart (Al-Kasimi 1977: 19))

In the *Groot Noord-Sotho Woordeboek* (1975) of Ziervogel and Mokgokong the user most probably suspects that the word will not be listed under the reflexive prefix *i*- itself and attempts to find the rules governing sound changes in the mini-grammar / introduction in order to identify the verbal root. He finds the heading under (d) on page 57 of the mini-grammar but is immediately referred to par. 8.6.3 — which does not exist (should have been par. 8.5.3). An attempt to look up the word directly is somewhat more successful. He finds *ikemišetša* and is referred to the stem *ema*. After having struggled through the article of *ema*, (almost four columns in length), he eventually finds *emišeditše* as a derivative of *emišetša*. In his view, all that is needed now, is adding the reflexive connotation expressed by *i*-. However, the meaning given here is 'be proud', 'look a gift horse in the mouth'. He runs the risk of concluding that + *i*- it means 'be proud of oneself'. In the (unlikely?) event of him looking further down, he will find the entry *ikemišetša*. This time the perfect suffix is missing, but he can correctly conclude that the meaning given here, 'intend' + 'past' must mean 'intended'.

The same line of argumentation could be followed for the highly used reflexive *itokišeditše* 'prepared oneself for'. It is highly unlikely that the target user of the *Pukuntšu* (1989) will be able to analyze this word, i.e. by 'reversing' the sound changes and figuring out its real meaning by combining the semantic values of the different affixes. Even if he does, he will conclude that *itokišeditše* means 'caused to become straight / correct for' instead of the actual meaning of 'prepared oneself for'.

An important interacting criterion should be seriously considered here, namely the format of the dictionary to be compiled: The smaller the dictionary format, the fewer the rules should be. In practical terms this means that for a small, or medium size dictionary, a balance should be created between a specific rule and the success of avoiding redundancies for that rule. It must be admitted that a very productive rule applied with a high frequency can avoid considerable redundancy even in a relatively small dictionary. But if a specific rule covers relatively few examples, an ad hoc approach should be taken by omitting the rule or even all the rules lemmatizing this limited number of entries. Thus, for this medium size dictionary at least 23 of the rules given in Table 3 could be deleted in exchange for redundancy of only a few extra pages. Consider for example the rule *i- + j- > itŝ* which affects not more than 10 entries in the entire dictionary.

## Conclusion

From the discussion above it should be clear that the solution to the problem of lemmatizing reflexives in Northern Sotho cannot be found in the traditional approach namely lemmatizing reflexives uncontrollably or randomly as the lexicographer comes across them, nor in the efforts to regulate the lemmatization thereof. It has been indicated that the success rate of the rule orientated approach for categories (b) and (d) is zero and for these categories the dilemma of what to include and what to leave out remains unsolved. As far as category (c) is concerned the price paid in terms of the amount of knowledge presupposed from the user, user friendliness, complexity of guidelines in the mini-grammar and in the text, impracticality, user uncertainty and frustration is too high.

I wish to argue that the ideal for a pocket size or medium size learners' dictionary would be to lemmatize those reflexives "most likely to be looked up by the target user" as quoted. For a student learning Northern Sotho it means the most frequently used reflexives. These reflexives should be lemmatized in such a user friendly way that it can be directly looked up under *i-* without prior reference to the mini-grammar or knowledge of difficult language rules. Zgusta (1971: 15) is in the right when he says that "*the lexicographer should know everything*". Although the lexicographer can go some way in relying on his *intuition*, he has to employ word frequency studies in order to ascertain the 'correct' corpus of reflexives for the specific type of dictionary to be compiled, e.g. pocket size or medium size. This is exactly what word frequency studies according to Prinsloo (1991) are all about: Selecting just the right corpus of words (reflexives) for a specific dictionary and secondly preventing the omission of essential words (reflexives). The relevance of this statement is clearly underlined in the introduction to the *Setswana English Afrikaans Dictionary* (1990), where Snyman and Shole honestly admit:

"The dictionary team is aware of the fact that *common and even essential words may easily be omitted* during the compilation of a dictionary. This can take place simply because *the lexicographer has not encountered such words.* We can only hope that there are not too many examples of this kind."

"The *decision what to include* in the dictionary still has to be made by the lexicographer himself, however, and this *depends in turn upon the nature and size of the dictionary* and its intended users. In this respect lemmatized *frequency lists can be a further help*." (Hartmann 1983: 81)

If the lexicographer takes into consideration that up to date less than 300 books have been published in Northern Sotho, and that advanced computer technology is available to him, then surely it is within his reach to extend word frequency studies on the reflexive from the current corpus of 15 books and magazines to cover all 300.

"... we have reached a stage where cooperation between man and machine is useful and perhaps indispensable in making better dictionaries." (Hartmann 1983: 87)

There is no doubt that this will be an immense task.

"... the worst criminals should neither be executed nor sentenced to forced labour, but should be condemned to compile dictionaries, because all the tortures are included in this work." (Scaliger (Zgusta 1971: 15))

## Notes

1    Italics in all quotations are mine.
2    See Prinsloo 1991 for a detailed discussion.

## References

Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries.* Leiden: E.J. Brill.
Bosoma, P.S. 1990. *Tshehlana ya ka.* Pretoria: Via Afrika.
Busane, Masidake. 1990. Lexicography in Central Africa: The User Perspective, with Special Reference to Zaïre. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa.* Exeter Linguistic Studies 15: 19-35. Exeter: University of Exeter Press.

El-Badry, Nawal. 1990. Arabic Lexicography in Northern Africa, with Special Reference to Egypt. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa*. Exeter Linguistic Studies 15: 36-43. Exeter: University of Exeter Press.

Gouws, R.H. 1989. *Leksikografie*. Pretoria: Academica.

Gouws, Rufus. 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa*. Exeter Linguistic Studies 15: 52-65. Exeter: University of Exeter Press.

Gove, P.B. (Ed.). 1961. *Webster's Third New International Dictionary of the English Language*. Springfield Massachusetts: Merriam-Webster.

Hoffmann, P.A. and P.W. van Heerden. 1987. *A re Bale Std. 1*. Pretoria: J.L. van Schaik.

Hartmann, R.R.K. 1983. *Lexicography: Principles and Practice*. London: Academic Press.

Johansson, S. and K. Hofland. 1989. *Frequency Analysis of English Vocabulary and Grammar*. Oxford: Clarendon Press.

Kekana, M.A. 1985. *Nonyana ya Tokologo*. Johannesburg: Educum.

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

Kriel, T.J. 1985. *New English-Northern Sotho Dictionary*. Johannesburg: Educum.

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

Kriel, T.J. and E.B. van Wyk. 1989. *Pukuntšu Woordeboek*. Pretoria: J.L. van Schaik.

Masola, I.S. 1987. *Maatla Mpulele*. Pretoria: J.L. van Schaik.

Mojapelo, G.P. and J.J.R. Masemola. s.a. *Ditaodišo*. Johannesburg: Bona.

Nkadimeng, H.M. (Ed.). 1979a. *Tšwelopele* 20(2). Johannesburg: Perskor.

Nkadimeng, H.M. (Ed.). 1979b. *Tšwelopele* 20(3). Johannesburg: Perskor.

Nkadimeng, H.M. (Ed.). 1979c. *Tšwelopele* 20(4). Johannesburg: Perskor.

Nkadimeng, H.M. (Ed.). 1979d. *Tšwelopele* 20(7). Johannesburg: Perskor.

Nkadimeng, H.M. (Ed.). 1979e. *Tšwelopele* 20(9). Johannesburg: Perskor.

Nkadimeng, H.M. (Ed.). 1985. *Tšwelopele* 29(9). Johannesburg: Perskor.

Palaki-Tsebe, D.G. s.a. *Noto ya masogana*. Johannesburg: APB.

Prinsloo, D.J. 1991. Towards Computer-assisted Word Frequency Studies in Northern Sotho. *S.A. Journal of African Languages* 11(2): 54-60.

Rammala, M.M. 1971. *Lukas Motšheletšhele*. Pretoria: J.L. van Schaik.

Seakamela, J.M. 1991. *Ke namane ya Morago*. Pretoria: Acacia.

Snyman, J.W. and J.S. Shole. 1990. *Setswana English Afrikaans Dictionary*. Pretoria: Via Afrika.

Tlooke, A.M. 1986. *Moremogolo*. Johannesburg: Educum.

Wells, R.A. 1973. *Dictionaries and the Authoritarian Tradition: A Study in English Usage and Lexicography*. The Hague: Mouton.

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

Ziervogel, D. and P.C. Mokgokong. 1979. *Klein Noord-Sotho Woordeboek*. Pretoria: J.L. van Schaik.

Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.