

A Critical Analysis of Multilingual Dictionaries

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

Abstract: This article evaluates the lexicographic value of multilingual dictionaries. Dictionaries covering three or more languages spoken in South Africa are taken as a case in point. An attempt will be made to reflect on their merits and shortcomings as reference works and learning tools but the focus will be on presumed shortcomings in the macro and micro structures of such dictionaries with special attention to the lemmatisation of common words, quality of the dictionary articles and consistency in presentation.

Keywords: MULTILINGUAL DICTIONARIES, MACRO STRUCTURE, MICRO STRUCTURE, AFRICAN LANGUAGES, THEMATIC ORDERING, ALPHABETIC ORDERING

Opsomming: 'n Kritiese evaluering van veeltalige woordeboeke. In hierdie artikel word die leksikografiese waarde van veeltalige woordeboeke geëvalueer. Woordeboeke wat drie of meer tale wat in Suid-Afrika gepraat word leksikografies bewerk, dien as voorbeeld. Daar sal gepoog word om hulle sterk punte en tekortkominge as naslaanbronne en aanleerhulpmiddels te omskryf. Klem sal gelê word op veronderstelde tekortkominge in die makro- en mikrostrukture van sulke woordeboeke met spesifieke verwysing na die lemmatisering van algemeen gebruikte woorde, kwaliteit van die woordeboekartikels en konsekwentheid in die aanbieding.

Sleutelwoorde: MEERTALIGE WOORDEBOEKE, MAKROSTRUKTUUR, MIKROSTRUKTUUR, AFRIKATALE, TEMATIESE ORDENING, ALFABETIESE ORDENING

1. Introduction

This paper evaluates the lexicographic value of multilingual dictionaries. Dictionaries covering three or more languages spoken in South Africa are taken as a case in point. An attempt will be made to reflect on their merits and shortcomings as reference works and learning tools. The focus will be on presumed shortcomings in the macro and micro structures of such dictionaries with special attention to the lemmatisation of common words, quality of the dictionary articles and consistency in presentation. The Multi-Language Dictionary (MLD), Multilingual Illustrated Dictionary (MID), Concise Multilingual Dictionary (CMD), and the Oxford Junior Primary Dictionary for Southern Africa (OJPD) will be analysed as typical examples. Reference will also be made to The English-Afrikaans-Northern Sotho-Tswana Aid (EANTA).

It has to be stated at the outset in terms of Gouws et al. (2014: 25) that no single dictionary can be everything for everyone.

Among the merits of these dictionaries count the impressive presentation

of specific everyday themes such as 'my body', 'family and friends', 'clothing', 'my home', thus in most cases a complementary thematic approach in which a number of typical situations in everyday life are presented with or without pictorial illustrations and the relevant words linked to an alphabetical section where they are presented as treated or un-treated lemmas.

Such a thematic approach naturally reflects appropriate contextualisation of lexical items that belong together which would otherwise have been scattered all over the dictionary if only an alphabetical ordering was followed. Furthermore, most of these dictionaries are appealing to the eye, attractive to use and encourage learning and dictionary use. So, for example, giving a picture of the face of a person indicating all the different parts in seven languages, encourages language learning, especially for young learners. Such dictionaries usually also score high marks in terms of comparison between different languages. The user can see at a glance how the translation equivalents in the other languages compare to his/her source and target language knowledge.

In contrast, however, it will be argued that the multilingual model underpinning the compilation of multilingual dictionaries is problematic on many levels, resulting in products of inferior lexicographic quality. This is aggravated by various degrees of inconsistency, the lack of cohesion in the treatment of the different languages and even many basic or sloppy errors in the dictionaries. Problematic aspects to be explored are (a) poor and inconsistent covering of lexical items, (b) impoverished/skeleton dictionary articles: lack of sense distinction, lack of parity between translation equivalents, absence of examples of usage, incomplete translation equivalent paradigms, etc. and (c) risk of misleading the user through inappropriate text/speech production or communicative guidance. A brief outline of the contents and layout of the dictionaries will first be presented as a basis for the discussion.

The preferred reference to the languages in this article is to give the names of the languages in the specific language, i.e. Afrikaans, English, Sepedi, Setswana, Sesotho, isiZulu and isiXhosa. The comparative references in MID, MLD, CMD, OJPD and EANTA are given in table 1.

	Afrikaans	English	Sepedi	Setswana	Sesotho	IsiZulu	IsiXhosa
MID	Afrikaans	English	Sepedi	Setswana	Sesotho	IsiZulu	IsiXhosa
MLD	Afrikaans	English	N(orthern) Sotho	Tswana	Sesotho	Zulu	Xhosa
CMD	Afrikaans	English/Engels	Northern Sotho/ Noord-Sotho	Tswana	Southern Sotho/ Suid-Sotho	Zulu/ Zoeloe	Xhosa
OJPD	Afrikaans	English	North Sotho	Setswana	South Sotho	-	-
EANTA	Afrikaans	English	Northern Sotho	Tswana	-	-	-

Table 1: Names of the languages used in MID, MLD, CMD, OJPD and EANTA

2. Dictionary versus text book as reference source

It could be argued that most multilingual dictionaries enable thematic as well as alphabetic lookup and viewed as such, this is a strong point. MID, EANTA and MLD attempt to strike a good balance between thematic and alphabetic lookup by means of a system of cross-referencing between the thematic sections and the alphabetical stretches. Using multilingual dictionaries such as MID, MLD and EANTA, be it through thematic or alphabetical lookup, also entails an element of "reading the dictionary". A substantial part of the learning potential of such dictionaries is to not only look up words in a specific theme, e.g. the human body, but then also to learn the translation equivalents for the other body parts. So, an issue to be considered at the outset for multilingual dictionaries is the question as to what extent the average user is interested in *reading* a dictionary versus *consulting* a dictionary for e.g. the meaning of a specific word. Getting maximum benefit from multilingual dictionaries implies/presumes a fair amount of reading. Martin, *Pharos Groot Woordeboek. Afrikaans en Nederlands* (ANNA) attempts to make a case for the value of reading a dictionary for recreation in addition to the standard procedure of consulting a dictionary.

Laaste maar nie die minste nie is 'n geamalgameerde woordeboek ook 'n leeswoordeboek: 'n woordeboek waarin 'n mens nie slegs vertaalekwivalente soek of verskille vind nie, maar een waarin jy ook net kan lees vir ontspanning. (ANNA: 23)

[Last, but not least, an amalgamated dictionary is also a dictionary for reading: a dictionary where you not only search for translation equivalents or contrasts, but one which one can read for recreation.]

This assumption, however, should be supported by actual user studies and falls outside the scope of this article.

3. Macrostructural considerations

3.1 Dictionary components of multilingual dictionaries

MID presents its table of contents in the seven languages covered by the dictionary, i.e. English (Contents), isiZulu (Okuqukethwe), Sesotho (Dikahare), isiXhosa (Isiqulatho), Setswana (Diteng), Afrikaans (Inhoud) and Sepedi (Diteng). It is not clear on what basis the sorting order for these languages was determined. So, for example, it would have been better to group the related African languages together, i.e. the so-called *Sotho languages* (Sepedi, Sesotho, Setswana) together and the Nguni languages (isiZulu and isiXhosa) together as has been done in MLD and CMD. (The other four official African languages of South Africa Siswati, isiNdebele, Tshivenda and Xitsonga are not included in MID.) A specific colour has been allocated to each language, e.g. purple for isiZulu, yellow for Setswana, green for isiXhosa, etc. which is applied throughout in a consistent manner. The user guide is also presented in these seven languages. This is followed by 13 theme sections, i.e. My body, Family and friends, Clothing, My home, Our community, Transport, Communication, Economy, Nature and us, Food and drink, Our spare time, South Africa and Useful words. The 13th

MLD's cover page indicates that it covers English, Afrikaans, "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu". In the table of contents it explains the colour keys for the different languages, how to use this book, know your grammar, visual dictionary followed by a section with English lemmas as the source language (L1) followed by Afrikaans, "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu", then a section Afrikaans as L1 followed by English, "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu" and then sections where the African languages lemmas in turn are given as L1, i.e. "Northern Sotho", Sesotho, "Tswana", "Xhosa" and "Zulu" with English and Afrikaans as target languages in each case. Consider the treatment of the stretch *abacus – accident* in Figure 2.

English	Afrikaans	N Sotho	Sesotho	Tswana	Xhosa	Zulu
abacus (n)	telraam	mmadiphetana	mmadibolwana	kgathatalama	isibali	i-abakhusi
abandon (v)	verlaat	-tlogela	-nyahlatsa	-tlogela	-ncama	-yeka
abattoir (n)	slagplaas	bohlabelo	selagapalo	bothabelo	isilarha	amadela
abdomen (n)	buik	mpa	mpa	mpa	isisu	isisu
abduct (v)	ontvoer	-thopa	-kwetela	-utswa	-thwala	-thwala
ability (n)	bekwaamheid	bokgoni	bokgoni	kgoni	ubuchule	ukwazi
able (a)	bekwaam	-kgona	-kgona	-kgonang	-kwazi	-nokwazi
abnormal (a)	abnormaal	-sa tlwaelegago	-sa tlwaelehang	-sa tlwaeegang	-ngaqhelekanga	-ngavamile
aboard (adv)	aan boord	ka gare ga	-palame	-palameng	-ngaphezulu	emkhunjini
abode (n)	woonplek	bodulo	bodulo	legae	indawo	isikhundla
abolish (v)	afskaf	-fediša	-fedisa	-khutisa	-bhangisa	-susa
aboriginal (n)	oorspronklike inwoner	semetafa	motala	monnafa	inzalwane	umuntu wondabu
abort (v)	aborteer		-nyopisa	-senyega	-phanza	-khipha isisu
abortion (n)	aborsie		nyopiso	tshenyegelo ya mpa	impanza	ukuphuphuma isisu
abortive (a)	misluk		-nyopiteng	-senyegetsweng	-phanzileyo	-gejile
about (adv)	rond	mo le mola	ka-	gomela	-ngokujikeleyo	nga-
about (prep)	oor		ka-	ka ga	nga-	phezu kwa-
above (prep)	bo	godimo ga	-hodima	mo godimo	-phezulu	enhla kwa-
abscess (n)	verswering	sekaku	seso	sekaku	igqitha	ithumba
abscess (n)	afwesigheid	tlhokego	bosiyo	tlhokafalo	ubungabikho	ukungabikho
absent (a)	afwesig	-hlokega	-ba siyo	-se yong	-ngekhoyo	-ngekho
absent-minded (a)	ingedagte	-ya le megopolo	-bua ka pelo	-bua ka pelo	-zungubala	-alukile
absolute (a)	absoluut		-hle	tota	kanye	ngempela
absolutely (adv)	absoluut	ruri	ho hang	tota	-ngenene	ngempela
absorb (v)	absorbeer	-hupa	-monya	-monyela	-funxa	-munca
abstain (v)	onthou	-ila	-sesefa	-ila	-yeka	-yeka
abstract (a)	afgetrokke		-sa bonweng	-kgopolo	-ngaphathekiyo	-okusengqondweni
absurd (a)	onsinnig	-segišago	-hloka kelelo	mo go se nang	-nobuhiba	-ngasile
abundance (n)	oorloed	bontši	nala	monono	ubuninzi	isibhidi
abundant (a)	oorvloedig	-ntši	-ngata	-tlatlalo	-ninzi	-xhaphakile
abuse (n)	misbruik		tlatlapo	tiriso e e sa siamang	ukusebenzisa kakubi	ukoniwa
abuse (v)	misbruik	-senya	-tlatlapa	-dirisa e seng ka fa siamang	-sebenzisa kakubi	-ona
acacia (n)	akasia	mohlare wa akasia	leoka	mmafu	umnga	umkhamba
academy (n)	akademie	aketemi	akhademi	maithuto	isikolo senkcubeko	imfundo ephakeme
accelerate (v)	versnel	-fa makhura	-akofisa	-okofatsa	-khawulezisa	-sheshisa
accelerator (n)	brandstofpedaal	sefamakhura	seakofisi	seokofatso	umcephe	isipuni
accent (n)	aksent	tobeletšo	tokodiso	kgateletšo	uhlobo lokuthetha	isigcizelelo
accept (v)	aanvaar	-amogela	-amohela	-amogela	-amkela	-amukela
acceptable (a)	aanvaarbaar	-amogelega	-amohelaha	-amogelesegang	-amkerekileyo	-amukelekayo
acceptance (n)	aanvaarding	kamogelo	kamohelo	kamogelo	uwamkelo	ukwamukela
accident (n)	ongeluk	kotsi	kotsi	kotsi	ingozi	ingozi

Figure 2: Layout and treatment of *abacus – accident* in MLD

Finally, MLD has a section devoted to Phrases with subthemes such as Time, Household, Travel, etc.

In contrast to MID a better grouping of the languages, is given, i.e. Sotho languages together and Nguni languages together. Under "Visual dictionary" MLD lists The human body, The house, The Office, The Car and Sport as main themes with subthemes, e.g. under Sport soccer, rugby and cricket are listed. Its stated aim is to help the user to achieve "a basic level of understanding in the seven major languages of South Africa" by means of (a) a visual dictionary, (b) a 5,000-word dictionary that translates English and Afrikaans into five African languages.

CMD does not offer any thematic sections but it covers English, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, Afrikaans.

A

English	Xhosa	Zulu	Northern Sotho	Southern Sotho	Tswana	Afrikaans
abandon	-lahla	-yeka	-tlogela	-nyahlatsa	-tlogêla	verlaat
abate	-damba	-nciphisa	-okobala	-kokobela	-reba	bedaar
abdomen	isisu	isisu	mpa	mpa	mpa	maag
abduct	ukuba umntu	-thwala	-thopa	-kwetela	-utswa	skaak
abide	-hlala	-hlala	-dula	-dula	-nna	bly
ablaze	-vutha	-vutha	-tuka	-tukang	tukang	in vlamme
abode	ikhaya	ikhaya	bodulo	bodulo	legaê	verblyfplek
above	ngaphezulu	phezu kwa-	godimo	hodimo	godimo	bo
abridge	-finyeza	-thothanisa	-khutsofatša	-kgutshufatsa	-khutswafatsa	verkort
absent	-ngekho	-ngekho	-hlokega	-ba siyo	-se yông	afwesig
absent, to be	ukungabukho	ukuba ngabikho	go hlokega	ho ba siyo	-tlhòkêgwa, go	afwesig te wees, om
accede	-vuma	-vuma	-dumela	-dumela	-dumela	toege
accident	ingozi	ingozi	kotsi	kotsi	kôtsi	ongeluk

Figure 3: Layout and treatment of *abandon* – *accident* in CMD

The dictionary consists of two sides, i.e. English followed by "Xhosa" "Zulu", "Northern Sotho" "Southern Sotho", "Tswana" and Afrikaans and a reverse side printed upside down (i.e. the user has to flip over the dictionary and start from the other end for Afrikaans as L1 followed by "Xhosa", "Zoeloe", "Noord-Sotho", "Suid-Sotho", "Tswana" and English. In the introduction section of CMD this macrostructural arrangement is praised as "English and Afrikaans word lists bound into one handy volume, with their inverse juxtapositions providing each of the books with equal recognition and prominence". However, unlike MLD, this "equality" is not carried through for the African languages since they are not listed as L1 nor are indexes provided to look up African language words as has been done in MID, see Figure 1 column 1 as an example of Sesotho and isiXhosa indexed lemmas with reference to the page where treatment is offered. No thematic pages or user guides are given but an approximately two-page mini-grammar is attempted for each of the African languages treated. Tables for days of the week, months of the year and numbers are given for English as L1 translated into the other languages. In the mini-grammars an effort is made to give grammatical information for "Xhosa", "Zulu", "Northern Sotho", "Southern Sotho" and "Tswana". It is clear that no attempt was

made to bring the data on a par for the other languages. Different issues and illustrations receive attention for the different languages. So, for example, for isiZulu, in addition to pronunciation guidance, detailed attention is given to the noun class system, the concordial system, pronouns and the verb but for isiXhosa, Sepedi, Sesotho, Setswana such important guidance in terms of the nominal class system is absent — the entire sections deal only with pronunciation.

OJPD's user's guide under the heading "to the teacher", states the fact that the dictionary gives the meanings of more than 1,500 words and highlights five characteristics of the dictionary, i.e. (1) the alphabet appears at the top of each page, (2) the defining vocabulary is controlled, (3) each word is used in an example sentence, (4) pictures help children to understand the meaning and (5) mother tongue translations help with meaning. The A-Z section with English as L1 represents more than 90% of the space allocation in the dictionary. Relatively few schematic illustrations are given, approximately one on every second page in the A-Z section. Five themes are presented in the form of schematic illustrations at the end of the dictionary, i.e. A Girl, A Boy, A Classroom, A City Street and A Kitchen, but these are given only in English. The user has to look up the English word in the alphabetical section. Tables are given for Days of the week and Months of the year with translations in "North Sotho", "South Sotho", Setswana and Afrikaans. This is followed by a Numbers theme 1 (one) to 50 (fifty) and 1st (first) to 50th (fiftieth) but not translated into the four target languages. Finally another range of theme pages of schematic illustrations of Animals and young animals, Inserts, Fruit and vegetables is given, also in English only.

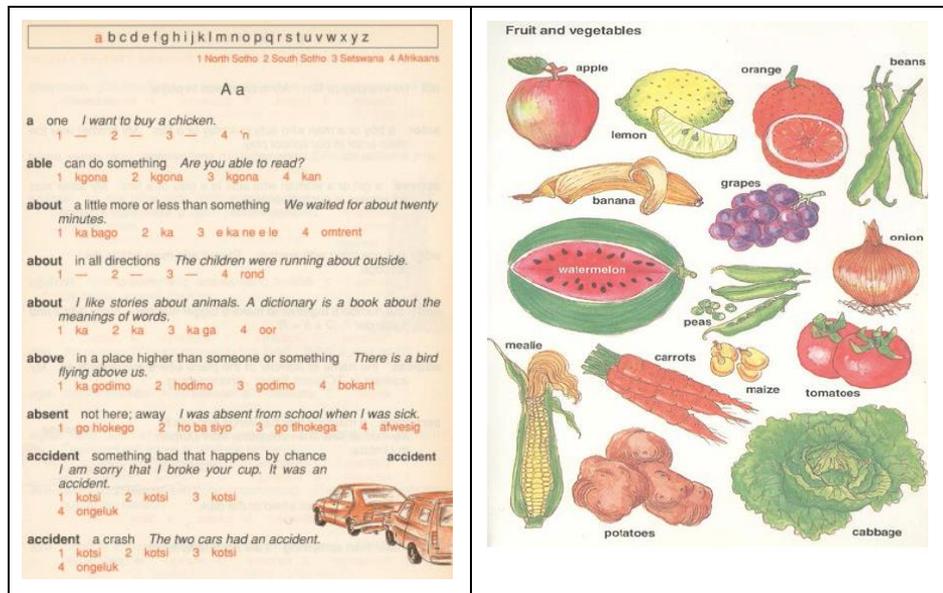


Figure 4: An extract from the alphabetical and thematic sections in OJPD

Minor mistakes noticed is that the English words for *days* and *months* are not correctly aligned with the translation table and that numbers are not translated, cf. Figure 5.

Days of the week					Numbers			
	North Sotho	South Sotho	Setswana	Afrikaans				
Sunday	Lamorena	Sontaha	Letsatsi la Tshipi	Sondag	1	one	1st	first
Monday	Mosupologo	Mantaha	Mosupologo	Maandag	2	two	2nd	second
Tuesday	Labobedi	Labobedi	Labobedi	Dinsdag	3	three	3rd	third
Wednesday	Laboraro	Laboraro	Laboraro	Woensdag	4	four	4th	fourth
Thursday	Labone	Labone	Labone	Donderdag				
Friday	Labohlano	Labohlano	Labothlano	Vrydag				
Saturday	Mokibelo	Moqebelo	Lamathatso	Saterdag				

Figure 5: Days of the week and Numbers in OJPD

OJPD is inconsistent in terms of reference to the Sotho languages. For Sepedi and Sesotho it uses the English terms North Sotho and South Sotho (which are linguistically speaking questionable) but refers to Setswana instead of Tswana which would linguistically be aligned with the terms North(ern) and South(ern) Sotho.

3.2 Coverage of lexical items

It is a reasonable expectation of the user to find the word (s)he is looking for in the dictionary.

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

Dictionaries for which the lemma lists are restricted to a relatively small number, e.g. 5,000 lemmas, naturally focus on the most common words of the language. The question is whether dictionaries such as MID, MLD, CMD and OJPD reflect a reasonable coverage of common words in the language and if they provide the required items in the comment on form and comment on semantics for the user. Formulated differently, for most users to "find the thing" in terms of Haas, simply means to find the lemma in the dictionary and especially the meaning of the word represented by the lemma sign, in a user-friendly way.

In MID the alphabetical indexes reflect the number of lemmas treated in the dictionary. For each language approximately 3000 words are listed with cross-reference to the page number where a picture is given and the translation equivalents in the other languages. cf. Figure 1 above.

MLD has more than 400 pages for the alphabetical section. This approach is user-friendly in the sense that the user can use any of the seven languages as direct access for lookup. There is no need, as in the case of MID, to consult an

index list in order to be cross-referred to the correct L1 lemma and page number to look it up, or as in the case of CMD only being able to look up Afrikaans and English words.

English/Afrikaans/Northern Sotho/Sesotho/Tswana/Xhosa/Zulu	50
Afrikaans/English/Northern Sotho/Sesotho/Tswana/Xhosa/Zulu	137
Northern Sotho/English/Afrikaans	225
Sesotho/English/Afrikaans	273
Tswana/English/Afrikaans	321
Xhosa/English/Afrikaans	369
Zulu/English/Afrikaans	417

Figure 6: The scope of translation equivalents given and page numbers for each L1 language in MLD

However, it drastically reduces the available dictionary space per section to approximately 86 pages for English; 87 for Afrikaans, 47 for "Northern Sotho", 47 for Sesotho, 48 for "Tswana", 47 for "Xhosa", and 47 for "Zulu". Furthermore, even though the African languages are in turn given lemma status as L1, only English and Afrikaans translation equivalents are given and no equivalents in the other African languages. Not giving translation equivalents in the other African languages enables the use of double columns per page which results in 50% space saving, but it means that the treatment of African languages as L1 is not on a par with English and Afrikaans.

Dividing the dictionary into seven alphabetical sections as in Figure 6 reduces the potential dictionary space for lemmas from 24,000 (400 pages x 60) lemmas per language to approximately 5,000 lemmas per language in order to give each language the opportunity to be lemmatised as L1. The question is whether an index system as in MID with cross reference to an English as L1 section would not have been a better option. This would be less user-friendly because Afrikaans and African language words will have to be looked up via alphabetical index lists but would enable a treated lemma list of e.g. 20,000 English L1 lemmas instead of the current 5,000 or less per language. Alphabetical index lists also require dictionary space but approximately six times less because indexed words can be presented in six to seven columns per page as was done in MID.

In CMD approximately 60 lemmas are given per page and a total of 155 pages could in principle render a lemma list of approximately 9,000. However, due to the fact that English and Afrikaans are presented in two separate sections as L1, this potential is divided by two i.e. the lemma list for English is reduced to circa 4,500.

For OJPD approximately 180 pages x 10 lemmas per page equals 1800. This is fairly limited and it is unlikely that the user will find the words he/she is looking for.

In all of these dictionaries the number of lemmas can be regarded as very limited and an attempt has been made to see to what extent they succeeded in lemmatising the most common words of the language. The 136 most frequently used words in the alphabetical stretch A, indicated with a three-star frequency band in the *Macmillan English Dictionary* (MED) was used as a measuring instrument. These words are compared to the most frequently used words in the Pretoria English Internet Corpus (PEIC) and to the lemma lists of MLD, OJPD, CMD and MID in Table 2.

MED ***	PEIC	MLD	OJPD	CMD	MID
and	32101	✓	✓	✓	✓
as	7794		✓	✓	✓
at	6384	✓	✓	✓	✓
all	3549	✓	✓	✓	
an	3060		✓		
any	1659	✓	✓		
about	1569	✓	✓		✓
after	1463	✓	✓	✓	✓
army	1399	✓		✓	
against	1072	✓	✓		✓
also	684	✓	✓		
among	560		✓	✓	
again	519	✓	✓		
away	483		✓	✓	
always	448	✓	✓		
another	442	✓	✓		
almost	405	✓	✓	✓	
already	395		✓		
above	387	✓	✓	✓	✓
able	342	✓	✓		
age	341	✓	✓	✓	
afterwards	333	✓		✓	
August	304	✓	✓		✓
attack	303	✓	✓		
ancient	289	✓			
according	278				
authority	276	✓			
along	259	✓	✓		✓
act	253		✓	✓	✓
art	243	✓			
anything	217		✓		
account	210	✓		✓	
alone	207	✓	✓	✓	
answer	183	✓	✓	✓	
around	174	✓			✓
attention	165		✓		
across	157	✓	✓	✓	✓
ask	147	✓	✓	✓	
ago	134		✓		
admit	129	✓			
attempt	122	✓		✓	
although	121	✓		✓	
arrival	121	✓			✓
April	119	✓	✓		✓
airport	110	✓	✓		✓
action	105				
afternoon	105	✓	✓		✓
air	100	✓	✓	✓	
assistance	94	✓			
add	88	✓	✓	✓	
assembly	88	✓			
afraid	85	✓	✓	✓	✓
amount	85		✓		
author	85	✓		✓	✓
ability	84	✓			
appearance	83	✓			
approach	83	✓		✓	
altogether	81				
allow	79	✓	✓		
appear	79	✓		✓	
argument	79	✓	✓		
active	73	✓			
arm	66	✓	✓	✓	✓
absence	63	✓			
accept	62	✓			
affair	59	✓			
ahead	59	✓			
absolutely	57	✓			
agree	55	✓	✓		
alive	54	✓	✓		
angry	52	✓	✓	✓	✓
artist	49				✓
assume	47				
anybody	46	✓	✓		
apply	43	✓			
article	43	✓			
accident	42	✓	✓	✓	
avoid	41	✓			
animal	40	✓			
address	39	✓	✓	✓	
attend	38				
apparent	37				
arrangement	37	✓			
afford	36	✓			
apparently	36				
adopt	35	✓			
audience	35				✓
actually	33	✓			
apart	33				
appeal	32	✓			
assist	32	✓		✓	
association	31				
annual	27				
application	27	✓		✓	
appointment	27	✓			

available	27	✓				
actual	26					
abroad	22					
anywhere	22	✓	✓			
arise	22	✓				
arrive	22	✓	✓	✓		
aspect	22					
adequate	19	✓				
agreement	19	✓				
affect	18	✓				
angle	18	✓				
arrange	18	✓				
attitude	18	✓				
access	17					
activity	16	✓				
appoint	15	✓		✓		
agent	14	✓		✓		
approve	14	✓				
approval	13	✓				
attract	13	✓				
aunt	13	✓	✓	✓	✓	
average	13	✓				
accuse	11	✓			✓	
aim	11	✓	✓			
associate	11					
alternative	8					
area	8	✓				
autumn	8	✓	✓	✓	✓	
attractive	7	✓				
achieve	6	✓				
argue	6	✓	✓	✓		
anyone	5				✓	
agency	4					
academic	3					
award	3					
aware	3					
achievement	2	✓				
actor	2	✓	✓	✓	✓	
analysis	2	✓				
anyway	2					
adult	1	✓			✓	✓

Table 2: MED three-star words compared to the PEIC, MLD, OJPD, CMD and MID

There are 136 three-star words in the alphabetic stretch A in MED. MLD has 102 (75.0%), OJPD 54 (39.7%), CMD 42 (30.9%) and MID 25 (18.4%) of these MED three-star words as treated lemmas. In PEIC the top 5,000 words occur with a frequency of 200 or more. 33 of these words are also three-star words in MED (indicated in bold in Table 2) and were also compared with the lemma lists of MLD, OJPD, CMD and MID. Of these 33 words OJPD lemmatised and treated 26 (78.8%), MLD 25 (75.8%), CMD 15 (45.5%) and MID 9 (27.3%).

Alphabetic stretch A	MLD	OJPD	CMD	MID
MED *** words	75.0%	39.7%	30.9%	18.4%
33 PEIC	75.8%	78.8%	45.5%	27.3%
Size of dictionary	495	198	150	365

Table 3: Size of MLD, OJPD, CMD and MID in relation to the overlap with MED three-star words and the PEIC

These percentages of overlap of MLD, OJPD, CMD and MID with MED three-star words and the PEIC are schematically illustrated in Figure 7.

From Figure 7 it is clear that none of the dictionaries captured more than 80% of the most frequently used words and e.g. that MID captured less than 30% of such words.

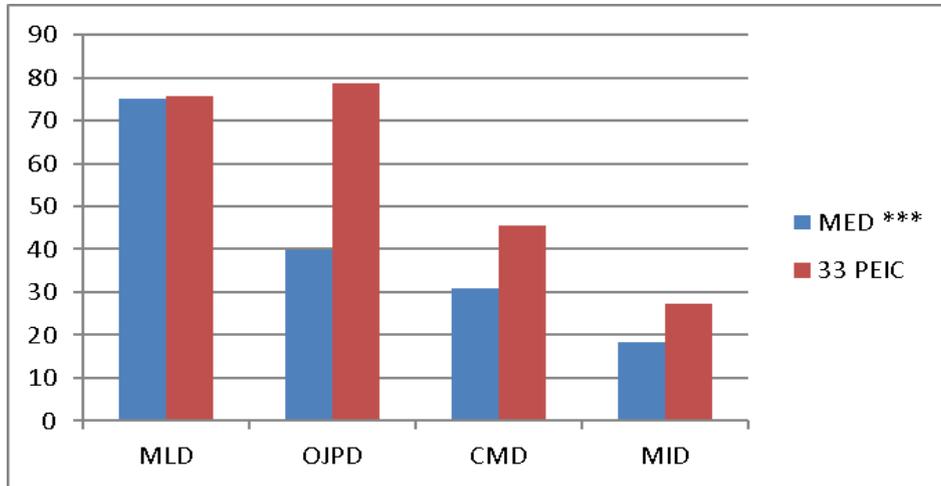


Figure 7: Size of MLD, OJPD, CMD and MID in relation to the overlap with MED three-star words and the PEIC

The percentages in Table 3 were then calculated as a percentage of the size of each dictionary and are illustrated in Figure 8.

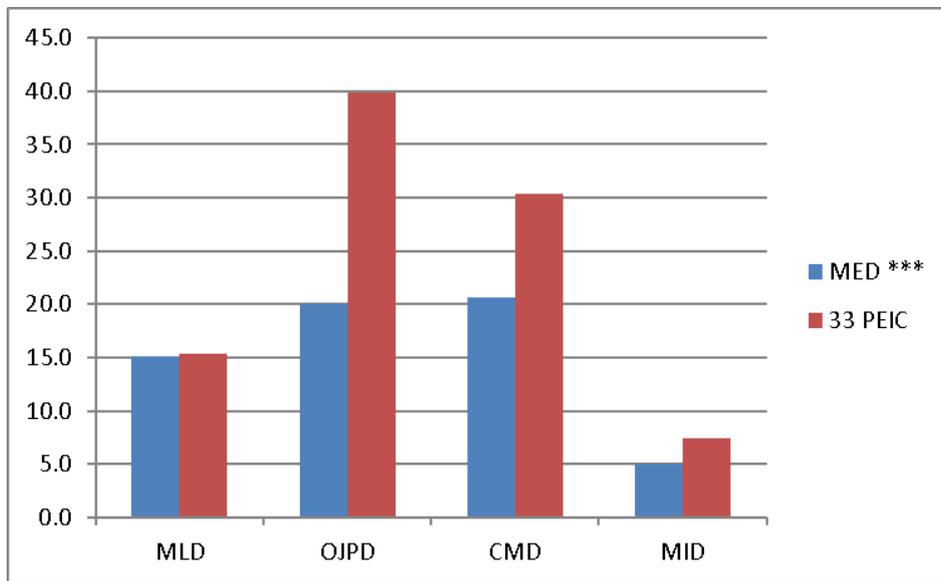


Figure 8: Percentage of top frequencies in MLD, OJPD, CMD and MID in relation to the size of the dictionaries

A slightly different situation emerged. OJPD and CMD scored higher than the other dictionaries because they achieved their percentages given in Table 3 in dictionaries consisting of less than 200 pages. MLD however, although it captured 75% of the frequently used words, scored worse because it is a much bigger dictionary, i.e. more than twice the size of OJPD and CMD. MID's scores remained low, i.e. a low percentage of frequently used words captured, as well as the fact that it is a relatively bigger dictionary than OJPD and CMD. Formulated differently: the four dictionaries did not do well in terms of capturing the most frequently used words, especially so for the two bigger dictionaries, MLD and MID. For example, the omission of common words such as *as, an, among, away, already* in MLD is less satisfactory for a relatively big dictionary, i.e. one consisting of 495 pages. Thus one would expect from the relatively bigger dictionaries, i.e. MLD and MID to reflect more of the most frequently used words.

4. Microstructural considerations

4.1 Comment on form

The root of the perceived deficiencies of multilingual dictionaries on the microstructural level is what could be referred to as impoverished or skeleton dictionary articles. Impoverished dictionary articles are an almost unavoidable result of the multilingual model because in dictionaries treating up to *seven* languages, there is hardly any space available beyond the listing of a single translation equivalent for each lemma. As indicated in Figure 3, on a single page in CMD more than 60 lemmas are treated for six languages, i.e. presenting approximately 400 articles per page. In CMD and MLD seven columns have to be fitted in on every page to reflect the seven languages thus limiting the space for the dictionary article to a default of 20 characters with the occasional use of a second line.

Such severe space limitation results in a number of deficiencies and even a combination of deficiencies ranging from mere omissions and inconsistencies to mistakes and the ultimate deficiency: misguidance of the user.

First, no other item besides the lemma is given in the comment on form. In some cases, comment on form is given in an inconsistent way or an incorrect label is given.

CMD employs a strategy of inserting a comma and guiding word(s) after the lemma. These guiding words which follow the comma are intended to inform the user that they should be used in front of the lemma. (CMD's use of leadword versus main word is not clear).

In order to facilitate arranging the English and Afrikaans leadwords in alphabetical order, for quick and easy reference, a comma (,) has been inserted — where thought necessary — after a leadword to indicate that the words following such comma must be utilised in front of the leadword to arrive at the correct shade of meaning of the main word. (CMD: v)

English	Xhosa	Zulu	Northern Sotho	Southern Sotho	Tswana	Afrikaans
abandon	-lahla	-yeka	-tlogela	-nyahlatsa	-tlogêla	verlaat
abate	-damba	-nciphisa	-okobala	-kokobela	-reba	bedaar
abdomen	isisu	isisu	mpa	mpa	mpa	maag
abduct	ukuba umntu	-thwala	-thopa	-kwetela	-utswa	skaak
abide	-hlala	-hlala	-dula	-dula	-nna	bly
ablaze	-vutha	-vutha	-tuka	-tukang	-tukang	in vlamme
abode	ikhaya	ikhaya	bodulo	bodulo	legaê	verblyfplek
above	ngaphezulu	phezu kwa-	godimo	hodimo	godimo	bo
abridge	-finyeza	-thoanisa	-khutsofatša	-kgutshufatsa	-khutswafatsa	verkort
absent	-ngekho	-ngekho	-hlokega	-ba siyo	-se yông	afwesig
absent, to be	ukungabukho	ukuba ngabikho	go hlokega	ho ba siyo	-tlhokêgwa, go	afwesig te wees, om
accede	-vuma	-vuma	-dumela	-dumela	-dumela	toege
accident	ingozi	ingozi	kotsi	kotsi	kôtsi	ongeluk
accompany, to	ukukhapha	-phelezela	-felegetša	ho felehetsa	-bulêdisa, go	vergesel
...						
account, to give an	ukubalisela	ukubalisisa	go hlaloša	ho hlalosa	supa tlôtlô, go	verslag te gee, om
account, give an	-balisela	-balisisa	-hlaloša	-hlalosa	supa tlôtlô	verslag gee
account, an	umlandu	i-akhawunti	tšhupamolato	akhaonte	tshupatlôtlô	'n rekening
accumulate (n)	-fumba	-andisa	-kgoboketša	-bokella	-koêla	ophoop
accuse	-tyhola	-beka icala	-bega	-qosa	latofatsa	beskuldig
accused, the	ummangalelwa	ummangalelw	mmeğiwa	moqosuwa	mosêkisiwa	beskuldigde, die
...						
acid (n)	iasidi	i-asidi	sedilana	esiti	êsêtê	suur (nw)
...						
act, an	isenzo	isenzo	tiro	ketso	tirô	'n daad
act, to	ukwenza	ukwenza	-dira	ho etsa	-dira, go	doen, om te
...						
afraid, be	-oyika	-esaba	-boifa	-tshaba	-bôifa	bang wees

Table 4: A selection of treated lemmas in CMD

The indication of part of speech, or the absence thereof, is problematic. For example, for the first ten lemmas, *abandon* – *absent* in Table 4, no part of speech guidance is given. The verb *accumulate* is incorrectly indicated as a noun by means of a part of speech label (*n*). There are three lemmas for *account*, all pertaining to a noun. One is followed by a comma plus "an" which could be interpreted as indicating its part of speech as a noun and which is indeed in line with all the nominals given as translation equivalents.

The approach of inserting a comma followed by so-called guide word(s) leads to many inaccuracies, inconsistencies, uncertainties and misinterpretations. Most users are likely to regard the use of the *,an* and *,to* conventions as implicit indicators of part of speech as nominal versus verbal, e.g. in the articles of *act, an* versus *act, to* which is correct for distinguishing between the nominal and verbal forms respectively. The pertinent statement that such words which follow the comma, e.g. *an, to, be*, etc. "must be utilised" in front of the lemma could even mislead the inexperienced user to believe that (*to*) *give an* must always be used in front of *account*, which is not the case.

For all the African languages the infinitive prefix is given in front of the

verb stem, but the comma convention is used in the Setswana translation equivalents. The treatment of *act*, *to* as reference to the infinitive form of the verb is inconsistent. The Setswana and Afrikaans equivalents follow the convention for English by giving *-dira* for Setswana preceded by a hyphen to indicate its status as a verb stem, followed by a comma and the infinitive prefix *go*, and Afrikaans *doen*, followed by a comma and the infinitive *om te* which is the equivalent of English *to*. The isiXhosa, isiZulu and Sesotho, however, do not follow the convention but give the infinitive forms as they would actually occur in text and speech i.e. with the infinitive prefixes preceding the verb. For Sepedi only the verb stem is given, with no infinitive guidance. To be consistent with the other languages it should have been either *go dira* to reflect actual occurrence or *-dira, go* to be in line with the English convention. Once again the user is ill-served because (s)he might conclude that *to act* is *go dira* in Setswana but *dira* in Sepedi whilst the two forms are exactly the same in Sepedi and Setswana.

Nouns are inconsistently handled as (a) without any part of speech indication, e.g. *accident* (b) with a part of speech label in brackets (*n*), e.g. *acid*, or (c) given with lead words such as *an*, or *the* as for *accused*. No label (*v*) is used for verbs.

CMD also claims that "the format adopted is straightforward and easy to understand; where the need arises for special connotations, a directive word is given between brackets, to indicate the particular shade of meaning associated with the particular leadword". This is useful in terms of homonym and sense distinction for instances such as *bow (in shooting)*, *gargle (mouth wash)*, *gargle (throat wash)*, etc. but less effective for Afrikaans (*language*) or alphabet (*letters of*) where the label does not make much sense, neither does it contribute to better understanding of the meaning of the word represented by the lemma or to delimitation of the range of application.

In CMD circumflexes are only indicated for Setswana. This creates the incorrect impression that, e.g. Sepedi and Setswana have the same word for *begin*, but that the pronunciation differs, i.e. *thoma* versus *thôma*. Circumflexes should be consistently and correctly used especially in cases where different meanings are conveyed, e.g. Sepedi *bola* 'speak' versus *bôla* 'rot'. Likewise, no tonal indication is given. Tonal indication in these dictionaries is at least required to distinguish between homonyms which have the same spelling but differ in tone such as *tlala* 'hunger' and *tlâla* 'become full'.

CMD also claims that using the dictionary does not require specialised knowledge of the languages covered by the dictionary because "all the troublesome grammatical complexities generally associated with the formation of the words have already been dealt with and incorporated into the words". Such "complexities" are indeed problematic for the user for lemma identification in order to look up isiZulu, isiXhosa, Siswati and isiNdebele words, cf. Prinsloo (2011) for a detailed discussion. In MLD the user has to identify the stem of the word before isiZulu and isiXhosa words can be looked up in the sections Xhosa/English/Afrikaans and Zulu/English/Afrikaans where these languages

are given as L1. CMD, however, does not provide lookup of isiZulu and isiXhosa words as L1, so CMD can hardly claim any credit for user-friendliness in terms of "not requiring specialised knowledge" or that "grammatical complexities ... have been dealt with" if it merely provides words from these languages as translation equivalents.

Finally spelling errors occur which is a serious mistake for any dictionary to make since dictionaries are often used to check spelling. So, for example, OJPD misspelled frequently used Sepedi words such as *woman* (**masadi* instead of *mosadi*), *wife* (**mogatsa* instead of *mogatšā*) found within a few randomly selected pages in the alphabetical section and *Monday* and *February* (**Mosupologo* instead of *Mošupologo* and **Dobokwane* instead of *Dibokwane*), respectively in the tables for Days of the week and Months of the year.

No pronunciation guidance is given in these dictionaries.

4.2 Comment on semantics

MLD admits that "shades of meaning and interpretations specific to the culture of one language are often lost in another" and continues that "this problem is obviously compounded when a word is translated through six other languages." (MLD: 6). It is not clear what is exactly meant by "shades of meaning". If it is taken to mean lack of minor differences in e.g. register or range of application it could be tolerable for a dictionary treating many languages simultaneously. If it refers to distinction between different senses of a word it constitutes a major deficiency, if it also refers to lack of distinction between homonyms it is an even bigger shortcoming of the dictionary. It will be illustrated in this section that the deficiencies in the dictionaries under discussion indeed go far beyond minor differences of meaning.

In consulting a dictionary the user expects a reasonable number of items in the *comment on semantics* in the article e.g. at least items giving basic data of the word represented by the lemma such as meaning, examples of use, collocations, different senses, idioms, etc. So, for example, treatment of the lemma *interest* in MID is limited to the single-word translations *inzalo*, *phaello*, *inzala*, *morokotso*, *rente* and *tswalo* in isiZulu, Sesotho, isiXhosa, Setswana, Afrikaans and Sepedi respectively. Sense distinction between *interest* (to be interested in something) and *interest* (profit, growth on investment) is not given. Lack of *sense distinction*, aggravated by inconsistent treatment and misguidance of the user is for example evident from the treatment of the lemma *with* (prep.) in MLD.

MLD

English	Afrikaans	Sepedi	Sesotho	Setswana	IsiXhosa	IsiZulu
with (prep)	met	le	le	ka	nga-	nga-

Table 5: MLD's treatment of *with* (prep)

Here only the sense of "together with/accompanying" is given for Sepedi and Sesotho and not the equally important sense of "with something, i.e. as an instrument" — thus an instance of an incomplete translation equivalent paradigm. The exact opposite applies for Setswana, isiXhosa and isiZulu in Table 5. *Ka* and *nga-* means "with something" but the translation equivalent for "together with" i.e. Setswana *le* and isiXhosa and isiZulu *na* is missing. This is correct for examples such as *O sepela le mosadi* 'He walks with the woman' but fails to add the other translation equivalent *ka* indicating with (an instrument), e.g. *O sepela ka maoto* 'He walks with (his) feet'. The user runs the risk of creating incorrect sentences such as **O sepela le maoto* 'he walks with (his) feet' in Sepedi or **Ukhuluma ngaye* 'he talks to her' in isiZulu. Although the translation equivalent paradigms in Table 5 are incomplete, at least the translations are correct. Even more serious mismatches are instances where homonyms (words with same spelling but different meanings) are confused in the treatment for the different languages. So, for example, is the lemma *might* (n) = power/strength in CMD translated for its nominal meaning in isiZulu and isiXhosa as *amandla* and in Sepedi and Sesotho as *maatla* and *matla* respectively but in Setswana as the modal verb *ka* 'can/may/might' instead of the nominal meaning *maatla*. The inexperienced user can incorrectly conclude that *ka* is the word for power in Setswana and produce **incorrect phrases such as **o tshwanetse go nna le ka* 'you should have power/strength'**.

Much more problematic for the multilingual model are instances of what Prinsloo and Gouws (2006) call grammatical divergence. An underlying detrimental characteristic of multilingual dictionaries is a "one fits all" approach and assumption. MID's treatment of pronouns results in numerous mistakes, inconsistencies and misguidance of the user.

English	IsiZulu	Sesotho	IsiXhosa	Setswana	Afrikaans	Sepedi
Pronouns:	Izabizwana:	Maemedi:	Izimelabizo:	Leemedi:	Voornaamwoorde:	Mainaina:
I	(mina) ngi-	nna	ndi-	nna	ek	nna
you	(wena) u-	wena	wena/u-	wena	jy	wena
he/she/it	(yena) u-/ (yena) u-/ (yona) i-	yena/yena/ yona	yena/yena/ yona	ena/ena/yona	hy/sy/dit	yena/yena/ yona
we	(thina) si-	re	si-	rona	ons	re
you	(nina) ni-	o	wena	wena	julle	le
they	(bona) ba-	ba	bona	bona	hulle	ba
me	mina	nna	mna	nna	ek	nna/ya ka
you	nawe	wena	wena	wena	jy/jou	wena
him/her/it	yena/yona	yena/yena/yona	yena/yena/yona	ena/ena/yona	hom/haar/dit	yena/yena/yona
us	thina	rona	thina/sithi	rona	ons	rena
you	wena	wena	wena/u-	wena	julle	wena
them	bona	bona	bona	bona	hulle	bona

Figure 9: Pronouns in MID: 303

This single table is fraught with different types of errors. For *I*, both the pronoun *mina* and the subject concord *ngi-*, which has a secondary pronominal function, is given for isiZulu but for Sesotho, Setswana and Sepedi only the pronoun *nna* and not the pronominal subject concord *ke-* is given. For isiXhosa only the subject concord *ndi-*, and not the pronoun, is given. The same information is offered for *you*, but now both the pronoun *wena* and the subject concord *u-* are given for isiXhosa. The same inconsistencies prevail for *we*, *you*, and *they*. In the case of *me* the equivalent *ya ka* is given which refers to *mine* (a possessive pronoun: *ya* = of, *ka* = me) for Sepedi but not considered for the other languages.

In the case of *he/she/it* use of the slash "/" convention is unclear. The presentation in Figure 9 gives the impression that *yena* is applicable to all cases for *he/she* and that *yona* for all possible referents of *it*. This is insufficient guidance for the African languages where third persons are subdivided into numerous noun classes resulting in different words for *he/she* and *it*.

Subject concords (Sc.); object concords (Oc.); demonstratives (Dem.); possessive concords (Poss.); emphatic pronouns (Ep.) and quantitative pronouns (Qp.).

Person or noun class	Example	Sc.	Oc.	Dem.	Poss.	Ep.	Qp.
1st Person singular	nna 'I'	ke	n-				
1st Person plural	rena 'we'	re	re				
2nd Person sing.	wena 'you' (singular)	o	go				
2nd Person plural	lena 'you' (plural)	le	le				
Class 1	monna 'man'	o/a	mo	yo	wa	yena	yohle
Class 2	banna 'men'	ba	ba	ba	ba	bona	bohle
Class 3	molato 'trouble, problem'	o	o	wo	wa	wona	wohle
Class 4	melato 'problems'	e	e	ye	ya	yona	yohle
Class 5	lesogana 'young man'	le	le	le	la	lona	lohle
Class 6	masogana 'young men'	a	a	a	a	ona	ohle
Class 7	selo 'object, thing'	se	se	se	sa	sona	sohle
Class 8	dilo 'objects, things'	di	di	tše	tša	tšona	tšohle
Class 9	ntlo 'hut'	e	e	ye	ya	yona	yohle
Class 10	dintlo 'huts'	di	di	tše	tša	tšona	tšohle
Class 14	bogobe 'porridge'	bo	bo	bjo	bja	bjona	bjohle
Class 15	go reka 'to buy'	go	go		ga		
Class 16	fase 'below'			fa			
Class 17	godimo 'above'	go	go		ga	gona	gohle
Class 18	morago 'behind'			mo			

Table 6: Sepedi noun classes, concords and pronouns

He/she/it is pronominalized depending on the noun class as *yena*, *wona*, *lona*, *sona*, or *yona* in Sepedi with similar paradigms for the other African languages.

On the same page (303) *he/she/it is* is translated in Sepedi as *o/o/e*. In the majority of cases *he/she/it is* is translated in Sepedi by the copulative particle *ke*, e.g. *he/she is a teacher*, '*ke morutiši*'. In following the information on this page the user could easily produce incorrect speech and text such as **(monna) o morutiši* 'he (the man) is a teacher'. The shortcomings are thus not limited to inconsistencies but guide the user to incorrect speech and text production.

The risk of the user being misguided by the above-mentioned shortcomings and inconsistencies is very high and refutes one of the basic principles of good lexicography, i.e. that the user should be guided by the dictionary not to make mistakes in especially text and speech production.

Finally the lack of examples of usage in these dictionaries is a significant shortcoming. Such examples help the user to understand the meaning of a word and to use it in text and speech production, and their value is not disputed in lexicography. For example, Kilgarriff et al. (2008) categorically state:

Users appreciate examples. If a dictionary entry includes an example which is a good match for the context in which the user has encountered a word, or for the context in which they want to use it, then the user generally gets what they want quickly and straightforwardly. Thus there is a case for including lots of examples, for lots of different contexts. Kilgarriff et al. (2008: 425)

Conclusion

In this article a critical evaluation of a number of South African multilingual dictionaries has been attempted. It is a well-known fact in lexicographic circles that no single dictionary can be everything for everyone. One should not be unreasonable in judging multilingual dictionaries from a single perspective, e.g. only on the number of treated lemmas while ignoring the many benefits and learning/reading potential of the thematically ordered sections. One also cannot expect detailed treatment of lemmas in each of the languages in a multilingual dictionary. However, as quoted from Haas (1962) the users expect to find the words they are looking for and therefore at least the most common words of the language should be included and correctly treated. As for any other dictionary the multilingual dictionary should also be a product of high lexicographic achievement. One cannot but come to the conclusion that the multilingual dictionaries used in this study are not products of high lexicographic achievement. In a sense it could be argued that multilingual dictionaries indeed try to be everything to everyone. It has been indicated that some shortcomings can be attributed to the model itself, e.g. the impossibility to satisfactorily treat multiple languages within the scope and physical limitations of a single paper dictionary. On the macrostructural level compilers of future multilingual paper dictionaries should ascertain coverage of the basic/common words of the language and make sure that at least the most common words of the languages are covered. On microstructural level compilers should pay much more attention to user guidance, sense distinction, different aspects of

consistency, parity between translation equivalents, conventions used, and to correct all kinds of errors. Consultation with linguistic experts of the African languages would also not be amiss.

Acknowledgement

This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant-specific unique reference number (UID) 8576. The Grantholder acknowledge that the opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF-supported research are those of the author, and that the NRF accepts no liability whatsoever in this regard.

References

Dictionaries

- (ANNA) Martin, Willy (Main ed.). 2011. *Pharos Groot Woordeboek. Afrikaans en Nederlands (Prisma Groot Woordenboek Afrikaans en Nederlands)*. Cape Town: Pharos.
- (CMD) Jennings, L.E. et al. 1995. *The Concise Multilingual Dictionary. English, Xhosa, Zulu, Northern Sotho, Southern Sotho, Tswana, Afrikaans*. Johannesburg: Ad Donker.
- (EANTA) Uys, Isabel. 2008. *The English, Afrikaans, Northern Sotho, Tswana Aid: Word Lists & Phrases in Four Languages*. Cape Town: Pharos.
- (MED) Rundell, M. (Ed.). 2007. *Macmillan English Dictionary for Advanced Learners*. Second Edition. Oxford: Macmillan.
- (MID) Bennett, J. et al. 2010. *Multilingual Illustrated Dictionary. English, IsiZulu Sesotho, IsiXhosa, Setswana, Afrikaans, Sepedi*. Cape Town: Pharos.
- (MLD) Reynierse, C. (Ed.). 1991. *South African Multi-Language Dictionary and Phrase Book. English, Afrikaans, Northern Sotho, Sesotho, Tswana, Xhosa, Zulu*. Cape Town: Reader's Digest Association South Africa.
- (OJPD) Goodwill, J.S. et al. 1991. *The Oxford Junior Primary Dictionary for Southern Africa with North Sotho, South Sotho, Setswana and Afrikaans Words*. Cape Town: Oxford University Press.

Other sources

- Gouws, R.H., D.J. Prinsloo and M. Dlali. 2014. A Series of Foundation Phase Dictionaries for a Multilingual Environment. *Stellenbosch Papers in Linguistics* 43(2014): 23-43.
- Haas, Mary R. 1962. What Belongs in a Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1962. *Problems in Lexicography*: 45-50. Bloomington: Indiana University.
- Kilgariff, A., M. Husák, K. McAdam, M. Rundell and P. Rychlý. 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda and Janet DeCesaris (Eds.).

2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 425-432. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

(PEIC) Gauton, Rachéle: The University of Pretoria English Internet Corpus.

Prinsloo, D.J. 2011. A Critical Analysis of the Lemmatisation of Nouns and Verbs in isiZulu. *Lexikos* 21: 169-193.

Prinsloo, D.J. and R.H. Gouws. 2006. Lexicographic Presentation of Grammatical Divergence in Sesotho sa Leboa. *South African Journal of African Languages* 16(4): 184-197.