



8<sup>th</sup> International Conference of the  
*African Association for Lexicography*

---

**AFRILEX 2003**  
*Bilingual Dictionaries*  
**Programme & Abstracts**

---

[To front and back cover of this booklet](#) (pdf 1.198KB)

Dates: 7-9 July 2003

Host: Department of Germanic & Romance Languages,  
University of Namibia, Windhoek, Namibia

Local Conference Organiser: Mr. Herman Beyer

Abstract Reviewers: Prof. Rufus H. Gouws, Prof. D.J. Prinsloo, Dr. Elsabé  
Taljard, Ms. Anneleen Van der Veken

Programme Committee: Mr. Herman Beyer, Mr. Gilles-Maurice de Schryver,  
Prof. D.J. Prinsloo

edited by  
Gilles-Maurice de Schryver  
Organiser: AFRILEX

---

Copyright © 2003 by the *African Association for Lexicography*  
ISBN 0-620-30795-1

Pretoria: (SF)<sup>2</sup> Press

[Cover](#) Screenshots by David Joffe: "From *TshwaneLex* to Online Dictionary"

([david.joffe@africanlanguages.com](mailto:david.joffe@africanlanguages.com) | <http://africanlanguages.com>)

[Cover](#) Artwork by Giovanni Plozner

([info@giovanniplozner.com](mailto:info@giovanniplozner.com) | <http://www.giovanniplozner.com>)

---

**A FEW WORDS FROM THE CHAIRPERSON**

Afrilex welcomes you to our 8<sup>th</sup> International Conference which also marks our 8<sup>th</sup> year of existence. We are proud to be a member of the international *-lex* family and to present you with this Conference Abstract Booklet, once again meticulously compiled and edited by Gilles-Maurice de Schryver.

I wish to thank you for attending the Conference and for your loyal support for our Association and lexicography in Africa.

Afrilex greetings

D.J. Prinsloo

## Table of Contents

### Programme

#### Keynote papers

- [Ulrich HEID](#) — The Handling of Collocations and Idiomatic Multiword Expressions: From Corpora to Dictionaries
- [Rufus H. GOUWS](#) — Outer Texts in Bilingual Dictionaries
- [Gwyneth FOX](#) — Corpus Research and Lexicography

#### Parallel sessions

- [Thierry AFANE OTSAGA](#) — Hybrid Dictionaries – The Future of Lexicography
- [Mariëtta ALBERTS](#) — Lexicography and Terminology Training at University Level
- [Herman L. BEYER](#) — Can We Quantify the Effects of Dictionary Use?
- [Emmanuel CHABATA](#) — Interviewer-Interviewee Interaction in Oral Interviews
- [Gilles-Maurice DE SCHRYVER](#) — Concurrent Over- and Under-treatment in Dictionaries — The *Woordeboek van die Afrikaanse Taal* as a case in point
- [James D. EMEJULU](#) — Revisiting Equivalence in Bilingual Lexicography
- [James D. EMEJULU, Yolande NZANG-BIE, Pierre ONDO-MEBIAME & D. Franck IDIATA](#) — Le rôle des dictionnaires bilingues dans le développement des langues Gabonaises: Le cas du fang
- [Rachéle GAUTON](#) — Bilingual Dictionaries, the Lexicographer and the Translator
- [Wilfrid H.G. HAACKE](#) — A Khoekhoegowab Dictionary in the Making: Some Lexicographic Considerations in Retrospect
- [Samukele HADEBE](#) — The Proposed Ndebele – Shona Dictionary: Prospects and Challenges
- [Kathy KAVANAGH](#) — English for New South African Bilingual Dictionaries
- [Langa KHUMALO](#) — From a General to an Advanced Ndebele Dictionary: An Outline

- [John M. LUBINDA](#) — The Incorporation and Handling of Metaphorical or Figurative Meaning in Bilingual Dictionaries
- [Matete MADIBA, Lorna MPHAHLELE & Matlakala KGANYAGO](#) — Capturing Cultural Glossaries. Case Study II: Medical Terms
- [Mandlenkosi MAPHOSA](#) — The Users' Perspectives on *Isichazamazwi SeSiNdebele*
- [Webster MAVHU](#) — Bilingual versus Monolingual: A Comparative Analysis of Two Trends in Shona Lexicography
- [Gift MHETA](#) — The Impact of Translation Activities on the Development of African Languages in Multilingual Societies: Shona – Ndebele – English Musical Terms Dictionary, a Case Study
- [Linkie MOHLALA, Gilles-Maurice DE SCHRYVER & Rachéle GAUTON](#) — The Lexicographic Treatment of the Feminine/Augmentative Suffix *-kazi* in isiZulu
- [Nomalanga MPOFU](#) — The ALRI Experience in the Compilation of a Dictionary of Biomedical Terms
- [Cornelias NCUBE](#) — Language Development or Language Corruption: A Case of Loanwords in *Isichazamazwi SeSiNdebele*
- [Salmina NONG & M.P. MOGODI](#) — The Lexicographic Treatment of the Demonstrative Copulative in Sesotho sa Leboa – An Exercise in Multiple Cross-referencing
- [Thapelo J. OTLOGETSWE](#) — Challenges to Representative and Balanced Corpora for African Lexicography
- [Annél OTTO & Nerina BOSMAN](#) — The User Perspective: Bible Reference Resources as Example
- [D.J. PRINSLOO](#) — The Lemmatisation of Adverbs in Northern Sotho
- [M.P. RAKGOKONG](#) — Are the Setswana Mockery Words that Objectionable?
- [Mariza STEYN & Liezl GOUWS](#) — *Woordeboek sonder Grense*: A Typological and Communicative Bridge
- [P.H. SWANEPOEL](#) — Dictionary Tailoring, SL Lexical Acquisition and Computer-Assisted Language Learning: The LINC Approach
- [Elsabé TALJARD](#) — On the Semi-automatic Extraction of Definitional Information: A Case Study for Northern Sotho
- [Dirk J. VAN SCHALKWYK](#) — Language Variation and the Lexicographer

## [Correspondence](#)

### Programme AFRILEX 2003

[To programme](#)

### Keynote papers

---

#### The Handling of Collocations and Idiomatic Multiword Expressions: From Corpora to Dictionaries

Ulrich HEID

*Institut für maschinelle Sprachverarbeitung – Computerlinguistik, Universität Stuttgart, Germany*

---

Corpus query tools, such as WordSmith Tools or Qwick (Birmingham University), come with a function to extract collocations of a given word from a corpus. As a result, they provide lists of word pairs, often together with a measure indicating how much the two elements belong together. Already years ago, a computational linguist told me in a discussion that, with these tool functions, the problem of collocations in corpus lexicography was solved. This talk is intended to show why this is not the case.

The abovementioned collocation tools are based on statistical association measures that determine statistically significant co-occurrences of words. Examples of such association measures include the t-test (Church & Hanks 1992), the log-likelihood ratio test (Dunning 1993), the Mutual Information measure, etc. They are all used to reorder lists of collocation candidates, possibly extracted beforehand by means of corpus query (e.g. for nouns and the verbs these nouns are objects of, as in “pay attention”, “ask a question”, etc.). Examples and a few well-known problems of the underlying statistics will be discussed; for example, Mutual Information unduly privileges low frequency words, and log-likelihood seems to be good in particular for the upper half of the frequency spectrum, however being quite dependent on frequency.

An analysis of some German and English data obtained in this way from corpora will show that the results of the statistical procedures, even though to some extent useful for lexicographic work, are far from homogeneous: they typically include a mixture of collocations and idiomatic word groups, as well as of trivial, lexicographically irrelevant, word combinations which may, for example, be artefacts of the corpus under analysis.

We thus need additional linguistic criteria to further classify the material, but also, more importantly, to discover additional morphosyntactic, syntactic and semantic properties of the word combinations identified so far only in terms of the lexemes involved. It is not sufficient to know that “pay” and “attention” go together, we must also know that “pay attention” has no article; or that “former” and “time” typically come as a plural expression, often with a preposition: “in former times”. These aspects contribute to the partial idiomatisation of collocations, and a learner of a foreign language must memorise them along with the collocation. For German and English noun+verb-combinations, an attempt will be made to provide a classified list of phenomena which need to be kept track of, beyond lexical co-occurrence, to make up for a detailed description of the respective multiword items. The claim we would like to make is that collocations and idiomatic multiword expressions must be lexicographically described in as much detail as any single-word lemma; this means that information about the components of the collocation, as well as about the collocation as a whole must be given with respect to morphosyntactic, syntactic (e.g. construction), semantic and pragmatic (e.g. style/register, frequency) properties. Furthermore, collocations tend to be combined, such that texts often include significant triples or quadruples of words (e.g. (pay+attention) + (careful+attention): pay careful attention). Along with the phenomena, a few suggestions for their corpus-based acquisition will be made (Heid & Zinsmeiser 2003).

In the third part, the question of the lexicographic data presentation will be discussed. Beyond the question of where to lemmatise collocations and idiomatic multiword groups, the detailed phenomena discussed above make the writing of an article somewhat more difficult, as they need to be kept track of. We look at this problem with bilingual (active) dictionaries in mind, printed as well as electronic. Inspiration for the article layout may come from experimental dictionaries such as Mel’cuk’s Explanatory Combinatorial Dictionaries, but also

from printed dictionaries for general users, such as the Van Dale series of bilingual dictionaries in the Netherlands. Sample entries in different “styles” will be briefly discussed.

[To Table of Contents](#)

---

## Outer Texts in Bilingual Dictionaries

**Rufus H. GOUWS**

*Department of Afrikaans and Dutch, University of Stellenbosch, South Africa*

---

Metalexigraphic research of the recent years has been characterised by a growing interest in and focus on various aspects regarding the structure of dictionaries. In this regard both the mutual features and dictionary-specific features have received attention. Dictionary research no longer only includes attempts to describe and analyse the contents of dictionaries and the different data types on offer, the different structural components of dictionaries also fall within the scope of this field of research. As a carrier of text types a dictionary is not only regarded as a source of information displaying a variety of data types in the central list. A new emphasis deviates the attention from a central list bias towards a more inclusive frame structure approach. This approach works with the assumption that the central list is complemented by front and back matter texts, constituting the outer texts of a dictionary.

Utilising the frame structure approach this paper focuses on the use of outer texts in bilingual dictionaries. The distinction between integrated and unintegrated outer texts is maintained and both these text types, their purpose and the role they play in devising the data distribution structure of a dictionary are examined. In using integrated outer texts it is shown that the data distribution does not have to focus exclusively on the default article in the central list although article stretches still accommodate the most typical data categories directed at the lemmata as guiding elements of articles and primary treatment units. It is shown how an interactive relation between the integrated outer texts and the central list can achieve an optimal realisation of the genuine purpose of a bilingual dictionary and can enhance the quality of dictionary consultation procedures.

As examples of unintegrated outer texts the use of alphabetically ordered equivalent registers, the listing of items representing the lemmata included in complex and synopsis articles as well as additional pedagogical data will be discussed. It is also shown how back matter texts can add a typological hybrid character to a dictionary by using alternative ordering systems, e.g. a thematic ordering as opposed to the alphabetical ordering of the central list. The way in which outer texts can ensure that a dictionary has a poly-accessible character that meets the needs of a user-driven project is also discussed. Looking at the user and usage situation the role of dictionary functions in the planning of the outer texts may never be underestimated and various aspects of the theory of lexicographic functions come to the fore in the discussion.

The successful use of outer texts demands a new look at the data distribution structure of bilingual dictionaries. Emphasis is yet again placed on the importance that each dictionary project should include a well-devised dictionary plan.

In this paper a dictionary is seen as a comprehensive container of knowledge and suggestions are made to improve the quality of the access structure to ensure an optimal retrieval of information by the intended target user.

[To Table of Contents](#)

---

## Corpus Research and Lexicography

Gwyneth FOX

*Macmillan Education: Publisher, Dictionaries*

---

Work with corpora over the past 20 years has shown us a great deal about how we use English. In particular, there have been many revelations about the ways in which vocabulary patterns are surprisingly predictable, and these findings are now being reflected in learners' dictionaries. This means that such dictionaries are probably the best record we have of the way in which English is now being used. Many examples will be given to justify this statement. But there is no reason why corpus research should not influence bilingual lexicography more than it presently seems to.

People are fascinated by language. And researchers have been studying it for centuries. But it is only in the past twenty years or so that we can be sure that the statements we make about the language are accurate. That is because the advent of computers has allowed us to build corpora, as large or as small as are appropriate for our particular needs, and analyse them for frequency, grammar, vocabulary, pragmatics, discourse functions, and so on. Perhaps the two areas where we have learned most are those of frequency and vocabulary.

Although we always knew that some words were more frequent than others, we now know which words these are, and how often they are used and in what contexts. This must be important information for learners of a language: they need to know which words are worth expending effort on!

We also realise that it is not enough just to look at words, however frequent they might be, in isolation. Collocation and colligation patterns stand out in the data, and force us to reassess the way in which we describe words, both in the classroom and in dictionaries. Collocation patterns range from the relatively fixed and difficult to decode, as in idioms and proverbs, through binomials and trinomials, through chunking, right down to those that are weak and perhaps not worth mentioning. The same is true of colligation. The phraseology of the language is much less random, much more predictable than we ever imagined.

Another vocabulary 'discovery' is that of semantic prosody. Why is it that some words have attracted to them other words, either positive or negative, so that it is almost impossible to use them in any other way? Some of these words are obvious, others much less so. How could a learner know about their prosody if it were not pointed out to them?

Corpus findings are now well known, and are expressed at their best in the new breed of learners' dictionaries produced in the UK in the past fifteen or so years. This makes these dictionaries the best, most up-to-date, most accurate record of English as it is presently being used. Some bilingual dictionaries are now being compiled with the benefit of two, often parallel, corpora; but it seems to me that they are not yet as good (or as helpful) descriptions of the language as you find in monolingual learners' dictionaries.

[To Table of Contents](#)

**Parallel sessions**

---

---

## Hybrid Dictionaries – The Future of Lexicography

Thierry AFANE OTSAGA

*Department of Afrikaans and Dutch, Stellenbosch, South Africa*

---

Dictionaries have been compiled for several thousand years. Their need arose when it became more difficult to read and understand religious texts. Therefore, dictionaries were invented in order to assist in the understanding of these texts that were actually written in a language that was no longer understood by the interested people. Nowadays, dictionaries are still produced because certain human linguistic and knowledge needs are observed in society and they are compiled to satisfy these needs. This basic characteristic is the main purpose of dictionaries.

In order to always satisfy user needs, lexicographers have been trying to compile different types of dictionaries, according to different aspects: the users' language competences, users' general culture and knowledge, users' respective field subjects, users' translation needs, etc. In general, they have to take into account the objectives of users when these users are using dictionaries. In that regard, various types of dictionaries have been compiled to be used by a specific target user group. Indeed, some dictionaries are directed at the extra-linguistic features of the items treated (encyclopaedic dictionaries), while other dictionaries focus on the linguistic and pragmatic aspects (linguistic dictionaries). Some dictionaries focus on the origin, history and development of the treated language (diachronic dictionaries), while still others focus on the lexicon of a language at a specific time in its development (synchronic dictionaries). In the category of linguistic dictionaries, monolingual dictionaries can aim at a scholar approach (school dictionaries), a learning approach (learners' dictionaries), a normative approach (standard dictionaries), or a comprehensive approach (comprehensive dictionaries). Conversely, bilingual or multilingual dictionaries can be compiled for a polyfunctional purpose (polyfunctional dictionaries), they can also be monoscopal or biscopal. All these various types of dictionaries were directed by the necessity to satisfy users' needs.

The main objective of lexicographical works is to satisfy the needs of the users. When dealing with the methodology and even with the planning of a dictionary, one must first define the target user; otherwise the compilation will not be efficient. However, in every lexicographical work the main interest is on the dictionary user. In modern lexicography, the role and the place of the user is more and more taken into account. The users are a great lobby and the publishing houses know it so well: even if a dictionary is compiled within a good methodology, if a user does not find the information he/she needs, this dictionary will not be sold or used. Thus, the user appears to be the focal point on which each element of the lexicographical process focuses. Because user needs are increasing and because most people want knowledge regarding different aspects of life, it is becoming increasingly difficult to satisfy user needs in one specific type of dictionary. At the same time, users do not want to spend more time and money by buying different dictionaries according to what they are looking for. The ideal solution for them could be to find most information they need in one single dictionary. On the other hand, it is important to specify that it is not possible to satisfy all the user needs in one dictionary, even in a multi-volume dictionary. Yet the lexicographer must try to come as close as possible to satisfying user needs. For that reason, the only solution could be the compilation of hybrid dictionaries. In fact, in modern-day lexicography hybrid dictionaries will be the solution of the future that will allow lexicographers to give to the users what there are looking for in a dictionary. In that regard, some dictionaries will not have one specific purpose, but could include two, three, four, and even five functions. A bilingual dictionary for instance will



not only give translation equivalents of lemmas, it will also give paraphrases of meaning in order to allow the users to utilise the same dictionary to solve not only their problem of translation, but also to be able to improve their knowledge in the same language. The main purpose of this paper is to show that as a result of new and increasing user needs, the best way for future lexicography will be the compilation of hybrid dictionaries. Dictionaries focusing on one unique and specific aspect will no longer satisfy a public who needs to have knowledge about various aspects and domains.

[To Table of Contents](#)

---

## **Lexicography and Terminology Training at University Level**

**Mariëtta ALBERTS**

*Manager: Lexicography and Terminology Development, PanSALB, South Africa*

---

The multilingual dispensation creates job opportunities for language practitioners. These language practitioners need training in various aspects regarding the language practice since lexicography, terminography, translation and editing (to name but a few) are practices that need highly skilled and knowledgeable practitioners.

Several of the focus areas of the *Pan South African Language Board* (PanSALB) concentrate to a certain extent on language development, such as terminology development, lexicography or aspects like translation and interpreting services. PanSALB is aware that all these language practices need skilled and highly trained personnel.

The Lexicography and Terminology Development (L&TD) focus area deals with the eleven National Lexicography Units (NLUs) and one national terminology office. The eleven national lexicography units were established and each is situated at a tertiary institution in the geolinguistic area where most of the mother-tongue speakers of the specific language are found. Unfortunately, there are only a few trained lexicographers available to work at these units. The only national terminology office in the country, the Terminology Coordination Section (TCS) is part of the National Language Service (NLS), Department of Arts and Culture (DAC). The terminologists receive in-house training on terminological and terminographical principles and practice. It is of the utmost importance to train language practitioners and students to be able to compile general as well as technical dictionaries for communication purposes.

The value of lexicography and terminology training cannot be stressed enough. The need might even be greater in South Africa than in other countries given the multilingual clause in the Constitution that provides for eleven official South African languages. Multilingual general as well as technical dictionaries are needed for proper communication between linguistic communities. Presently there are very few trained lexicographers and terminologists, especially in the African languages. Language practitioners, who are going to work on lexicographical or terminographical projects in future, need training as soon as possible.

This paper addresses the current situation regarding lexicography and terminology training. Suggestions are made regarding the utilisation of Schools for Languages as training venues for lexicography and terminology courses. The benefits for the Schools of Languages are spelled out. The value to other departments and faculties at the given university, the benefit to other students at other universities in the country and worldwide and to language offices or language



units receives attention. The process as described would train students in the theory, principles and practice of lexicography and terminology. It would be to the advantage of the NLU's as well as the TCS and the to be established language units to appoint trained personnel rather than to devote time on in-house training. Production of general dictionaries as well as various technical dictionaries would show progress.

The various tertiary institutions such as the universities and technikons would benefit because they would train students and there would be positive and worthwhile outcomes.

The Human Language Technology virtual network would benefit by receiving multilingual general words and multilingual, polythematic terms into its database for dissemination to linguistic communities.

The language community would benefit since they would have words and terms available for better communication. Minority languages would be developed to become functional languages in the higher echelons of science and technology. Finally, the South African languages would be available as functional world languages on the Internet.

[To Table of Contents](#)

---

## Can We Quantify the Effects of Dictionary Use?

**Herman L. BEYER**

*Department of Germanic & Romance Languages, University of Namibia, Windhoek, Namibia*

---

This paper aims to give an overview of the empiric research into the possibility of quantifying the effects of dictionary use among school learners, which has been conducted as a pilot study at the University of Namibia. The initial processes and results are explained, which provides insight into how the project may be amended to continue meaningfully.

The first instances of data captured in this project took place in 1997 while the researcher was a language teacher in Swakopmund, employed by the Ministry of Basic Education and Culture of Namibia. The working hypothesis was to determine whether the use of dictionaries by school learners would result in improved linguistic performance. One linguistic skill, that of spelling, was chosen for the experiment. The respondents comprised of two classes of Grade 11 learners who took Afrikaans as a first language. One class group was labeled the test group, the other the control group. Both groups were given a series of four unannounced spelling tests, the intervals ranging from three days to as much as two months. Each test consisted of the same 25 items, chosen on the basis of the potential spelling difficulties they might pose for learners. The learners were not informed that the test would be repeated. They were, however, on each occasion advised that the tests did not contribute to their continuous assessment mark and were not designed to measure any aspect of intelligence. By doing this, it was hoped that conditions resembling as closely as possible to normal class conditions could be created.

The first spelling test was written by both groups under similar conditions: normal test conditions without the benefit of a dictionary.

During the second test each member of the test was provided with a dictionary on his/her desk. The respondents were given the freedom to look up any item in the dictionary to make sure of its spelling, provided that they would indicate dictionary use. This would enable the researcher to identify those items that a particular respondent chose to look up. The control group wrote the second test under conditions identical to those during the first test, i.e. without

the benefit of a dictionary. Unlike the test group, however, the control group members were given immediate feedback on their tests by having them marked after exchanging the scripts among the respondent (i.e. a respondent would not mark his/her own test). Respondents were instructed to clearly indicate mistakes on their fellow respondents' scripts and to write down the correct form in full each time. After the respondents received their tests back, they were given about 30 seconds to take a look at the results, including the corrections made by their fellow respondents. The test group was given no feedback of any nature on their tests.

The third and fourth tests were conducted under the same conditions as the first, i.e. normal test conditions without the benefit of a dictionary.

This experimental procedure provided the researcher with extensive data, from which it is hoped the following questions could be approached with quantitative support:

- Does a respondent who looks up a word for spelling purposes remember its correct spelling later? If yes, for how long? If no, are any consistencies identifiable that may allow us insight into the reasons for the perceived failure to learn and perhaps into spelling rehabilitation?
- Is a respondent who looks up a word for spelling purposes more likely to remember its correct spelling than a respondent who does not utilise a dictionary but who is provided with rehabilitative feedback in the 'traditional' way? If yes, what is the role of the dictionary in this case? If no, why has learning seemingly not taken place?

The above questions underlie the basic research question that this project aims to address: Does dictionary use result in quantifiable improved linguistic performance?

[To Table of Contents](#)

---

## **Interviewer-Interviewee Interaction in Oral Interviews**

**Emmanuel CHABATA**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

The intended presentation will be an analysis of language used by an interviewer and that of the interviewee during an oral interview. It will focus on the language of penetration by the interviewer, that is, the language somebody usually uses when he/she approaches a person for an interview in search of specific information. It will also look at the respondent's language when he/she responds to different types of questions as well as that used by the people concerned in their subsequent conversation. The presentation will also look at the factors that may shape the respondent's answers as well as the interviewer's follow-up questions. It will furthermore look at the element of 'misfiring' by either of the parties and its consequences.

The intended presentation will focus on the strategies that an interviewer may use when he/she tries to get information from a respondent. In doing this, the presenter will be guided by the principle that each interview and each interview setting is different and needs different skills and also that each situation involves expectations and assumptions. He/She will also be guided by the assumption that whenever the sender of information, in this case the interviewer, sends a question, he/she hopes to be understood by the receiver/interviewee. However, the message may or may not go through. To see whether it has gone through or not, one has to assess the feedback that the sender gets. The presenter will also look at the interviewer's challenges, some

of which will include respondent's attitude towards interviewer or the subject under discussion, the environment of the interview, misfiring by the respondent as well as lack of knowledge by the interviewee.

The presentation will also focus on what an interviewer needs to do before he/she gets out to conduct an interview. For example, the interviewer has to be thoroughly prepared. Being prepared means that one has to formulate one's questions before starting an interview. One has to come up with questions that can incite the respondent to say what he/she knows about the subject under discussion. For example, the questions have to be structured in a way that is most effective and friendly. Preparedness also entails getting the right person to interview. Depending on the purpose or subject of the interview, the interviewer has to get somebody who can supply the desired information. Besides knowing the subject, the person has to be willing to spare time for answering questions. This is an important dimension, especially given the fact that most people are usually busy. Thus, one may expect to obtain better results if one interviews a person who is prepared to give out information. The presenter will look at the common strategies that interviewers usually use to cultivate interest in the respondent.

The presenter will also devote some time to the qualifications one should possess as a good interviewer. For one to be effective in getting information, one has to have the skill to ask questions. The assumption to be adopted here is that a skilled interviewer is better than one who is not. But this assumption also triggers a few questions. For example, how does one become skilled? Is it through training or not? How does personal character determine the end result?

In trying to understand exactly what goes on between an interviewer and an interviewee, an analysis of their respective body languages will be part of the investigation. In this case, the assumption to be adopted is that verbal communication should match what is implied by body language. The assumption is based on the fact that verbal and non-verbal messages are intertwined, with the non-verbal symbols usually complementing the verbal ones. However, the analyses to be made will not be blind to the fact that sometimes non-verbal symbols may substitute verbal ones and also that non-verbal symbols may be inconsistent with verbal ones. Body language is considered important in oral interviews because it has a direct impact on what either of the persons involved will say after observing the gesture(s).

The intended presentation was inspired by the writer's experiences as an oral interviewer during data collection for the Shona linguistic corpus. As a result of this, some of the illustrative examples to be used in the presentation will be drawn from the oral Shona corpus, that is, from audiocassettes that were recorded during the mentioned exercise. Other examples will come from general observation, as well as from analyses of what one usually sees in television interviews.

[To Table of Contents](#)

---

**Concurrent Over- and Under-treatment in Dictionaries — The *Woordeboek van die Afrikaanse Taal* as a case in point**

**Gilles-Maurice DE SCHRYVER**

*Department of African Languages and Cultures, Ghent University, Ghent, Belgium*  
*Department of African Languages, University of Pretoria, Pretoria, South Africa*

---

In Prinsloo & De Schryver (2002) a so-called multidimensional lexicographic Ruler was introduced. With this powerful instrument measurements and predictions can be made on various macro- and microstructural dictionary levels. Three levels received thorough treatment so far, viz. considerations regarding the relative size of each alphabetical stretch, the corresponding number of lemma signs, as well as compilation-time aspects. In this paper the interplay between these levels is studied with a focus on ‘moving’ average article length, and the correlated aspects of inclusion versus omission of lemma signs.

In its most basic form, a Ruler is simply an instrument to guide the relative alphabetical breakdown in semasiological dictionaries. As such, each alphabetical category is assigned a certain percentage, reflecting the relative size of that category. Different languages, and even different types of dictionaries for a specific language, have different Rulers. The Rulers themselves are built from statistics derived from electronic corpora, as well as from existing dictionary data. Just as physical rulers with which one measures, they can be made as fine-grained as one wishes, by simply breaking down the alphabetical categories further into smaller sections. Just as the human rulers who govern us, a multidimensional lexicographic Ruler can be called in to manage a project. To date, general-language Rulers for isiNdebele (De Schryver 2002), Afrikaans (Prinsloo & De Schryver 2003a), and Sesotho sa Leboa (Prinsloo & De Schryver 2003b), as well as for Tshivenda, Xitsonga, Setswana and Sesotho have been designed.

During the presentation it will be indicated that the very same Ruler for a specific language can now also be used with regard to *average article length*. The value of this new dimension can be successfully illustrated when one analyses the huge multi-volume overall-descriptive *Woordeboek van die Afrikaanse Taal* (WAT), in compilation for the past three-quarter century and published up to the letter **O** (volume XI). Comparing WAT with a so-called ‘Afikaans AO-Ruler’ immediately reveals extreme inconsistencies with regard to average article length. For the letters **A** and **B**, for instance, it is clear that both number of pages and number of lemma signs are heavily under-treated in WAT. The under-treatment in terms of space allocation, however, is much more severe, which results in a very low average article length. Up to the letter **J**, the relative allocation to space is always smaller than the relative allocation to the number of lemma signs. From **K** onwards, a sudden reversal in this pattern occurs, and this remains so up to **O**. Throughout **K**, both space allocation and number of lemma signs are *extremely* heavily over-treated compared to the AO-Ruler. It should not come as a surprise that, after having spent *almost 30 years* on the compilation of **K**, the editors at WAT decided to drastically reconsider their compilation strategies, and entered a ‘new’ era (cf. Botha 1994: 423). Page-wise the compilers indeed moved closer to the AO-Ruler, with **L** and **M** slightly above and **N** and **O** under the AO-Ruler. As far as the number of lemma signs is concerned, however, these have been consistently under-treated, with **O** an all-time low.

Although everyone will agree that the compilation of **K** was unfortunate for WAT, a new negative trend might have started with the completion of **L** and then **M**, where one observes a *concurrent over- and under-treatment*, in terms of space allocation and number of articles respectively. One should guard against the temptation to move ever faster through the alphabet, as seems to be the case in the last volume, where space allocation is now also under-treated, the number of articles even more so, yet where this is masked by an ever-increasing average article length.

In order to substantiate the latter claim, an in-depth comparison between WAT and the desktop *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT) will be presented for the

category **O**. Given that the *entire* HAT is smaller than the *single category O* in WAT, it is logical to assume that every single lemma sign in HAT should in principle also be entered in WAT. Upon comparison, however, one has to conclude that as many as 499 o-initial lemma signs from HAT have not been lemmatised in WAT. Just one of these 499 has been treated as a sub-lemma in WAT, 40 can only be found as untreated sub-lemmas and 175 as untreated run-ons, while 211 have not been lemmatised and do not occur anywhere in the WAT text either. The remaining 72 have not been lemmatised in WAT – either as lemmas, as sub-lemmas or as run-ons – despite the fact that those very same items *are used throughout* the WAT text itself. Especially problematic are those missing items that are not only highly frequent in a 10-million-word Afrikaans corpus, but are moreover *cross-referred* to from other items in WAT. Numerous examples of such cases will be discussed.

Monitoring the compilation of especially a multi-volume dictionary project with an average article length Ruler is crucial if one is to avoid such major inconsistencies. A concurrent over- and under-treatment in terms of space allocation and number of articles respectively, must alert compilers of an overall-descriptive dictionary that they are starting to miss out on too many lemma signs.

[To Table of Contents](#)

---

## Revisiting Equivalence in Bilingual Lexicography

James D. EMEJULU

*Groupe de Recherche en Langues et Cultures Orales, Université Omar Bongo, Gabon*

---

The problem of equivalence is based in the fact about which there exists interdisciplinary consensus: the lexical-semantic structures of the lexicon of a particular language are language-specific and therefore partly unique. That implies that the lexical-semantic structures of two (or more) languages are not isomorphous. — Wiegand (2002)

The postulate of non-isomorphism of languages is an underlying factor governing translations and translation dictionaries, and poses the crucial question of equivalence. Even with kindred languages that have in common some linguistic and anthropological affinities, the problem of equivalence always thwarts meanings. This is all the more subtle in quality when one is dealing with languages of different linguistic families and/or of variant cultural levels. Several linguistic and lexicographic theories do show serious differences for the very concept of equivalence. Translation theories are not all unanimous on the theoretical perception and even not on the practical treatment of the concept of equivalence. Regarding (meta)lexicography, Wiegand (2002) criticised some salient conceptual discrepancies that he judiciously called *grave differences of opinion that have led to a whole range of misjudgements about the features of equivalent relationships in bilingual lexicography*. Arguing from the Saussurian distinction, language system vs. parole system, he suggested some conceptual changes that reserved the term *correspondence* for the language system and *equivalence* for the parole system. Though subtle, these suggestions do little to clarify these *misjudgements*.

If one can postulate that *equivalence* in languages is all about conveying meaning in a language-to-language communication situation, and that bilingual lexicography (or to be more precise, translation-oriented lexicography) is by definition called upon to provide compatible



referential interfaces and tools to make meanings work, then language should be perceived in its totality when one is retaining and treating *equivalence* in lexicography and dictionary research. Language *per se* is a communal conventional construct, a theoretical representation of *the real*, hence the Saussurian bicameral *signifier vs. signified* perception. Another postulate is that language is dynamic and within time and space no language is absolutely and reflexively homogenous. This fact undermines the symmetrical relations between the signifier and the signified across dialects. The objective consequence of this is that the non-reflexivity and non-symmetry thus observed do handicap effective translation into another language. Translation *per se* is transitivity. The question of equivalence is therefore posed on two levels: the language system level and the reality system that are always dissymmetrical among languages. Here, real conceptual clarity of the theoretical constructs of *equivalence* and their logical adequacy programming are needed. In a computer environment, *equivalence* as a theoretical translanguing matrix should support a logical array of compatible data that can generate required sets of meanings across the languages present.

Gabon is a multilingual systemic maze with over 40 Gabonese heritage languages (GHLs) of the Niger-Congo family. The predominance is Bantu according to Guthrie's definition. Paradoxically, the school and development environments are officially monolingual, based on French that is of the Latin and Indo-European stock. The sociolinguistic dynamics of this situation are all the more complicated in favour of French as a result of the political, social and economic set-ups that exclude the official use of any of the GHLs. The official status of French makes it a compulsory and privileged mediation in all social and official communications on the macro-level. It is insidiously permeating the micro-level and gradually eroding interpersonal communications. That is to say that all communicational exchanges, be it economical, political, judicial and the like, must go through French. One cannot overemphasize the excruciating problems that this poses to the less lettered. Hence, the problems of the cognitive development of the Gabonese child are keenly associated with the systemic linguistic imbalance between two language families with unequal social status. However, diverse sociolinguistic patterns have been identified at the micro-level. The urban tendency is towards French monolingualism. The rural areas present divergent patterns of GHL monolingualism, inter-GHL bilingualism and/or GHL-French bilingualism. Multilingual cases of various GHLs and GHLs-French combinations have also been attested. All these combinations, that are theoretically sociolectal, pose some tough cognitive and *equivalence* questions when it comes to translating French texts into GHLs for use as instruments of knowledge and crafts acquisition.

This paper is a modest contribution to the ongoing debate on the concept of equivalence and its applications in bilingual (meta)lexicography. Its approach is rather concerned with experience derived from Gabonese bilingual lexicography where the source language (SL) is Indo-European and the target language (TL) is of Greenberg's Niger-Congo. Conclusions from this study are expected to shed new light on the debate.

[To Table of Contents](#)

---

**Le rôle des dictionnaires bilingues dans le développement des langues Gabonaises: Le cas du fang**

Pour rendre compte du rôle du (des) dictionnaire(s) bilingue(s) dans le développement du parler fang du Gabon, notre propos commencera par montrer en quoi, sur le plan général, les dictionnaires participent-ils au développement des langues. Outil didactique de référence, le dictionnaire est l'ouvrage par excellence du transfert des connaissances. Il peut être considéré comme le patrimoine collectif, culturel, technologique, social, de la communauté à laquelle il est destiné. Il est perçu comme étant la norme à laquelle la communauté linguistique doit, en toutes circonstances, pouvoir se référer.

A ce jour, constatons-nous, de nombreux peuples ont réalisé l'importance de l'outil dictionnaire. Ce n'est malheureusement pas le cas des peuples africains dont les langues manquent, la plupart du temps, d'être accompagnées par cet outil. Au Gabon, en l'occurrence, l'élaboration des dictionnaires se trouve encore au stade embryonnaire ; l'activité, pensons-nous, doit être assortie d'une culture de cet outil, si l'on veut percevoir son importance et partant, garantir le développement des langues. C'est cette vision que nous voulons porter sur le parler fang.

Nous présenterons ensuite le parler et les populations fang, ainsi que les espaces dans lesquels on les rencontre. Nous indiquerons en effet que le fang est parlé dans un espace constitué par :

- Une partie de l'extrême sud du Cameroun ;
- La moitié est de la Guinée Equatoriale ;
- Une portion de la partie nord-ouest du Congo Brazzaville ;
- La moitié Nord du Gabon.

Nous poursuivrons avec l'inventaire et la présentation succincte des différentes propositions lexicographiques qui ont été faites sur le fang du Gabon. C'est suite à cela que nous aborderons le rôle des dictionnaires dans le développement du fang. Nous y montrerons d'abord l'apport des textes existants en indiquant ce à quoi ils ont servi, s'ils ont été consultés ou exploités, et s'ils le sont encore à ce jour. Nous exposerons ensuite les perspectives d'avenir, en montrant que le développement actuel de la science linguistique au Gabon peut aider à améliorer la qualité des propositions anciennes, et nous suggérerons les types de dictionnaires fang que l'on peut élaborer à court, moyen et long terme, pour promouvoir le développement de ce parler.

Nous terminerons par la proposition d'éléments portant sur la relation entre la vitalité du fang et le dictionnaire, pour montrer qu'il faut préalablement que le fang soit utile et utilisé, pour que le dictionnaire participe à son développement. En d'autres termes, l'importance du dictionnaire dans le développement du fang, réside dans l'importance que l'on pourrait lui accorder en tant que médium de communication. L'adéquation vitalité linguistique et dictionnaire assurerait d'une part, un développement durable du fang, et d'autre part, sa meilleure intégration dans le système éducatif. Ce développement peut, de la sorte, soutenir sa standardisation, de même que les programmes d'alphabétisation.

Nous montrerons en définitive, que le dictionnaire se présente comme un atout majeur pour la conservation et la promotion du fang parmi ses différentes couches socioculturelles.



---

## Bilingual Dictionaries, the Lexicographer and the Translator

**Rachéle GAUTON**

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

---

This paper focuses on the problems, advantages and disadvantages of the bilingual dictionary from both the lexicographer and the translator's point of view, with specific reference to bilingual Zulu dictionaries.

Clearly the fundamental problem regarding the bilingual dictionary from both the lexicographer and translator's points of view is the basic lack of equivalence or anisomorphism between languages.

According to Nida (Al-Kasimi 1983: 58), for the lexicographer, the semantic problems involved in bilingual dictionaries are different from, and also more complicated than, those encountered in the compilation of monolingual dictionaries. The reason for this is that whereas monolingual dictionaries are prepared for users who participate in and understand the culture being described, bilingual dictionaries describe a culture that differs in various proportions from that of the users'.

The non-equivalence between languages is also the root cause of the difficulties that the translator or user of the bilingual dictionary has to contend with. The problems experienced by translators, therefore, overlap to a great extent with those problems that the lexicographer experiences in compiling a bilingual dictionary.

For the translator, the bilingual dictionary could be a dangerous tool. It is therefore imperative that the user should be aware of what a bilingual dictionary is, and what it is not.

Manning (1990: 159) indicates that the bilingual dictionary is the translator's basic tool, and that it is the bridge that makes interlingual transfer possible. Pinchuck (1977: 223) warns, however, that the bilingual dictionary is an instrument that has to be used with caution and discernment. Pinchuck (1977: 231) further cautions:

The bilingual dictionary has a particular importance for the translator, but it is also a very dangerous tool. In general when a translator needs to resort to a dictionary to find an equivalent he will do better to consult a good monolingual dictionary in the SL and, if necessary, one in the TL as well. The bilingual dictionary appears to be a short cut and to save time, but only a perfect bilingual dictionary can really do this, and no bilingual dictionary is perfect.

Swanepoel (1989: 202–203) agrees that it is a misconception to assume that the general bilingual dictionary is sufficiently sophisticated to be an ideal translator's aid for the professional translator. It is merely a useful, albeit a limited, aid.

This paper will show that there are clear criteria that the lexicographer can follow in compiling the bilingual dictionary, which would then enable the user to disambiguate the recorded information with great(er) success.

[To Table of Contents](#)

---

# A Khoekhoegowab Dictionary in the Making: Some Lexicographic Considerations in Retrospect

Wilfrid H.G. HAACKE

*Department of African Languages, University of Namibia, Windhoek, Namibia*

---

The present paper deals with certain lexicographic issues that had to be addressed in deciding on the editorial policy for the compilation of *A Khoekhoegowab Dictionary* with an English – Khoekhoegowab index, which appeared in December 2002. Khoekhoegowab is the revived name for the language formerly known as i.a. Nama/Damara.

The dictionary project had a dual aim: firstly practical, to provide a comprehensive bilingual dictionary for general usage; secondly academic, to record the lexicon of this last surviving language of the Khoekhoe branch of Central Khoesaaan for comparative and other linguistic purposes. Hence certain compromises had to be made in an attempt to meet the widely diverging demands of the target users. Aims of corpus planning to counteract the further recession of an endangered language on the one hand, and scholarly interests in documenting this declining language in as much lexical and tonological detail as possible on the other hand, may require conflicting strategies. While the compilers strived to be descriptive by only documenting lexicon that was actually encountered, without attempting to fill lacunae by coining equivalents for English concepts, the dictionary is prescriptive with regard to orthography, as it employs the officially approved orthography (with some systematic adaptations in order to accommodate tonal diacritics). (Near-)obsolete catchwords are included – occasionally with cultural elaboration – for the dual purpose of (re-)introducing Khoekhoe speakers to cultural concepts about to be forgotten, and of providing comparative linguists with data that may be crucial in reconstructing a proto-lexicon of Central Khoesaaan.

The paper elaborates on how the choice of target users and specific purposes of the dictionary determined the lexicographic strategies that had to be adopted, and eventually resulted in the publication of two separate works: a simplified bidirectional glossary without tone marking for the less discerning user and schools (1999), and a unidirectional comprehensive Khoekhoegowab – English dictionary with a glossary-type English – Khoekhoegowab index for the more demanding user (2002).

The main issues that are discussed are: prescriptiveness versus descriptiveness; redundancy (what to include; what to omit at the expense of user-friendliness by relying on predictability); the arrangement of catchwords in articles for tonological reasons; alphabetisation according to (sometimes polygraphic) phonemes instead of letters; and the kind of linguistic information supplied. Technical aspects like the retrieval facilities of the electronic database and the customised programme for conversion to text format are also briefly touched upon.

*An earlier version of this paper was published in Schladt, M. (ed). 1998. Language, Identity, and Conceptualization among the Khoisan (Quellen zur Khoisan-Forschung 15). Cologne: Rüdiger Köppe. p. 35-64. At the Afrilex conference the present paper is to be augmented by an exhibition of materials and posters.*

[To Table of Contents](#)

---

## The Proposed Ndebele – Shona Dictionary: Prospects and Challenges

Samukele HADEBE

*Department of African Languages and Literature, University of Zimbabwe, Harare, Zimbabwe*

---

The ALLEX Project master plan includes a bilingual Ndebele – Shona dictionary in its proposed dictionary projects. According to this master plan, the bilingual Ndebele – Shona dictionary would be compiled after the completion of the general Shona and the general Ndebele dictionaries. The two dictionaries have since been completed and published in 1996 and 2001 respectively. The stage has been set for the compilation of the bilingual dictionary. At the time of writing, this proposed bilingual dictionary project has not begun as the lexicographers at the *African Languages Research Institute* (ALRI) are still working on other projects like the trilingual dictionary of musical terms, the dictionary of linguistic and literary terms and the advanced Ndebele dictionary. Bilingual and multilingual dictionaries are prevalent in Zimbabwe, yet the proposed Ndebele – Shona dictionary raises some interesting challenges, especially for those who intend to compile it.

First, dictionary making in Zimbabwe more-or-less reflects the language development needs of the nation. In this paper I intend to outline how different dictionary types for both Ndebele and Shona reflected the intentions of those responsible of language planning in Zimbabwe. For instance, the early dictionaries compiled by missionaries were bilingual, that is, English to Ndebele/Shona and vice versa. These bilingual dictionaries mainly targeted Ndebele and Shona speakers learning English and Europeans who wanted to learn African languages. The recently compiled monolingual ALLEX dictionaries are targeted at the mother tongue speakers of African languages and attempt to redress the inadequacy of reference books in African languages. One can also link the types of dictionaries with historical periods. The colonial period saw mainly bilingual dictionaries where English was always one of the languages. The post-independence period saw mainly the monolingual ALLEX dictionaries. The compilation of a trilingual dictionary of musical terms and more significantly the proposed bilingual Ndebele – Shona dictionary will set a new trend back to bilingual dictionaries. Nonetheless, it is a different type of bilingual dictionaries where English is not always one of the languages. It would be two African languages.

Second, this proposed bilingual dictionary will be corpus-based like other ALLEX dictionaries. This will require a parallel Ndebele – Shona corpus. This corpus is still in its infancy and its structure also poses very interesting challenges. Using the corpus has its own challenges too. In the paper I intend to highlight possible advantages and setbacks in using corpora for such a dictionary.

Third, there is the issue of the potential target users to take into consideration. The various monolingual ALLEX dictionaries have clearly defined user groups. Defining the target users raises both sociolinguistic and political questions. So far there is little research on user needs and reference skills in Zimbabwe. The lexicographers for the proposed bilingual Ndebele – Shona dictionary will have to be clear on who the target users will be. Language debates in Zimbabwe usually trigger political concerns. Some of the language controversies in Zimbabwe have been on whether the two languages, that is, Ndebele and Shona, have to be compulsorily taught to everyone. Questions have been asked whether pupils could write both Ndebele and Shona at O Level for instance, and have these counted as different subjects. In short, who needs the bilingual Ndebele – Shona dictionary? In addressing some of these sociolinguistic and lexicographic questions, I hope to bring out the challenges that have to be seriously

considered when this project of national significance is pursued. I also intend to propose some solutions and approaches to these challenges that could be useful not only to the prospective compilers of the bilingual Ndebele – Shona dictionary, but also to other lexicographers facing similar situations. Finally, I will show how the proposed bilingual Ndebele – Shona dictionary reflects the language planning needs of Zimbabwean society.

[To Table of Contents](#)

---

## English for New South African Bilingual Dictionaries

Kathy KAVANAGH

*Dictionary Unit for South African English, Rhodes University, Grahamstown, South Africa*

---

A number of bilingual dictionaries are in various stages of preparation in South Africa. Others are being planned. Most of these dictionaries include English and an African language. The English is contained in the headword list, in, for example, an English – isiNdebele dictionary, whereas in an isiNdebele – English dictionary the English consists of translations of the African-language headwords. The quality of either type of dictionary will be greatly influenced by the sources of the English used by the lexicographers.

The scope of any dictionary depends on the needs of the target users. This will dictate the approximate number and range of headwords to be included, the macro- and microstructure, and also the physical size of the dictionary. Compilation of the headword list will take account of all these factors. In the English – African language type dictionary there are several possible ways of developing the English headword list. Words may be drawn from corpora or databases, or from published dictionaries, perhaps supplemented by lists of specialist terms and words picked up from the media. It is also possible to obtain an English headword list by ‘reversing out’ the translations used in an African language – English format dictionary. The pros and cons of these very different approaches, and some of the pitfalls likely to be encountered, are discussed in detail.

Dictionaries and other sources of English, which may be used as the basis for or to supplement a headword list, are evaluated. Dictionaries aimed at first-language speakers and those written for learners of English are compared in this context. Lists of specialist terms for science or mathematics may be of value in some circumstances. British and American dictionaries contain little, if any, South African English, which will need to be obtained from South African material. New World English words, which have become current since the publication of the source dictionary, may also need to be gathered.

Frequency information given in some learners’ dictionaries may be a useful guide, especially when deciding which headwords to include in a pocket dictionary, but must be treated with discretion. For instance, the *Collins COBUILD English Dictionary* indicates the highest level of frequency with five diamonds. Words of this frequency level include **and**, **house**, and **old**. Many words with lower levels of frequency are also essential in any bilingual dictionary, words such as **bridge** (4 diamonds), **medicine** (3 diamonds) and **ice cream** (2 diamonds). Words of even lower frequency, such as **geometry** and **invoice**, will be vital to students and businesspeople respectively and will need to be included in dictionaries whose target audience includes such people. Frequency information is attached to lemma not to sense, so there is no guidance as to how many senses of a polysemous word should be included in a headword list. The main purpose of an English – African language type dictionary is to help

users understand English texts. English has a huge lexicon, and selection of headwords is a challenge.

In an African language – English type dictionary the English may often take the form of a single-word English translation or several synonyms. Some of the problems associated with the latter approach are briefly mentioned. The quality of translation and number and appropriateness of synonyms chosen will have implications for ‘reversing out’, if that method is used to form an English headword list.

Bilingual dictionaries are complex works and benefit from collaborative effort between speakers of both languages. It is suggested that collaboration is more valuable if it occurs throughout a dictionary project. A checking process by first-language speakers that occurs only after all translation work has been completed is of limited use. It may pick up spelling and typographical errors, some inconsistencies, and some problems relating to synonymy, but does not permit a detailed assessment of the headword list or systematic checks on the treatment of related headwords. The quality of expression in both languages needs to be of the highest order, and the *Dictionary Unit for South African English* looks forward to collaborating with other lexicographical units in order to achieve this.

[To Table of Contents](#)

---

## **From a General to an Advanced Ndebele Dictionary: An Outline**

**Langa KHUMALO**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

This paper seeks to discuss the prospective *Advanced Ndebele Dictionary*, henceforth the AND. The paper will be divided into two major parts. The first part of the paper will discuss the forerunner of the AND. The AND comes after the Ndebele team of editors produced *Isichazamazwi SeSiNdebele*, henceforth the ISN, which is the first-ever monolingual dictionary in the Ndebele language, published in the year 2001. The title of the dictionary itself indicates that it is a dictionary of Ndebele in Ndebele, in which the resources of the language are used for the first time lexicographically to analyse and describe itself. Until then, Ndebele people were using a bilingual Ndebele-English dictionary by J.N. Pelling entitled *A Practical Ndebele Dictionary*, with a total of about four thousand headwords. The ISN is a medium-sized, general-purpose dictionary designed to be inexpensive and easy to handle. The dictionary has a total of twenty thousand and eighty (20,080) headwords. Because the ISN is a general dictionary, that is, a general-purpose reference work, it appeals to a wide spectrum of users. It is principally targeted at secondary school teachers and students’ classes to assist them understand and teach the structure of their language through the provision, for the first time, of a technical terminology in Ndebele, dealing with its linguistic features. Teachers and students of Ndebele are more likely to need to consult a dictionary than others are and to make use of its contents in the course of their daily lives as well as to mediate its contents to others. For the ordinary reader, such a reference work can provide, with ease and understanding, the meaning, use, and function of words. This would not be so easily or fully grasped if conveyed in, and then translated from, a foreign language, as has hitherto been the case. The animating heart of the project is to promote the status and use of the language. The dictionary, it is hoped, will help to make people use it appositely in widening areas of life, and to value it as conferring self-respect and the means towards a better and developed standard of life. It is therefore a



contribution to a change of policy, for Ndebele to be recognised as an official language in Zimbabwe and to allow the Ndebele people to carry out their affairs in all spheres of life in their mother tongue. Hitherto it has remained inferior to English, which is the only official language in Zimbabwe.

The second part of the paper will discuss the structure and content of the advanced monolingual Ndebele dictionary that is in its infancy. The target groups of this dictionary are first and foremost high schools and tertiary institutions and other specialised disciplines. It will be argued that the AND is not just going to be a bigger volume than the ISN, but will be advanced in terms of its depth and scope relative to both the lexical items and definitions. Whereas the ISN was based on a corpus of about a million running words, the AND is envisaged to be based on a corpus of about five million words. The paper will discuss how additional targeted quality fieldwork will improve the size of the corpus and provide appropriate context and content for an advanced dictionary of this nature. Unlike its forerunner the AND will have additional grammatical information that includes a phonetic transcription of each lexical item, tone marking and etymology. Etymology will be of two types, i.e. semantic etymology and lexical etymology. The paper will also discuss how the semasiological fields of the lexical items in the AND will be different from those in the ISN. Finally the paper will discuss how the scope of headword selection will be broadened to accommodate modern or international terms, mathematical terms, scientific terms, cultural terms and other specialised terms. The paper will demonstrate as a way of concluding that the prospective AND will not just be a volume with just more headwords than its forerunner, but its target audience will be different, its depth and scope will be greatly improved and hence worth its title, “An Advanced Ndebele Dictionary”. The Ndebele language will have a special reference work that should change or improve both the status and use of the language.

[To Table of Contents](#)

---

## **The Incorporation and Handling of Metaphorical or Figurative Meaning in Bilingual Dictionaries**

**John M. LUBINDA**

*Department of French, University of Botswana, Gaborone, Botswana*

---

The presentation of the meaning(s) of recorded lemmas in a dictionary is one of the essential tasks of the lexicographer and one that certainly requires very careful consideration in practical lexicography. There are several different types of lexical meaning and all these ought to be taken into consideration in a dictionary that purports to give a full account of word meaning. Semanticists (e.g. Leech 1974) have long recognized that, in addition to its conceptual (denotative) meaning, a word may have several other meanings, e.g. figurative or connotative meaning. One criticism that one may be justified to make against some bilingual dictionaries, especially compact pocket dictionaries, is their tendency to neglect figurative and other ‘secondary’ meanings that words may have in particular contexts. Thus, the dictionary user is given only a partial semantic account of the lemmas presented in the central list. We know, from practical linguistic experience with languages that we are familiar with, that metaphorical use of words is extremely common. Many words are in fact used more often in their figurative sense. Consequently, language users will come across the use of these words more often in their figurative sense than in their denotative one.

Dictionaries vary considerably in size and scope. They also differ (in some cases quite substantially) in their approach to semantic presentation. However, regardless of whether they are of the explanatory or translation type, some dictionaries offer only what could at best be described as the barest minimum in terms of meaning specification. They supply only the basic conceptual meanings of lemmas of the type:

**head** *noun* upper part of the body above the neck that contains the brain and on which are located the eyes, ears, nose and mouth. [in a monolingual dictionary of English]

**head** *noun* tête (n, f). [in an English – French dictionary]

**head** *noun* tlhogo. [in an English – Setswana dictionary]

On the other hand, there are other dictionaries that provide much more than this minimal conceptual meaning by also indicating the various polysemic distinctions that the lemma **head**, for example, may have figuratively, colloquially or collocationally, depending on the context of usage. A dictionary that provides such extensive semantic information will usually, as a matter of necessity and sound lexicographical convention, make use of usage labels to indicate contextual restrictions of usage of particular senses of the lemma as well as numbers to delineate the various semantic values. It will also provide explicit examples, in the form of illustrative sentences to capture the different distinctions of meaning presented in the definition or translation equivalent.

The paper argues that it is this type of dictionary that the foreign or second language learner and the language practitioner (such as the translator / interpreter, script writer / editor, language teacher, etc.) is most interested in. The proposed conference paper sets out to highlight the different types of meanings as defined in semantics literature and then goes on to compare and contrast the practices of semantic presentation followed in two bilingual dictionaries. David Crystal (1987: 108) points out that *'the best way to evaluate the coverage of a dictionary is to compare the words and senses it includes with another dictionary of about the same size'*. The paper draws attention to the shortcomings of the type of dictionary that registers only the conceptual meaning of lemmas while neglecting the rest – a sort of 'glorified wordlist'. It pleads the case for a lexicographical practice that seeks to present a full semantic account of lemmas with special reference to figurative/metaphorical meaning. Issues of semantic broadening (polysemic expansion), meaning transfer and dialectal variations in meaning are briefly discussed. The paper further highlights the problems and pragmatic limitations to be expected (encountered?) in terms of scope, organising principle and size of the dictionary in trying to follow this practice.

[To Table of Contents](#)

---

## Capturing Cultural Glossaries. Case Study II: Medical Terms

**Matete MADIBA & Lorna MPHAHLELE**

*Technikon Northern Gauteng Pretoria North, South Africa*

**Matlakala KGANYAGO**

*Nkoshilo High School, South Africa*

---

This work is a continuation of a project that was initiated in 2002 and presented by the same authors as a case study at the 7<sup>th</sup> International Conference of AFRILEX, Rhodes University,



South Africa. The project aims at capturing cultural glossaries within an authentic context of a school setting in a rural area in the Limpopo Province. Of particular interest is the potential projects of this nature have to capture and record cultural words that would otherwise be lost. The previous presentation concentrated on a cultural glossary of cooking terms in Northern Sotho. The present work, considered as Case study II, is dedicated to medical terms, gleaned from the preparations and execution of medicinal processes in the Northern Sotho culture. The authors do not claim to present a comprehensive glossary, instead, they hope to share what a school project was able to uncover; with the wish that other ventures and bigger projects will focus on more comprehensive products.

Working on projects like these also investigates how these glossaries can help realising and implementing innovative lexicographic methodologies and concepts such as ‘simultaneous feedback’ (De Schryver & Prinsloo 2000), and ‘hybrid dictionaries’, to support major dictionary work in South Africa, as suggested in the previous presentation. As with the previous project, it is also interesting to note that the glossary is a ‘secondary’ and not a ‘primary’ product of the project, in the sense that the project had a different target. The main target is the teaching and learning of Northern Sotho as a first language within the Outcomes-Based Education (OBE) environment. The project is also an acknowledgement that the OBE approach has stimulated thinking about activities for teaching and learning that were previously not thought of. Mother-tongue teaching and learning in African languages, Northern Sotho in particular, was far less engaging for both teachers and learners in the past. It is this distinctive feature (of being a ‘secondary’ product) that also has to be investigated for further implications.

The case study approach is found to be more suitable to a project like this as it will be easier to formulate lessons learnt in the process of compiling this brief glossary. It is the exploration of these lessons that is considered another step in the process to work towards a possible and authentic model for the collection of other glossaries of this nature.

This work also hopes to provide ways to supplement the corpus-based approach in the writing and producing of dictionaries for African languages. It is also seen as a project with enormous potential to contribute towards initiatives in Indigenous Knowledge Systems (IKS). One more benefit is that the capturing of such cultural terminologies within an authentic setting helps to provide contextual information for related idioms and proverbs. The meanings of the idioms and proverbs become more transparent. It is also worth mentioning that the contextual capturing moreover helps in providing ‘encyclopaedic information’ that would otherwise not be captured, a challenge that this work would also like to raise for other African languages.

[To Table of Contents](#)

---

## **The Users’ Perspectives on *Isichazamazwi SeSiNdebele***

**Mandlenkosi MAPHOSA**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

This paper seeks to discuss the findings that came to the fore after the *Isichazamazwi SeSiNdebele* (ISN) editorial team embarked on a feedback exercise in the Ndebele speaking provinces of Zimbabwe. The paper will discuss the objectives of the exercise, the outline of the fieldwork, the research frame, the seminars that were conducted on aspects of the dictionary and the users’ responses. The whole exercise was carried out with the aim of getting feedback

from the users of the ISN, which is the first-ever monolingual dictionary of Zimbabwean Ndebele. However, the exercise was not meant to benefit the editorial team only as it also involved the enculturation of users into the dictionary world since the dictionary was the first of its kind that was expected to be widely used by the Ndebele language community. It should also be stated that the exercise was a way of laying the foundation for the advanced Ndebele dictionary and the revised ISN.

The fieldwork exercise was carried out in the Ndebele speaking provinces of Zimbabwe and it covered schools and teachers' colleges. The exercise looked at various aspects of lexicography that cover the very foundation on which the dictionary is based, i.e. the corpus. The corpus was explained in detail since it had a bearing on the final product in many ways. The corpus was defined and this was a way of summing up the main source of lexical entries of the ISN and to show the extent to which the speakers of the language had themselves contributed to the final output. A general background on the structural aspects of the dictionary was also presented enunciating the major structural aspects of the dictionary and their roles emphasising on how they complemented each other. Headword selection was also covered explaining in detail to the users how the headwords that they saw in the dictionary found their way there. This also brought to light the issue of lemmatisation. The team also explained the defining formats that were used in the dictionary and the grammatical information that they could find in the dictionary. The aim of doing all these presentations on lexicography was to enlighten the users on the process of compiling dictionaries so that they could give their feedback from an enlightened position.

Having received the 'lexicographic enlightenment' the users then gave their perspectives on the dictionary. These perspectives will be the major aspect of this paper. The field research provided a variety of responses some of which were expected by the editors whilst some were a complete surprise. Some of the responses bordered on lexicographic ignorance whilst some bordered on nationalistic assertions on their language. It is these perspectives that this paper wants to discuss; to indicate the divide that the lexicographer encounters of user needs versus lexicographic principles. The paper will also deal with the contents of the discussions that took place in the seminars. As such headword selection will take centre stage especially with regard to loanwords.

[To Table of Contents](#)

---

## **Bilingual versus Monolingual: A Comparative Analysis of Two Trends in Shona Lexicography**

**Webster MAVHU**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

The intended paper will focus on lexicographic trends in Shona, an indigenous Zimbabwean language that is made up of five dialects and that is spoken by about seventy-five percent of the country's total population. The Shona language has a lexicographic tradition that stretches backwards for nearly one and a half centuries, to 1856 when W.H.I. Bleek published the first lexicographic work on the language, a work that was bilingual in nature. From that date, more bilingual dictionaries, comprising mainly Shona and English, continued to be produced by missionaries. Up to 1923, the bilingual dictionaries appeared in a dialect that each compiler favoured. It was only after Doke suggested, in 1931, that the five Shona dialects were supposed

to be unified, that bilingual dictionaries were compiled that represented the five varieties. The bilingual trend continued for some time until it was ‘broken’ in 1996.

Indeed, in 1996 the *African Languages Lexical* (ALLEX) Project, which was housed in the Department of African Languages and Literature at the University of Zimbabwe, published the first-ever Shona monolingual, synchronic, medium-sized and general-purpose dictionary, *Duramazwi ReChiShona* (DRC). The publication was followed by yet another monolingual dictionary that can be regarded as an advanced version of the same dictionary and whose title is *Duramazwi Guru ReChiShona* (DGC) in 2001. A six-member team of mother-tongue speaker-writers of the Shona language, of which the present writer is part, produced the latter monolingual Shona dictionary.

Just like in many other African countries, there is a poor dictionary culture in Zimbabwe. Most people, some of them notable scholars, are not able to see how Shona bilingual dictionaries are different from monolingual ones. In fact, one scholar proclaimed in 1996 that Dale (1981) and not DGC (1996) is ‘the first Shona dictionary’ (SAPEM 1996/97: 34). By this statement he intended to imply that Dale’s dictionary (which is bilingual) would be the first monolingual dictionary, which is not true. Even though Dale reduces the English component quite drastically in his dictionary, the work is still bilingual. It is from the realisation that there are such misunderstandings as the one that is mentioned above that the intended paper arose.

In the intended paper, the present writer will offer a comparative analysis of the bilingual and monolingual lexicographic practices in Shona lexicography. This will be done through the use of selected bilingual dictionaries: Hannan (1959, 1974 & 1984) and Dale (1981); and the current monolingual Shona dictionaries: Chimhundu et al. (1996 & 2001). The focus will be on how the above-mentioned dictionaries present both the lemma and lexical meaning in their entries. The lemma is that part of a dictionary entry that gives information revolving around the lexical unit itself (Zgusta 1971: 249). The information includes headword identity and spelling, tone indication, noun class numbering, etymological data, and the marking of dialectal origin of the lexical item. Lexical meaning, on the other hand, is the main part of a dictionary entry. The basic instruments for the description of lexical meaning are the plural field of the headword, the lexicographic definition, exemplification, and the location of the headword in the system of synonyms and illustrations.

From the discussion of the above-mentioned items, the present writer will attempt to clearly show how problems, issues and advances in monolingual dictionary making are distinct from those in bilingual dictionary making. The researcher hopes that the paper will help scholars in general, and speaker-writers of the Shona language in particular, to clearly understand the different approaches in these two lexicographic practices. He also hopes that the paper will suggest the way forward for further advances in improving lexicographic traditions in Africa in general and in Zimbabwe in particular.

[To Table of Contents](#)

---

**The Impact of Translation Activities on the Development of African Languages in Multilingual Societies: Shona – Ndebele – English Musical Terms Dictionary, a Case Study**

**Gift MHETA**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

The paper will examine the impact of translation activities on the development of African languages. It will analyse Shona musical terms that were created through term-creation and translation processes and strategies such as borrowing, blending, coining, compounding, clipping, derivation and paraphrasing. Focus will be on how such strategies are contributing or hindering the development of the Shona language. The importance of such strategies and processes will be discussed in the broader context of empowering African languages. Recommendations on the best strategies to employ when dealing with specific musical terms will be given with a view to creating a uniform body of musical terms. Non-linguistic recommendations that can contribute to the development of the Shona language will also be offered in the presentation.

The main aim of the paper is to reveal the inconsistencies inherent in the music discipline with regard to the creation of terms used for music education and appreciation. Term-creation has been going on in this field and the main problem is that it has largely been unplanned and uncoordinated. In Zimbabwe, there are no language or term banks, which can come out with consistent methods of term-creation and standardisation of created terminology. The chaotic situation in term-creation is not confined to the music field but is a general problem affecting nearly all sectors in the Zimbabwean community. According to Chimhundu (1987: 142), term-creation is a growing phenomenon, particularly in the post-independence era in Zimbabwe. It is proliferating in business, central and local government, commerce, industry, mining, agriculture, broadcasting, education and other spheres of life.

To reveal the above-mentioned problems the paper will focus on Shona, the main indigenous language in Zimbabwe. Shona musical terms that have been collected by the *African Languages Research Institute (ALRI)* for the compilation of a Shona, Ndebele and English Dictionary of Musical Terms, will be analysed in the proposed paper.

In Zimbabwe, music is a well-established discipline that uses specialised terms in the analysis and teaching of sound components. Like in most specialised fields, it is taught in English and some Shona terms that exist are equivalents created from English through translation. Chimhundu (1996: 449) aptly notes that:

a main trend in translation between international languages or languages of wider communication (LWCs) and indigenous languages or national official languages (NOLs) is unidirectional transfer from the LWCs as SLs to NOLs as TLs during the translation process. Both ideas and words are transferred as African societies modernize and change.

Examples of such terms created through translation in the musicology branch of music are tabulated below:

<b>English form</b>	<b>Shona form</b>
alto	aruto
bass	bhesi
beat	bhiti
chorus	korasi
soprano	sopurano
verse	vhesi

In analysing translations from which some Shona terms in the music field emanate, the present researcher will use Chimhundu's 'scan and balance theory'. Chimhundu (1996: 452) proposes that part of the process that involves searching for equivalence or creating new terms may be viewed as moving in and out of each language and culture with a scanner (i.e. brain) to identify equivalent terms and expressions. When these have been found or created, the translator compares their senses or ranges of meaning, usage, appropriate registers and impact, and then makes selections accordingly. This, according to Chimhundu, is the balancing part that comes after the initial scanning phase of the translation process, hence the term scan and balance theory. This theory emphasises the translator's creativity when dealing with cultural and linguistic differences in the SL and TL texts. This makes the theory readily applicable to issues of language development, especially for languages of limited diffusion (LLDs) such as Shona, which have limited technical terminology.

The paper will analyse the term creation strategies, most of which have been mentioned in the first paragraph, in terms of how they affect language development. It will highlight phonological, morphological and semantic processes and changes undergone by terms in the creation process.

The paper will discuss the problems evident in the above-mentioned grammatical categories. It will focus on:

- problems created by the Shona alphabet;
- problems created by consonant and vowel combinations in Shona;
- problems created by dialectal variations.

It will offer solutions that emanate from acceptable linguistic analyses of existing musical terms. Such solutions are hoped to encourage uniformity in the creation of terms. This is very important, as it is a way of contributing to the ongoing standardisation of African languages. It is also a way of enhancing the communicative power of Shona, a goal quite in line with the resurgent African renaissance.

The paper will also present recommendations, not on the development of musical terms per se, but on how to develop term-creation activities in general. It will focus on how such linguistic activities can be supported by various stakeholders in Zimbabwe. The roles to be played by the government, to-be term bank committees, language committees and representatives from fields that use specialised terms will be highlighted.

[To Table of Contents](#)

---

## **The Lexicographic Treatment of the Feminine/Augmentative Suffix *-kazi* in isiZulu**

**Linkie MOHLALA**

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

**Gilles-Maurice DE SCHRYVER**

*Department of African Languages and Cultures, Ghent University, Ghent, Belgium*

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

**Rachéle GAUTON**

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

---

According to Barnhart (1953: ix), a good dictionary is a guide to *usage* much as a good map primarily shows you the nature of the terrain. It is this view that underpins the current investigation into the usage of the isiZulu feminine/augmentative suffix *-kazi*. Usage, as clarified by Allen (1964: 272), “is the relationship between what goes on inside a language and the context of speaker, audience, time, place, and the occasion in which it occurs”.

Modern technology has made it possible for lexicographers and linguists to revisit long-held notions on usage in order to provide for more accurate descriptions. In this regard we specifically refer to the recent availability of an electronic isiZulu (text) corpus, known as the University of Pretoria Zulu Corpus (PZC). PZC presently stands at 5.0 million running words (or tokens), and can be analysed with software such as WordSmith Tools. As a matter of fact, by testing the prevailing views regarding the suffix *-kazi* as found in standard isiZulu reference works against PZC, one is in a position to arrive at a description that is conditioned not by preconceived notions, introspection or anecdotal data, but a description that is based on a vast storehouse of actual language usage. Such descriptions can then be used to prepare truly modern dictionary articles.

A total of 11,857 occurrences of the suffix *-kazi* affixed to nouns are found in PZC, which amounts to 92% of all examples containing this nominal suffix. The corpus study confirms what is implicit to Taljaard & Bosch (1988: 144 *et seq.*) and Doke’s (1973<sup>6</sup>: 70 *et seq.*) discussions of this suffix, and explicitly stated by Wanger (1917: 138) and Van Eeden (1956: 725 & 726), namely that the primary significance of the suffix *-kazi* is the expression of the feminine form, with the augmentative significance as secondary. There are, however, a few claims found in standard isiZulu sources that are proven incorrect when tested against the corpus data. Poulos & Msimang (1998: 112), for instance, claim that *-kazi* is never used to derive feminine forms from nouns denoting wild animals. In the corpus, as many as 700 instances are found of the suffix *-kazi* in exactly this environment. As another example, Wanger (1917: 139) states categorically that the feminine suffix *-kazi* does not occur with nouns ending in *-o*. The corpus data disproves this claim, as there are 206 instances of the feminine *-kazi* suffixed to *o*-final nouns. Furthermore, when studying the corpus, certain aspects of the suffix *-kazi* come to the fore that have been under-emphasised, inadequately treated and/or omitted in standard works on the isiZulu language. Such sources tend to define augmentatives as primarily indicating ‘bigness’ or ‘greatness’ (cf. Poulos & Msimang 1998: 110). Yet, it would seem from the corpus data that when *-kazi* is used as an augmentative suffix, it primarily serves to indicate ‘added value’, ‘importance’ or ‘intensity’ (sometimes in a neutral context, but often in either a positive or a negative context), as opposed to an increase in size.

The main aim of this paper, then, is to revisit the lexicographic treatment of the feminine/augmentative suffix *-kazi* in existing isiZulu dictionaries. Our investigations have firstly shown that many lexicographers apparently do not regard this suffix as a word category that merits thorough treatment in the central text of a dictionary, especially when it comes to its meaning and usage. Indeed, whereas most dictionaries (e.g. Roberts 1942, Dekker & Ries 1958, or Nkabinde 1985) do not treat this suffix at all, those that do (such as e.g. Bryant 1905, Nyembezi 1992), often do so in the front matter of the dictionary only. It is unfortunately well known that very few dictionary users consult front and back matter material. Only a handful of dictionaries (e.g. Doke & Vilakazi 1953<sup>2</sup>) do treat *-kazi* in the central section of the dictionary itself.



All these dictionary treatments of *-kazi* will be carefully scrutinised, compared to one another, placed next to the grammars, and contrasted to the fresh corpus data briefly outlined above. Moreover, in order to ascertain current, spoken mother-tongue usage of this suffix and its relative frequency, the results of various recordings centring on the use of this suffix will be presented, for, as stated by Matthews (1925: 1173):

... language is never in the exclusive control of scholars. It does not belong to them alone, as they are often inclined to believe; it belongs to all who have it as a mother-tongue. It is governed not by elected representatives, but by direct democracy, by the people as a whole ...

In conclusion, various model articles will be drawn up that summarise the findings and illustrate how the feminine/augmentative suffix *-kazi* should be treated in a modern, user-friendly isiZulu dictionary.

[To Table of Contents](#)

---

## **The ALRI Experience in the Compilation of a Dictionary of Biomedical Terms**

**Nomalanga MPOFU**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

The paper will seek to highlight the experience and the challenges that were faced by a team of four researchers in the *African Languages Research Institute* (ALRI) at the University of Zimbabwe, in the compilation of the first bilingual Shona-English dictionary of biomedical terms, *Duramazwi reUrapi neUtano* (A Dictionary of Biomedical Terms). This dictionary project was started in August 2001 and is set to be completed in May 2003. The product of this research, a bilingual dictionary of biomedical terms, will aim at improving efficiency of communication between doctors and patients. The dictionary is composed of terms from both modern and traditional medicinal practices.

The project was initiated by the *Institute of Continuing Health Education* (ICHE), which is based at the University of Zimbabwe's Medical School. The Dictionary of Biomedical Terms is being compiled with the aim of providing a tool for communication between the doctor and his/her patient. There seemed to be a need for doctors and patients to communicate better so that patient expectations would be fulfilled after he/she has been to the doctor. The present scenario that has been observed and that has acted as a barrier to communication between doctor and patient in Zimbabwe, is that doctors are trained in English while the majority of the people they will be dealing with use indigenous languages, in this case Shona. It was thus seen that there would automatically be a communication problem because of the different languages and levels at which the two people in contact use language. Quite often, there is also a generation gap between the doctor and the patient, since some of the doctors are young people fresh from medical school. As a result, there are cultural nuances that are loaded in the language that are usually missed by the younger generation of doctors. This dictionary would thus serve to address the needs of doctors to understand the terms and expressions used by the patients and to standardise terms that are used by different age groups in different parts of the country.



The dictionary is divided into two sections. The first section comprises of Shona headwords with English equivalents. The headwords in this section are defined in Shona. The second section has a reversal of the words in section one. In the second section the English headword is the main entry followed by the Shona equivalent. There are no definitions in this section. The reversal is in alphabetic order.

The paper will look at: (i) the method(s) that led to the production of this dictionary; (ii) the presentation of entries in the dictionary as well as some sample entries; and (iii) the challenges encountered in the compilation process, namely to develop Shona medical terminology in a cultural context and the problems of equivalence between English and Shona biomedical terms.

[To Table of Contents](#)

---

## **Language Development or Language Corruption: A Case of Loanwords in *Isichazamazwi SeSiNdebele***

**Cornelias NCUBE**

*African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe*

---

This paper identifies and analysis loanwords in *Isichazamazwi SeSiNdebele*, henceforth referred to as ISN. In particular the paper looks at the acceptability and/or non-acceptability of loanwords by the target users of the first monolingual Ndebele dictionary. The Ndebele language of Zimbabwe has not been immune to the phenomenon of language contact and its resultant effect of cultural borrowing and dialect borrowing. In Zimbabwe the language shares the same linguistic environment with English, Shona and official minority languages such as Kalanga, Tonga and Nambya. The language also has a historical heritage that links it with its Nguni sister dialects such as Zulu and Xhosa spoken in South Africa. The Ndebele language of Zimbabwe draws some of its lexicon from these languages. In some cases Afrikaans words have found their way into the language through Zulu. In selecting headwords for the ISN the *Ndebele Lexicography Unit* (NLU) mostly used the frequency list derived from, and the lemmatised headwords found in, the corpus. This method inevitably gave leeway to the adoption of loanwords in the ISN with resultant public outcry.

The paper will be divided into two broad sections. The first section gives a general review of comments from users of ISN about the inclusion of such ‘passport words’, also referred to as loanwords in the dictionary. The NLU conducted its outreach programme in 2002 to solicit views from Ndebele language speakers about the user-friendliness of the dictionary. The team received criticism from the target users for having included loanwords in the dictionary. It must be noted however that acceptance of loanwords in the ISN varies with different age groups. The younger generation freely accepts the loanwords as part of the Ndebele lexicon as opposed to the older generation. The second section analyses the justification by editors for lemmatising loanwords against views by the target users. This section will show that the editors’ justification is at variance with the expectations of the target users because of the latter’s reasons which go beyond lexicographic principles. The paper will prove that reservations against loanwords in ISN go beyond principles of dictionary making. At the forefront is the users’ attitude towards the source language. Language attitude in Zimbabwe is by and large a result of socio-political and economic power that characterise the different tribal groups in the

country. It also draws from the historical tribal relations in Zimbabwe before and after the country's independence. As a result, Ndebele lexicographers find themselves torn in-between adhering to principles of corpus-based dictionary making and language conservationists championing 'language puritism' and 'language emancipation'. Suffice it to say that 'language puritism' and 'language emancipation' are forms of protest by speakers of the borrowing language against domination (of any form) by speakers of the lending language. 'Language puritism' and 'language emancipation' are a nostalgic longing for the defunct historical 'prestige status' of the Ndebele people over other tribal groupings before the pre-colonial period in Zimbabwe. The paper concludes by discussing possible solutions to the problem of loanwords to be adopted in the forthcoming *Advanced Ndebele Dictionary* (AND).

[To Table of Contents](#)

---

## **The Lexicographic Treatment of the Demonstrative Copulative in Sesotho sa Leboa – An Exercise in Multiple Cross-referencing**

**Salmina NONG**

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

**M.P. MOGODI**

*Sesotho sa Leboa National Lexicography Unit, Pretoria Branch, Pretoria, South Africa*

---

The main aim of this paper is to expound on some of the procedures used during the compilation of *SeDiPro*, i.e. the Sesotho sa Leboa Dictionary Project. Just like any other modern dictionary, also this dictionary is corpus based. This basically means that high-frequency items are treated first, whilst lesser-frequency ones are left for a later phase of the project. However, for the African languages it is sometimes advisable, in fact even *crucial*, to treat all items that belong to a given *paradigm as a group*. This will be illustrated for the demonstrative-copulative paradigm.

It is well known that nouns are grouped into noun classes in the African languages. Depending on the class, each series of demonstrative copulatives will have a different form. Current grammars indicate that there are 6 different types of demonstrative copulatives in Sesotho sa Leboa, for 15 classes, or thus 90 demonstrative-copulative forms in all. Our research, however, soon revealed that there are many more; we recovered 152 forms so far. Obviously, some of these are many times more frequent and thus more important than others, while still others occur only exceptionally. Taken at face value, frequency counts could and should thus be the arbiter in order to decide on inclusion or omission. On the other hand, it feels awkward not to provide a complete paradigm – for what basically is a single concept – in a dictionary. The research question thus revolves around the issue of how to reconcile these two opposite aims. As it turns out, it is an exercise in applying various cross-referencing devices.

The first step in the research process was to draw up a matrix, consisting of all the noun classes and their different demonstrative copulatives, according to the positions that are distinguished for each class. During the second phase of the research our aim was to verify and record the frequency counts of each demonstrative copulative to enable us to make an informed and lexicographically sound decision as to which items should be treated in the central lemma-sign

list and which ones should be treated in tabulated form in the front or back matter, with appropriate cross-references.

A Sesotho sa Leboa corpus, consisting of 6.1 million running words chosen from 350 texts, was used as an electronic database to do this research. Several problematic issues had to be addressed during this phase of the research, many of them on a purely practical level. Since the corpus used is an untagged one, homonymy posed a real obstacle. Contrary to what one may think, demonstrative copulatives being agreement morphemes, some are indeed homonymous to other lexical items. For example, the demonstrative copulative of classes 8 and 10, position I, *šedi* ‘here they are’ is homonymous with the class 9 noun *šedi* ‘care; attention’. A blind concordance search therefore did not produce satisfactory results. Instead, all concordance lines had to be read through in order to isolate and to identify the true occurrences of the demonstrative copulatives.

Another problem that is equally relevant to the issue of frequency is the fact that some demonstrative copulatives of different classes are morphologically similar; for instance, those of classes 1 and 3, or those of classes 4 and 9. This has a direct bearing on the decision as to whether items should be treated fully in the central list or else only sketchily through cross-referencing.

Dialectal variation is another challenging issue. In some cases it was found that there are dialectal variants which are frequently used in everyday spoken language, but do not appear in the corpus. The reason for this is that these forms are regarded as non-standard, with their usage as a result being discouraged in written language.

Since dictionary making continues to move from prescriptiveness to descriptiveness, these issues have to be addressed in order to enable us to compile a dictionary that is not only lexicographically sound, but also answers the need for user-friendliness. In this case, corpus-informed cross-referencing was used as a powerful device to link over 150 members of a highly complex paradigm.

[To Table of Contents](#)

---

## Challenges to Representative and Balanced Corpora for African Lexicography

**Thapelo J. OTLOGETSWE**

*Department of English, University of Botswana, Gaborone, Botswana*

---

Some of the latest corpus-based lexicography researches consider issues of representation and balance (Ooi 1998) as marks of standards of authenticity and robustness in corpus construction. A language corpus must be balanced and representative of the language from which it is extracted. By representativeness is meant *‘the extent to which a sample [text] includes the full range of variability in a population’* (Biber 1993: 243) and as Summers (1993: 186) maintains *‘unless the corpus is representative, it is ipso facto unreliable as a means of acquiring lexical knowledge’*. Therefore for the corpus to be representative it must reflect the typical cross-spectrum of language use of a defined language community or period (Ooi 1998: 49). But we would return to Summers’ (1993) claim since it raises considerable difficulties, particularly for corpus building in many African contexts and to certain linguistic theories.

The problem of what constitutes balanced and representative corpora still remains controversial. The selection of language from different genres to include in the language

database is largely unresolved. The compilation of text must finally capture language from a specified population from which a sample is taken, which reflects how that particular language community uses language. This is significant since as Summers (1993: 186, 190) points out, the results of corpora analysis must be generalised to the general language community from which the samples were abstracted. It is in a way clear that issues of balance and representativeness of corpora are related. A representative corpus must reflect a representation of different genres of language use in a language community; while a balanced corpus should attempt to capture those different percentage levels or ratios in the way they occur in the specified language community. This is obviously difficult to achieve, mainly because it is difficult to know precisely all the text types and their proportions of use in a population with its ever-changing dimensions. The difficulties are compounded when one faces the building of a corpus of spoken language. This is the case, since as Kilgarriff (1997) points out, dialectal varieties stand at different ratios to one another and should be represented within a corpus that attempts to accurately capture the language characteristics as a whole. We must also contend with whether spoken text can be accurately sampled and represented along the same lines as written text. How many words are we looking for and what percentage of the spoken language do such words constitute? Whether spoken text can be sampled in a representative manner is greatly questionable. Although we can have a sample of the Sengwaketse dialect or Sekwena or Sekgatla, establishing an acceptable representative percentage of the spoken form of these dialects poses great difficulties, since speech is a flood that refuses to be adequately accounted for numerically. Even as we attempt to quantify it, more of it is being produced.

Atkins proposes a way of getting around this problem of an ever-changing corpus. Atkins (1997) proposes an ambitious approach of maintaining a corpus by using it, then identifying its strengths and weaknesses, and then adding or deleting material from it to enhance it. This approach of continuous checking of a corpus reveals how difficult it is to have a reliable corpus since there is an ever-flowing text that gets added to the language corpus on a daily basis. Atkins doesn't say whether proportions of language representation should be checked frequently to ensure that language patterns reflect the proportional language change. That is, should the percentage of newspaper data be checked against that of novels and other genres? Perhaps Atkins' approach may be feasible in the construction of a corpus of written texts, but it is difficult to see how it would be successful in the construction of a spoken corpus.

With this background in mind, this paper attempts to investigate the problems associated with the construction of corpora for dictionary building in many African contexts. It argues that some of the challenges facing the construction of robust corpora to be used in language research are the poverty of data, the lack of text to construct corpora that can be representative of the different instances of language usage in a specific speech community. High illiteracy levels in African countries poses great challenges to researchers who hope to collect written text read by populations. Added to that is the fact that even where levels of literacy have increased, the literate members of the society read and write text written in English or French and not in their native languages. Even where such texts could be found in African languages such texts are mostly of a certain genre, like novels, plays and poetry, to the exclusion of another genre, like newspapers and academic texts. Even if we attempted to use such data, we would have to contend with 'sanitised' data, purified by the editorial policies and stylebooks of many publishing houses and newspapers, calling into question its authenticity as original and credible text. The problem of representing speech still stands as one of the great challenges not

only to African lexicographic research but also to research in many western countries and is handled in this paper.

## [To Table of Contents](#)

---

### **The User Perspective: Bible Reference Resources as Example**

**Annél OTTO & Nerina BOSMAN**

*Department of Afrikaans, Vista University, Port Elizabeth, South Africa*

---

The fact that more research on user needs should be done has been stressed in the lexicographical literature. Compare the following statement by Stein: *'Dictionaries are obviously written for their users. We therefore need much more research on the dictionary user; his needs, his expectations, and his prejudices'* (1984: 4). Hartmann (1987: 12) distinguishes between four groups of research on dictionary use:

- research on the categories of information which are provided in dictionaries (dictionary typology);
- research on specific dictionary user groups (user typology);
- research on the contexts of dictionary use (needs typology);
- research on dictionary consultation strategies (skills typology).

Hartmann (1987: 27) concludes that research on dictionary use has only been done on a small scale and that this research is often non-representative, non-comparable, non-correlational and non-repeatable.

According to Hatherall (1984: 184) the inherent restrictions of research based on questionnaires should also be taken into account. These restrictions may be especially valid when research with regard to look-up procedures are being investigated. At that point closer direct observation would certainly provide better results. When respondents should merely indicate which resources they use and which information types they would like to have in a particular type of resource, then the use of questionnaires may be valid, despite possible inherent restrictions.

Hatherall (1984: 189) also indicates in which direction research should move in order to make progress. This may include *'closer direct observation by means of protocol, film and audio recordings as well as personal interviews, plus computer tests involving the logging-in and processing of data through video screens (for the latter, cf. Fox et al. 1980)'* (Hartmann 1987: 22-23).

Since these findings/problems stated by Hartmann and Hatherall, several studies have been done and articles and books published on this topic, e.g. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*, edited by Atkins (1998). In this book there are for instance several reports on how different users used different dictionaries to perform various project-specific linguistic exercises. Questionnaires are also being used during fieldwork when user preferences need to be established, e.g. the study about loanwords versus indigenous words in Northern Sotho by Nong, De Schryver & Prinsloo (2002). Despite the possible shortcomings of questionnaires, it is still an acceptable way to determine user needs and expectations, especially when supplemented with other methods.

In this survey the needs and expectations of respondents regarding the use of Bible reference resources have been investigated. The survey has been conducted among 100 randomly



selected persons, who are either ministers or other persons reading and/or studying the Bible. These respondents are from different age and gender groups, different denominations, places of residence, etc.

A first step was to determine which types of users actually use Bible reference resources and for which purposes, and to find out whether current Bible reference resources meet the needs and expectations of different types of dictionary users, based on the information types that they indicated as either essential, desirable or superfluous and the purposes for which the dictionaries are used.

The findings of these questionnaires are supplemented by a list of Bible dictionaries found on the Internet together with a list of characteristics which these dictionaries contain and which are (i) saying what the editors think the positive aspects of the dictionaries are, and (ii) how the different dictionaries are rated by customers as well as their opinions with regard to the usefulness or not of the dictionaries. It is being argued that valuable information can be gleaned from these customer reviews.

The next step was to see if there is any correlation between the needs and expectations of the users and the implied needs and expectations of the editors.

In line with Tarp (2000: 199) it is being argued that a ranking of functions and user types, giving first priority to some of them, second priority to others and third priority to still others, is needed. That means at least that you are sure that you are making a homogeneous quality product that meets the functions and serves the user types that you regard as most important for this particular dictionary. For the second and third categories of functions and user types, the dictionary may not be perfect, but it provides at least some kind of assistance to the users.

[To Table of Contents](#)

---

## **The Lemmatisation of Adverbs in Northern Sotho**

**D.J. PRINSLOO**

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

---

To date, Northern Sotho metalexigraphers have focused their attention on lemmatisation problems in respect of the so-called main or primary part-of-speech (POS) categories, viz. nouns and verbs. No attention has been given to lemmatisation of the *adverb*, which is regarded as a secondary POS. Adverbs in Northern Sotho appear thousands of times in the *Pretoria Sesotho sa Leboa Corpus* (PSC). These enormous overall counts clearly indicate not only that they should be included as lemmas but also that exhaustive treatment is required/justified especially for the encoding needs of inexperienced target users.

The aim of this paper is to offer solutions to the lemmatisation problems regarding adverbs in Northern Sotho and to propose guiding entries for paper and electronic dictionaries that could serve as models for future dictionaries. The current treatment of adverbs in Northern Sotho dictionaries will also be critically evaluated, especially in terms of frequency of use and target users' needs.

A prerequisite to a successful lemmatisation strategy for, and treatment of, adverbs, is a thorough understanding of the nature of adverbs in Northern Sotho. Lombard (1985: 166, 167) rightfully states that, morphologically, adverbs are heterogeneous, i.e. a specific form, or specific structural characteristics, cannot be attached to adverbs. Van Wyk et al. (1992: 118)

simplify the issue for first-year students in dividing adverbs into three categories namely basic adverbs, derived adverbs and adverbs which have been adopted from other word categories:

- Basic adverbs: **ruri** ‘really’, **fela** ‘just, only’, **ntshe** ‘there’
- Derived adverbs: **gagolo** ‘mostly’, **gantši** ‘often’, **gatee**, ‘once’
- Adverbs adopted from other word categories: **bošego** ‘at night’, **mošate** ‘in the capital’, **gae** ‘at home’

When consulting other basic grammars as well, the learner soon gets entangled in the terminology when additional/alternative terms and phrases are used in the description and classification of adverbs such as ‘developed’, ‘common adverbs’ and ‘used as adverbs without becoming purely adverbs’.

The distinction between basic, derived and adopted is of special importance to the lexicographer. They are problematic in terms of decisions regarding inclusion in, or omission from, the dictionary. Although all being ‘adverbs’, it will be argued that different approaches towards inclusion versus omission should be followed:

- Only a limited number of *basic adverbs* exist in Northern Sotho and since they are all frequently used, they should all be lemmatised.
- Since the number of *derived adverbs* is unlimited/open-ended, it is not possible to lemmatise all forms separately. A strategy for inclusion versus omission has to be found. The need for such a strategy is for example clearly indicated in the case of numeral adverbs. It simply boils down to the question: If *once, twice, ... ten times* are lemmatised, why not *twenty times, hundred times*, etc. Frequency of use can be used as criterion for inclusion or omission, supported by proper description of the policy in the guidelines to the users and more elaborate patterns in the back matter.
- In the case of *adopted adverbs*, frequency could once again be used for decisions on dual POS labelling or even single versus multiple entries. For example, should nouns which are used more frequently with an adverbial than a nominal function be entered twice, or only once with a dual label, or should it be assumed that *all* nouns can be used as adverbs and thus not be labelled as adverbs? A sound application of the metalanguage could be to order the POS-labels according to the dominant function, i.e. n./adv. if the nominal function is more frequent, and adv./n. if the word is more frequently used as an adverb. This has to be clearly explained in the front matter of the dictionary.

What should be avoided is a situation where the same adverb is labelled differently in different dictionaries, or even in different editions of the same dictionary, or clearly ‘related’ adverbs labelled differently in the same dictionary. The treatment of the three nouns listed by Lombard (1985: 167) as adverbs that developed from class 6 nouns, i.e. **maabane**, **maloba** and **mantšiboa** will be considered as a case in point, and suggestions for improvement in paper and electronic dictionaries will be offered.

It will be concluded that compiling user-friendly dictionaries of a high lexicographic standard for African languages poses a great challenge to prospective lexicographers. They often are the mediators between complicated grammatical structures and the decoding and encoding needs of their target users. Adverbs should not be lemmatised haphazardly as they cross the compiler’s way. They should be carefully researched and lemmatised in a structured way. On the macrostructural level, candidates for inclusion (or omission) should carefully be



considered, preferably based on corpus data. On the microstructural level, data should be presented in such a way that it satisfies both the needs of encoding as well as decoding users.

## [To Table of Contents](#)

---

### Are the Setswana Mockery Words that Objectionable?

M.P. RAKGOKONG

*Setswana National Lexicography Unit, Mmabatho, South Africa*

---

“Hai! wena o dimpa di **matogo**, mpiletse mosimane yo o **digoro** a tle go ja **semanya** se sa bogobe se. Kana ke raya wena o ditsebe di **makgela**, kgotsa o a bo o gopotse ‘mmataago’ yo o tlhogo e **letlapirwa** yole?”

*“Hey you with a bulging stomach, call me that boy with in-curving legs to come and eat this stodgy porridge (normally without relish). I am talking to you with thick ears or are you thinking of your friend with a big head?”*

The words in bold above used to be heard in abundance among the Setswana speakers of yesterday. It was at the time when the language was still ‘pure’ – when it was not yet invaded by other languages. It was at a time Setswana was rich with idiomatic and poetic expressions. A million-dollar question to be answered is why these words are palpably marching into oblivion?

Indeed, providing an answer to this question is no easy task to perform. The main reason may be that they are used in a mockery, derogatory, humiliating or denigrating manner and thus unpalatable or offensive; they are avoided at all cost, especially in dialogue. As the avoidance goes on for a certain time, the words become tabooed and automatically call for euphemising.

Another possible reason may be that, unlike in the past, the Setswana language, like other African languages, does no longer enjoy the status it used to. This may partly be blamed on the colonial and apartheid era when only English and Afrikaans were accorded the status of ‘official languages’. Most unfortunate and ironical too, after gaining independence and democracy we became first-class citizens and proclaimed our languages official languages, but government now says “away with humanities in favour of science, technology and commerce”. In their motivational speeches politicians and government officials brazenly tell high school students that they should move away from the ‘softies’ – referring to humanities – as the government is in dire need of people who can develop a sustainable economy for the country. While one is not disinclined to agree with this move, the risk is that it is exaggerated. It strongly encourages the youth to disrespect the values, customs and norms of their communities as these are embedded in, and find expression in, language. The present situation is terrifying, and the future horrifying. Maybe it is high time that we learn Cingo’s words of wisdom in *Duminy* (1967: 137):

When a nation loses this intimate vehicle “which runs like a golden thread through the warp and hoof of its very existence, that nation will cease to exist in the exalted sense of a real nation.” A nation, then, which wishes to preserve its identity and its language heritage for posterity, and which wishes to enrich humanity with its special contribution must take steps to preserve its language.

Yes, the plain truth is that mockery words are somewhat objectionable because they are emotionally disturbing. The paper argues that whatever function these words have and whatever connotations they have, they remain words and as words they are part and parcel of the Setswana language even if we may wish them away – just as we did with swear words which finally found their way into our dictionaries. On the importance of inclusion of this type of words Rawson (1979: 11) cautions: *'keep in mind Shakespeare's advice (Henry IV, Part 2, 1600): It is needful that the most immodest word be looked up and learned'*.

The general aim of this research is to collect Setswana mockery words from the Batswana, as well as from written materials in Setswana, in order to access the range of these words currently available and to explore (via content analysis) the meanings and attitudes they convey. For the purposes of this paper a reasonable collection, good enough to serve as a sample, will be assembled. In other words, this is a pilot study. The research will be of great significance for the advocacy of the inclusion of mockery words in Setswana dictionaries.

Data will be collected randomly from men and women, aged 30 to 80, in the districts of Bafokeng and Molopo respectively, by means of a questionnaire and interviews. The results of the two districts will be compared. Where necessary a tape recorder will be used.

Existing dictionaries and manuscripts will be examined to determine the extent to which they have included the Setswana mockery words.

The words collected will be subjected to semantic and content analysis.

Recommendations, based on the findings, will be made vis-à-vis inclusion or exclusion of mockery words in Setswana dictionaries.

[To Table of Contents](#)

---

## ***Woordeboek sonder Grense: A Typological and Communicative Bridge***

**Mariza STEYN & Liezl GOUWS**

*Unit for Afrikaans, Language Centre, University of Stellenbosch, South Africa*

---

Research in pedagogical lexicography has gained momentum over the past fifty years. Recently, research on the influence of language learning and language acquisition theories, as well as the incorporation of the mother tongue in learners' dictionaries, have come under the spotlight. The dictionary practice has centred on advanced learners' dictionaries within the British lexicographic tradition. These dictionaries have been praised for their technological advances and innovative and creative presentation and structure. A good deal of criticism has also been expressed about the complexity of the presentation and the level of reference skills expected from the users of these dictionaries.

There are two important theoretical issues/problems within pedagogical lexicography that are relevant for this presentation/paper. Kernerman (2000: 922) identifies the first problem with existing learners' dictionaries: *'This will give rise to dictionary research for beginners and intermediates and a new generation of English learners' dictionaries designed specifically for lower levels'*. Present-day learners' dictionaries focus mostly on the advanced learner, at the expense of learners at pre-high school levels. Secondly, lexicographers experience problems

with the typological classification of learners' dictionaries on account of the insertion of the learners' mother tongue. An example of the typological confusion is the switching between terms like for instance 'bilingualised' and 'semi-bilingual' for hybrid learners' dictionaries.

The same problems can be identified within the South African lexicographical context. No provision has been made for beginners and intermediate learners. The typological classification of learners' dictionaries also creates problems for South African lexicographers. A new learners' dictionary attempts to address these problems. *Woordeboek sonder Grense* is written for learners from grades four to seven who have Afrikaans as a second, third or even fourth language. The dictionary aims at assisting the learners in everyday communication and usage in the classroom. *Woordeboek sonder Grense* is a hybrid dictionary that can be classified as a monolingual dictionary with a bilingual feature. Translation equivalents of the lemma are added in the example sentences and as a consequence the dictionary functions as a language bridge. Text reception is also facilitated and accelerated.

Learners using monolingual dictionaries also experience the following problem, as formulated by Atkins (1985: 21): '*Users of a monolingual L2 dictionary can access the material in it only by means of a foreign language headword. It might be just that word that they do not know*'. *Woordeboek sonder Grense* includes an equivalent register as outer text, thus helping the learner to find the lemma via an English equivalent. This outer text functions as a communicative bridge whereby learners are referred from a foreign language, English, to the object language, another foreign language, Afrikaans. The dictionary bridges the boundary between different dictionary classes, because it is primarily monolingual with one bilingual feature, namely translation equivalents.

*Woordeboek sonder Grense* differs from other learners' dictionaries on account of the following two reasons: The dictionary forms part of a series of established Afrikaans handbooks for primary school students, namely *Nuwe Afrikaans sonder Grense*. The dictionary therefore agrees in title and look with the series. This connection has major implications for the database and macrostructure of the dictionary and will be discussed in the latter part of the paper. A second advantage that goes along with the first point is the extensive knowledge available about the users. Hartmann's desideratum that the success of a dictionary depends on the product's suitability for the particular needs of the users can be realised optimally. Knowledge about the primary, secondary and reference skills of the learners also has immediate and extensive implications concerning the microstructure. These implications will be discussed and illustrated.

Like most other dictionaries, *Woordeboek sonder Grense* started with a dictionary plan consisting of five phases. During the first phase, the general preparation phase, a style guide was formulated, decisions about the frame structure were made and the database was compiled. The compilation of the database from the specific primary references ensured that relevant and familiar words were included. The general preparation was followed by the gathering and preparation of material and finally the processing of the specific material. It is especially the micro- and article structures and the micro-architecture as part of the processing phase that will be dealt with in this paper. The next phase concerns the preparation for publication.

*Woordeboek sonder Grense* functions as a bridge between different typological classes within the learners' family in order to assist the inexperienced users/learners. Being integrated within a handbook and workbook series, the dictionary can be used as an optimal communication instrument in the classroom.

---

**Dictionary Tailoring, SL Lexical Acquisition and Computer-Assisted Language Learning: The LINC Approach**

**P.H. SWANEPOEL**

*Department of Afrikaans & Theory of Literature, University of South Africa, South Africa*

---

In information design, tailoring refers to the practice of presenting information to clients in such a way that it meets their immediate needs, interests or concerns and not in some generic way, forcing them to sift through irrelevant information to get to what they are interested in. In this sense, information tailoring is not altogether foreign to the world of dictionary design, as is testified to by the various kinds of specialised dictionaries that are on the market for different languages aimed at the linguistic needs of different kinds of users. Despite these efforts, we are well aware of the problems SL or FL dictionary users encounter, for example, when consulting pedagogical dictionaries, despite some innovative redesigns of some of the most well-known ones.

In this paper I will focus on the principles that underlie the design of a minimalist dictionary, tailored to meet the demands set for SL lexical acquisition in an interactive CD-ROM SL language acquisition package. What is of broader interest is the fact that the design of the SL acquisition package, including the design of the minimalist dictionary, was based on current theories and theory-driven empirical research on SL lexical acquisition.

The architecture of the package itself is rather simple: 10 lessons, each beginning with a short video, a set of exercises, a self-reference grammar, a WWW-link to a tutor and a dictionary. In designing the dictionary, however, a myriad of possibilities existed. For example, one could have opted for any one of the existing hard-copy pedagogical dictionaries of the languages for which LINC caters, or any one of those available on-line; one could have commissioned a smaller dictionary based on vocabulary-levels, or one could have opted for monolingual or bilingual dictionaries, or for no dictionary at all and could have used annotations instead, etc.

In designing the dictionaries for LINC however, the approach adopted was to match exposure to the input materials (video and exercises) and their cognitive processing to the theoretical requirements set for lexical acquisition and to adopt the design of the dictionary to complement these processes. The result was a dictionary design that provides minimal linguistic information when the dictionary is consulted and which, unlike most other dictionaries, reduces cognitive load, leaving more cognitive resources for the lexical acquisition process itself.

The empirical question is, of course: Did the design lead to better SL lexical acquisition? In the final part of this section of the paper, I will discuss the results of the research of my partners at the University of Antwerp. I will also focus on the drawback of ‘tailoring’: personalised service is always labour-intensive and ways are needed of capitalising on existing lexical resources such as dictionaries.

Within the field of lexicography however, this project has two broader implications that I would like to stress in the rest of the paper:

- the need for theory-driven research on dictionary (re)design;

- the need for theory-driven research to support dictionary use as a sub-field of inquiry in the field of lexicography.

[To Table of Contents](#)

---

**On the Semi-automatic Extraction of Definitional Information: A Case Study for Northern Sotho**

**Elsabé TALJARD**

*Department of African Languages, University of Pretoria, Pretoria, South Africa*

---

Pearson (1998: 1) states that the use of corpora in general lexicography is a well-established practice, but that corpora have not been used for specialised lexicography, i.e. terminology, in the same way as they have been used for general-language lexicography. This is also true for the South African context, and specifically for the South African Bantu languages. A number of reasons could be cited for this state of affairs, the main one being the hitherto unavailability of suitable specialised corpora, thus denying terminologists the opportunity to base their terminological work on authentic special-field texts. Also, the perception that terms are context-independent has for a long time dominated terminological work and it is only recently that the emphasis has moved to *usage* of terms and making use of real texts as a primary source of data. It is furthermore generally accepted that the input of special-field experts is indispensable in the identification and definition of terms. Unfortunately, there seems to be a lack of commitment on the part of special-field experts who are mother-tongue speakers of the South African Bantu languages to develop terminology in these languages. South African terminologists therefore have no option but to investigate other avenues to overcome these obstacles.

The increasing availability of special-field texts, many of them in electronic format, enables terminologists to build their own corpora for special purposes. Access to user-friendly and affordable software such as *WordSmith Tools* further opens the door for terminologists to base their work on authentic special-field texts. It has already been illustrated by Taljard & De Schryver (2002) that it is indeed possible to extract terms semi-automatically from Bantu-language corpora composed of special-field texts, thus reducing (but of course not eliminating) the dependence of the terminologist on the co-operation of the subject-field specialist. The next logical step in this process would therefore be to investigate the possibility of not only extracting terms from special-purpose corpora, but to also extract definitional information semi-automatically from these corpora.

Three issues are addressed in this paper. In the first instance, Pearson (1998: 5) states that authors writing within certain specified communicative settings are likely to provide explanations of at least some of the terms they use. This hypothesis is tested with regard to Northern Sotho, using linguistic texts as authentic data. Secondly, the possibility of extracting definitional information in a semi-automatic way from these texts is investigated. For this purpose, 50 linguistics terms have been identified, the main aim being to retrieve definitional information on as many of these terms as possible. It has to be borne in mind that the texts which are currently available are unmarked and untagged, thus restricting the scope of the study to the identification of mainly lexical and orthographical markers as indicators of definitional information. A third aspect which is investigated is to ascertain whether or not

there is a connection between the strategy being used for identification of definitional information on the one hand, and the kind of information being provided in the text on the other.

The aim of this paper is therefore to indicate that a corpus-based approach to terminology is not only a possibility for the South African Bantu languages, but indeed an imperative and that terminologists stand much to gain in making use of such an approach.

[To Table of Contents](#)

---

## **Language Variation and the Lexicographer**

**Dirk J. VAN SCHALKWYK**

*Bureau of the Woordeboek van die Afrikaanse Taal (WAT), Stellenbosch, South Africa*

---

### **Variation in language**

There is no language community where all speakers exhibit the same linguistic behaviour. No two individuals in the same language community exhibit identical linguistic behaviour. Each individual has his own so-called idiolect.

The variation in the idiolects of the speakers of a language is sometimes minimal. As soon as this variation is larger in scope, varieties of a language and even dialects develop.

### **Varieties and dialects**

A variety of a language consists, just as a dialect does, of the sum of the idiolects of all speakers who speak the variety or dialect.

The distinction between a variety and a dialect is to a certain degree artificial, as it is at the very least difficult, probably impossible, to distinguish between a variety and a dialect. Since a pejorative value is often given to the term 'dialect', the term 'variety' is used in this paper.

Different varieties may be the result of geographical, historical or social factors.

### **The extent of language variation**

Variation may affect all aspects of language, e.g. phonology, morphology, syntax and the lexicon. In this paper, however, only the lexicon will be discussed.

### **Varieties of Afrikaans**

Afrikaans developed from Dutch and kept its Dutch character in several respects. Many different forces impacted upon Dutch at the Cape since the time Jan van Riebeeck established his way station in 1652 at the Cape of Good Hope. The result was a new language, i.e. Afrikaans.

South Africa is a vast country and the influences on Dutch and later on Afrikaans did not apply everywhere or not everywhere to the same degree. As a result, different varieties of Afrikaans developed, e.g. Cape Afrikaans, Malay Afrikaans, Afrikaans as spoken in the Swartland, the Boland, in Namaqualand and Boesmanland, and Afrikaans as spoken by the Griquas and the people from Rehoboth in Namibia.

### **Language variation and dictionaries**



The degree to which variation is included in the macrostructure of a dictionary depends on the kind of dictionary being compiled.

Standard dictionaries, school dictionaries and multilingual dictionaries do not reflect variation fully, but comprehensive dictionaries such as the *Woordeboek van die Afrikaanse Taal* (WAT) cannot evade this responsibility.

In order to report on lexical variation in a language in a dictionary, many requirements must be met.

### **The demands language variation in Afrikaans make on the Bureau of the WAT**

It falls within the brief of a comprehensive dictionary to report exhaustively in a lexicographical manner on the specific language. It makes stiff demands on a lexicographic project like the Bureau of the WAT, because:

- both the written and the spoken form of the Afrikaans language must be recorded, and
- in addition to the standard variety of Afrikaans, all other varieties must be recorded.

The data collection policy of a comprehensive dictionary project must be directed in such a way that the collected data meets the requirement of comprehensiveness as far as possible.

In order to carry out such a policy in the best possible way, a reliable data collection network must be established.

### **Dealing with lexical variation in the *Woordeboek van die Afrikaanse Taal***

Lexical variation in the WAT is dealt with in the form of variants (*wisselvorme*). Variants are a mention or reference to one or more equivalent or non-equivalent lexical items that are variants of the specific lemma with respect to pronunciation and spelling.

Variants are equivalent if they are as commonly or almost as commonly used in a language and non-equivalent when one or more variant(s) is/are more common than any other(s).

When mentioning a variant, a formula with “Ook” is used if the variant is equivalent, followed by an italic entry: **bosgasie** Ook *boskasie* (a wild bush of hair). In the case of non-equivalent variants, formulas such as “Ook soms”, “Selde ook” or variations thereof are used, followed by the variant in italics. In the case of a reference, a “Sien” formula is used, followed by an entry in small capitals: Sien BOSGASIE. The fact that the user is referred to **bosgasie** at **boskasie**, shows the user that **bosgasie** is used more often than **boskasie** and that further information, e.g. meaning, etc., can be found there.

As **boskasie** has not been provided with a label that shows that it is seldom used, the user may deduce that both of the variants are used in the standard variety of Afrikaans.

Sometimes there may be many variants. At the lemma **kasaterwater**, for example, 39 forms have been listed. If the details given at **kasaterwater** and at all variants are carefully studied, the user not only receives information on how often all variants are used, but also on the geographical area(s) where they are used and the variety (-ies) within which they occur.

### **Lexicographic dilemmas caused by variation**

As long as no dialectal dictionaries or dictionaries describing a specific variety have been compiled and no extensive and carefully compiled databases exist, it is almost impossible to record all variants within a language accurately in order to determine precisely where a certain variant is used, whether it is commonly used and whether it belongs to the standard variety or another variety.

## Correspondence

>

### **AFRILEX**

*African Association for Lexicography*

Department of African Languages

University of Pretoria

Pretoria, 0002

South Africa

T: + 27 12 420 2320

F: + 27 12 420 3163

E: [prinsloo@postino.up.ac.za](mailto:prinsloo@postino.up.ac.za) (Chairperson: Prof. D.J. Prinsloo)

I:

<http://www.up.ac.za/academic/libarts/afrilang/homelex.html>

(Home Page AFRILEX)

>

Mr. Thierry **Afane Otsaga**

24 Weidenhof Street

Weidenhof Lodge

Stellenbosch, 7599

South Africa

T: (021) 886 7161 (H) | 082 798 7467 (Cell)

E: [thiafane@yahoo.fr](mailto:thiafane@yahoo.fr) | [afane@lantic.net](mailto:afane@lantic.net)

I: [www.chez.com/afanotsaga](http://www.chez.com/afanotsaga)

>

Dr. Mariëtta **Alberts**

Manager: Lexicography and Terminology Development

*Pan South African Language Board (PanSALB)*

Private Bag X08

Arcadia, 0007

South Africa

T: +27 (0)12 341 9638 | 083 306 9924 (Cell)

F: +27 (0)12 341 5938

E-mail: [marietta@pansalb.org.za](mailto:marietta@pansalb.org.za)

>

Mr. Herman L. **Beyer**

Department of Germanic & Romance Languages

University of Namibia

Private Bag 13301

Windhoek

Namibia

T: (+264 61) 206 3850  
F: (+264 61) 206 3863  
E: hbeyer@unam.na

>

Dr. Nerina **Bosman**  
Departement Afrikaans  
Mamelodikampus  
Universiteit Vista  
Privaatsak X1311  
Silverton, 0127  
South Africa

T: 012 842 3500  
F: 012 842 3633

>

Mr. Emmanuel **Chabata**  
*African Languages Research Institute (ALRI)*  
University of Zimbabwe  
PO Box 167  
Mount Pleasant  
Harare  
Zimbabwe

T: (263-4) 303298 (W)  
F: (263-4) 333674 | (263-4) 333407  
E: echabata@arts.uz.ac.zw

>

Mr. Gilles-Maurice **de Schryver**  
Residentie Wellington  
F. Rooseveltlaan 381  
B-9000 Gent  
Belgium

E: gillesmaurice.deschryver@UGent.be |  
schryver@postino.up.ac.za

I:

<http://www.up.ac.za/academic/libarts/afrilang/elcforall.htm>

>

Prof. James D. **Emejulu**  
Directeur: *Groupe de Recherche en Langues et Cultures*  
*Orales (GRELACO)*  
Faculté des Lettres et Sciences Humaines  
Université Omar Bongo  
BP 20241  
Libreville  
Gabon

T: +241 73 16 42  
F: +241 76 00 95  
E: jimacs@eudoramail.com | jdemejulu@compaqnet.fr

>

Ms. Gwyneth **Fox**  
Macmillan Education: Publisher, Dictionaries  
Macmillan Oxford  
Between Towns Road  
Oxford OX4 3PP  
United Kingdom

T: +44 (0)1865 405 700  
F: +44 (0)121 242 2279  
Direct line: +44 (0)121 454 3980  
E: g.fox@macmillan.com  
I: [www.macmillandictionary.com](http://www.macmillandictionary.com)

>

Prof. Rachéle **Gauton**  
Department of African Languages  
University of Pretoria  
Pretoria, 0002  
South Africa

T: (012) 420 3715 (W) | (012) 361 3355 (H)  
F: (012) 420 3163  
E: [rgauton@postino.up.ac.za](mailto:rgauton@postino.up.ac.za)

>

Ms. Liezl **Gouws**  
University of Stellenbosch  
South Africa

T: 0721322655 (Cell)  
F: 021 8083815  
E: 13112635@sun.ac.za

>

Prof. Rufus H. **Gouws**  
Department of Afrikaans and Dutch  
University of Stellenbosch  
Private Bag X1  
Matieland, 7602  
South Africa

T: 021 808 2164  
F: 021 808 3815  
E: rhg@akad.sun.ac.za

>

Prof. Wilfrid H.G. **Haacke**  
Department of African Languages  
University of Namibia  
Private Bag 13301  
Windhoek  
Namibia

T: +264 61 206 3845  
F: +264 61 206 3806  
E: [whaacke@unam.na](mailto:whaacke@unam.na)

>

Mr. Samukele **Hadebe**  
Department of African Languages and Literature  
University of Zimbabwe  
PO Box 167  
Mount Pleasant  
Harare  
Zimbabwe

E: [samukeleh@yahoo.co.uk](mailto:samukeleh@yahoo.co.uk)

>

Dr. Ulrich **Heid**  
*Institut für maschinelle Sprachverarbeitung –*  
*Computerlinguistik (IMS-CL)*  
Universität Stuttgart  
Azenbergstrasse 12  
D - 70 174 Stuttgart  
Germany

T: +49 711-121-1373  
F: +49 711-121-1366  
E: [uli@ims.uni-stuttgart.de](mailto:uli@ims.uni-stuttgart.de)

>

Dr. D. Franck **Idiata**  
*Groupe de Recherche en Langues et Cultures Orales*  
*(GRELACO)*  
Faculté des Lettres et Sciences Humaines  
Université Omar Bongo  
BP 9985  
Libreville  
Gabon

T: +241 76 07 84 / 73 28 02 (W) | +241 73 59 08 (H) |  
+241 61 15 24 (Cell)  
F: +241 76 39 09  
E: [idiata@yahoo.fr](mailto:idiata@yahoo.fr)

>

Ms. Kathy **Kavanagh**  
Executive Director, *Dictionary Unit for South African  
English* (DSAE)  
Rhodes University  
PO Box 94  
Grahamstown, 6140  
South Africa

T/F: 046 603 8107  
E: k.kavanagh@ru.ac.za

>

Ms. Matlakala **Kganyago**  
Nkoshilo High School  
South Africa

>

Mr. Langa **Khumalo** (FCCS)  
Head: *Ndebele Lexicography Unit*  
*African Languages Research Institute* (ALRI)  
University of Zimbabwe  
PO Box 167  
Mount Pleasant  
Harare  
Zimbabwe

T: (263-4) 303298 (W)  
F: (263-4) 333674 | (263-4) 333407  
E: langa@arts.uz.ac.zw

>

Dr. John M. **Lubinda**  
Department of French  
University of Botswana  
Private Bag 00703  
Gaborone  
Botswana

T: (267) 3552975  
F: (267) 585098  
E: lubindaj@mopipi.ub.bw

>

Ms. Matete **Madiba**  
Senior Academic Development Practitioner  
Department: Teaching and Learning Development  
Technikon Northern Gauteng  
Private Bag X07  
Pretoria North, 0116  
South Africa



T: 012 799 9293 | 0823970191 (Cell)  
F: 012 799 9167  
E: madiba.m@tng.ac.za | matetenti@hotmail.com

>

Mr. Mandlenkosi **Maphosa**  
*African Languages Research Institute (ALRI)*  
University of Zimbabwe  
PO Box 167  
Mount Pleasant  
Harare  
Zimbabwe

T: (263-4) 303298 (W)  
F: (263-4) 333674 | (263-4) 333407  
E: mandlamaphosa@arts.uz.ac.zw

>

Mr. Webster **Mavhu**  
*African Languages Research Institute (ALRI)*  
University of Zimbabwe  
PO Box 167  
Mount Pleasant  
Harare  
Zimbabwe

T: (263-4) 303298  
F: (263-4) 333674 | (263-4) 333407  
E: webma@arts.uz.ac.zw | vhez2000@yahoo.com

>

Mr. Gift **Mheta**  
*African Languages Research Institute (ALRI)*  
University of Zimbabwe  
PO Box 167  
Mount Pleasant  
Harare  
Zimbabwe

T: (263-4) 303298 (W) | (263-4) 573294 (H) | 091 917  
782 (Cell)  
F: (263-4) 333674 | (263-4) 333407  
E: gmheta@yahoo.com

>

Ms. M.P. **Mogodi**  
*Sesotho sa Leboa National Lexicography Unit*  
Pretoria Branch  
Department of African Languages  
University of Pretoria  
Pretoria, 0002

South Africa

T: (012) 420 3076 (W) | 083 751 8466 (Cell)

F: (012) 420 3163

E: pmogodi@postino.up.ac.za

>

Ms. Linkie **Mohlala**

Department of African Languages

University of Pretoria

Pretoria, 0002

South Africa

T: (012) 420 3076 (W) | (012) 803 3911 (H) | 083 596

2329 (Cell)

F: (012) 420 3163

E: s9514561@mx1.up.ac.za

>

Ms. Lorna **Mphahlele**

Department: Teaching and Learning Development

Technikon Northern Gauteng

Private Bag X07

Pretoria North, 0116

South Africa

E: mphahlele.1@tng.ac.za

>

Ms. Nomalanga **Mpofu**

*African Languages Research Institute (ALRI)*

University of Zimbabwe

PO Box 167

Mount Pleasant

Harare

Zimbabwe

T: (263-4) 303298 | 263 11 806019 (Cell)

F: (263-4) 333674 | (263-4) 333407

E: nmpofu@arts.uz.ac.zw | nomalanm@yahoo.com

>

Mr. Cornelias **Ncube**

*African Languages Research Institute (ALRI)*

University of Zimbabwe

PO Box 167

Mount Pleasant

Harare

Zimbabwe

T: (263-4) 303298

F: (263-4) 333674 | (263-4) 333407  
E: corneliasbncube@yahoo.co.uk

>

**Ms. Salmina Nong**  
Department of African Languages  
University of Pretoria  
Pretoria, 0002  
South Africa

T: (012) 420 3076 (W) | (012) 998 3015 (H) | 082 343  
9723 (Cell)  
F: (012) 420 3163  
E: snong@postino.up.ac.za

>

**Dr. Yolande Nzang-Bie**  
*Groupe de Recherche en Langues et Cultures Orales*  
(GRELACO)  
Faculté des Lettres et Sciences Humaines  
Université Omar Bongo  
BP 20241  
Libreville  
Gabon

T: +241 73 16 42 / 73 28 02 (W) | +241 72 68 50 (H) |  
+241 36 12 07 (Cell)  
F: +241 76 39 09  
E: yolnzang@yahoo.fr

>

**Dr. Pierre Ondo-Mebiame**  
*Groupe de Recherche en Langues et Cultures Orales*  
(GRELACO)  
Faculté des Lettres et Sciences Humaines  
Université Omar Bongo  
BP 20241  
Libreville  
Gabon

T: +241 73 16 42  
F: +241 76 00 95

>

**Mr. Thapelo J. Otlogetswe**  
Department of English  
University of Botswana  
Private Bag 0022  
Gaborone  
Botswana

T: (+267) 355-5081 (W) | (+267) 71959452 (Cell)

F: (+267) 3585-098

E: otlogets@mopipi.ub.bw

>

**Dr. Annél Otto**

Department of Afrikaans

Vista University

Private Bag X613

Port Elizabeth, 6000

South Africa

T: 041-4083172 (W) | 041-3608141 (H)

E: otto-an@pelican.vista.ac.za

>

**Prof. D.J. Prinsloo**

Department of African Languages

University of Pretoria

Pretoria, 0002

South Africa

T: (012) 420 2320

F: (012) 420 3163

E: prinsloo@postino.up.ac.za

I:

<http://www.up.ac.za/academic/libarts/afrilang/elcforall.htm>

>

**Mr. M.P. Rakgokong**

*Setswana National Lexicography Unit*

Kgetsana ya Sephiri 2046

Mmabatho, 2735

South Africa

E: c/o O.J. Mokakale: mokakaleoj@uniwest.ac.za

>

**Ms. Mariza Steyn**

Unit for Afrikaans

Language Centre

University of Stellenbosch

South Africa

T: 021 8082905 | 0827799955 (Cell)

F: 021 8083815

E: ms@sun.ac.za

>

**Prof. P.H. Swanepoel**

Department of Afrikaans & Theory of Literature

University of South Africa

PO Box 392  
Pretoria, 0003  
South Africa

E: swaneph@unisa.ac.za

>  
Dr. Elsabé **Taljard**  
Department of African Languages  
University of Pretoria  
Pretoria, 0002  
South Africa

T: (012) 420 2494 (W) | (012) 332 1357 (H) | 082 353  
6906 (Cell)  
F: (012) 420 3163  
E: etaljard@postino.up.ac.za

>  
Dr. Dirk J. **van Schalkwyk**  
Editor-in-Chief, Bureau of the *Woordeboek van die  
Afrikaanse Taal* (WAT)  
PO Box 245  
Stellenbosch, 7599  
South Africa

T: 021 887 3113  
F: 021 883 9492  
E: [wat@wat.sun.ac.za](mailto:wat@wat.sun.ac.za)  
I: <http://www.sun.ac.za/wat/index.htm> (Home Page  
WAT)

Back to [HOME](#)