
Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective*

Salmina Nong, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (snong@postino.up.ac.za)*,
Gilles-Maurice de Schryver, *Research Assistant of the Fund for Scientific Research — Flanders (Belgium) and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (gillesmaurice.deschryver@rug.ac.be)*, and
D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (prinsloo@postino.up.ac.za)*

Abstract: The aim of this article is to investigate, from a lexicographic perspective, the preferences of Northern Sotho mother-tongue speakers for loan words versus so-called 'traditional' or 'original' counterparts in the language. Results obtained from a survey conducted among 100 randomly selected mother-tongue speakers from different age and gender groups, backgrounds, places of residence, etc. will be analysed. It is shown that although the overwhelming preference of the respondents lies with the use of (more) indigenous words in comparison to loan words, lexicographers should be alerted to possible, even rapid, changes in this preference pattern. The results from the survey are compared throughout with frequency counts derived from a corpus as well as with current dictionary treatment.

Keywords: LEXICOGRAPHY, DICTIONARY, LEMMATISATION, NORTHERN SOTHO (SEPEDE), LOAN WORD, SOTHOISED WORD, INDIGENOUS WORD, QUESTIONNAIRE, CORPUS, DESCRIPTIVENESS, PROSCRIPTIVENESS, PRESCRIPTIVENESS, PREFERENCE PATTERN

Senaganwa: Maadingwa ge a bapetšwa le Mantšu a Setlogo go Sesotho sa Leboa — Kgopolo ya Bangwalapukuntšu. Maikemišetšo a taodišwana ye ke go nyakišiša, go ya ka kgopolo ya bangwalapukuntšu, ka fao baboleli ba Sesotho sa Leboa ba dirago kgetho ya mantšu magareng ga maadingwa le mantšu a setlogo polelong ye. Dipolelo tše di hweditšwego go tšwa go bakgathatema ba e lego baboleli ba Sesotho sa Leboa, banna le basadi, ba lekgolo (100) ba mengwaga ya go fapana, maemo a a fapanego a thuto, ba ba dulago mafelong ao a fapafapanego, bj.bj. di tla fetlekwa. Go ipontšha gore le ge dipolelo tša nyakišišo ye di laetša gore bontši bja bakgathatema bo kgetha go šomiša mantšu a setlogo go ena le maadingwa, bangwadi ba

* An earlier version of this article was presented as a paper at the Seventh International Conference of the African Association for Lexicography, organised by the Dictionary Unit for South African English, Rhodes University, Grahamstown, 8–10 July 2002.

dipukuntšu ba swanetše go phafošwa mabapi le diphetogo tše di ka bago gona pateroneng ya kgetho ya tšhomišo ya mantšu. Dipolelo tša nyakišišo ye di bapetšwa le ka moo mantšu a tšwelelago kgafetšakgafetša go tšwa khophaseng gammogo le ka fao dipukuntšu tše di šetšego di le gona di šomišitšego mantšu ao ka gona.

Mantšu a bohlokwa: GO HLAMA PUKUNTŠU, PUKUNTŠU, TSELA YA GO NGWALA MANTŠU KA PUKUNTŠUNG, SESOTHO SA LEBOA (SEPEDİ), LEADINGWA, LENTŠU LE LE SOTHO FADITŠWEGO, LENTŠU LA SETLOGO, LENANEOPOTŠIŠO, KHOPHASE, GO HLALOŠA POLELO, GO HLALOŠA POLELO KA GO FA KGETHO, GO LAELA KA GA TŠHOMIŠO YA POLELO, PATERONE YA KGETHO YA MANTŠU

Loan words versus their (more) indigenous counterparts

The use of loan words versus their (more) indigenous counterparts is studied in various disciplines such as science and technology, socio-linguistics, syntax and semantics, morphology and phonology, and not the least, lexicography. From a lexicographic perspective the issue to be investigated links well with one of the fundamental dichotomies in dictionary compilation, namely prescriptiveness versus descriptiveness. The coinage of loan words by language boards, for instance, is a prescriptive activity in itself. Within a descriptive approach towards dictionary compilation, it is imperative to know to what extent loan words in contrast to their so-called 'traditional' or 'original' counterparts are actually and actively used, and to study preferences of the target-user group of an envisaged dictionary in this regard. Not only should the lexicographer strive to lemmatise and lexicographically treat those words most likely to be looked for by the target users, the lexicographer should also be sensitive towards potential changes in preferences regarding the use of loan words versus their (more) indigenous counterparts.

In order to study this preference for Northern Sotho, and this with the compilation of a Northern Sotho dictionary in mind, a survey was conducted among 100 randomly selected mother-tongue speakers of varying ages, professions, regions, etc.¹ From the above it should be clear that only loan words with Northern Sotho counterparts are studied in this article, with the main focus on their macrostructural treatment.²

Prescriptiveness versus descriptiveness, and proscriptiveness

Traditionally dictionaries were regarded as a source of linguistic information. A good example is Samuel Johnson's *A dictionary of the English language* published in 1755. Wells (1973), Al-Kasimi (1977), Gouws (1989), Landau (2001) and many others agree that this dictionary is generally regarded as the reference work which normalised the English language. For centuries the lexicographer was seen as judge or ruler of the language who was to prescribe good and

to point out unacceptable use. According to this view the dictionary should rectify and cleanse the language, preserve its purity, lengthen its duration, correct or ban improprieties and absurdities, sensor faulty usage and repress anomaly. This authoritarian tradition, however, collapsed mainly due to one simple principle, viz. *language change*. In the words of Philip Gove (1961³: 4a), the Editor-in-Chief of *Webster's Third*:

English like other living languages is in a metabolic process of constant change. The changes affect not only word stock but meaning, syntax, morphology and pronunciation.

Indeed, in contrast to the prescriptive approach stands the descriptive approach with the focus on actual language usage. Gove (1961a: 13, quoted in Al-Kasimi (1977: 84)) writes in a letter to *Life Magazine*:

The responsibility of a dictionary is to record the language, not set its style. For us to attempt to prescribe the language would be like *Life* reporting the news as its editors would prefer it to happen.

Prinsloo (1992: 10) rightfully emphasizes that the then Northern Sotho Language Board made an invaluable contribution towards the clarification, systematisation, standardization and coining of new terms for Northern Sotho in, for example, religion, news broadcasting, mathematics, general science, etc. The last Terminology and Orthography for Northern Sotho, for short T&O, produced by this language board was published in 1988 (Departmental Northern Sotho Language Board 1988⁴). The Language Board adopted a sensible approach in being prescriptive in the coinage and approval/disapproval of terminology on the one hand, while still placing a high premium on actual usage as criterion for acceptability on the other hand. The Language Board (T&O: 3) allows for more than one option rather than attempting to enforce just one term while suppressing others:

It is anticipated that practical usage of the terms offered will prove that some of them can be replaced by other more commendable ones.

The same holds true for the Language Board's attitude towards certain entries in existing dictionaries (T&O: 1):

[C]ertain new terms and concepts are included which appear in some dictionaries but which are not generally accepted in the language yet. In such instances the Northern Sotho Language Board devised a term of their own which in their opinion is more appropriate.

Let there be no doubt that *actual usage* and not the sentiments of a language board will eventually determine whether a word should be included into or

omitted from a dictionary. Zgusta (1971: 187) strongly argues:

Lexicographers can coin new expressions, they can normalize their form and meaning, they can systematize and clarify the old ones, they can help in an endless number of such exceedingly useful and necessary tasks. The real life of a language, however, is in its use; and the definitive, full-fledged stabilization of the standard national language is brought about by its being really and extensively used in literature and in oral communications of all types.

It is interesting to note, for example, that as far as the months of the year are concerned, the then Language Board prescribes the use of "Sothoised" terms like **Matšhe** 'March' instead of **Hlakola**, **Aprele** 'April' instead of **Moranang**, etc. If the actual usage of the traditional terms would gain in importance or even prevail, a normalising board would have to back down and take a more descriptive approach.

As will be indicated below, the names for the months of the year seem to be an exception in the general preference pattern, namely that in this case loan words are preferred to the traditional words. Note, however, that there might be other factors influencing the choice in respect of months of the year, such as the lack of a one-to-one correlation between the traditional name and the actual month to which it refers. As a matter of fact, Ziervogel and Mokgokong (1975: 828, 1022, 1039) state that **Moranang** can refer to both April and June, **Phato** to both August and October, **Pherekgong** to both January and March, etc. which is of course unacceptable in real life situations where activities are punctually scheduled in terms of date and time. In simple terms it means that two people can agree on meeting each other on, say, the 1st of **Pherekgong** but then miss each other by two months.

More recently, metalexigraphers such as Bergenholtz (2002: 12) have introduced the term proscriptiveness:

I wish to suggest a specification and the introduction of a new term, proscription, which in actual fact is only new as a term, since the phenomenon itself is known in many dictionaries around the world. What is meant is the suggested use of a certain variant based on an exact analysis of an empirical basis without prohibiting other existing variants.

Within a proscriptive framework, the paradox that even a descriptive dictionary has a prescriptive effect on the target users is taken into account (compare Bergenholtz 2001). With reference to the current study, the task of the proscriptive compiler of dictionaries for Northern Sotho in terms of loan words versus their (more) indigenous counterparts, is thus to reflect user *preferences* in the selection of lemma signs on the macrostructural level as well as in the extent of treatment on the microstructural level, while still allowing for other existing variants. Consequently, within a proscriptive approach towards the lemmatisation of loan words in contrast to their 'traditional' or 'original' coun-

terparts, it is imperative for the lexicographer to know what the preferences of the target-user group are in this regard. This is exactly what will be pursued in the following paragraphs.

The survey

The scope of the research was limited to a random sample of 100 respondents, all mother-tongue speakers of Northern Sotho. The survey was conducted in May 2002, and the breakdown in terms of gender, age, birthplace and education/job is presented in condensed format in (1).

(1) Loan word survey — Basic respondents' data (N = 100)

Gender	Male: 47 Female: 53
Age	16: 1 17: 2 18: 8 19: 6 20: 5 21: 6 22: 6 23: 7 24: 5 25: 7 26: 4 27: 7 28: 5 29: 8 30: 4 31: 2 32: 5 34: 1 35: 2 36: 1 37: 1 38: 1 48: 1 53: 1 54: 2 65: 2
Birthplace	Atteridgeville, Bochum, Botlokwa, Burgersfort, Buschbuckridge, Ga-Dikgale, Ga-Kgoroshi, Ga-Ledwaba, Ga-Maja, Ga-Mamabolo, Ga-Masemola, Ga-Mashabela, Ga-Mashashane, Ga-Matlala, Ga-Molepo, Ga-Moraba, Ga-Mphahlele, Ga-Nkwana, Ga-Rankuwa, Ga-Sekhukhune, Giyani, Glencowie, Kgapanne, Lebowa-Kgomo, Lenenye, Magatle, Mahwelereng, Mohodi, Moletjie, Mphakane, Pietersburg, Potgietersrus, Pretoria, Seshego, Siyabuswa, Steilloop, Turfloop
Education/Job	Accountant, Assistant Lecturer, Bank teller, Business person, Cashier, Consultant, Co-ordinator, Dental Assistant, Dietician, Driver, Engineer, Fashion designer, Housewife, Lawyer, Learner, Lexicographer, Librarian, Marketing official, Nurse, Professor, Student, Teacher, Unemployed

A total number of 64 single words were presented in pairs to the respondents, thus 32 pairs each containing a loan word and a (more) indigenous counterpart, e.g. **radio** versus **seyalemoya** 'radio', or **dimonamonane** versus **malekere** 'sweets', etc. Respondents were asked to mark the alternative(s) which they would like to see included in a Northern Sotho dictionary. A third column was added for comments and suggestions of other words considered to be still better than the two choices offered. Respondents were also invited to report spelling errors or to suggest improvement of spelling, and even to motivate why a word should be included or excluded from the dictionary. Finally, an informal conversation was conducted with each respondent in order to obtain additional information and an overall impression. A typical example of a completed questionnaire is reproduced in the Appendix. Compare also the translated version of the questionnaire in (2).

- (2) Loan word survey — Translation of the questionnaire (with in Columns **A** and **B**: **L** = loan word, and **I** = (more) indigenous word; and in Column **C**: the English translation)

Hello!		
Choose from A and B those words which, according to you, should be included in a Northern Sotho dictionary. You may choose either A or B , or both, and in C you may write any relevant comment, such as: any other word that you think is better than the two words already given in A and B , if you see a spelling mistake for the words given you may provide the correct spelling, or if you wish you may give reasons why you say the word you have chosen or given is the one to be included in the dictionary, or, conversely, you may give the reason why you say the other words should not be included in the dictionary, etc.		
A	B	C
L radio	I seyalemoya	<i>a radio</i>
I sefatanaga	L mmotoro	<i>a car</i>
I mmila	L seterata	<i>a street</i>
L watšhe	I sešupanako	<i>a watch</i>
L terene	L setimela	<i>a train (distracter)</i>
I Pherekgong	L Janeware	<i>January</i>
L dikšinare	I pukuntšu ³	<i>a dictionary</i>
I šoma	L bereka	<i>to work</i>
I fofo	L teye	<i>tea</i>
L sekerete	I motsoko	<i>a cigarette; tobacco (distracter)</i>
L khomphutha	L computer	<i>a computer (distracter)</i>
L amalantshe	L ambulantshe	<i>an ambulance (distracter)</i>
L tonki	I molahlwaleboya	<i>a donkey</i>
L sebara	I molamo	<i>a brother-in-law</i>
I nkadingala	L sankoma	<i>a witchdoctor</i>
L nalete	I lemao	<i>a needle</i>
L sekoloto	I mokitlana	<i>debt</i>
I bookelo	L sepetlela	<i>a hospital</i>
I letsikangope	L lefasetere	<i>a window</i>
L tšhoko	I motaga	<i>chalk</i>
I phapoši	L kamora	<i>a room</i>
I dimonamonane	L malekere	<i>sweets</i>
L reila	I otlela	<i>to drive</i>
I befelwa	L kwata	<i>become angry</i>
forminal	sephekgo	<i>(distracter)</i>
I ntlanya	L paesekela	<i>a bicycle</i>
L pata	I tsela	<i>a road</i>
L sepete	L garafo	<i>a spade (distracter)</i>
I sephuthelwana	L phasela	<i>a parcel</i>
L polane	I leano	<i>a plan</i>
L ranta	I lesome	<i>Rand</i>
L newspaper	L kuranta	<i>a newspaper (distracter)</i>
We assure that all the information you have provided will be processed anonymously. You are kindly requested to provide us with the following for research purposes:		
Age (e.g.: 29)		
Home language (e.g. Selobedu)		
Birthplace (e.g. Ga-Matlala)		
THANK YOU FOR YOUR COOPERATION!		

Note that the pairs were not presented in a fixed order. In some cases the loan word **L** is given first followed by the (more) indigenous word **I**, sometimes it is the reverse. The loan words themselves are all adopted from Afrikaans, English or Zulu — occasionally Sothoised, at other times direct borrowings. The corresponding (more) indigenous forms are either new coinages in Northern Sotho, shifts in meaning from existing Northern Sotho words, or simply traditional/original words.

Analysing the survey

Basically, there are three levels on which the questionnaires were analysed. Firstly, each respondent's input was analysed in isolation. This resulted in 100 user profiles (Level 1). Secondly, these 100 profiles were summed, upon which general tendencies became evident (Level 2). Thirdly, each **L-I** pair was studied separately, which provided interesting and highly specific data for each specific **L-I** pair. For each of these pairs, frequency counts derived from a corpus as well as the treatment in all currently available dictionaries were also taken into account (Level 3). Space considerations unfortunately do not allow us to present the outcome on these three levels exhaustively. Rather, some representative findings will be singled out.

The 100 filled-in questionnaires were processed in spread-sheet format, reflecting each pair and the full statistical response for every pair — viz. a score for **L**, a score for **I**, and a score for both — and this for Respondent 1, 2, ... up to Respondent 100. See (3).

(3) Analysing the loan word survey — Levels 1 and 2

	L-I	1	2	3	...	36	...	98	99	100	only L only I both	L or both I or both
radio	L		1			1	1	17%	42
seyalemoya	I	1	1	1	...	1	...	1	1	1	57%	82
		b	ab	b	...	b	...	b	ab	ab	25%	
...
malekere	L	1	1	1	...	1	...			1	24%	50
dimonamonane	I	1	1		1	1	1	45%	71
		ab	ab	a	...	a	...	b	b	ab	26%	
...
Age		22	32	27	...	18	...	29	37	24		
Home language		S	S	S	...	S	...	S	S	S		
Birthplace		Siy	Ml	Mt	...	Mt	...	Mj	Sk	Pta		
Gender		M	F	M	...	F	...	F	M	F		
Education/Job		S	L	S	...	S	...	U	T	S		
Address		M	S	U	...	Pta	...	Ptb	Ptb	Pta		

From (3) one can for instance see that 57% of the respondents opted for **seyalemoya**, 17% for **radio**, and as many as 25% for both. (Only one respondent did

R slightly favour **L** (42% vs. 38%). C seemingly (cf. below) only has **I**, and D agree on **I**. However, **L** appears as **Janaware** (19/0) in New E, T&O, Pop. and Grb., which is thus the form to be included in the D (and not **Janeware** (0/0)). The plural form for **Janaware** is **diJanaware** and for **Pherekgong** it is **diPherekgong**, yet C does not contain any of the two plurals. 16% of R agree on both **L** and **I**, while 4% suggest nothing. The preferred form by both R and C is thus **L**, but with spelling **Janaware**. D are therefore very wrong, as they mainly focus on **I**. Klein has nothing. T&O (p. 23): "The names of the months of the year are Sothoised instead of using the Sotho names, e.g.: *Janaware* (instead of *Pherekgong*) ...". One of the respondents, aged 21, who is a student, female and a Northern Sotho speaker wrote: "Go ya ka nna re swanetše re hlalose mantšu ao a šomišwago tšatši ka tšatši re se ke ra šomiša Sepedi sa kgale seo e le go [sic] gore bana ba matšatši a ba ka se se kwešiše, go swana le bo 'Pherekgong' [sic]." ("According to me we should only include those words that are used every day and avoid using old Sepedi words like **Pherekgong**, which children of today do not understand.") 15% of the respondents say the correct spelling is **Janaware** (19/0) not **Janeware** (0/0) which means that the correct word to be included in D is indeed **Janaware**. **Pherekgong** may also be included, but with a note pointing out the potential confusion (January vs. March), and a cross-reference to **Janaware**.

The extract shown in (5) is taken verbatim from the survey analysis. Even from this single (cryptic) **L–I** sketch the proscriptive approach should be evident: the target users and the corpus favour the loan word **L** (descriptiveness), and this happens to correspond with the Language Board's suggestion (prescriptiveness). However, as the target users and the corpus also indicate that the traditional word **I** has a right to be included in the dictionary, this 'variant' should also be treated on both the macro- and microstructural levels, albeit with the necessary cross-references and cursory notes (proscriptiveness).

Analysing the survey — Distracters

The **Janeware–Pherekgong** pair is one example of an **L–I** pair where the one-to-one correlation does not truly hold, yet where there is still a large overlap. Another example is shown in (6).

(6) Analysing the loan word survey — **mmila–seterata** pair

Items	%	Σ	Fsg Fst	Fpl -	Z&M	New E	Klein	T&O	V Wyk	P&S	P et al.	Pop.	Grb.
mmila	53	79	226	57	✓	✓	✓	-	✓	✓	✓	✓	✓
seterata	16	42	9	8	✓	✓	✓	✓	✓	✓	✓	✓	✓
both	26	-	-	-	-	-	-	-	-	-	-	-	-
∅	5	5	-	-	-	-	-	-	-	-	-	-	-

Taken at face value, the two options seem semantically different with **mmila** 'footpath' versus **seterata** 'street'. However, when the respondents' considerations are studied, it becomes clear that **mmila** and **seterata** can indeed be semantically linked via 'road' since some respondents and dictionaries seem to

imply a semantic continuum for the end points **mmila** versus **seterata** in foot-path ↔ road ↔ street. This overlap is obvious in the treatment of **mmila** and **seterata** in, e.g., Ziervogel and Mokgokong (1975):

(7)(a) **MMILÁ** ... footpath, trail (of game), road, street, side-walk

(7)(b) **-TÉRÁTA, se-** ... street

Pairs such as **Janeware–Pherekgong** and **mmila–seterata** were purposely inserted into the survey in an attempt to trigger comments that would enable the deduction of the overriding sentiment with regard to loan words versus (more) indigenous words. For the latter pair, several respondents aged between 21 and 36, from different areas in both the Limpopo and Gauteng provinces, state that **seterata** is equivalent to **mokgotha** (13/47). This is (partly) true and clearly points in the direction of a preference for indigenous words.

Apart from pairs which only partly overlap, carefully selected *distracters* were also built into the questionnaire. These have been marked as such in (2) above. Their main aim was to verify the quality of the respondents' feedback. Some distracter-pairs consist of two loan words **L**, for others the two options have very different meanings, and a third category was added just to find out if the respondents themselves were consistent. Consider (8) and (9) as examples of the first category.

(8) Analysing the loan word survey — **newspaper–kuranta** pair

Items	%	Σ	Fsg Fst	Fpl -	Z&M	New	E	Klein	T&O	V	Wyk	P&S	P et al.	Pop.	Grb.
newspaper	1	7	0	0	-	-	-	-	-	-	-	-	-	-	-
kuranta	82	88	168	84	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
both	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
∅	11	11	-	-	-	-	-	-	-	-	-	-	-	-	-

(9) Analysing the loan word survey — **khomphutha–computer** pair

Items	%	Σ	Fsg Fst	Fpl -	Z&M	New	E	Klein	T&O	V	Wyk	P&S	P et al.	Pop.	Grb.
khomphutha	61	68	18	4	-	-	-	✓	-	-	-	-	-	✓	✓
computer	10	17	0	0	-	-	-	-	-	-	-	-	-	-	-
both	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
∅	22	22	-	-	-	-	-	-	-	-	-	-	-	-	-

In these examples the respondents were presented with a direct borrowing (not the purpose of the investigation) and a (Sothoised) loan word. In (8) just one respondent preferred **newspaper** as the only option and 6 allowed for both, while in (9) 10 opted for **computer** and another 7 for both. These distracter-pairs thus managed to discriminate well. Compare also the following remark by one of the respondents:

Dealing with two non-words, still half of the respondents opted for what they perceived to be an original word.⁵ At first glance this might question the reliability of the feedback of a huge number of the respondents. However, it is a clear confirmation of the respondents' general preference for indigenous words over loan words to be treated in a dictionary. Compare the following telling remark from one of the respondents in this regard:

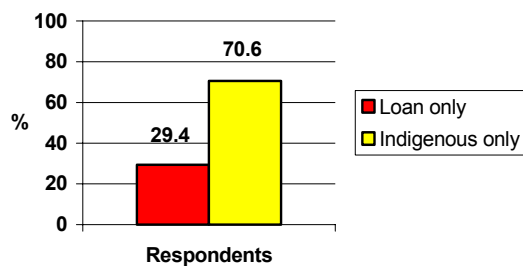
The word **sephkego** is unknown to me. I chose it because I want to know what it means and I think that if it is included in the dictionary, most of the people who are like me and don't know anything about the word will know and understand it once it is explained in the dictionary. I therefore request that you include it.

General findings of the survey

The general findings of the survey will now be summarised. For the statistics presented below, the data for the randomly-interspersed distracters have been excluded.

Firstly, one can calculate the percentage of respondents in favour of loan words as opposed to the respondents in favour of their (more) indigenous counterparts. The results are shown in (13).

(13) Respondents' preference for loan words vs. their (more) indigenous counterparts



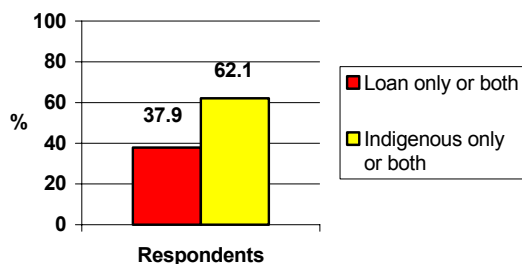
This graph clearly indicates that an overwhelming more than two-thirds majority of those who only accept one option are in favour of the indigenous word. The following remark by one of the respondents is quite revealing in this respect:

I personally believe that our language will lose value if more and more words from other languages are accepted in Northern Sotho. If we have original Northern Sotho words, why do we have to loan words from other languages? I don't see the necessity for us to loan words from other languages if we have our own original words. Only those words which do not exist in our language could be

loaned from other languages, for example AIDS. The Zulus have managed to formulate their own word for AIDS and called it **nxolazi** [sic], why can't we the Northern Sotho people do the same and stop loaning from other languages?

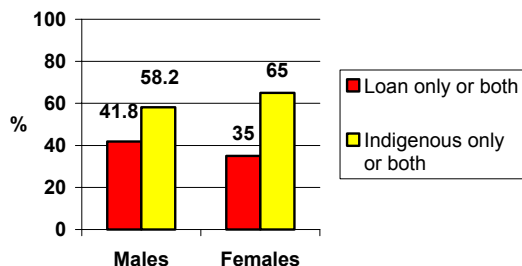
If one studies the overall pattern for 'loan + both' versus 'indigenous + both', then one arrives at the pattern shown in (14).

(14) Respondents' preference for 'loan words or both' vs. their '(more) indigenous counterparts or both'



Secondly, differentiating between preferences of males versus females results — with the data shown in (14) as a point of departure — in the percentages shown in (15).

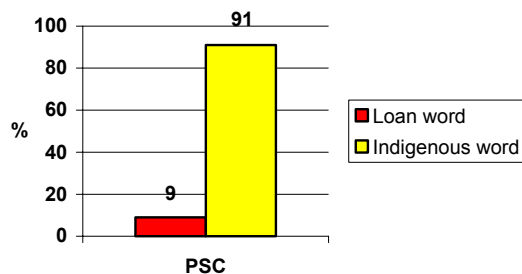
(15) Males' and females' preference for 'loan words or both' vs. their '(more) indigenous counterparts or both'



From (15) one can clearly see that males tend to prefer the inclusion of loan words in dictionaries more than females do, as they suggest 6.8% more of them on average.

Thirdly, the above findings should be compared with the occurrence of loan words versus (more) indigenous counterparts in the corpus. The latter is done in (16).

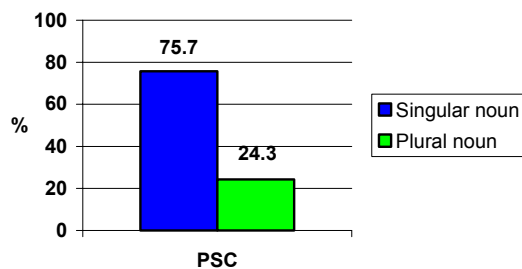
(16) Loan words vs. their (more) indigenous counterparts in PSC



The Pretoria Sepedi Corpus (PSC) is based for 100% on written sources. Ten times more indigenous words appear in PSC than their loan counterparts. This is a significant and clear indication that indigenous words take overwhelming preference over loan words in written texts. If a dictionary were solely based on corpus data, chances would be good that loan words would turn out to be undertreated compared to the respondents' preferences. The difference between the respondents' preferences for loan words and the corpus attestations for those same loan words is as high as 20.4% (= 29.4 – 9.0). In order for a Northern Sotho dictionary to reflect the true needs of the community, it is thus clear that this research reveals an important hiatus in an approach *solely* based on corpus data.

The graph shown in (17) indicates that the nouns that were the topic of this study appear more often in their singular form than in their plural form in PSC.

(17) Singular vs. plural nouns in PSC

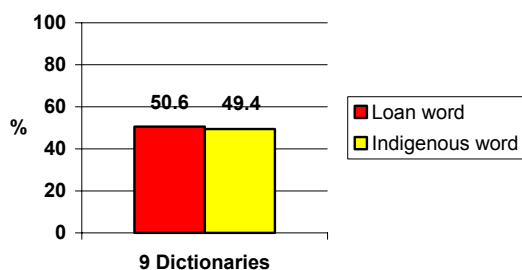


This simply means that, as far as the microstructure of a Northern Sotho dictionary is concerned, the studied nouns should rather be treated and exemplified in the singular. A clear exception is of course **malekere** / **dimonamonane**

'sweets' discussed above.

Fourthly, the respondents' input and the corpus data should also be compared with the overall treatment in the 9 currently-available dictionaries. The latter is summarised in (18).

(18) Loan words vs. their (more) indigenous counterparts in 9 Northern Sotho dictionaries



From (18) one can see that, on average, the 9 available dictionaries for Northern Sotho treat both loan words and their (more) indigenous counterparts on a par. In order to see whether or not this situation would suit future dictionary users, the next series of analyses needs to be presented first.

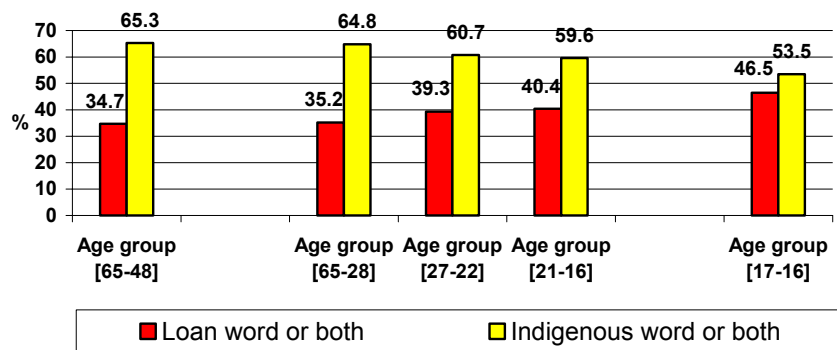
Fifthly, and lastly, the sum of the user profiles should also be brought into account. As far as the profiles for rural versus urban birthplaces are concerned, the results are unfortunately inconclusive, probably as a result of the fact that most respondents commute frequently between, and live in, both places. This is an area which should be researched further. When the age profiles are summed, however, a clear pattern emerges. Indeed, the most significant finding of the survey is obtained when the respondents' answers are broken down according to age groups. In the shaded columns of (19) the respondents' feedback is divided into three major age groups, namely 16-21, 22-27 and 28-65 (each containing roughly the same number of respondents). On both sides of this shaded block the preferences for L versus I of the extreme age groups 16-17 and 48-65 are also indicated.

(19) Distribution across age groups of the respondents' preference for 'loan words or both' vs. the respondents' preference for their '(more) indigenous counterparts or both' — 1. data

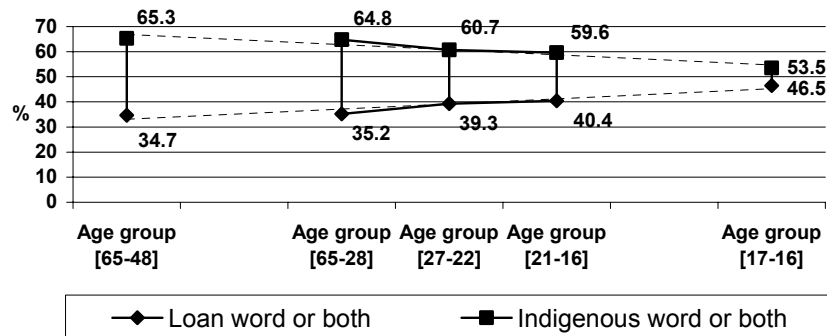
	<i>upper extreme</i> Age group [65-48]	Age group [65-28]	Age group [27-22]	Age group [21-16]	<i>lower extreme</i> Age group [17-16]
Loan word or both	34.7	35.2	39.3	40.4	46.5
Indigenous word or both	65.3	64.8	60.7	59.6	53.5

Thus, read from left to right, the preference for loan words gradually increases from 34.7% for the age group 48-65 to 46.5% for the age group 16-17. Conversely, the preference for their (more) indigenous counterparts decreases from 65.3% to 53.5% across the same spectrum. This is shown graphically in (20) and (21) by means of respectively histograms and trendlines.

(20) Distribution across age groups of the respondents' preference for 'loan words or both' vs. the respondents' preference for their '(more) indigenous counterparts or both' — 2. histograms



(21) Distribution across age groups of the respondents' preference for 'loan words or both' vs. the respondents' preference for their '(more) indigenous counterparts or both' — 3. trendlines



Graph (21) clearly indicates that the gap between preference for indigenous words and preference for loan words narrows as the age of the respondents decreases. This discovery is of crucial importance to dictionary compilers, especially for the compilation of dictionaries for younger mother-tongue speak-

ers of Northern Sotho as target users. At the same time this result renders justification for the current dictionary situation as indicated in (18) above, where an almost equal treatment of loan words versus indigenous counterparts was found.

A typical manifestation of this phenomenon could be the preference pattern for **teye** versus **fofo**, both meaning 'tea', where it is clear that most of the respondents who chose the indigenous **fofo** are aged between 40 and 53, whereas those who opted for the loan word **teye** belong to the younger generation (compare in this regard the findings of Slabbert and Finlayson 1999).

Conclusion

In this article, loan words versus their (more) indigenous counterparts in Northern Sotho were studied from a lexicographic perspective. Within a prescriptive approach to dictionary compilation, the selection of the main variant is based on the analysis of a sound empirical basis, whilst that empirical basis also provides the lesser-important items to be treated. It was shown that the by-now standard empirical basis consisting of data derived from an electronic corpus is not good enough for the treatment of loan words, and that fieldwork is imperative. The latter was achieved by means of a survey conducted among 100 mother-tongue speakers of Northern Sotho.

By studying the respondents' comments as a whole, it is clear that they prefer the (more) indigenous words to be treated in dictionaries, and that loan words should only be used if there is no good alternative in Northern Sotho. Quite a number of respondents even suggest that words should be coined in order to have a Northern Sotho word instead of an adoptive from other languages. Where offered a direct borrowing and a (Sothoised) loan as only options, the (Sothoised) loan is preferred to the direct borrowing. This thus suggests the following preference hierarchy: *indigenous word* → *Sothoised loan word* → *direct borrowing*.

The most important finding of this study is that younger respondents seem to accept loan words much more easily than the older generation. This might be a result of the intensified influence from other languages in both rural and urban areas, as well as a direct consequence of the fact that most teenagers are no longer enrolling for Northern Sotho as a school subject. Also, older people tend to favour so-called 'old words' that are no longer known by the youth of today.

All in all, however, first preference is still given to (more) indigenous words over loan words whenever there is a choice. This pattern is rapidly changing, and today's dictionaries should definitely pay more attention to loan words than the dictionaries compiled half a century ago. A watchful eye will have to be kept on this evolving preference pattern.

Endnotes

1. Nowadays, one often finds so-called politically correct phraseologies such as "primary language" and "home language" for "mother tongue".
2. We are aware of the fact that a great deal of linguistic research has been devoted to what linguists variously call 'loan words', 'borrowings', 'adoptives', 'transliterations', etc. In this article, however, the focus is on lexicography, and we only use those terms *in contrast* to their traditional/original counterparts. We refer to the latter as '(more) indigenous counterparts'.
3. Strictly speaking **pukuntšu** is a combination of **puku** 'book' (a loan word) and **lentšu** 'word' (an indigenous word).
4. The noun **setimela**, e.g., has a class prefix of class 7 **se-**, while **terene** has no class prefix.
5. The non-words **forminal** and **sephkego** were 'created' in such a way that they resemble genuine English and Northern Sotho words respectively. Actually, **forminal** was derived from a permutation of sections of the English word 'informal'. Unfortunately, an Internet search with Google (<http://www.google.com/>) returns 56 hits, revealing that **forminal** is a technical neurology term used in for instance 'Forminal Stenosis of Cervical Spine' or 'Extra Forminal Non-discogenic Lumbar Nerve Entrapment as a Cause of Sciatica'. We can however safely assume that no respondent knew this technical sense. The non-word **sephkego** seems to be a Northern Sotho noun belonging to class 7 **se-**. Fortunately, no Internet pages were found containing the non-word **sephkego**.

List of Dictionary and Corpus Abbreviations

Grb	Gerber 2000
Klein	Ziervogel and Mokgokong 1988 ⁴
New E	Kriel 1976 ⁴
P et al.	Prinsloo, Sathekge and Kapp 1997
P&S	Prinsloo and Sathekge 1996
Pop.	Kriel, Prinsloo and Sathekge 1997 ⁴
PSC	Pretoria Sepedi Corpus (5.8 million running words)
T&O	Departmental Northern Sotho Language Board 1988 ⁴
V Wyk	Kriel, Van Wyk and Makopo 1989 ⁴
Z&M	Ziervogel and Mokgokong 1975

References

- Al-Kasimi, Ali M. 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Bergenholtz, Henning. 2001. Proskription, oder: So kann man dem Wörterbuchbenutzer bei Textproduktionsschwierigkeiten am ehesten helfen. Lehr, Andrea, Matthias Kammerer, Klaus-Peter Konerding, Angelika Storrer, Caja Thimm and Werner Wolski (Eds.). 2001. *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*: 499-520. Berlin: de Gruyter.

- Bergenholtz, Henning.** 2002. Towards a User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. De Schryver, Gilles-Maurice (Ed.). 2002. *AFRILEX 2002, Culture and Dictionaries, Programme and Abstracts*: 12-13. Pretoria: (SF)² Press.
- Departmental Northern Sotho Language Board.** 1988^a. *Northern Sotho Terminology and Orthography No. 4 / Noord-Sotho terminologie en spelreëls No. 4 / Sesotho sa Leboa mareo le mongwalo No. 4*. Pretoria: Government Printer.
- Gerber, Hendri H.** 2000. *Woordeboek Afrikaans–Noord-Sotho / Pukuntšu Seburu–Sesotho sa Leboa*. Eldoraigne: Arbeidsprestasie BK.
- Gouws, Rufus H.** 1989. *Leksikografie*. Pretoria: Academica.
- Gove, Philip B.** (Ed.). 1961³. *Webster's Third New International Dictionary of the English Language*. Springfield: Merriam-Webster.
- Gove, Philip B.** 1961a. Letter to the Editor. *Life Magazine*, 17 November 1961: 13.
- Johnson, Samuel.** 1755. *A dictionary of the English language: in which the words are deduced from their originals, and illustrated in their different significations by examples from the best writers; to which are prefixed, A history of the language, and an English grammar*. London: Printed by W. Strahan, for J. and P. Knapton.
- Kriel, Theunis J.** 1976^a. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English*. Johannesburg: Educum Publishers.
- Kriel, Theunis J., D.J. Prinsloo and Bethuel P. Sathekge.** 1997^a. *Popular Northern Sotho Dictionary, Northern Sotho–English, English–Northern Sotho*. Cape Town: Pharos.
- Kriel, Theunis J., Egidius B. van Wyk and S.A. Makopo.** 1989^a. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.
- Landau, Sidney I.** 2001. *Dictionaries: The Art and Craft of Lexicography (2nd edition)*. Cambridge: Cambridge University Press.
- Prinsloo, D.J.** 1992. Prescriptiveness versus Descriptiveness — an Approach for Northern Sotho Lexicography. *South African Journal of African Languages* 12(1): 7-11.
- Prinsloo, D.J. and Bethuel P. Sathekge.** 1996. *New Sepedi Dictionary, English–Sepedi (Northern Sotho), Sepedi (Northern Sotho)–English*. Pietermaritzburg: Shuter and Shooter.
- Prinsloo, D.J., Bethuel P. Sathekge and Lizeth Kapp.** 1997. *Nuwe Sepedi Woordeboek, Afrikaans–Sepedi (Noord Sotho), Sepedi (Noord Sotho)–Afrikaans*. Pietermaritzburg: Shuter and Shooter.
- Slabbert, Sarah and Rosalie Finlayson.** 1999. The Future of the Standard African Languages in the Multilingual South African Classroom. Rosalie Finlayson (Ed.). 1999. *African Mosaic: A Festschrift for J.A. Louw on his 75th Birthday*: 235-249. Pretoria: Unisa Press, University of South Africa.
- Wells, Ronald A.** 1973. *Dictionaries and the Authoritarian Tradition. Study in English Usage and Lexicography*. Janua Linguarum. Series Practica, 196. Berlin: Mouton de Gruyter.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. The Hague: Mouton.
- Ziervogel, Dirk and Pothinus C.M. 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.
- Ziervogel, Dirk and Pothinus C.M. Mokgokong.** 1988^a. *Klein Noord-Sotho woordeboek, N.-Sotho–Afrikaans–English, Afrikaans–N.-Sotho, English–N.-Sotho*. Pretoria: J.L. van Schaik.

Appendix: Reproduction of a randomly-selected filled-in questionnaire (# 36)

36

Thobela!

Kgetha go tšwa go A le B mantšu ao o naganago gore a loketšwe go ba gona ka mo gare ga pukuntšu ya Sesotho sa Leboa. O ka no kgetha ka mo go A le ka go B gomme o ngwale se sengwe ka mo go C, bjalo ka; lentšu leo o naganago gore le kaone go phala a mabedi ao a filwego ka go A le B, ge o bona phošo ya mopeleto mo manišung ao a filwego ka go A le B o ngwale mopeleto wa nnete wa mantšu ao, ge o rata o ka no fa mabaka a gore ke ka lebaka la eng o re lentšu leo le loketšwe go ba ka gare ga pukuntšu goba ga la lokelwa, bj. bj.

A	B	C
radio	seyalemoya ✓	Radio ke lentšu leo le atlingwilewego.
sefatanaga ✓	mmotoro	mmoto le tšwa go English.
mmila ✓	seterata	Tsela e ka ba le lengwe leo le ka lokelwa.
watšhe	sešupanako ✓	
terene	setimela ✓	Tšhutšumakgale le tona e ka šomišwa
Pherekgong ✓	Janeware	
dikšinare	pukuntšu ✓	
šoma ✓	bereka	
fofo	teye ✓	Teye fofo ke sona sesotho cupša e gatho-aelega.
sekerete	motsoko ✓	
khomphutha ✓	computer	Computer ke English ga se sesotho.
amalantshe	ambulantshe	
tonki	molahlwaleboya ✓	
sebara	molamo ✓	
nkadingala ✓	sankoma	
nalete	lemao ✓	
sekoloto	mokitlana	Ka moka a ka šomišwa.
bookelo ✓	sepetlela	
letsikangope ✓	lefasetere	
tšhoko	motaga	Nke e tšeke (a e gona ko gae).
phapoši ✓	kamora	
dimonamonane	malekere ✓	
reila	otlela ✓	
befelwa ✓	kwata	Kwata ga se sesotho se nnete
forminal	sepheko ✓	
ntlanya	paesekela ✓	Le ge e e gore Paesekela ke lead ingwana.
pata	tsela ✓	
sepete	garafo	Garafo e swanetše go ba kgarafo gape.
sephuthelwana ✓	phasela	sepete le tona ga e sišwanane.
polane	leano ✓	
ranta ✓	lesome	lesome ga le kwagala ka go re R ke currency yo sa
newspaper	kuranta ✓	

Re go netefaleša gore tšohle tše o di ngwadilego fa e tla ba sephiri. O kgopelwa gore o tlatše ka mo fase: Mengwaga (Mohlala: 29) 1.8
 Leleme la ka gae (Mohlala: Selobedu) Sepedi
 O belegetšwe kae (Mohlala: Ga-Matlala) Ga-Matlala (C.phofo)

RE LEOGA GO KGATHA TEMA GA GAGO!