
Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: A Viability Study*

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

Abstract: The aim of this article is to investigate the viability of the compilation of a single bidirectional dictionary with a single lemma list for the Sesotho sa Leboa, Setswana and Sesotho → English side and a simultaneous treatment of the three Sotho languages in the articles of the English lemmas in the English → Sesotho sa Leboa, Setswana and Sesotho side of the dictionary. Specific attention will be given to selected macrostructural and microstructural aspects of such a compilation.

Keywords: SOTHO LANGUAGES, NGUNI LANGUAGES, BILINGUAL DICTIONARIES, COMMUNICATIVE EQUIVALENCE, CORPORA, USER PERSPECTIVE

Opsomming: Die samestelling van 'n tweerigtingwoordeboek wat Engels en die Sothotale oorbrug: 'n Lewensvatbaarheidstudie. Die doel van hierdie artikel is om die lewensvatbaarheid van die samestelling van 'n enkele tweerigtingwoordeboek met 'n enkele lemmalys vir die Sesotho sa Leboa, Setswana en Sesotho → Engelse kant en 'n gelyktydige bewerking van die drie Sothotale in die artikels van die Engelse lemmas in die Engels → Sesotho sa Leboa, Setswana en Sesotho kant van die woordeboek te ondersoek. Besondere aandag sal aan geselekteerde makrostrukturele en mikrostrukturele aspekte van so 'n samestelling gegee word.

Sleutelwoorde: SOTHOTALE, NGUNITALE, TWEETALIGE WOORDEBOEKE, KOMMUNIKATIEWE EKWIVALENSIE, KORPUSSE, GEBRUIKERSPERSPEKTIEF

Introduction

The aim of this article is to study the viability of a bidirectional dictionary bridging English and the Sotho languages[†]: Sesotho, Setswana and Sesotho sa Leboa. The focus will be on the advantages and disadvantages of such a single

* This article is based on a paper presented at the Tenth International Conference of the African Association for Lexicography (AFRILEX), organised by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

dictionary compared to three comparative bidirectional bilinguals: English–Sesotho/Sesotho–English, English–Setswana/Setswana–English and English–Sesotho sa Leboa/Sesotho sa Leboa–English, and the additional value it would have in the absence of bilingual dictionaries bridging African languages with each other. This will be an important achievement since publishers generally do not regard the compilation of separate dictionaries bridging the African languages with each other as economically viable. Compiling a dictionary with a single lemma list for the Sotho languages at this stage can indeed become the forerunner to such an eventual goal, i.e. true bidirectional dictionaries bridging the African languages with each other. It also has the potential to pave the way for an English ↔ Nguni languages dictionary.

The analysis and design of the macrostructure and microstructure will be based on existing bilingual dictionaries bridging English and a Sotho language and will be aimed at the same target users. The viability study will firstly be performed for a combined article for the Sotho languages usable mainly for basic receptive information, i.e. treatment limited to a translation equivalent or two, and secondly for combined articles where a more exhaustive treatment is given. The bilingual dictionaries analysed are *The New English–Northern Sotho Dictionary*, *English–Northern Sotho*, *Northern Sotho–English* (NEN) (Kriel 1976) for Sesotho sa Leboa, *Dikišinare ya Setswana English Afrikaans* (DS) (Snyman et al. 1990) and *Setswana–English–Setswana Dictionary* (SESD) (Matumo 1993) for Setswana and *Southern Sotho–English Dictionary* (SSED) (Mabille and Dieterlen 1988) for Sesotho.

The results of this study will hopefully enable prospective compilers to decide whether it is worthwhile to compile such a dictionary and to provide guidelines and examples for such a compilation. It is not possible to do a detailed analysis of all relevant lexicographic aspects within the limitation of a journal article and the discussion will therefore be limited to a number of key microstructural and macrostructural aspects.

The compilation of such a dictionary will require the combined skills of mother-tongue speakers of all four languages and corpora for these languages.

Impact and range of application for the Sotho and Nguni languages

A bidirectional English → {Sesotho, Setswana and Sesotho sa Leboa}, {Sesotho, Setswana and Sesotho sa Leboa} → English dictionary is comparable to *three* bidirectional bilingual dictionaries, English–Sesotho, Sesotho–English, English–Setswana, Setswana–English, and English–Sesotho sa Leboa, Sesotho sa Leboa–English, thus two directions for the envisaged model versus six sides for separately bridging English and a Sotho language. It could be argued that separate bilinguals bridging English and each of the Sotho languages do exist but in most cases they are out of print or in need of revision. The situation for bridging Sotho languages with each other is much less promising. At this stage in the development of South African lexicography publishers' interest is virtually

non-existent for bridging African languages with each other, thus little hope for bi-directional Sesotho sa Leboa ↔ Sesotho, Sesotho sa Leboa ↔ Setswana and Sesotho ↔ Setswana dictionaries.

A successful bidirectional English ↔ Nguni languages dictionary will be even more significant for the official South African Nguni languages because four languages are involved. A bidirectional English → {siSwati, isiXhosa, isiNdebele and isiZulu}, {siSwati, isiXhosa, isiNdebele and isiZulu} → English dictionary is comparable to *four* bidirectional bilingual dictionaries, English–siSwati, siSwati–English, English–isiXhosa, isiXhosa–English, English–isiZulu, isiZulu–English, and English–isiNdebele, isiNdebele–English. Thus two directions for the envisaged model versus eight sides for separately bridging English and a Nguni language. For bridging Nguni languages with *each other* six bilinguals, thus twelve directions should be compiled, i.e. isiZulu–isiXhosa, isiXhosa–isiZulu, isiZulu–siSwati, siSwati–isiZulu, isiZulu–isiNdebele, isiNdebele–isiZulu, isiXhosa–siSwati, siSwati–isiXhosa, isiXhosa–isiNdebele, isiNdebele–isiXhosa, and siSwati–isiNdebele, isiNdebele–siSwati.

Orthographic words versus lemmas in the Sotho languages

As a result of the disjunctive orthographies and the word-lemmatisation strategy followed in the Sotho languages, the relation between orthographic words versus lemmas is almost 1–1. In corpus-based lexicography this simplifies the compilation significantly since limited lemmatisation has to be performed upon the types generated from the corpora for the Sotho languages. Consider, for example, the 100 most frequently used words in Sesotho sa Leboa in ranking order on high to low frequency count:

- (1) a, moo, le, gago, go, bjale, ka, bolela, ba, tseba, o, mme, ke, dira, e, morena, ya, wena, re, monna, ge, taba, se, kua, wa, lego, gore, tšona, ga, fao, sa, mola, di, rena, mo, kgoši, be, bana, tša, mongwe, la, gwa, bona, leo, ye, bao, tla, ao, gona, bile, tše, ra, gagwe, šetše, ile, eng, yo, tsebe, na, woo, yena, tloga, yeo, kwa, fela, no, motho, mang, gomme, lena, goba, gape, bjalo, morago, tlo, swanetše, batho, mokgwa, bja, selo, nna, ngwana, yona, fihla, tšeo, mosadi, pele, banna, bego, thoma, wo, pelo, gobane, modimo, moka, eupša, bo, nako, seo, tee

A lemma list for the words in (1) will consist of between 95 and 100 lemmas depending on the lemmatisation strategy.

Using the Dutch–Afrikaans dictionary as a design model

The envisaged English ↔ Sotho languages dictionary is a multifunctional dictionary where a dictionary consultation environment is created in which, in

terms of Martin and Gouws (2000: 788), 'both differences and similarities become apparent in an efficient and contrastive way'. Reflecting differences and similarities will indeed be the key factor in simultaneous treatment of the Sotho languages.

A second important observation made by Martin and Gouws (2000: 790) is that 'the combinatory data represents the core of the lexicographic presentation'. Compare in this regard the following example of the approach of Martin and Gouws (2000: 790) where these principles are honoured for non-contrastive combinations, contrastive combinations and idiomatic expressions in the article of the lemma *bril* 'spectacles':

<p>bril 1 [om te kijken] # – een <u>bril</u> hebben/dragen; hij heeft zijn <u>bril</u> niet op; ... • <inf.> een <u>bril</u> moeten 'n <u>bril</u> moet kry ... ♦ elk ziet door zijn eigen <u>bril</u> <i>elkeen kyk deur sy eie <u>bril</u></i> door 'n roze <u>bril</u> kijken <i>deur 'n rooskleurige <u>bril</u> kyk</i> ▶ <i>iemand 'n <u>bril</u> op die neus sit</i> iemand te grazen nemen</p>
--

Non-contrastive combinations are marked by '-' and only a Dutch example is given. Contrastive combinations are marked with '•' and '♦' marks idiomatic expressions. In this way different search zones (in a fixed order) are clearly marked in a user-friendly way and differences and similarities are clearly illustrated.

The envisaged English ↔ Sotho languages dictionary reflects a striking resemblance to, but also clear differences with, the Dutch–Afrikaans dictionary described by Martin and Gouws (2000). Among the similarities on macrostructural level count the compilation of a single central list, i.e. a single access structure, and consideration of different lemma types. The study for the Sotho languages differs from the model of Martin and Gouws in that the lemma list is for three and not for only two languages and that a full bridging with English is done.

The macrostructure

Size and impact of a single lemma list

The prospective compiler of a dictionary with a single lemma list should firstly decide on the size of the lemma lists for both sides of the dictionary in order to compile a dictionary that would cover a reasonable percentage of use of the four languages in question. As a point of departure an assessment was made of the size of the lemma lists of dictionaries bridging English and Sotho languages as well as the top frequencies in English dictionaries such as *Collins COBUILD English Dictionary* (COBUILD2), *Macmillan English Dictionary for Advanced Learners* (MED) and *Longman Dictionary of Contemporary English* (LDOCE).

For English, the data given in COBUILD2 regarding the impact of the frequency bands give useful guidelines to the size and nature of an English lemma list for the envisaged dictionary.

Table 1: Summary of frequency band values in COBUILD2

Number of filled diamonds	Lemmas per category	Totals	% of all written and spoken English
5	700		
4	1 200		
(Total 5+4)		1 900	75
3	1 500		
2	3 200		
1	8 100		
(Total 3+2+1)		12 800	20
(Total 5+4+3+2+1)		14 700	95

From Table 1 it is clear that the top 14 700 lemmas represent an astonishing 95% of the tokens or running words in a given English text.

The words in the five frequency bands are of immense importance to learners because they make up 95% of all spoken and written English. (COBUILD2 1995: xiii)

The sizes of the lemma lists for the Sotho languages are reflected in Table 2.

Table 2: Number of lemmas in NEN, DS, SESD, SSED

Dictionary	Lemmas: English → Sotho languages	Lemmas: Sotho languages → English
Sesotho sa Leboa: NEN	11 640	13 900
Setswana: DS	10 600	15 600
Setswana: SESD	7 258	21 700
Sesotho: SSED	—	11 900

The impact of a single lemma list for the Sotho languages will be studied taking the top 10 000 words in the Pretoria corpora for each of the three languages as a point of departure.

Table 3: Sources, types and tokens in the Pretoria Sotho languages corpora

	Sesotho sa Leboa	Setswana	Sesotho
Sources	327	235	77
Tokens	5 957 553	5 466 438	3 159 568
Types	157 391	151 701	

The number of types given in the final row of Table 3 reflects 100% of the use of the languages given in terms of tokens in the second row. This simply means, for example, that a Sesotho sa Leboa dictionary containing 150 000 lemmas would account for each of the 5.9 million words in the 327 texts that make up this corpus. The same analysis is applicable to Setswana and Sesotho from the figures given in Table 3. The question is what the impact of a lemma list consisting of only 10 000 lemmas for each of the languages will be in terms of token coverage. The comparable statistics for the Sotho languages are given in Table 4.

Table 4: Percentage of tokens represented by the top 10 000 types

Language	Tokens for the top 10 000 types	Tokens in the entire corpus	% of words covered by the top 10 000 types
Sesotho sa Leboa	5 462 558	5 957 553	91.7%
Setswana	4 948 959	5 466 438	90.5%
Sesotho	2 868 451	3 159 568	90.8%

From Table 4 it is clear that, as in the case of English, lemma lists compiled for the top 10 000 tokens in each of the Sotho languages represent more than 90% of the use of the language. It can therefore be argued that the selection of 10 000 lemmas for English, Sesotho sa Leboa, Setswana and Sesotho is viable in terms of considerable coverage of all four languages.

Lexical overlap in the Sotho languages

The second aspect studied on macrostructural level in the consideration of a consolidated lemma list is the percentage of words that the languages have *in common*, simply referred to as *overlap*. It stands to reason that the greater the overlap the better the chances of success for such a dictionary will be.

As a point of departure the overlap between Dutch and Afrikaans was studied since, as reported above, that project is regarded as a viable one. A comparison between Dutch and Afrikaans corpora reveal an overlap of 20%. Consider in this regard a selection of such mutual lexical items with high occurrence frequencies per million running words in Table 5.

Table 5: Afrikaans compared to Dutch: mutual lexical items, with frequencies per million running words (Gouws et al. 2004: 798)

Item	Afrikaans	Dutch	Item	Afrikaans	Dutch	Item	Afrikaans	Dutch
<i>van</i>	19 688	18 091	<i>niet</i>	46	7 358	<i>als</i>	26	4 011
<i>het</i>	15 111	16 655	<i>te</i>	8 876	7 025	<i>aan</i>	4 338	3 846
<i>een</i>	1 892	16 223	<i>die</i>	58 733	6 520	<i>of</i>	6 653	3 811

<i>en</i>	22 516	13 663	<i>met</i>	7 139	6 268	<i>nog</i>	1 930	3 507
<i>in</i>	15 619	13 483	<i>voor</i>	1 316	6 220	<i>wel</i>	276	3 146
<i>is</i>	15 628	11 539	<i>maar</i>	3 862	5 121	<i>om</i>	7 323	3 074
<i>dat</i>	4 841	9 565	<i>ook</i>	3 068	4 414	<i>wat</i>	5 268	2 918
<i>op</i>	7 132	8 125	<i>dan</i>	907	4 032	<i>kan</i>	2 069	2 488

For the English → Sotho languages side of the dictionary a flying start exists, (technically speaking a 100% overlap), since the lemma list will only be 10 000 English lemmas and not three times 10 000 lemmas as for three separate dictionaries.

For the Sotho languages a comparison of the top 10 000 words in Sesotho, Setswana and Sesotho sa Leboa reveals that the three languages have 1 943 (19.4%) words in common. Sesotho sa Leboa and Setswana share 3 276 (32.7%) words. Sesotho sa Leboa and Sesotho have 2 689 (26.9%) words in common and Setswana and Sesotho share 3 441 (34.4%) words. This results in a single lemma list of 22 537 compared to a 30 000 lemma list in three separate dictionaries, thus a reduction of almost 30%.

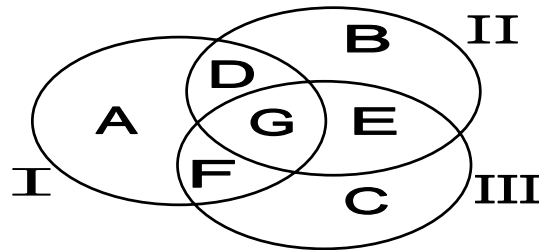


Figure 1: I = Sesotho sa Leboa; II = Setswana; III = Sesotho

(2) Formula and values for I, II and III:

Single lemma list =
 [top 10 000 Sesotho sa Leboa lemmas (A+F+D+G) +
 top 10 000 Setswana lemmas (B+E+D+G) +
 top 10 000 Sesotho lemmas (F+E+C+G)] –
 [[two times the Sesotho sa Leboa/Setswana/Sesotho lemmas in common
 (2xG)] +
 [Sesotho sa Leboa/Setswana lemmas in common (D)] +
 [Sesotho sa Leboa/Sesotho lemmas in common (F)] +
 [Setswana/Sesotho lemmas in common (E)]
 A = 5 978; B = 5 226; C = 5 813; D = 1 333; E = 1 498; F = 746; G = 1 943

Consider as a concrete example a mini-lemma list compiled on the basis of the overlap between the top 100 types in Sesotho sa Leboa, Sesotho and Setswana. The trilingual overlap, i.e. words that all three languages share within the top 100 is 32%, i.e. the 32 words in (3).

- (3) motho, bo, bona, se, sa, re, di, pele, e, nna, ngwana, tla, na, o, mosadi, morena, monna, mo, mme, yona, a, le, la, ke, ka, nako, bana, wa, wena, batho, ba, ya

The bilingual overlap between Sesotho sa Leboa and Setswana within the top 100 is 47%, i.e. the words in (3) plus the 15 words in (4).

- (4) ga, dira, eng, pelo, morago, mongwe, fela, gago, kwa, gore, yo, go, gape, bile, gagwe

The bilingual overlap between Sesotho sa Leboa and Sesotho within the top 100 is 45% or 45, i.e. the words in (3) plus the 13 words in (5).

- (5) ile, ra, tseba, fihla, moo, yena, lena, modimo, leo, seo, mang, taba, be

Finally, the bilingual overlap between Sesotho and Setswana within the top 100 is 46%, i.e. the words in (3) plus the 14 words in (6).

- (6) ntse, nngwe, tswa, ne, neng, teng, ena, utlwa, tsa, tsena, tse, tle, bua, rona

Sesotho sa Leboa has 40 unique words and Setswana and Sesotho 39 and 41 unique words respectively. This renders a single lemma list of 194 lemmas. For this section it can be concluded that lemma lists based upon the top 10 000 tokens in English and the Sotho languages will render sufficient coverage of these languages and that the amount of overlap in the Sotho languages and the resulting single lemma list suggest that the compilation of an English ↔ Sotho languages dictionary is viable on macrostructural level. The prospective compiler, however, has to keep in mind that words which have the same orthographic form in the Sotho languages but different grammatical functions will have to be entered as more than one lemma depending on the lexicographic approach. For example, *nna* in (3) as a pronoun of the first person singular in the Sotho languages but also as a verb in Setswana.

The microstructure

On microstructural level preliminary tests indicate that the average article length in the envisaged English → Sotho languages side of the dictionary would vary between one-third and two-thirds of the combined article length of the English → Sesotho/Setswana/Sesotho sa Leboa sides of three separate dictionaries, thus a 30% – 60% reduction. Compare the following randomly selected lemmas where unmarked forms such as *pula* 'rain' and *motho* 'a person' reflect a complete overlap between the three Sotho languages while double subscripts, e.g. *gagwe* 'his/her' and *phela* 'live', mark similarities between two languages and single subscripts, such as for *hae* 'his/her' and *jang* 'how', uniqueness in one language only.

- (7) pula
- motho
- gagwe [SsL/Set]
- phela [Ses/SsL]
- hae [Ses]
- jang [Set]

The real challenge lies in the successful compilation of the Sotho articles in the English → Sotho languages side of the dictionary. Failure to do so will simply result in articles reflecting the mere stacking of translation equivalents of English lemmas in Sesotho sa Leboa, Setswana and Sesotho, without consideration of crucial aspects of *differences, similarities and combinatory data* as highlighted in terms of Martin and Gouws (2000) above. There will thus be no gain in reduction and comparison. Consider Table 6 as an extract from the *Concise Multilingual Dictionary* (CMD) as a case in point.

Table 6: *Concise Multilingual Dictionary*

English	Xhosa	Zulu	Northern Sotho	Southern Sotho	Tswana	Afrikaans
carry	-thwala	-thwala	-rwala	-rwala	-rwala	dra
close (v.)	-vala	-vala	-tswalela	-kwala	-tswala	toemaak (ww.)
hot	-shushu	-shisa	-borutho (nom. rel.)	-tjhesa	bolelo	warm
pretty (water)	amahle	amanzi amahle	-bose (adj.)	metsi a matle	monate	lekker (water)

Firstly, translation equivalents are simply chronologically stacked for each language with considerable repetition in both the Sotho and Nguni languages without any attempt towards reduction. Secondly, a complete lack of communicative equivalence poses a great risk to the user to incorrectly use the equivalent(s). For example, *tswalela* has a limited range of application in Sesotho sa Leboa and cannot be used in all contexts as an equivalent of 'close', thus misleading the user. He/She is further misled by gross inconsistencies, e.g. in the final row where the compilers failed to add the word for *water* in isiXhosa, Sesotho sa Leboa and Setswana. The user would conclude that *monate* means 'pretty water' in Setswana while it only means 'nice, pretty'.

In spite of its shortcomings, it could be argued that an English → Sotho languages/Nguni languages/Afrikaans dictionary of this magnitude is a useful contribution in the complete absence of dictionaries bridging African languages with each other and could be improved by simultaneous treatment of the target languages. Consider the following attempt to improve CMD's articles for the lemmas *coffee, page* and *verb* in Table 7 versus example (8).

Table 7: *Concise Multilingual Dictionary*

English	Northern Sotho	Southern Sotho	Setswana
coffee	kofi	kofi	kofi
page	letlakala	leqephe	tsebe
verb	lediri	leetsi	lediri

- (8) **coffee** kofi
page letlakala _[SsL], leqephe _[Ses], tsebe _[Set]
verb lediri _[SsL/Set], leetsi _[Ses]

The aim of this viability study, however, is to compile more comprehensive articles, at least of the same size as existing bilingual dictionaries such as NEN, DS, SESD and SSED. The articles compiled in (9) for the English → Sotho languages and Sotho languages → English sides of the dictionary are still relatively speaking restricted to receptive use but they do go some way towards productive use and communicative equivalence.

- (9) **rain** *n.* pula
motho *n.* a person, a human being
his *pp.* gagwe _[SsL/Set], hae _[Ses]
live *v.* phela _[Ses/SsL], tshela _[Set] ~ **(survive)** phela _[Set]
how? *adj.* jwang? _[Ses] o phela jwang?; jang? _[Set] o tshela jang?; bjang? _[SsL]
o phela bjang? **how are you?**
pelo (peló); N Cl. 9/10; **dipelo** heart; ~**ng** in the heart; **pelwana** little heart; **fela** ~ give up; **beta** ~ be brave; ~**bohloko** _[SsL/Ses]; ~**botlhoko** _[Set]; ~**mpe**, ~**nolo**; ~**telele**; ~**tlhomogi** _[SsL/Set]; ~**tlhomohi** _[Ses]

Sensible reduction in these examples is achieved in terms of, among others, tonal indication, grammatical information, and translation equivalents. They should, however, be submitted to target users and the feedback obtained should be carefully studied.

In the Sotho languages → English side, the mediostructure (system of cross-referencing) can be fruitfully utilised to link the lemma with its equivalent in the other Sotho language(s), thus further strengthening the aspect of bridging Sesotho sa Leboa, Setswana and Sesotho with each other.

- (10) **hore** _[Ses] *conj.* that, in order that ... cf. gore _[SsL/Set]

Selecting suitable dictionary conventions will be a crucial aspect in order to present user-friendly search zones. In the examples above, subscripts were used to mark the distinctions between Sesotho, Setswana and Sesotho sa Leboa. Similar layouts should be tested utilising a combination of different colours and standard conventions such as bold, underline and italics as in Figure 2

where the different languages are marked in the dictionary in colour and coloured shadings (which can unfortunately not be reproduced in this article) with the aid of running footers:

rain *n.* pula
 motho *n.* a person, a human being
 his *pp.* gagwe, hae
 live *v.* phela, tshela; ~ (survive) phela
 how? *adj.* jwang? o phela jwang?; jang? o tshela jang?; bjang? o phela bjang?
how are you?
 {Running footer:}
 Ses; Set; SsL; SsL/Set; Ses/SsL; Ses/Set

All three languages in common = black
 SsL (Sesotho sa Leboa) = red
 Set (Setswana) = green
 Ses (Sesotho) = blue

Figure 2: Using colour and colour shading in Sotho languages articles of English lemmas

Conclusion

In this article macrostructural and microstructural aspects were analysed in terms of the viability of an English ↔ Sotho languages dictionary with a single lemma list for the Sotho languages and simultaneous treatment of the Sotho languages in the English → Sotho languages side. It can be concluded that such a compilation will be successful (a) for lemma lists of a reasonable size taken from all four languages as a point of departure, (b) because substantial lexical overlap exists between Sesotho sa Leboa, Setswana and Sesotho, (c) because treatment in terms of similarities, differences and combinatory data is possible, and (d) because user-friendly articles comprehensive enough for receptive and limited productive use can be compiled.

Endnote

† The term *Sotho languages* refers to Sesotho sa Leboa (Northern Sotho), Setswana (Tswana), Sesotho (Southern Sotho), and the term *Nguni languages* refers to isiZulu (Zulu), siSwati (Swazi), isiXhosa (Xhosa) and isiNdebele (Ndebele), cf. Van Wyk (1966: 17, 21, 36-37), Guthrie (1970), etc.

References

Dictionaries

- CMD = Jennings, L.E. et al. 1995. *The Concise Multilingual Dictionary*. Johannesburg: Ad Donker.
 COBUILD2 = Sinclair, J. (Ed.). 1995. *Collins COBUILD English Dictionary*. London: HarperCollins.

- DS = Snyman, J.W. et al. 1990. *Dikišinare ya Setswana English Afrikaans Dictionary Woordeboek*. Pretoria: Via Afrika.
- LDOCE = Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English*. Harlow: Longman.
- MED = *Macmillan English Dictionary for Advanced Learners*. 2002. Macmillan Publishers Limited.
- NEN = Kriel, Theunis J. 1976⁴. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English*. Johannesburg: Educum Publishers.
- SESD = Matumo, Z.I. 1993 *Setswana–English–Setswana Dictionary*. Gaborone: Macmillan.
- SSED = Mabile, A. and H. Dieterlen. 1988. *Southern Sotho–English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.

Other literature

- Gouws, R.H., D.J. Prinsloo and G.-M. de Schryver.** 2004. Friends will be Friends — True or False. Lexicographic Approaches to the Treatment of False Friends. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 797-805. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- Guthrie, M.** 1970. *Comparative Bantu Vol. III. An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg Press.
- Martin, Willy and Rufus Gouws.** 2000. A New Dictionary Model for Closely Related Languages: The Dutch–Afrikaans Dictionary Project as a Case in Point. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, 8–12 August 2000*: 783-792. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Van Wyk, E.B.** 1966. *Die Bantoetale: 'n Beknopte algemene oorsig*. Pretoria: Van Schaik.