The Compilation of Electronic Dictionaries for the African Languages

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

Abstract: Lexicographers increasingly acknowledge the enormous potential of electronic dictionaries. The great capacity and speed characteristic of electronic products, combined with enhanced query and data retrieval technology, pave the way to a new generation of dictionaries unimagined in the paper-dictionary era. It is amazing to see how many of the lexicographer's greatest obstacles disappear in the electronic dictionary. This article will, firstly, attempt to give a perspective on typical features of electronic dictionaries. Secondly, electronic-dictionary entries will be designed as a solution to some of the most burning lemmatization problems encountered by lexicographers for African languages in paper dictionaries.

Keywords: ELECTRONIC DICTIONARY, LEXICOGRAPHY, DATA RETRIEVAL, LEMMATIZATION, CD-ROM, ACCESS ROUTE, POP-UP FUNCTIONALITIES, POP-UP SCREENS, EDUTAINMENT, CROSS-REFERENCING, INFORMATION RETRIEVAL, ENCODING, DECODING, AFRICAN LANGUAGES, NAVIGATION BAR.

Opsomming: Die samestelling van elektroniese woordeboeke vir die Afrikatale. Leksikograwe erken in toenemende mate die enorme moontlikhede van elektroniese woordeboeke. Die groot vermoë en spoed wat kenmerkend is van elektroniese produkte, tesame met die bykomende tegnologie van soektogte en dataopsporing, berei die weg voor na 'n nuwe generasie woordeboeke wat ondenkbaar was in die era van gedrukte woordeboeke. Dit is verbasend om te sien hoeveel van die leksikograaf se grootste struikelblokke verdwyn in die elektroniese woordeboek. Hierdie artikel sal eerstens probeer om 'n oorsig te gee oor die kenmerkende eienskappe van elektroniese woordeboeke. Tweedens sal inskrywings vir die elektroniese woordeboek ontwerp word as oplossing vir sommige van die dringendste lemmatiseringsprobleme wat leksikograwe van Afrikatale by gedrukte woordeboeke teëkom.

Sleutelwoorde: ELEKTRONIESE WOORDEBOEK, LEKSIKOGRAFIE, DATAOPSPORING, LEMMATISERING, CD-ROM, TOEGANGSROETE, OPSPRINGFUNKSIONALITEITE, OPSPRINGSKERMS, OPVOEDKUNDIGE VERMAAK, KRUISVERWYSING, INLIGTINGSHERWINNING, ENKODERING, DEKODERING, AFRIKATALE, NAVIGASIESTAAF.

Introduction

The electronic dictionary, whether on CD-ROM, online, or hand-held[1], will supersede the paper dictionary in ways unimaginable in the paper-dictionary dimension, just as the computer has completely superseded the typewriter. Many lexicographers are of the opinion that, generally speaking, the paper dictionary has attained its maximum potential. In reference to learners' dictionaries, Bolinger (1990: 144) says that hard-copy dictionaries have reached their capacity and that "any really dramatic advance would burst the covers". Nesi (1999: 56) emphasizes the fact that many useful features such as indexes and cross-reference symbols have been added in the paper-dictionaries to assist the user in the navigation of multiword searches and that coding systems have been refined. However, she (Nesi 1999: 55-56) concludes:

> The more information the paper-based dictionary contains, the harder (and more time-consuming) it will become for learner users to find exactly what they need to know, without first having to negotiate a quantity of information that they do not need to know, or cannot process.

Meijs (1992: 152, as reported by Nesi 1999: 65) predicts "the imminent demise of the dictionary as a book":

> In a decade or so, on-line dictionaries on disk or CD-ROM will no doubt be the norm rather than the exception.

Atkins (1996: 515-516) takes a less subtle stand:

> At last we are liberated from the straitjacket of the printed page and alphabetical order.

It is amazing how many of the lexicographer's major problems of the past centuries simply disappear in the electronic dictionary dimension. For example, the selection problem, what to put in the dictionary and what to leave out, traditionally believed to be one of the lexicographers greatest difficulties, is solved. Limitations on the size of the dictionary are no longer major concerns, especially in online dictionaries. No difficult decisions on lumping or splitting, one of the greatest vexations of the African language lexicographer, have to be made. The dictionary comes alive in a way that is unimaginable in the paper-dictionary dimension. For the first time the user can control data accessing from the dictionary. He/She can decide, for example, to access only those information types that are relevant to him/her at that very moment, for example, definitions, translation equivalents, etymology, examples of use, collocations, citations, and so forth, by simply switching them on or off (as in Table 3).

In exploring the electronic version of the *Grote Van Dale*, Geeraerts (2000: 76) distinguishes three basic functions of that electronic dictionary: *semasiological*, *onomasiological* and *edutainment* functions. He points out that these functions are combined in a single dictionary and that this fact illustrates the enhanced multifunctionality that can be achieved in an electronic dictionary.

It could be argued that the power of an electronic dictionary is rooted in its *storage capacity*, the *speed of information retrieval* and *sophisticated search and* *querying strategies*.

> The storage capacity of a CD-ROM is about 600 MegaBytes — enough to contain, for example, the 44 million-word database of the Britannica CD™. The twenty-volume Oxford English Dictionary, and the twelve-dictionary compilation of The Sanseido Wordhunter (Sharpe 1995: 48) are both sold on single CD-ROMs. (Nesi 1999: 59)

> No need to *compress text* into cramped paragraphs — each definition and example sentence can start on a new line. No need for *obscure codes* — fuller descriptions become possible (e.g. parts of speech and grammar descriptions no longer need to be abbreviated to save space), with hyperlinks to more detailed information. No need to *exclude possible inclusions or extra example sentences* on grounds of space. (Harley 2000: 85)[2]

> A computer database is *almost infinitely extensible*, and so there is more scope for the *inclusion of extra material on any item*. … Far *more examples can be stored* and presented to the user. … The … dispute about the inclusion or exclusion of encyclopaedic information also disappears. These data can be retained in the database without a balancing reduction in the space used for "purely" dictionary material. (Dodd 1989: 91)

As for speed, regular users of electronic dictionaries will know that the average retrieval time of an entry is but a second or two.

One should also not think of an electronic dictionary as a fixed object or as fixed as a printed dictionary is. Rather one should think of an electronic dictionary in terms of an *organic*, *changing database* through which a variety of searches and even artificial intelligence queries may be conducted, or, in the words of Dodd, as a "dictionary service":

> We are not far from the point at which the *dictionary will cease to be merely a product*, such as a book, or a somewhat more sophisticated substitute for a book, for example, a CD-ROM, which remains as fixed in its contents as a book is, and *will also become a service*. … Instead of multiple identical copies of a dictionary, sold to users, there would be *a single version of a database, from which clients of the dictionary service obtained the information they required*, much as professionals of various sorts already get abstracts and similar data "on-line". (Dodd 1989: 87)

Consider in this regard the online version of the Oxford English Dictionary (*OED Online*) in which efforts are made to update the dictionary continuously "offering online subscribers a unique opportunity to chart the course of the English language on an ongoing basis" (*OED Online*, Introduction).

Towards multiple access routes to the dictionary[3]

In contrast to the paper dictionary a variety of sophisticated search and querying strategies become possible in the electronic dimension. In fully exploiting such strategies the compiler of an electronic dictionary can offer the user an exciting new range of data access routes to the dictionary and at the same time can add a new dimension to traditional paper-dictionary access routes.

> Access to and between sources on CD-ROM is quickly *made by a variety of routes. Dictionary, grammar and usage information in COBUILD, LIED and LIAD may be accessed directly, by clicking on a section menu.* OALD on CD-ROM likewise provides a choice of five menus, for the A-Z dictionary, pictures, maps, appendices and games. In LIED and LIAD *the user searching one section of the database will also be alerted if extra information about a search word is contained elsewhere*, while in COBUILD *all sections of the database are accessed during an initial word search*; the number of 'hits' in each area are then displayed on the screen and the user can choose whether to access the Dictionary, usage, Grammar or Word Bank. All the dictionaries also allow *searches by 'chaining' or 'hyperlinking'*, a search mechanism by which a double click on a word on screen will call up a dictionary entry for that word. (Nesi 1999: 61)

> The major advantage of a computer-based dictionary is that it *permits a whole range of new routes to the data stored in it.* A printed dictionary is condemned to a single method of gaining access to the information it holds, usually an alphabetic route. … In a truly dynamic dictionary, it should be possible to gain access to any entry by means of any of the pieces of information composing it. *Potential routes are thus limited only to the frontiers of what is contained in the dictionary, combined with possible manipulations or intersections of these items of data.* (Dodd 1989: 87, 88)

Dodd (1989: 89) envisages a series of typical potential search routes such as:

> "sounds like A"; "rhymes with B"; "is spelt like C"; "has an etymology of D"; "dates from year/century E"; "is used in style of F"; "is used in technical field G"; "is an antonym of H"; "is a synonym of I"; "is a hyponym of J "; "is a superordinate of K"; "includes the word(s) L in its definition"; "is of grammatical class M", "has syntactic valency or pattern N".

Perhaps the single most promising new search route/strategy currently being developed in electronic dictionaries is one that enables users to find words based on their semantic specifications in contrast to the traditional way of looking up alphabetically ordered lexemes. Formulated differently it means that an information retrieval access route meaning/definition → lemma has been developed in contrast to the traditional route, lemma → meaning/definition.

> Disk-based dictionaries are beginning to cast off the constraints of hardcopy and *to move away from a linear approach based on form rather than meaning*. Rogers (1996: 84) points out that electronic replications of paperbased publications are *still 'word-based rather than meaning-based'* even though they offer better search and retrieval facilities. She proposes a *semantically organized dictionary which would take the user from the definition to the word*, and which would deal with such queries *as 'find me*

> *the name of the thing* which is a kind of boat and which is flat-bottomed and travels on canals and rivers'. (Nesi 1999: 63)

Prospective electronic-dictionary compilers should, however, realize that an electronic dictionary is not merely "a paper dictionary on computer", just as television is not a "picture-radio" or a computer is not an electronic typewriter. The user would not look up the word on-screen in the same way as he/she would look it up in a paper dictionary. Sharpe (1995: 48) rightfully complains that most electronic bilingual dictionaries "seem to use the content of printed dictionaries as their database without making any additions or alterations" thus not utilizing the potential of the electronic dictionary to the full. Atkins (1996: 515-516) agrees:

> Dictionaries of the present … may even come to you on a CD-ROM rather than in book form, but *underneath these superficial modernizations lurks the same old dictionary.* … It is up to us to take up the real challenge of the computer age, by asking not how the computer can help us to produce old-style dictionaries better, but how it can help us to create something new.

Consider the following example which is an extract from *SeDiPro 1.0* (Prinsloo and De Schryver 2000), a monodirectional Northern Sotho – English dictionary compiled with a basic word-processing package. The paper version can only be used in the direction Northern Sotho → English. However, the basic built-in search function of a typical word processor is useful in looking up Northern Sotho words, and this function can even be used to look up English words, thus simulating the existence of an English → Northern Sotho dictionary.

thuša help, assist, **~go** useful, handy, serviceable, profitable, who helps
thušanô mutual aid/assistance, co-operation
thušô help, assistance, aid, **~ya potlakô/tlhaganêlô/pele** first aid
thušôpele first aid
thušwa be helped/assisted/aided
thušwe, thušwê must be helped/assisted/aided; **..ga/sa/se..~** not be helped/assisted/aided

The user can type any English word (or part of it) into the "Find/ Search" box, for example, "help", "assist", "co-operation", and so on, and find the Sepedi lemma-sign in which article the requested English word is included. However useful this "added value" to the user might be, it does not mean that the dictionary can be regarded as an electronic dictionary. It still offers little more than searchable versions of the printed text.

Dodd (1989: 87, 88) emphasizes the limitations in this regard in reference to the *Oxford English Dictionary*. Also the very useful *Pharos Dictionaries 5 in 1* (2000) which is marketed as an electronic dictionary, is to a large extent merely five paper dictionaries on CD-ROM. A true electronic dictionary should offer much more than such "value-added" features. It should reflect a totally new design that exploits all available electronic features.

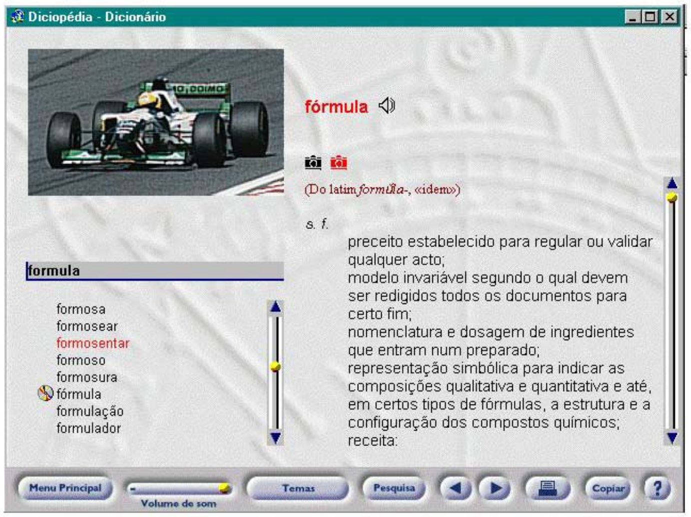
> Electronic formats have big advantages over paper formats, a fact that has too often been sadly ignored when electronic dictionaries have provided little more than searchable versions of the printed text. … The important thing to note is that it is a new design process. (Harley 2000: 85)

In the following paragraph a brief overview of typical functions of electronic dictionaries is given especially for the benefit of the reader who might not be so familiar with electronic dictionaries. For a detailed discussion of the numerous advantages of electronic dictionaries and their functions, see De Schryver (forthcoming).

Enhanced information retrieval processes

Table 1 represents the typical layout of an entry in the electronic Portuguese dictionary *Diciopédia, Grande Dicionário Enciclopédico Multimédia* (1997). It reveals a neatly integrated layout of pronunciation, two clickable camera icons, a full treatment of the lemma, an alphabetical scroll bar, a general navigation bar, and so on.

Table 1



formula

formosa
formosear
formosentar
formoso
formosura
fórmula
formulação
formulador

fórmula

(Do latim *formŭla-*, «idem»)

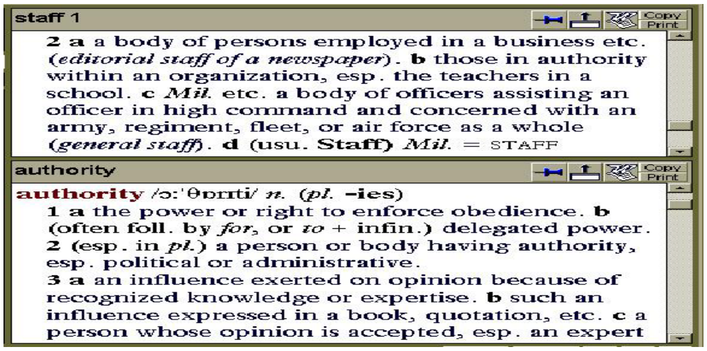
s. f.

preceito estabelecido para regular ou validar qualquer acto;
modelo invariável segundo o qual devem ser redigidos todos os documentos para certo fim;
nomenclatura e dosagem de ingredientes que entram num preparado;
representação simbólica para indicar as composições qualitativa e quantitativa e até, em certos tipos de fórmulas, a estrutura e a configuração dos compostos químicos;
receita:

Probably the most exciting feature for a first-time user who is learning to master the pronunciation of words in the target language, is the function that enables the user to actually hear the pronunciation of words through the speakers attached to his/her computer. This is normally achieved by simply (double-)clicking on the phonetic transcription of the lemma, or, as in this case, on a specially designed icon. No phonetic orthographical presentation, be it a formal orthography such as IPA or an informal respelling of the word, can even come close to hearing the lemma pronounced by live male or female voices. Paper dictionaries have probably reached their maximum potential as far as guiding the user in pronunciation is concerned and the electronic pronunciation function is a good example of what is meant by "the electronic dictionary superseding the paper dictionary".

The second exciting innovation is user control over cross-referencing in the electronic dictionary. In the past the user had to page manually from an entry at the beginning of the dictionary to a cross-reference at the back, meanwhile forgetting along the way what he/she was looking for. In the electronic dictionary dimension the user can now read a definition and if any word is unclear he/she can immediately access a full cross-reference by simply double-clicking on the unclear word. Consider the following example from the electronic version of the *Concise Oxford Dictionary* (ninth edition on CD-ROM).

Table 2



staff 1

2 a a body of persons employed in a business etc. (*editorial staff of a newspaper*). **b** those in authority within an organization, esp. the teachers in a school. **c** *Mil.* etc. a body of officers assisting an officer in high command and concerned with an army, regiment, fleet, or air force as a whole (*general staff*). **d** (usu. **Staff**) *Mil.* = STAFF

authority

authority /ɔːˈθɒrɪti/ *n.* (*pl.* **-ies**)
1 a the power or right to enforce obedience. **b** (often foll. by *for*, or *to* + infin.) delegated power.
2 (esp. in *pl.*) a person or body having authority, esp. political or administrative.
3 a an influence exerted on opinion because of recognized knowledge or expertise. **b** such an influence expressed in a book, quotation, etc. **c** a person whose opinion is accepted, esp. an expert

While looking up the word *staff* in this electronic dictionary, the user is confronted with the word *authority* which is also unknown to him/her. However, by immediately double-clicking on *authority* a definition of this word is presented. In this dictionary the user has the added advantage that both definitions can be displayed on screen simultaneously, thus bringing both reference entry and reference address visually together.

The user can even adjust the font size of the different information types relative to each other in order to easily retrieve information.

Some electronic dictionaries such as the *Oxford English Dictionary* (second edition on CD-ROM, OED II), allow the user to customize the dictionary by offering him/her limited options of switching information categories on or off.

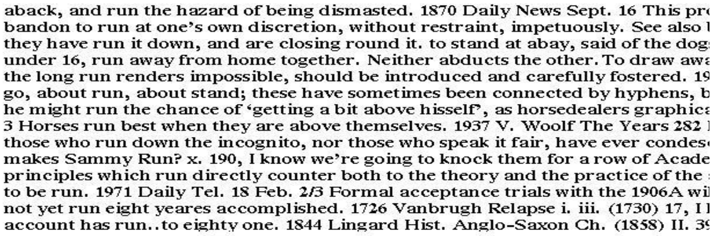
In OED II the user has the choice to switch the categories *etymology*, *definition* and *quotations* on or off. Compare, for example, four different information offerings which can be achieved in this way for the word *Miliola*.

Table 3

1 *etymology:* *OFF* *definition:* *OFF* *quotations:* *OFF*	‖ **Miliola** (mɪ'laɪəʊlə). *Zool.* Pl. **-æ**.
2 *etymology:* *ON* *definition:* *OFF* *quotations:* *OFF*	‖ **Miliola** (mɪ'laɪəʊlə). *Zool.* Pl. **-æ**. ⌈mod.L., dim. of L. *milium* millet.⌉
3 *etymology:* *ON* *definition:* *ON* *quotations:* *OFF*	‖ **Miliola** (mɪ'laɪəʊlə). *Zool.* Pl. **-æ**. ⌈mod.L., dim. of L. *milium* millet.⌉ An important genus of imperforate foraminifera; an animal of this genus.
4 *etymology:* *ON* *definition:* *ON* *quotations:* *ON*	‖ **Miliola** (mɪ'laɪəʊlə). *Zool.* Pl. **-æ**.. ⌈mod.L., dim. of L. *milium* millet.⌉ An important genus of imperforate foraminifera; an animal of this genus. **1836** Buckland *Geol. & Min.* (1837) I. 385 The *Miliola*, a small multilocular shell, no larger than a millet seed, with which the strata of many quarries in the neighbourhood of Paris are largely interspersed. **1879** Carpenter in *Encycl. Brit.* IX. 376/2 The shells of the *Miliolæ*..are at present found in the shore sands of almost every sea.

Electronic dictionaries such as the OED II enable the user to run full text searches of the lemma which amount to the generation of concordance lines. Consider the following selection from a full text search of *run* in OED II.

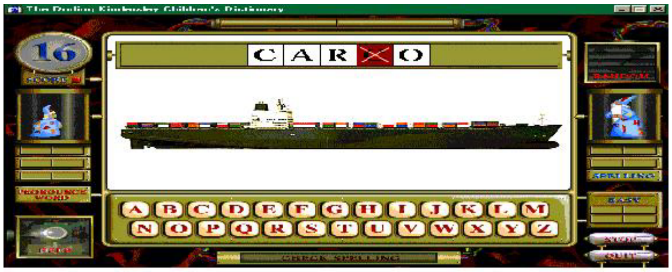
Table 4



```
aback, and run the hazard of being dismasted. 1870 Daily News Sept. 16 This pr
bandon to run at one's own discretion, without restraint, impetuously. See also
they have run it down, and are closing round it. to stand at abay, said of the dog
under 16, run away from home together. Neither abducts the other. To draw aw
the long run renders impossible, should be introduced and carefully fostered. 19
go, about run, about stand; these have sometimes been connected by hyphens, b
he might run the chance of 'getting a bit above hisself', as horsedealers graphic
3 Horses run best when they are above themselves. 1937 V. Woolf The Years 282
those who run down the incognito, nor those who speak it fair, have ever condes
makes Sammy Run? x. 190, I know we're going to knock them for a row of Acade
principles which run directly counter both to the theory and the practice of the
to be run. 1971 Daily Tel. 18 Feb. 2/3 Formal acceptance trials with the 1906A wi
not yet run eight yeares accomplished. 1726 Vanbrugh Relapse i. iii. (1730) 17, I
account has run..to eighty one. 1844 Lingard Hist. Anglo-Saxon Ch. (1858) II. 3
```

Learning the language while using the electronic dictionary can be great fun for children. Electronic dictionaries offer a lot in terms of *edutainment* by allowing the user to click on words and pictures and play spelling or word games. Consider the following example from *The Dorling Kindersley Children's Dictionary* (1996).

Table 5



In this example taken from the spelling game, the user has been presented with the picture of cargo on a ship but has given the incorrect final consonant. This mistake is then clearly marked and he/she is given the opportunity to correct it. Correct and incorrect efforts are reacted to by a variety of sound effects.

A perspective on crucial lemmatization problems in the African languages

Lexicographers generally agree that lemmatisation, especially of nouns and verbs, in the African languages is extremely problematic. The lexicographer is, in the words of Prinsloo and Gouws (1996: 103), the *mediator between linguistics and the everyday dictionary user*. For the African languages it means that the lexicographer must find lemmatisation strategies that result in a user-friendly end product. However, as will be indicated in terms of Van Wyk (1995) below, instances even occur where neither the user nor the lexicographer knows for sure what the exact lemma-sign should be, especially in so-called "stem dictionaries" where lexical items are lemmatized by their stem forms, e.g. **mosadi** 'woman', under the stem **-sadi**. The core of the problem lies in a complicated derivational system and a multitude of irregular forms in the language. Such problems are multiplied if the language has a conjunctive orthography.

> The treatment of lexical entries in the dictionary has been one of the main difficulties in Swahili lexicography, as in that of many other Bantu languages. *The central problem is particularly the method of arranging the nominal and verbal items of the language, emanating from the complex morphological structure common to Bantu languages, of a nominal classification system categorising nouns by means of prefixes and a verbal derivation system forming new verbs by means of derivational affixes.*

u	subject
me	present tense perfect aspect
tu	object
kut	verbal base ("meet")
an	reciprocal
ish	causative
a	declarative

> The main task for the lexicographer will be to decide on which lexical forms should be listed in the dictionary and in what manner. (Bwenge 1989: 5, 6)

Furthermore, lexicographers report a constant struggle in the paper dimension against the redundancy factor when trying to increase user-friendliness.

> Het punt is dat "een connectief + **balùme**" lemma-status heeft en dus gelemmatiseerd zou moeten worden in de macro-structuur. Dit zou echter betekenen dat men deze informatie ofwel op 15 (aantal *verschillende* connectieven) plaatsen herhaalt; ofwel 14 verwijzingen voorziet naar één bepaald connectief. Daarbij komt nog dat men deze werkwijze dan ook zou moeten herhalen voor alle lemma's beginnend met een connectief! Een *ontoelaatbare redundantie* dus. (De Schryver and Kabuta 1998: xiii, original italics)

> Such complexities can lead to major problems in a normal printed dictionary, obliged to be of much greater bulk than otherwise, through including the varying forms at least as cross-references to the normal headword. The alternative is *to force the unsophisticated user to wearisome plodding through potential entries, with no certainty of success.* (Dodd 1989: 90)

As far as the lemmatization of *nouns* on the macrostructural level is concerned, compilers generally fail in their efforts to lemmatize them satisfactorily within the physical limitations of a printed dictionary while still producing a user-friendly product, or they err by including nouns unlikely to be looked for by the target user at the expense of essential ones.

Physical limitations on volume (generally based on the number of pages and therefore on the amount of entries that can be accommodated in a specific dictionary or subdictionary) have a far greater impact on lemmatization in African languages than one would expect. It amounts to the need for a *strategy of word selection* (that is, a strategy to determine which words are to be chosen and which words can be left out of the dictionary) and the absence of such a strategy.

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

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

> The decision what to include in the dictionary still has to be made by the lexicographer himself, however, and this depends in turn upon the nature and size of the dictionary and its intended users. In this respect lemmatised frequency-lists can be a further help … We have reached a stage where co-operation between man and machine is useful and perhaps indispensable in making better dictionaries. (Martin et al. 1983: 81-2, 87)

> Lexicographers constantly have to make pragmatic decisions on what to include

> in a dictionary to conform to the dictates of space available. (Walter 1996: 640)

This fact, namely the need for selection in the sense of which nouns to include or exclude from the dictionary, whether to lemmatize singular and plural forms or singular forms only, whether to lemmatize noun stems, and the procedure for handling irregular forms, is an impediment for many compilers of African language dictionaries. (See Prinsloo and De Schryver (1999) for a detailed analysis of problematic aspects regarding the lemmatisation of nouns and an evaluation of different lemmatisation strategies.)

Van Wyk (1995: 89) emphasizes the fact that the prefix morphology of nouns, apart from being irregular, is also subject to fairly complex morphophonological rules. He elaborates on this issue in great detail and analyses existing dictionaries where nouns are lemmatised on their stem forms. It is clear that in many instances (such as stems containing the *nasal prefix* of class nine or *aspirated* and *non aspirated* noun stems) it is simply not possible for *neither the user or the lexicographer to determine what the stem is*. This results in great inconsistency and even the abandonment of the dictionary's editorial policy. This immense difficulty for lexicographers can be summarised in terms of the following quotation from Van Wyk (1995: 90):

> This forces the lexicographer to choose between four options. (1) He may lemmatize all such stems with aspirated consonants, irrespective of whether they occur in this form or not, e.g. *impala* 'impala', *intaba* 'mountain', *impilo* ['health'] and *inkosi* ['king'] as **-phala**, **-thaba**, **-thombi** and -**khosi** respectively. (2) He may enter all stems with unaspirated consonants, i.e. **-pala**, **-taba**, **-tombi** and **-kosi**. (3) He may enter those stems which also occur with aspirates under the aspirates, and stems which do not under the unaspirated explosives, e.g. **-thombi**, **-khosi**, **-pala** and -**taba**. (4) Or he may enter stems with corresponding aspirates under the aspirates, and those without under the relevant nasal compounds, e.g. -**thombi**, **-khosi**, **-mpala** and **-ntaba**. … DV [Doke and Vilakazi 1948], and virtually all other stem lexicographers, opt for the fourth option. This puts the onus on the user to know which nouns have stems to which the de-aspiration rule applies and which not. The result is quite confusing, as the following examples show.

NOUN	LEMMA
impala 'impala'	**-mpala (impala)**
impilo 'health'	**-philo (impilo)** [<phila]
intaba 'mountain'	**-ntaba (intaba)**
intombi 'girl'	**-thombi (intombi)** [<thomba]
ubuntombi 'maidenhood'	**-ntombi** [<intombi]
inkosi 'king'	**-khosi (inkosi amakhosi)**
inkabi 'ox'	**-nkabi (inkabi)**

In the case of verbs, numerous derivations have to be considered by the lexicographer and will be illustrated by the Sepedi verb *reka* "to buy".

In the case of verbs, Ziervogel and Mokgokong (1975) offer a complicated tiered layout of 18 modules, each representing the root, with or without one or more suffixes. For each module, "standard modifications" (as Prinsloo (1994: 96) calls them) are added. These "standard modifications" (the *perfect*, *passive* and *perfect plus passive*) are given in the third column of Table 6. Finally numerous nominal derivations within each of the modules are given. These are listed in the fourth column of Table 6.

Table 6

#	***structure***	***derivations***	***deverbatives***
1	root + standard modifications	*reka, rekile, rekwa, rekilwe*	*direkarekane, lereko, mareko, moreki, bareki, sereki, direki, sereko, direko, theko, ditheko*
2	root + reciprocal + standard modifications	*rekana, rekane, rekanwa, rekanwe*	*barekani, thekano, dithekano*
3	root + reciprocal + causative + standard modifications	*rekantšha, rekantšhitše, rekantšhwa, rekantšhitšwe*	*morekantšhi, barekantšhi, serekantšhwa, direkantšhwa, thekantšho, dithekantšho*
4	root + alernative causative + standard modifications	*rekanya, rekantše, rekanywa, rekantšwe*	*morekanyi, barekanyi, serekanywa, direkanywa, thekanyo, dithekanyo*
5	root + neutro passive + standard modifications	*rekega, rekegile*	
6	root + applicative + standard modifications	*rekela, reketše, rekelwa, reketšwe*	*borekelo, morekedi, barekedi, morekelwa, barekelwa, serekelo, direkelo, thekelo, dithekelo*
7	root + applicative + reciprocal + standard modifications	*rekelana, rekelane, rekelanwa, rekelanwe*	*barekelani, thekelano, dithekelano*
8	root + causative + standard modifications	*rekiša, rekišitše, rekišwa, rekišitšwe*	*morekiši, barekiši, serekišwa, direkišwa, thekišo, dithekišo*
9	root + causative + reciprocal + standard modifications	*rekišana, rekišane, rekišanwa, rekišanwe*	*barekišani, thekišano, dithekišano*
10	root + causative + neutro passive + standard modifications	*rekišega, rekišegile*	
11	root + causative + applicative + standard modifications	*rekišetša, rekišeditše, rekišetšwa, rekišeditšwe*	*borekišetšo, morekišetši, barekišetši, thekišetšo, dithekišetšo*
12	root + causative + applicative + reciprocal + standard modifications	*rekišetšana, rekišetšane, rekišetšanwa, rekišetšanwe*	*barekišetšani, thekišetšano, dithekišetšano*
13	root + reversive transitive + standard modifications	*rekolla, rekolotše, rekollwa, rekolotšwe*	*morekolli, barekolli, serekollwa, direkollwa, thekollo, dithekollo*
14	root + reversive transitive + reciprocal + standard modifications	*rekollana, rekollane, rekollanwa, rekollanwe*	*barekollani, thekollano, dithekollano*
15	root + reversive transitive + applicative + standard modifications	*rekollela, rekolletše, rekollelwa, rekolletšwe*	*morekolledi, barekolledi, thekollelo, dithekollelo*
16	root + reversive transitive + applicative + reciprocal + standard modifications	*rekollelana, rekollelane, rekollelanwa, rekollelanwe*	*barekollelani, thekollelano, dithekollelano*
17	root + reversive transitive + causative + standard modifications	*rekolliša, rekollišitše, rekollišwa, rekollišitšwe*	*morekolliši, barekolliši, thekollišo, dithekollišo*
18	root + reversive transitive + causative + reciprocal + standard modifications	*rekollišana, rekollišane, rekollišanwa, rekollišanwe*	*barekollišani, thekollišano, dithekollišano*

Ziervogel and Mokgokong (1975) deserve some credit for their exhaustive treatment of verbs, but they produced an extremely unfriendly product in the end. (See Prinsloo (1994) for a detailed discussion.)

Towards a solution in the electronic dimension of lemmatization problems in the African languages

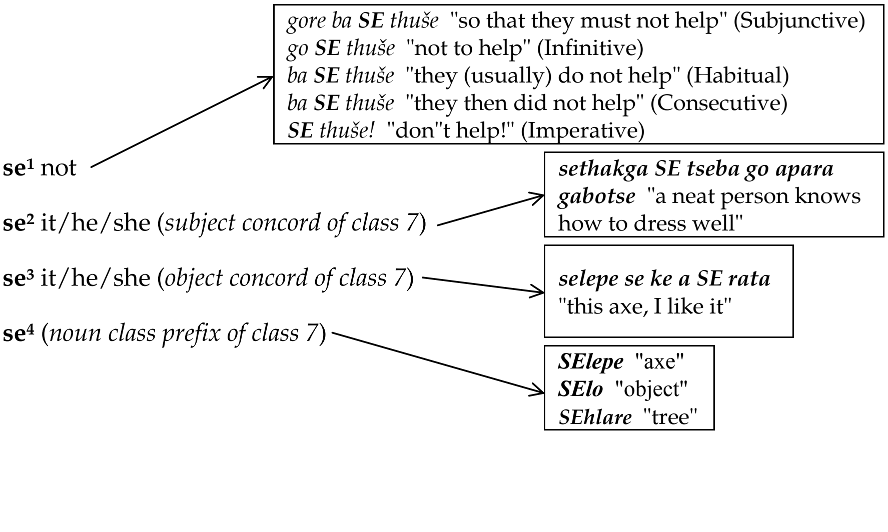
After the presentation of the principal problems with lemmatization in the African languages, it will now be attempted to design typical prototypes of (mainly) pop-up screens to solve these "major" lemmatization problems that are encountered in the paper dictionary.

> Adding information is facilitated by the pop-up functionalities of the electronic product. Information that would take up too much space in the classical dictionary, can now be made accessible from within an entry without disturbing the layout of the article. (Geeraerts 2000: 77)

Prospective lexicographers could adapt the following prototypes for any African language or use them as models for lemmatization problems other than those for nouns and verbs considered here. Most of these designs are deliberately proposed within the *capabilities and limitations of an ordinary word processor*. This means that sophisticated multimedia capabilities are not, per definition, a prerequisite for "adding electronic value" to the dictionary.

In principle a variety of approaches is possible. The lexicographer could for instance decide to compile a dictionary entry which resembles the typical layout in a conventional paper dictionary as in Table 7, linking the pop-up screens directly with selected words in the article.

Table 7



se^{1} not

gore ba ***SE*** *thuše* "so that they must not help" (Subjunctive)
go ***SE*** *thuše* "not to help" (Infinitive)
ba ***SE*** *thuše* "they (usually) do not help" (Habitual)
ba ***SE*** *thuše* "they then did not help" (Consecutive)
SE *thuše!* "don"t help!" (Imperative)

se^{2} it/he/she (*subject concord of class 7*)

sethakga SE tseba go apara gabotse "a neat person knows how to dress well"

se^{3} it/he/she (*object concord of class 7*)

selepe se ke a SE rata "this axe, I like it"

se^{4} (*noun class prefix of class 7*)

SElepe "axe"
SElo "object"
SEhlare "tree"

In this example the inexperienced user has to deal with different kinds of *se* in Northern Sotho which can vary in function from being concords and negation morphemes to being affixes. Text boxes are *directly linked to the lemma-sign* in each case. In moving the cursor over *se*[1] to *se*[4] different boxes open up, each containing vital semantic, morphological and syntactic information. In the case of *se*[1] the user is guided towards distinguishing between different moods of the verb in which this negative morpheme occurs. The information given for *se*[2] and *se*[3] are short and straightforward examples of the typical use of the subject and object concords respectively. In the case of *se*[4] a few examples of nouns in Class 7 consisting of this *se* plus a nominal stem are given. The only effort required of the user is to rest the cursor momentarily on the lemma-sign; he/ she does not even have to click the mouse!

Such dictionary articles can be short but multifunctional in that they simultaneously serve the decoding needs of the more experienced user, as well as the encoding needs of the less experienced user. For decoding purposes the skeleton entry might suffice and the user does not have to consider additional information. For the inexperienced user the text boxes offer encoding capabilities which simply cannot be offered within the physical limitations of the paper dictionary. One of the major advantages is that the gap between dictionary and grammar which is generally believed to be "unbridgeable" is starting to close.

> In this way, the *differences between dictionary and grammar begin to diminish*: the dictionary entries are linked to a grammatical description of the language that offers more detail than the grammatical compendium that is sometimes included with paper dictionaries. (Geeraerts 2000: 77)

Instead of, or in addition to, linking text boxes directly to the lemma-sign a *navigation bar*, marked with a special symbol, ∩, can be introduced as shown in Tables 8 and 10.

Table 8

leratô love, affection, passion; passionate; **~ng** in love

∩ **structure**; pronunciation; combination; frequency; concords; idioms; expressions; picture

Class 1 ***mo**nna*	Class 7 ***se**lepe*
Class 2 ***ba**nna*	Class 8 ***di**lepe*
Class 3 ***mo**swe*	Class 9 ***n**ku*
Class 4 ***meswe***	Class 10 ***di**nku*
Class 5 ***le**rato*	Class 14 ***bo**gobe*
Class 6 ***ma**sogana*	

In the case of nouns, the *noun class system* could be presented in an innovative way. In Table 8 the user looks up the word *lerato* and finds the translation equivalents "love", "affection", "passion", and so on. If the user now puts the cursor on ***structure*** in the navigation bar, a text box opens, not only reflecting the *total scope of the noun class system*, but also *putting the word itself within its appropriate position in the noun class system*, namely class five. In addition, a variety of information regarding *structure, pronunciation, typical combinations and collocations, frequency of use, concords, idioms, expressions* and so on, can be accessed from the navigation bar (see also Table 10).

All this is achieved by moving the mouse over different sections of the navigation bar. Thus, information boxes only appear if the user wants to see them. This innovation is impossible in the paper-dictionary dimension where the lexicographer has to decide whether to include or exclude information, and where information is normally excluded since space is severely limited and all presentations in a paper dictionary are at surface level — there is no opportunity to switch information categories on and off.

> COBUILD takes matters further, by *excluding all overt grammatical classification from entries, relegating it to a special margin along with synonyms and antonyms*, although a careful choice of definition formats permits the reader to deduce the word class of an entry. *What is true of word class and syntactic information also applies to the synonyms and antonyms just mentioned, to phonetics, to illustrative citations, in brief to all the potential elements of information that constitute a dictionary*. In a sense this is an attempt to *achieve the dream of the perfect dictionary, which should have in it one entry — just the word the user is seeking at the moment in question, with just the information needed and no other.* (Dodd 1989: 92)

In the case of verbs, we can recall the lemmatisation problems that resulted from the complicated derivation structure discussed in the previous paragraph. In Table 6, 18 different modules were distinguished for every verb (module one is repeated for the ease of reference):

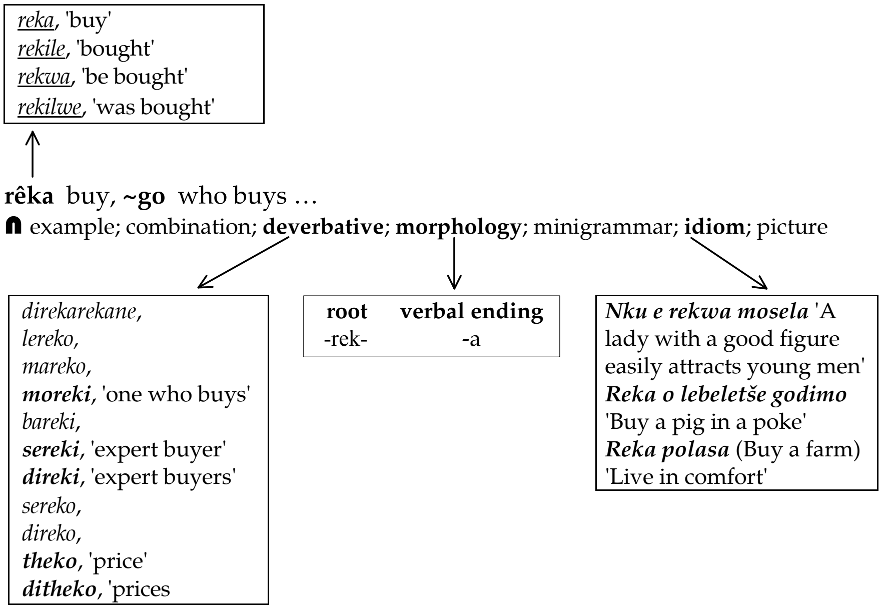
Table 9

1	root + standard modifications	*reka, rekile, rekwa, rekilwe*	*direkarekane, lereko, mareko, moreki, bareki, sereki, direki, sereko, direko, theko, ditheko*

A suggested treatment for module 1 can be found in Table 10.

Here the lexicographer could opt for linking the "standard modifications" (*perfect*, *passive* and *perfect* plus *passive*) to the lemma-sign. The user can also access all the nominal derivations of *reka* from the reference marker *deverbative* in the navigation bar. The value of this information to the user is threefold. Firstly, it gives him/her a quick summary of the different deverbatives with their basic meanings. Secondly, words with a high frequency of use are presented in boldface. Thirdly, boldface is also an implicit cross-reference to a reference address where *full treatment* of these frequently used derivations can be found. Depending on the sophistication of the software, the full treatment can be accessed directly, by simply clicking the mouse, in the same way that normal cross-referencing is done.

Table 10



<u>*reka*</u>, 'buy'
<u>*rekile*</u>, 'bought'
<u>*rekwa*</u>, 'be bought'
<u>*rekilwe*</u>, 'was bought'

rêka buy, **~go** who buys …
∩ example; combination; **deverbative**; **morphology**; minigrammar; **idiom**; picture

direkarekane,
lereko,
mareko,
moreki, 'one who buys'
bareki,
sereki, 'expert buyer'
direki, 'expert buyers'
sereko,
direko,
theko, 'price'
ditheko, 'prices

root	verbal ending
-rek-	-a

Nku e rekwa mosela 'A lady with a good figure easily attracts young men'
Reka o lebeletše godimo 'Buy a pig in a poke'
Reka polasa (Buy a farm) 'Live in comfort'

As far as grammatical information is concerned a distinction could be made between giving a basic morphological analysis and giving a cross-reference option to a full minigrammar. Constant reference to the minigrammar can be so crucial to some users that De Schryver and Kabuta (1998) repeat it as a footnote on every other page of their dictionary.

Table 11

Affix	**1pe**	**1pm**	**2pe**	**2pm**	**1**	**2**	**3**	**4**	**5**	**6**
NP	mu-	ba-	mu-	ba-	mu-	ba-	mu-	mi-	di-	ma-
SC	"N-	tu-	"u-	nu-	ù/"à-	bà-	"ù-	"ì-	dì-	"à-
OC	-N-	-tù-	-ku-	-nù-	-mu-	-bà-	-"ù-	-"ì-	-dì-	-"à-
PP	u-	\`bà-	u-	\`bà-	u-	\`bà-	\`ù-	\`ì-	\`dì-	\`à-

Affix	**7**	**8**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
NP	ci-	bi-	lu-	ka-	tu-	bu-	ku-	pa-	ku-	mu-
SC	cì-	bì-	lù-	kà-	tù-	bù-	kù-	pà-	kù-	mù-
OC	-cì-	-bì-	-lù-	-kà-	-tù-	-bù-	-kù-	-pà-	-kù-	-mù-
PP	\`cì-	\`bì-	\`lù-	\`kà-	\`tù-	\`bù-	\`kù-	\`pà-	\`kù-	\`mù-

From *idiom* in the navigation bar the user can activate typical idioms and expressions in which the verb being treated occurs. What has been done for module one can then be repeated for the other modules, producing in the end a well-structured but simple layered representation.

> Optimized microstructural representation is achieved through the *layered representation of entries.* In a dictionary of the size of the GVD [Grote van Dale], entries may be quite long, with a complex internal structure. In order to facilitate finding one's way through the entries, the entries may be accessed at different levels. There are four levels:
>
> — the headword level, containing information about spelling, pronunciation, hyphenation, grammatical and morphological characteristics, and etymology
>
> — the level of senses, i.e. the definitions constituting the semantic backbone of the dictionary
>
> — the level of nuances and phraseological units (collocations, idioms, proverbs etc.) that belong with a given sense
>
> — the quotations and examples that illustrate senses, or nuances, or phraseological entities.
>
> (Geeraerts 2000: 78)

The great virtue of the layout is that the article is a simple and fixed point of reference. The user does not have to leave the article to find more information.

Finally, an attempt will be made to solve the problem of identification of the lemma especially in conjunctively written languages, as discussed by Van Wyk (1995). An inexperienced learner of isiZulu could easily be confronted by one or more of the following derivations of the word *-hamba* "go":

> i**hamb**e, uku**hamb**a, kayi**hamb**i, aye**hamb**a, sebe**hamb**a, ngangili**hamb**a, nginga**hamb**a, **hamb**ani, eku**hamb**eni, ube**hamb**ele, ngizi**hamb**ela, owaye**hamb**ele, wam**hamb**isa, ayengasa**hamb**eli, zi**hamb**ayo, ngi**hamb**ile, kaba**hamb**anga

In most cases he/she will be unable to isolate the stem and will thus not be able to look up the word in a paper dictionary. However, in an electronic dictionary the whole paradigm for *-hamba* could be incorporated and linked to the lemma-sign. Depending on the sophistication of the software and the level of markup of the words, a variety of options is possible. In sophisticated applications the user can be automatically rerouted from the orthographic word to the lemma-sign and thus find a full treatment. This goal can even be achieved in a basic word processor by entering the full range of derivations as running text with manual or automatic pointers[4] to the lemma-sign.

> Within this area of new routes to information could easily be included a grammatical analyser that would be able to take any given morphologically marked word and find the corresponding headword, yielding, in addition, an "audit trail"

> that would demonstrate what the morphological markings that were detected indicated in terms of case, number, tense, or whatever. (Dodd 1989: 89)

> *Even the perennial problem of what should constitute a headword*, in particular with regard to the thorny problems of polysemy and homonymy, *ceases to be relevant*, as it no longer matters whether the set of letters "ROW", to give an example, demands one entry, two, or five, on the basis of its single spelling, two pronunciations, and five senses (as verb meaning "propel with oars", noun meaning "voyage in a boat propelled with oars", verb meaning "argue", noun meaning "argument" and noun meaning "file, rank, or series"). (Dodd 1989: 91)

Conclusion

The *multimedia dimension* opens new horizons for dictionary compilation especially for languages with complicated morphological structures such as the African languages. This enormous potential coupled with the power of *electronic corpora* should be maximally *exploited* to enhance traditional access routes for information retrieval and to develop new accessing strategies which are only possible within the electronic environment.

It is important that the newly established South African National Lexicography Units for the African languages should plan ahead in terms of electronic dictionaries. One could predict that the end product for each African language (a 20 to 30 volume dictionary) may never appear on paper but will only be accessible in electronic media because it is estimated that a 20 volume dictionary might cost R20 000 to R30 000 in the year 2010. Thus, if we speak about dictionaries for the African languages beyond 2000 we must maintain a clear perspective on the advantages of electronic dictionaries versus paper dictionaries, even at this relatively early stage of lexicographic planning.

Some argue that electronic sources of reference are playthings of the rich and out of reach of the poor. However, this situation is changing dramatically since the emphasis is no longer on ownership of expensive computer equipment, but on *access* to such equipment. Today even pupils in some primary schools in the poorest rural areas of South Africa have *access* to computers at school.

> The printed dictionary will not disappear overnight, nor perhaps ever, given the durability and relative inexpensiveness of books, but *the increasing advantages of a dictionary in a dynamic machine form will swing the balance in favour of words retrieved from a constantly changing database and displayed rather than words fixed for ever on paper.* (Dodd 1989: 87)

> The past is print dictionaries; the present is print dictionaries with some electronic versions of the same text; the future must be print dictionaries *and truly electronic dictionaries, compiled afresh for the new medium, enriched with new types of information the better to meet the needs of the multifarious users*. (Atkins 1996: 515)

Endnotes

1. Electronic dictionaries can be stored and accessed in a number of different ways. They can be built into a hand-held device, or inserted in a hand-held device via an 8cm CD-ROM or an IC (Integrated Circuit) card. Alternatively, they can be stored on a hard disk or a 12cm CD-ROM for use with a desktop computer. An electronic dictionary may be purchased and used by just one individual, but it may be possible for a dictionary on disk to be accessed from all the computers on a local area network, while a dictionary linked to a World Wide Web site can be accessed by users all around the world. (Nesi 1999: 55, 56)
2. Unless indicated otherwise, emphasis in all quotations is added.
3. Detailed discussions of access structures in paper dictionaries are given in Hausmann and Wiegand (1989: 337-339), Gouws (1996: 19-25) and Louw (1999 and 2000). In this article it will not be attempted to formulate access routes to the electronic dictionary in terms of such criteria since a different set of rules apply which cannot necessarily be described in existing theories.
4. Depending on the sophistication of the program, the user, on entering a search word or phrase, can either be prompted to *type in* the suggested lemma or, if sophisticated enough, the software can *directly display* the suggested entry in full.

References

Atkins, B.T.S. 1996. Bilingual Dictionaries: Past, Present and Future. Gellerstam, Martin et al. (Eds.). 1996. *Euralex '96 Proceedings: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515–546. Göteborg: Göteborg University.

Bolinger, D. 1990. Review of the *Oxford Advanced Learner's Dictionary of Current English. International Journal of Lexicography* 3(2): 133–145.

Bwenge, Charles. 1989. Lexicographical Treatment of Affixal Morphology: A Case Study of Four Swahili Dictionaries. James, G. (Ed.). *Lexicographers and their Works*: 5–17. Exeter: University of Exeter.

Concise Oxford Dictionary, Ninth Edition. 1996. Oxford: Oxford CD-ROM, Oxford University Press.

De Schryver, Gilles-Maurice. Forthcoming. Lexicographers' Dreams in the Electronic Dictionary Age.

De Schryver, Gilles-Maurice and Ngo S. Kabuta. 1998. *Beknopt woordenboek Cilubà – Nederlands en Kalombodi-mfùndilu kàà Cilubà (Spellingsgids Cilubà): Een op gebruiksfrequentie gebaseerd vertalend aanleerderslexicon met decodeerfunctie bestaande uit circa 3.000 strikt alfabetisch geordende lemma's en Mfùndilu wa myakù ìdì ìtàmba kumwèneka (De orthografie van de meest gangbare woorden).* Ghent: Recall.

Diciopédia, Grande Dicionário Enciclopédico Multimédia. 1997. Porto Codex: Priberam. Poerto Editora Multimedia.

Dodd, W.S. 1989. Lexicomputing and the Dictionary of the Future. James, G. (Ed.). *Lexicographers and their Works*: 83–93. Exeter: University of Exeter.

Doke, C.M. and B.J. Vilakazi. 1948. *Zulu–English Dictionary*. Johannesburg: Witwatersrand University Press.

Dorling Kindersley Children's Dictionary (The). 1996. London: Dorling Kindersley Multimedia.

Geeraerts, D. 2000. Adding Electronic Value: The Electronic Version of the *Grote van Dale*. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 75–84. Stuttgart: Stuttgart University.

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

Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14–31.

Harley, Andrew. 2000. Software Demonstration: Cambridge Dictionaries Online. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 85–88. Stuttgart: Stuttgart University.

Hausmann, F.J. and H.E. Wiegand. 1989. Component Parts and Structures of Monolingual Dictionaries. Hausmann, F.J. et al. (Eds.). 1989–1991: 328–360.

Hausmann, F.J. et al. (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexicographie. / Dictionaries. An International Encyclopedia of Lexicography. / Dictionnaires. Encyclopédie internationale de lexicographie*. Berlin: Walter de Gruyter.

Louw, Phillip. 2000. An Integrated Semasiological and Onomasiological Presentation of Semantic Information in General Monolingual Dictionaries as Proposed in H.E. Wiegand's *Semantics and Lexicography*. *Lexikos* 10: 119–137.

Louw, Phillip Adriaan. 1999. Access Structures in a Standard Translation Dictionary. *Lexikos* 9: 108–118.

Martin, W.J.R., B.P.F. Al, and P.J.G. van Sterkenburg. 1983. On the Processing of a Text Corpus from Textual Data to Lexicographical Information. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 77–87. London: Academic Press.

Meijs, W. 1992. Computers and Dictionaries. Butler C. (Ed.). *Computers and Written Texts*: 141–65. Oxford: Blackwell.

Nesi, Hilary. 1999. A User's Guide to Electronic Dictionaries for Language Learners. *International Journal of Lexicography* 12(1): 55–66.

Oxford English Dictionary, Second Edition (OED on CD-ROM). 1989. Oxford: Oxford University Press.

Oxford English Dictionary (OED Online). 2001. Oxford: Oxford University Press. http://www.oed.com.

Pharos Woordeboeke/Dictionaries 5 in 1. 2000. Johannesburg: Pharos and Logos Information Systems.

Prinsloo, D.J. 1994. Lemmatization of Verbs in Northern Sotho. *S.A. Journal of African Languages* 14(2): 93–102.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *S.A. Journal of African Languages* 19(4): 258–275.

Prinsloo, D.J. and Gilles-Maurice de Schryver (Eds.). 2000. *SeDiPro 1.0, First Parallel Dictionary Sepêdi–English*. Pretoria: University of Pretoria.

Prinsloo, D.J. and R.H. Gouws. 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *S.A. Journal of African Languages* 16(3): 100–7.

Rogers, M. 1996. Beyond the Dictionary: The Translator, the L2 Learner and the Computer. Anderman, G. and M. Rogers (Eds.). 1996. *Words, Words, Words: The Translator and the Language Learner*: 69–95. Clevedon: Multilingual Matters.

Sharpe, P. 1995. Electronic Dictionaries with Particular Reference to the Design of an Electronic Bilingual Dictionary for English-Speaking Learners of Japanese. *International Journal of Lexicography* 8(1): 39–54.

Tomaszczyk, J. 1983. On Bilingual Dictionaries: The Case for Bilingual Dictionaries for Foreign Language Learners. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 41–51. London: Academic Press.

Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82–96.

Walter, Elizabeth. 1996. Parallel Development of Monolingual and Bilingual Dictionaries for Learners of English. Gellerstam, Martin et al. (Eds.). 1996. *Euralex '96 Proceedings: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 635–41: Göteborg: Göteborg University.

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