AFRILEX

7th International Conference of the
African Association for Lexicography

AFRILEX 2002
Culture and Dictionaries
Programme & Abstracts

Dates:	8-10 July 2002
Hosts:	Dictionary Unit for South African English, Rhodes University, Grahamstown, South Africa (Kathryn Kavanagh, Dotty Mantzel, Madeleine Wright, Jill Wolvaardt)
Venue:	Eden Grove, Rhodes University
Exhibitors:	Macmillan, Maskew Miller Longman, Oxford University Press, Pharos
Abstract Reviewers:	Mariëtta Alberts, Sonja E. Bosch, Rachélle Gauton, Rufus H. Gouws, Laura Löfberg, D.J. Prinsloo, Elsabé Taljard
Programme Committee:	Gilles-Maurice de Schryver, Kathryn Kavanagh, D.J. Prinsloo, Jill Wolvaardt

edited by
Gilles-Maurice de Schryver

Welcome Dear Conference Delegate!

This booklet comprises the programme and abstracts of the papers scheduled to be presented at the 7th International Conference of the *African Association for Lexicography* (AFRILEX).

This is the first time we at AFRILEX compile such a collection, and we hope that it will prove to be useful. This first AFRILEX Programme & Abstracts booklet comes at a good time indeed, as our annual Conference has never been this popular. No less than 49 presenters will cover 40 papers. At the time of writing, already over a hundred attendees confirmed their registration. Especially strong delegations from Gabon and Zimbabwe will be welcomed, as well as presenters from as far away as Tanzania, Belgium, Germany, the United Kingdom, Denmark, Sweden, Finland and Hong Kong.

A record-number of languages will also be covered, ranging from Zulu (isiZulu), Ndebele (isiNdebele), Swati (SiSwati), Northern Sotho (Sepedi, Sesotho sa Leboa), Southern Sotho (Sesotho) and Tswana (Setswana), to Shona (ChiShona), Zimbabwean Ndebele, Lunyole, Swahili (Kiswahili), the Gabonese languages, and finally to German, Dutch, English, French, Portuguese and Chinese. The conference theme, *Culture and Dictionaries*, is given the attention it deserves, and metalexicographic presentations are also well-balanced. Each day of the three-day conference will begin with a keynote address. Together with the regular parallel sessions and the two special sessions (one on dictionary funding and one on morphological analysers), attendees should have ample options to choose from.

The Conference Hosts invite us to visit their *Dictionary Unit for South African English* (DSAE), and might convince us to see more of Grahamstown (Egazini Tour). Finally, the Vice-Chancellor of Rhodes University will entertain us with a cocktail party, while *Pharos* will once again contribute generously towards the Conference Dinner.

No doubt, we might be heading for our most successful conference so far. Enjoy!

Pretoria, June 2002

Gilles-Maurice de Schryver
Organiser: AFRILEX.

Table of Contents

Programme

Keynote addresses
- Henning **Bergenholtz** — Towards a User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography
- A.C. **Nkabinde** — Cultural Implications on Lexicography

Special Session 1: Fundraising for Dictionary Publishing

- Mariëtta **Alberts** — Funding of Technical Dictionaries
- Dirk J. **van Schalkwyk** — Financing of the *National Lexicography Units* (NLUs) in a New Lexicographic Dispensation
- Jill **Wolvaardt** — Dictionaries, A Cultural Investment? Some Thoughts on Fundraising for Dictionary Projects

Special Session 2: Morphological Analysers for the Bantu Languages
- Arvi **Hurskainen** (**Keynote address**) — New Advances in Corpus-based Lexicography
- Sonja E. **Bosch** & Laurette **Pretorius** — Using Finite-State Computational Morphology to Enhance a Machine-Readable Lexicon
- Gilles-Maurice **de Schryver** — First Steps in the Finite-State Morphological Analysis of Northern Sotho
- Mandlenkosi **Maphosa** — Word Division and Orthography as Some of the Factors Posing Challenges in the Development of the Ndebele Grammatical Parser
- Daniel **Ridings** & Webster **Mavhu** — Problems and Challenges Encountered when Developing a Morphological Parser for the Shona Language
- Justus C. **Roux** & Sonja E. **Bosch** — Human Language Technologies and the *National Lexicography Units*

Parallel Sessions
- Emmanuel **Chabata** — Language Development in a Multilingual Society
- Emmanuel **Chabata** & Webster **Mavhu** — Torn Between Calling a Spade a Spade and Being Euphemistic: The Dilemma of a Lexicographer Defining Offensive Headwords in Shona Monolingual Lexicography
- Rachélle **Gauton** — Bilingual Zulu Dictionaries and the Translation of Culture
- Rufus H. **Gouws** — Using a Frame Structure to Accommodate Cultural Data
- Karen **Hendriks** — The Treatment of Culture-Specific Lexical Items in Bilingual Dictionaries
- Gregory **James** & Bronson **So** Ming Cheung — Culture and a Dictionary: Evidence from the First European Lexicographical Work in China
- Kathryn **Kavanagh**, Philisiwe **Manyisa**, Disebo **Moeti** & Neo **Mpalami** — Work in Progress: First Steps in Using a Bilingual Dictionary Framework
- Langa **Khumalo** — History, Language Contact and Lexical Change: A Lexicography/Terminography Interface in Zimbabwean Ndebele
- Diapo N. **Lekganyane** — On Corpora and the Process of Selecting High Function Words and Their Treatment in Currently Available Dictionaries for Sepedi
- Matete **Madiba**, Lorna **Mphahlele** & Matlakala **Kganyago** — Capturing Cultural Glossaries: A Case Study Presentation of N. Sotho Cooking Terms
- P.S. **Malebe** & Gilles-Maurice **de Schryver** — From Corpus Data to the First isiNdebele Dictionaries
- Esau **Mangoya** — Divergent Approaches to Corpus Processing: The Need for Standardisation
- Godfrey Baile **Mareme** — Prejudice and Reality in a Setswana Monolingual Dictionary – The Systematic and Deliberate Biasing of Cultural Issues
- P.A. **Mavoungou**, T. **Afane Otsaga** & G.-R. **Mihindou** — Dictionaries Compiled with French and the Reproduction of the Gabonese Cultures

- [Ron **Moe**](#) — Compiling Dictionaries Using Semantic Domains
- [M.P. **Mogodi**](#) — Layered Definitions: A Northern Sotho Case Study
- [V.M. **Mojela**](#) — The Etymological Aspects of the Idiomatic and Proverbial Expressions in the Lexicographic Development of Sesotho sa Leboa – A Semantic Analysis
- [L.E. **Mphasha**](#) — Cultural Relativism in Dictionaries: Is this the Right Direction?
- [Nomalanga **Mpofu**](#) — Cultural Aspects in the Shona Monolingual Dictionary: *Duramazwi Guru reChiShona*
- [Vezumuzi kaDayisa **Ndlovu**](#) — A Corpus-based Approach to Terminography: An Analysis and Evaluation of the ALRI Corpora
- [Salmina **Nong**](#) — The Lexicographic Treatment of Loan Words in Northern Sotho
- [D.J. **Prinsloo**](#) — The Lemmatization of Copulatives in Northern Sotho
- [Peter A. **Schmitt**](#) — Culture-specific Concepts in Technical Dictionaries
- [Penny **Silva**](#) — The OED as Cultural Icon
- [Elsabé **Taljard** & Gilles-Maurice **de Schryver**](#) — Semi-Automatic Term Extraction for the African Languages, with special reference to Northern Sotho
- [Pieter N. **van der Westhuizen**](#) — A Southern African Lexicographer's Working Definition of the Term Culture
- [Linda **Van Huyssteen**](#) — The Value of Culture in the Development of Medical Corpora in Zulu
- [Jacques A.J. **Van Keymeulen**](#) — Making a Dictionary of an Oral Tradition
- [Advice **Viriri**](#) — Missionary Influence on Shona Lexicography, with Special Reference to Father Hannan's Problem of Translation

[Correspondence](#)

Programme

AFRILEX 2002

40 papers
49 presenters

[To programme](#)

KEYNOTE ADDRESS (1)

Towards a User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography

Henning BERGENHOLTZ
Centre for Lexicography, The Aarhus School of Business, Denmark

There is much uncertainty and confusion as to the real differences between prescription and description. Is introspection part of the empirical basis, i.e. a part of a descriptive process? Or are introspective verdicts always a part of the prescriptive process? Both conceptions are expressed in existing linguistic dictionaries. It is also uncertain whether prescription must always contain statements which differ from descriptive statements. Finally, it is uncertain whether you can distinguish clearly between description and prescription, as several levels of descriptive accuracy are pointed out. You could presume the existence of a transitional zone between a descriptively low accuracy and the use of a very small empirical basis respectively, and a prescription without major differences to actual usage. In conclusion, you may say that admittedly the dispute between *usus tyrannus* and *usus imperans* has lasted for at least 300 years, but it is still of current interest. Is usage a tyrant or is it the ruler?

This uncertainty has carried on to lexicography where there is much confusion as to the real differences between prescriptive and descriptive dictionaries. In general, the majority of existing accounts can be summed up to this: Descriptive relates to the empirical basis; accordance between the empirical data and the dictionary is required. Prescriptive relates to the genuine purpose of the dictionary; the dictionary is meant to help with problems concerning text production and will thus affect usage. This asymmetrical understanding would imply prescriptive and descriptive in practise being false contrasts. This is also related to the paradox which Wiegand (1986) has pointed out: The statements of a descriptive dictionary have a prescriptive effect on the users. Or in other words: When maintaining a certain usage in a dictionary this description obtains an oracular status. The descriptive dictionary also has an effect on usage, often a conversational one. The oracular status actually corresponds with the expectations of the dictionary user in the event that he or she has a problem which relates to text production. The user has a problem and seeks help with a specialist. Naturally, he or she will trust the statements given by the dictionary unreservedly. This applies both to descriptive and prescriptive dictionaries.

However, this does not mean that I will argue in favour of abandoning the distinction between descriptive and prescriptive dictionaries in my lecture. On the contrary, I wish to suggest a specification and the introduction of a new term, proscription, which in actual fact is only new as a term, since the phenomenon itself is known in many dictionaries around the world. What is meant is the suggested use of a certain variant based on an exact analysis of an empirical basis without prohibiting other existing variants. Coincident with this, a specification of both the new term and the two hitherto used terms will be suggested which will allow for both the nature and the use of the empirical basis:

(1) introspection,
(2) analysis of a linguistic survey,
(3) the involvement of descriptions in existing dictionaries, grammars, monographs, articles, etc.,
(4) the analysis of a number of examples which have been randomly chosen from random texts (corresponding with the practice of dictionary making before the age of computers),
(5) the analysis of a specifically constructed text corpus,
(6) the analysis of usage found in texts in the examined language in all available websites on the Internet.

Furthermore, you must allow for the nature of usage recommendations:

(1) a specific linguistic variant is explicitly prohibited,

(2) one or more linguistic variants are explicitly prescribed, thus prohibiting all other non-mentioned variants,

(3) a specific linguistic variant is explicitly prescribed; as opposed to prescription (2) this involves a new word, new spelling, new pronunciation, a new inflexion or a neologism, cf. Wiegand (1996).

Here, I suggest a more consistent terminology which allows for both the function of the dictionary and the relation of the dictionary to the empirical basis:

	empirical basis	accordance with empirical basis	wishes to influence the user
descriptive dictionary	+	+	−
proscriptive dictionary	+	+	+
prescriptive dictionary	±	±	+

KEYNOTE ADDRESS (2)

Cultural Implications on Lexicography

A.C. NKABINDE

A speech community's origin, history, mythology, exploits, legendary, wisdom lore and world view are reflected in its language. Similarly the arts, crafts, and other activities together with phenomena in nature and the environment are expressed or described by means of language. Language is inextricably bound with the culture of a people. Language captures the changes and developments that occur in society from generation to generation.

Linguistics offers a useful tool for analysing the structure, function and meaning of words in a language. It, however, does not always provide the necessary background to the meanings of words, particularly in unwritten languages where historical linguistics is based on hypotheses rather than factual evidence.

The challenge confronting the lexicographer is how to deal with cultural material in an organised and consistent manner in the compilation of a dictionary. She/he must walk a tight rope of defining words without straying into other fields of knowledge such as ethnography, sociology, medicine, science, anthropology, etc. in which she/he has no training nor expertise.

This article attempts to identify some of the problem areas in the accommodation of culture in lexicography and raises some questions or makes tentative proposals on how to deal with problems. These are:

- "standard" and "non-standard" variations of a language including lexical variations of words in a language,
- use of a corpus in unwritten languages or languages with a limited written tradition,
- figurative use of language,
- different kinds of dictionaries,
- euphemisms, taboo and **"hlonipha"**,
- tone dialects,
- translation and monolingual dictionaries,
- concepts and functional mobility of a language,

- socio-economic, political, and historical influences on language.

SPECIAL SESSION 1: **Fundraising for Dictionary Publishing** (1)

Funding of Technical Dictionaries

Mariëtta ALBERTS
Department of Arts, Culture Science and Technology (DACST), South Africa

Already in the early fifties, the Government started with the funding of terminology projects and with the publication of the technical dictionaries that resulted from these endeavours. At that stage the focus was on the compilation of English/Afrikaans technical dictionaries because of the bilingual policy of the then government. The terminology projects were funded in the sense that terminologists were paid salaries to compile these dictionaries. The terminologists who were employed by the *Department of National Education*, the forerunner of the present *Department of Arts, Culture Science and Technology* (DACST), mainly did the terminology work. The Government Printer published all the technical dictionaries compiled by the Terminology Section.

Other Government departments such as the then *Department for Defence*, the *South African Railways and Harbours* and the *Department of Education* (to mention just a few) were also involved in the compilation of technical dictionaries in their special fields of interest. They all published their respective dictionaries on their own.

Bodies such as the *Suid-Afrikaanse Akademie vir Wetenskap en Kuns* (SAAWK, 'South African Academy for Science and Art'), the *Afrikaanse Taal- en Kultuurvereniging* (ATKV, 'Afrikaans Language and Cultural Association'), municipalities, etc., also devoted time to terminology work and published their respective dictionaries. They all employed language practitioners to compile these dictionaries.

Besides the Government and the above-mentioned bodies, publishers also typically commission subject specialists to compile dictionaries on various subject areas. In these cases the dictionary makers do not receive any funding whilst compiling. Instead, royalties are paid to such compilers of technical dictionaries.

Since 1994 the Government devotes time to the provision of African-language term equivalents in a variety of subject areas. The Terminology Co-ordination Section of the *National Language Service* compiles terminology lists in the eleven official languages. The terminologists utilise the MultiTerm program of TRADOS to capture terminological information. Draft terminology lists can be printed directly from the MultiTerm program for distribution to collaborators, PanSALB structures (*National Language Bodies* (NLBs), *National Lexicography Units* (NLUs), *Provincial Language Committees* (PLCs)), etc. Once the Terminology Co-ordination Section has received feedback from the various collaborators, the terminological data can be finalised in the Termbank and the multilingual terms can be disseminated to the language users, the subject specialists and the National Lexicography Units (NLUs) for inclusion in their wordbanks.

Since the multilingual polythematic terminologies form part of the Termbank, they will also be available on the Internet in the future. The terminological data can also be made

available and disseminated on CD-ROM. Whether the terminological data will also be published in traditional dictionaries remains to be seen.

In the past there were always private initiatives where an individual felt the need for a dictionary on a specific subject area. These people would then compile such a term list or ask someone else to do it on their behalf and publish it on their own. The problem with this kind of work is that the compiler has to work on a term list without earning a salary. The compiler will only receive some form of benefit from his/her endeavour once the technical dictionary has been published. If one takes royalties compared to the amount of work involved in the compilation of a dictionary into account, it is really not worth the while to compile a dictionary in private capacity.

SPECIAL SESSION 1: **Fundraising for Dictionary Publishing** (2)

Financing of the *National Lexicography Units* (NLUs) in a New Lexicographic Dispensation

Dirk J. VAN SCHALKWYK
Bureau of the Woordeboek van die Afrikaanse Taal (WAT), South Africa

The history of lexicographic projects throughout the world shows that they always have too little money and therefore too few staff members to finish their assigned task effectively and within a reasonable length of time. The history of the *Woordenboek der Nederlandsche Taal* is a good example. The *National Lexicography Units* (NLUs) for the official languages of South Africa will not be able to avoid these problems.

The NLUs are financed by government. These government funds are channelled from the *Department of Arts, Culture, Science and Technology* (DACST) to the *Pan South African Language Board* (PanSALB) for allocation to the NLUs. Owing to Government's numerous responsibilities these funds will not provide in all needs of the NLUs. Therefore the responsibility lies with the NLUs to become involved in fund-raising.

The NLUs can obtain funds by generating funds and/or by fund-raising. The units can generate funds with lexicographic products and services. Dictionaries written with the needs of the dictionary users in mind, will be good marketing tools. A language query service, translation service, language editing service or training on how to use dictionaries are examples of possible services. Funds can be raised at provincial authorities, municipalities or town councils, at the tertiary institutions where units are established, but also in the private sector. The users of a specific language ought to be the best possible donors for the language, as would businesses and companies to which individuals who are sensitive to language are affiliated. Certain trust funds earmark their funds for language and language development. They are good potential donors for the NLUs.

It is the responsibility of the Editor-in-Chief or Executive Director to generate and raise funds. The question that arises is whether the Editor-in-Chief or Executive Director can take on the comprehensive and specialized task of the generation of funds and fund-raising in addition to his or her lexicographic responsibilities and functions arising from the performance areas of a lexicographic unit.

According to the *Articles of Association* of the National Lexicography Units the functions of the Editor-in-Chief may be reduced to managing the unit and reporting to the Board of Directors. When these functions are analysed carefully, however, it is clear that several focus areas are relevant. It includes the whole process of dictionary making from needs analysis, building a database, determination of the macro- and microstructure of the dictionary, the development of a style guide and its computerisation, lexicographic processing and editing of the data, typesetting, printing and binding of the lexicographic products, and the marketing of the products. Yet the finances and staff members of the unit, physical facilities, research, editorial and administrative support services, etc., are also his/her concern.

The establishment of a trust for each of the National Lexicography Units should be considered to help generate and raise funds. This will lighten the load of the Editor-in-Chief considerably, providing the Trust has its own staff members. If not, the Editor-in-Chief will in effect obtain an extra job.

The Pan South African Language Board is aware of the financial situation of the National Lexicography Units. Therefore the *Subcommittee: Lexicography and Terminology Development* has established an Ad hoc committee for Fund-raising. The aim of this Ad hoc committee is to raise funds for the NLUs.

In order to ensure a satisfactory financial dispensation for the NLUs, it is important that the financial responsibilities of Government, of PanSALB, as well as of the NLUs and their trusts are properly identified and synchronised.

PanSALB and the National Lexicography Units must know in time every year what funding they will receive from government and PanSALB and the trusts must agree how fund-raising will take place and how potential donors will be treated. This will prevent potential donors from becoming irritated and annoyed.

SPECIAL SESSION 1: **Fundraising for Dictionary Publishing** (3)

Dictionaries, A Cultural Investment?
Some Thoughts on Fundraising for Dictionary Projects

Jill WOLVAARDT
Dictionary Unit for South African English (DSAE), South Africa

This session is intended for staff of National Lexicography Units, their board members, advisers, and others interested in ensuring that speakers of all South African languages will have access to a well-designed range of dictionary products adequate to their needs.

National Lexicography Units (NLUs) for each of SA's official languages have been established as Section 21 Companies, that is as non-profit organisations, subsidised by national government via the Pan South African Language Board (PanSALB). Their principal objective is to write definitive monolingual dictionaries for their respective speech communities. So much we know. However, it is already apparent a) that the funds that PanSALB has at its disposal are only sufficient to maintain the minimal/basic functions of each lexicography unit and b) that most language groups are expressing a more urgent need for bilingual dictionaries. The discussion in this session will focus on how, given this context, NLUs can respond effectively to the demands of their target users.

It should be recognised from the outset that dictionaries are rarely commercially profitable products. This is all the more so in South Africa where the book-buying public is small, and public institutions such as libraries and schools have limited resources. Publishers may, therefore, be reluctant to take on our products. If, however, we can identify organisations prepared to underwrite some of the production costs, our dictionaries may become a more attractive proposition for publishers. In the light of this, the presentation will suggest that in order to subsidise the production of much-needed dictionaries for our speech communities, we should look for potential investors amongst those who are interested in the cultural dividends of our work, rather than in the financial return on their investment. That is, NLUs should – like other non-governmental organisations – consider seeking funds for their work from the so-called "donor community".

To do this, it is important that each dictionary be considered as a separate 'project' which should be formulated with the same structure and discipline as, for example, a project for a local clinic which is formulated by a community health organisation. By the same token, NLUs will have to be prepared to market and promote the cultural and educational importance of each of their proposals, in order to compete successfully alongside the many other worthy enterprises seeking support from the donor community.

The discussion will review what preparing a dictionary project within these parameters might entail. It will outline the preliminary processes to be considered long before the first entry is even drafted, highlighting some of the basic elements required for formulating a publishing proposal. Fortunately, these are largely similar to the components that the funder of any development project will be looking for when considering a proposal for support. So by undertaking this sort of preparatory work before embarking on each dictionary, NLUs will be equipping themselves to access both publishers and the necessary funding.

In this session, National Lexicography Units will be encouraged to see themselves as an active part of a process which unites those who need dictionaries with those who can make them available. The premise is that rather than working as cloistered researchers, we should find pro-active ways of linking with our speech communities. By building a relationship in this way, we can ensure that we are working to provide our target users with appropriate materials; equally, we will be in a position to assure prospective publishers that there will be a market for our product. And finally, we will be in a position to assure potential investors of the benefits of what we aim to produce. The workshop will discuss the characteristics of each of the elements in this cycle: user group – publisher – investor/donor, and first steps we might take to identify these elements, develop a relationship, and incorporate them in our plans for forthcoming projects.

SPECIAL SESSION 2: **Morphological Analysers for the Bantu Languages** (1)

KEYNOTE ADDRESS (3)

New Advances in Corpus-based Lexicography

Arvi HURSKAINEN
Institute for Asian and African Studies, University of Helsinki, Finland

In this paper I shall point out and demonstrate how language analysis tools can be maximally utilised in dictionary compilation based on text corpora.

1. Requirements for analysis tools

With the help of comprehensive language analysis tools it is possible to automate several labour-intensive phases in dictionary compilation. Such tools have to be able to:

    a.   identify the lemma form of each word
    b.   give full linguistic analysis of each word-form
    c.   solve ambiguity in analysis
    d.   optionally give glosses in target language
    e.   find examples of use for each key-word from corpus

Below I shall describe a set of tools and their development environments which, when applied together, fulfil these requirements.

2. General description of tools

a. Morphological analyser

The morphological analyser is the first and also the most labour-intensive of the components in the analysis system. It lays the foundation for other modules, and special attention has to be paid to its accuracy. The Finite-State calculus, advocated by Xerox, has so far been the most successful method, especially in analysing agglutinating languages. It is closely related to the more traditional two-level morphology, which also utilises finite states. A more recent approach that utilizes regular expressions has been used by Conexor in language management systems.

b. Disambiguator

In disambiguation there are two major approaches. One of them relies on probabilities in choosing the correct interpretation. Another method uses linguistic rules. Although the success rate in probabilistic disambiguation has been reported to be fairly good, it has two major disadvantages. It is not a knowledge-based, or intelligent, system, and the danger of wrong guesses is remarkable. It should be fairly clear that the knowledge-based system is the preferable one. This is particularly obvious with Bantu languages, where the concord system lays the linguistic foundation for writing disambiguation rules. Heuristic rules are applied only if there is no information available for writing knowledge-based rules.

c. Semantic analyser

Semantic analysis can be performed in two ways. Semantic information may be written directly into the morphological lexicon, or it can be done later with the help of a special external semantic lexicon. In the former case, the lexicon becomes large and its maintenance is burdensome. The latter method keeps individual modules more manageable, and their use is more flexible. What is particularly useful in the latter method is that semantic tagging can be performed to the morphologically disambiguated text. Semantics adds again ambiguity, of course, but this can be done in a more manageable environment, when morphological disambiguation has already been carried out earlier.

d. Syntactic analyser

In syntactic parsing, there are currently two successful methods available. Constraint Grammar is fairly good and for many applications sufficient. If the aim of the system is to develop into a

genuine language translation system, then more is needed from syntax. Functional Dependency Grammar, already applied to several languages by Conexor, seems to provide a 'full' syntactic analysis of text, and by this a major problem in knowledge-based translation systems is solved.

All phases of analysis will be demonstrated with a system applied to Swahili.

SPECIAL SESSION 2: **Morphological Analysers for the Bantu Languages** (2)

Using Finite-State Computational Morphology to Enhance a Machine-Readable Lexicon

Sonja E. BOSCH & Laurette PRETORIUS
Department of African Languages & Department of Computer Science and Information Systems, University of South Africa, South Africa

Introduction
The technological/computational treatment or natural language processing of morphologically complex languages, such as those belonging to the Bantu language family, requires the existence of a machine-readable lexicon, in other words a list of all word roots in the language. Although a first version of such a lexicon can be obtained from an existing dictionary, ideally such a lexicon needs to be supplemented with new word roots occurring in large collections of corpora on a regular basis, in order to reflect the dynamic nature of the language.

The aim of this presentation is to explain how a finite-state computational morphological analyser/generator can be used as a tool to enhance the machine-readable lexicon of a Bantu language such as Zulu. In particular, we consider the following questions:

What is Finite-State computational morphology?
After a brief introduction to finite-state methods and tools in general, we focus on their suitability for natural language processing, and specifically for computational morphological analysis and generation. We emphasise the importance of modelling natural language as accurately as possible and show how the Xerox tools may be used for this purpose.

What do we understand by a machine-readable lexicon?
For the purposes of this paper we understand the notion machine-readable lexicon to mean a list of all word roots in the language, stored in some convenient electronic format. We distinguish this notion from that of electronic dictionary/lexicon, which is the technical term currently used to refer to the handheld electronic device that often replaces the paper dictionary.

The reason why we focus on word roots is that words in Bantu languages are formed by productive affixations of derivational and inflectional suffixes to roots or stems. So, the constant core element in Zulu words is the root. A single verb root in Zulu for instance, may have hundreds of thousands of different inflected/derived forms. It is clearly very cumbersome and inefficient to add such a root plus all its forms to a wordlist. Unlike the case of a language such as English, a machine-readable lexicon for Zulu is therefore not a word list of complete words as they would appear in a Zulu text, but rather a list of word roots.

What do we mean by enhancing a machine-readable lexicon?

By enhancing a machine-readable lexicon we mean extending the lexicon by extracting new word roots from large collections of texts and adding them to the lexicon. In conjunctively written languages such as Zulu these new roots are embedded in the words in the texts, and need to be identified and extracted.

In order to perform this task of maintaining and updating the lexicon in a systematic and exhaustive way, the process should be automated. In particular, we need to automate the identification and extraction of new roots. Then, having done this, we require the facility to also modify the existing machine-readable lexicon by adding these new roots to it.

In our approach the identification and extraction of new roots are performed by the computational morphological analyser. We show why and how the morphological analyser represents the constant nature of the morphotactics and the alternation rules, while also reflecting and enabling the growth that takes place due to the evolution of the language and the subsequent expansion of the machine-readable lexicon. Moreover, in an agglutinating language such as Zulu, including a single new root in the lexicon, adds large numbers of different word forms to the language, most of which cannot be found in dictionaries or word lists compiled from corpora, but will be catered for by the morphological analyser/generator, based on the modified machine-readable lexicon!

How do we use a computational morphological analyser to enhance a machine-readable Zulu lexicon?

By means of a simple example we illustrate the procedure of enhancing an existing lexicon, based on a short natural language Zulu text.

In particular, we apply the morphological analyser to the given text, we extract/identify all the new word roots in the text, we consider these new word roots for inclusion in the next version of the lexicon, we add them to the machine-readable lexicon, and then we finally rebuild the morphological analyser, based on the modified lexicon.

How is this useful for lexicography?

We answer this question by mentioning three basic problems/needs that lexicographers often face, and which may be readily solved by means of finite-state computational morphology:

- the seemingly unbridgeable gap between dictionary and grammar in the context of machine-readable dictionaries vs. paper dictionaries (Prinsloo 2001: 152);
- the addition of new word roots by enabling the lemmatisation (that is, morphological analysis) of a corpus (De Schryver & Prinsloo 2000: 95);
- obtaining frequency counts of word roots as one of the basic outputs of electronic corpora (De Schryver & Prinsloo 2000: 98).

Indeed, the approach that we follow solves these problems by

- emphasising and exploiting the intrinsic connection between the words in a language ("dictionary") and their morphological structure ("grammar");
- facilitating the systematic, exhaustive identification and inclusion of new word roots that occur in language corpora;
- providing an accurate way of determining frequency counts of existing word roots.

First Steps in the Finite-State Morphological Analysis of Northern Sotho

Gilles-Maurice DE SCHRYVER

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

In lexicography, one can come a long way with just introspection and a series of good grammatical descriptions. Better is when these can be supplemented with data obtained by means of informant elicitation and corpus consultation. The great majority of the dictionaries the world over have been compiled in this way. This is not different for African-language dictionaries.

Over the past few decades the notion of "corpus" in the language sciences has shifted from 'huge collections of paper slips' to 'running text available electronically'. When analysed with versatile corpus-query software these electronic corpora provide unprecedented insights into how languages *really* work. Again, also here African-language lexicographers have attempted to follow the international trend, with the first corpus-based dictionaries for languages such as Northern Sotho, Cilubà, ChiShona, Zimbabwean Ndebele and Kiswahili already on the market. Nonetheless, it is an open secret that most corpora built and queried for these lexicographic projects were not very different from so-called "raw corpora". Most African-language corpora to date are indeed actually plain running text without any linguistic annotations or text markup whatsoever. One notable exception in this regard is the tools developed by Hurskainen (1992, and later) to tag Kiswahili texts.

Although no one will dispute the value of the reference works based on raw corpora, looking ahead means realising that modern electronic corpora for all languages – and thus not only for say English, French or Spanish – will (have to) be annotated linguistically. The first type of tags one generally adds to corpora are those for *morphology*. In text-based computational linguistics one can then proceed to word disambiguation (part-of-speech tagging), shallow or robust parsing (chunking), syntactic parsing, summarisation, information extraction, question answering, and finally to machine translation. Components in speech-based computational linguistics include text-to-speech generation, speech recognition, question answering and automatic interpretation. Morphological analysis can be considered as the first step, if not the core, of any such system, and all serious future electronic dictionaries, for instance, that are not simply the electronic variant of the hardcopy original, will contain a built-in morphological analyser.

Given this, the first question one obviously has to answer is how to go about the computational morphological analysis of a language, *in casu* an African language. The few early African-language projects in this regard (for Kiswahili, ChiShona and Zimbabwean Ndebele) all use finite-state tools, albeit revolving around the somewhat dated two-level model. The newcomers in the field (for isiZulu and Northern Sotho) use the Xerox finite-state programming languages *xfst* and *lexc* to create finite-state networks that perform morphological analysis.

The main aim of this paper is to show that finite-state morphological analysis is not as esoteric as it might sound. To illustrate this, a small prototype finite-state transducer (FST) for Northern Sotho will be presented. With an FST, not only the *analysis*, but also the *generation* of linguistically correct strings can be performed. Although the prototype to be discussed

contains only a thousand "root forms" (verb roots, noun stems, concords, etc.) in the lexicon compiler **lexc**, and only a hundred "alternation rules" in the Xerox finite-state transducer **xfst**, both the analysis and generation potential are already impressive. This will be exemplified with a discussion of the 'recall' and the degree of 'ambiguity' of a randomly selected text fed into the prototype FST. Other lexicographically relevant issues that will receive attention are the degree of disjunctiveness / conjunctiveness and its implications, across-word sound adjustments (e.g. ***mo bôna → mmôna**), the lexicographic inclusion *versus* the orthographic deletion of the circumflexes for **ê** and **ô** *versus* **e** and **o** respectively, the use of composition filters to treat dialectal forms, etc. Finally, it will also be pointed out that, already in this early stage, planning and discovery go hand in hand. In other words, the fact that existing descriptions and dictionaries for Northern Sotho are inaccurate and incomplete, is automatically brought to light through the creation of a morphological analyser. Thinking about morphological analysers thus leads to new fieldwork and ever-better dictionaries.

SPECIAL SESSION 2: **Morphological Analysers for the Bantu Languages** (4)

Word Division and Orthography as Some of the Factors Posing Challenges in the Development of the Ndebele Grammatical Parser

Mandlenkosi MAPHOSA

ALRI (African Languages Research Institute), Zimbabwe

The Ndebele corpus, like any language corpus, is undergoing many forms of processing in order to produce a number of useful language products. One such product is the grammatical parser which is currently being developed using a two-level model which is dubbed PC-KIMMO after its inventor. The grammatical parser is aimed at comprehensively describing the Ndebele language. As such, through working on it, some linguistic details come to the fore.

This paper aims to first explore the stages that have been followed in the development of the Ndebele grammatical parser. Through this descriptive approach the challenges that have been faced in the development of the parser will be exposed. It will become apparent that the major challenges that have come about are those that result from word division and orthographic problems that exist in the Ndebele language. This is because a substantial part of the Ndebele corpus is oral in nature. Oral corpora by their nature contain language that does not strictly adhere to the word division and orthography rules of a particular language. Oral corpora are full of fast speech, shortened word forms and noun compounds. As such the paper will highlight some of the orthographic and word division issues that came to the fore as a result of the oral nature of the corpus and how these factors were or are still a challenge in the development of the grammatical parser.

However, word division and orthography problems are not confined to the oral corpus only. The Ndebele language in its written form is still fraught with a substantial part of orthographic and word division deficiencies which will also form part of the discussion. Though the major part of this paper will be on exploring the challenges posed by word division and orthography in the development of the parser, it will nonetheless explore other factors that proved challenging, among those: noun classification and grammatical categorisation. Some of the problems to be discussed are not inherent in the linguistic patterns of the language under

study but are the creations of man. The major problem in this regard is methodological. When the task of compiling the Ndebele corpus began, the product that the compilers had in mind was the dictionary. As such the tags that were used in tagging the corpus materials were inclined towards the dictionaries. This inevitably brought about some challenges in the development of the parser. The other problem concerning human error which brought about problems in the development of the parser was that of lack of proper proof-reading of corpus materials which resulted in an unclean corpus. The paper will thus show how this impacted negatively on the development of the parser. Having looked at the challenges that were brought about by the afore-mentioned factors the paper will proceed to look at the solutions that were adopted and also highlight those that are yet to be attended to.

SPECIAL SESSION 2: **Morphological Analysers for the Bantu Languages** (5)

Problems and Challenges Encountered when Developing a Morphological Parser for the Shona Language

Daniel RIDINGS & Webster MAVHU
ALRI (African Languages Research Institute), Zimbabwe

The intended paper arises out of the present writers' involvement in the process of developing a morphological parser for the Shona language. The present writers are members of the African Languages Research Institute (ALRI). ALRI was formerly the African Languages Lexical (ALLEX) Project and was housed in the Department of African Languages and Literature at the University of Zimbabwe. It is now a non-faculty semi-autonomous language research unit that is affiliated to the University of Zimbabwe. ALRI's major aim, as stated in its mission statement, is to carry out research that enhances the development of the indigenous languages of Zimbabwe. The institute mainly focuses on research in corpus work, computational lexicography and language technology. These three areas are the institute's basic and essential research activities on which all other services depend as their tools and facilities.

The ALRI team's research activities as ALLEX (1992-99/2000) have so far culminated in the development of corpora for two of Zimbabwe's main languages: Shona and Ndebele. The corpora have since been used to produce two monolingual Shona dictionaries, that is, the General Shona Dictionary and the Advanced Shona Dictionary plus one monolingual dictionary on Ndebele, the General Ndebele Dictionary. ALRI members intend to produce more works, not only in these two languages, but also in all the other indigenous languages of Zimbabwe. In addition to compiling dictionaries, grammar books and other reference works for the indigenous languages of Zimbabwe, ALRI intends to create some language technology applications for them. These applications include grammatical parsers, syntactic analysers and spellcheckers. Currently, the institute is engaged in the process of developing grammatical parsers for Shona and Ndebele. The first step has been, however, to develop morphological parsers for them.

The morphological parser for the Shona language is now at an advanced stage and can recognise at least seventy percent of unrestricted text from the Shona corpus which currently stands at about 2.6 million running words. It is hoped that by the end of 2002 the morphological parser should be in use. The exercise of developing a morphological parser for

the Shona language began with the creation of a morphological lexicon of Shona. The initial process of creating a morphological lexicon of Shona was based on frequency counts in the Shona corpus created for ALLEX. The most frequent verbs of the first person singular appearing in the present habitual were isolated: **ndinoda** 'I like', **ndinofunga** 'I think', **ndinoziva** 'I know', etc. **Kufunga** 'to think' was chosen as the first verb to be fully analysed. This verb occurs in 810 inflectional forms. This was ascertained by using Unix programs such as "egrep" to isolate all forms in the frequency list that ended in **funga** 'think'. These 810 word types represent 5,508 tokens in the total corpus. Existing grammars were then used to "fill in the slots" for subject concords, tempus, auxiliaries, object concords and extensions. This was done until we felt satisfied with the success we had in analysing all word types that contained **funga** 'think'*,* **fungisa** 'make one think' and **fungira** 'think for'.

We then populated the morphological lexicon using heuristics. A simplistic example is as follows; we isolated all word types beginning with the form **ndino** 'I'. We took the remainder, e.g. **ziva** 'know' in the case of **ndinoziva** 'I know' and searched for the form **kuziva** 'to know'. If we found it, we made a preliminary assumption that **ziva** 'know' could be used as a verb root. We then did tests to verify our assumptions and then reiterated the process based on our newly won knowledge. In this process Ridings assigned linguistic tags to each morpheme. These tags were designed in such a way as to facilitate disambiguating ambiguous forms using the methodology made popular by Fred Karlsson.

The next step was the analysis in maximum detail, of some selected texts from the corpus. What followed this step was the marking of morpheme boundaries on the words in those texts. After that, tags were inserted in the texts. Ridings then incorporated the tags into a two-level model for morphological analysis. The two-level model that is being used on the Shona language is designed along Kimmo Koskenniemi's methodology as implemented in PC-KIMMO, software from the Summer Institute of Linguistics which is dubbed KIMMO after its inventor. ALRI researchers are currently involved in the incremental population of the model's lexicon files with lexical items. There are problems and challenges that are associated with each step that has been mentioned above. The intended paper will point out these problems and challenges. It will also mention the solutions that have been adopted to counter them. We also intend to illustrate the above-mentioned methods in detail, in a cookbook manner, so that they can be reused on other languages in the region.

SPECIAL SESSION 2: **Morphological Analysers for the Bantu Languages** (6)

Human Language Technologies and the *National Lexicography Units*

Justus C. ROUX
Department of African Languages, University of Stellenbosch, South Africa &
Sonja E. BOSCH
Department of African Languages, University of South Africa, South Africa

The development of software tools such as morphological analysers and syntactic parsers for different languages is a significant step not only for the creation of appropriate dictionaries, but also for entering a particular sector within the Information Society. It is important to realise that dictionaries, especially those in electronic format, play an essential role in the development of

Human Language Technologies (HLTs). HLTs are enabling technologies which are implemented in systems which allow humans to interact with computer systems in different modes (through text or speech) by using natural, everyday language. Among other developments, appropriate lexicons have to be constructed to enable the development of the following types of interactive systems operating in, for instance, African languages:

- **Multilingual telephone based information systems**
  - Tourism & Travel: Hotel booking systems; train, air, bus schedules; road conditions, weather reports, travel packages
  - Health services: First-level medical help lines, Aids hotlines, TB hotlines
  - Public services: Applications for pensions, travel documents, car registrations; telephone accounts, telephone number enquiries
  - Business: Bank balance requests, mobile shopping
  - Leisure: Automated booking systems for theatres, sports events, voice SMSs

- **Multilingual multimedia information systems**
  - Education: Language learning, voice-based training systems
  - For the blind: "Speaking" books, newspapers – making Braille obsolete
  - For the deaf: Screens on telephone converting speech of the caller into text
  - For paraplegics: Voice-activated systems for support, e.g. typing text on a computer using voice in any applicable language
  - For non-literates: Vocally communicating with a computer to obtain relevant information

- **Multilingual automatic / machine-aided translation systems**
  - State services: Official documents, Hansards in national, provincial, local governments
  - Education: Developing multilingual teaching material
  - Business: Translation of technical manuals, instructions on the use of products, etc.

A Steering Committee of the *Department of Arts, Culture, Science and Technology* (DACST) and PanSALB devised a strategic plan for HLT development in SA in 1999. This plan included the establishment of a National Resource Centre for electronic text and speech which is linked to the *National Lexicography Units* (NLUs) through the Internet. The HLT plan is not dependent on the participation of the NLUs, but it provides a vast range of opportunities to the NLUs to assist them in their primary tasks of producing dictionaries. Participation by NLUs in the HLT initiative could "fast track" many of the activities of the Units because:

  - they will have *continuous access* to electronic text and speech data in a language of choice, which will cut down on time spent on scanning or typing appropriate texts and which will allow them to focus on the *core business*, i.e. constructing dictionaries with all the complexities involved,
  - they will have recourse to *technical backup* with respect to software and hardware issues,
  - they will be in a position to build capacity by sending staff to attend carefully designed and monitored *training programmes* in HLT and lexicography,
  - they will be assured that their work meet *international standards* and display good practice results.

The Minister has appointed an Advisory Panel which is to report to him in August 2002 on ways and means to implement the proposed strategic plan. Members of this panel are in the

process of making contact with the NLUs who now have the opportunity to state their position with respect to possible participation.

PARALLEL SESSION (1)

Language Development in a Multilingual Society

Emmanuel CHABATA

ALRI (African Languages Research Institute), Zimbabwe

Mono-lingualism is a very rare phenomenon the whole world over. Most societies are multi-lingual for there is usually more than one language spoken within the confines of each and every community. It is also a common feature that within these communities some languages are smaller or bigger than others when it comes to the number of speakers in each speech community. It is also generally agreeable to those people who care about language that all languages of the world need to be developed as a way of either modernising and/or protecting them from extinction. This is especially the case when it comes to 'community' languages that have been neglected for a very long time or that have suffered from unequal development ever since. These include those languages that are spoken by minority groups. Different societies have taken different approaches in order to address issues of language development, but whatever the approach, the result is the same: some languages are developed either earlier or later than others.

The proposed presentation will focus on issues of language development as they come as a challenge to language planners and researchers. The presenter hopes to do so by looking at challenges that are encountered when dealing with issues such as language selection for development, the people that should be qualified to get involved in particular developmental projects, as well as the role that the government should play in such activities. The presenter will also look at how factors like history, geography, attitude, politics, the economy and others can influence issues of language development. He will also examine how different groups of people may view developmental initiatives on language as well as the different attitudes language planners and researchers are bound to face or experience during the process of carrying out their duties.

The Zimbabwean situation will be taken as a case study. The reasons for choosing Zimbabwe as an example case are that, like many other countries, it is multi-lingual; it has about seventeen known languages that are spoken within its borders. Most of these languages are spoken by small groups of people and have suffered from little or no development over the years. The other reason is that it is the Zimbabwean situation that the presenter has experienced during the past ten years as a language researcher. In fact, the presenter will comment on some of the problems and challenges that he and other language researchers have faced during the process of trying to develop some of Zimbabwe's local languages. He will draw examples from projects that he has participated in, especially those that deal with dictionary making, corpora building and others. He will also examine efforts by earlier researchers. It is also hoped that the Zimbabwean situation is applicable to situations obtaining in many other African countries.

In the same presentation, the researcher will also look at some practical solutions to some of the common problems and challenges that researchers usually meet in carrying out

research projects that have to do with language development.

PARALLEL SESSION (2)

Torn Between Calling a Spade a Spade and Being Euphemistic: The Dilemma of a Lexicographer Defining Offensive Headwords in Shona Monolingual Lexicography

Emmanuel CHABATA & Webster MAVHU
ALRI (African Languages Research Institute), Zimbabwe

The intended presentation will focus on the challenges that a lexicographer faces when defining offensive words in Shona monolingual dictionaries. Offensive words will here be taken to refer to those terms that are used to refer to people and other things in a commentary, derogatory or insulting manner. Some such words are vulgar and impolite. In Shona, examples of such terms include words that refer to the coloured community, migrant labourers, and the albinos as well as to people who are crippled. They also include those that refer to private parts of the body, which can be used to hurt somebody's feelings. The use of these words is not normally acceptable in Shona culture, especially in public. Although most speakers of Shona may be aware of the existence of these words, they may not use them freely to refer to events, activities, objects or people that they are known to refer to. This is because offensive words are generally sensitive; they may not be used without arousing bad feelings for people who would have been referred to. In Shona language and culture, sensitive words are not normally used for fear of hurting other people's feelings. Instead, people would prefer to use euphemisms in place of such words. Although they would still refer to the same objects and events that the offensive words would, the use of euphemisms would do so in a culturally more polite way.

The dilemma for the Shona lexicographer is whether or not to include offensive terms in a monolingual dictionary. Excluding these words would compromise the dictionary's representativeness of the language that it would intend to describe. However, if one decides to include them, the challenge would be on how to describe their meanings, that is, whether to be explicit or to be euphemistic. When deciding on either of these two approaches, the lexicographer has to realise that one of the purposes of providing a definition in a dictionary is to give an accurate description of meanings of all the headwords contained. The question to be asked is thus whether or not euphemisms can capture the exact meanings of some offensive terms.

The proposed presentation will look at ways in which the editors of Duramazwi Guru reChiShona handled offensive terms by looking at how they tried to balance the equation of either being explicit or euphemistic. Duramazwi Guru reChiShona is a general, medium-sized monolingual Shona dictionary which was published by a team of researchers at the African Languages Research Institute, based at the University of Zimbabwe. The presenters of the proposed paper are also part of this research team. In fact, the presentation arises out of the researchers' experiences during the defining stage of this dictionary. Unlike in Duramazwi reChiShona, the forerunner to this more advanced dictionary, where offensive words were avoided as much as possible, these terms were included in the later dictionary which was supposed to be more comprehensive. The editors felt that offensive terms were supposed to be

included in Duramazwi Guru reChiShona because they form part of the Shona language. Evidence of their use is their appearance in the Shona corpus.

PARALLEL SESSION (3)

Bilingual Zulu Dictionaries and the Translation of Culture

Rachélle GAUTON
Department of African Languages, University of Pretoria, South Africa

This paper focuses on bilingual dictionaries and the translator, with specific reference to the translation of culture in bilingual Zulu dictionaries.

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 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." (Pinchuck 1977: 231)

Pinchuck (1977: 233) asserts that the bilingual dictionary should only be used as a last resort, and should not be the first aid that is sought. He contends that the first dictionary a translator should consult, must be a terminological one if available. Next a technical dictionary dealing with the subject field in question, should be consulted. Should these dictionaries not suffice and the problem be one of general vocabulary, a monolingual dictionary should be explored. According to Pinchuck (1977: 233-234), the aforementioned methodology is more likely to lead the translator to the concept underlying the lexical item and its associations, than the use of a bilingual dictionary. He does state, however, that if this methodology cannot be followed, only a good bilingual dictionary should be consulted, as a bad bilingual dictionary will be dangerous.

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. Swanepoel argues that the bilingual dictionary is limited for the following two reasons:

- it does not contain sufficient information for the user; and
- it cannot be a substitute for the user's competence in the SL and TL. The process of translation involves the user's total communicative competence, which also includes a grasp of the text's sociocultural context.

Swanepoel concludes that the bilingual dictionary is nothing more than an aid to the professional translator in cases where his/her acquired knowledge of the TL is lacking.

In this paper, the reasons for this state of affairs will be elucidated by indicating:

- which problems are experienced by the lexicographer in the compilation of the bilingual dictionary, with specific reference to the translation of cultural concepts in a variety of Zulu bilingual dictionaries; and
- which problems are experienced by the translator when attempting to find suitable translation equivalents by consulting the bilingual dictionary.

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.

With reference to a selection of bilingual Zulu dictionaries, this paper will also show how the various compilers of these dictionaries have allowed their own cultural biases to influence the choice of translation equivalents in the TL.

PARALLEL SESSION (4)

Using a Frame Structure to Accommodate Cultural Data

Rufus H. GOUWS
Department of Afrikaans and Dutch, University of Stellenbosch, South Africa

A dictionary can be regarded as a carrier of text types including a variety of different texts. Different positions in the dictionary as a "big text" are allocated to these texts. A textual approach to lexicography emphasises the need for an unambiguous identification of the function and the nature of the different text types prevailing in dictionaries.

Recent research in the field of metalexicography has focused increasingly on the structural components of dictionaries. An innovative approach in this regard has been the introduction of the data distribution structure. This component of a dictionary determines the way in which data types are presented and different texts are positioned in the dictionary. The central list of a dictionary is no longer the only venue for texts to occur. Although the central list remains an important and compulsory text, data can also be presented in texts preceding and texts following the central list. These outer texts, occurring in the front matter and the back matter of a dictionary, complement the central list to constitute the frame structure of a dictionary. Although the idea of front and back matter texts is not new the emphasis on the functional value of the frame structure has led to a renewed interest in the inclusion of text types not traditionally regarded as part of a dictionary.

Especially in a multilingual environment dictionaries do not only function as linguistic instruments but also as cultural instruments. The traditional approach to the central list of dictionaries with its strong linguistic bias has not allowed a proper transfer of cultural information. This applies to both the selection of lemmata and the treatment presented in a typical dictionary article. An optimal utilisation of a frame structure allows the lexicographer a much more diverse approach to the data distribution and the types of texts to prevail in a dictionary. In many dictionaries the outer texts are still dominated by data primarily supporting the linguistic treatment presented in the articles of the central list. However, more and more lexicographers realise that the frame structure offers them the opportunity to diversify the lexicographic treatment in terms of both the nature and the extent of data and text types to be included in dictionaries.

This paper focuses on ways in which lexicographers can utilise the frame structure to present cultural data. Provision will be made for the use of both integrated and unintegrated outer texts to enhance the transfer of cultural information. New developments in metalexicography have resulted in a much more detailed analysis of the frame structure. These developments will be discussed and evaluated in terms of their potential to improve the presentation and treatment of culture-specific lexical items. Attention will not only be given to the outer texts but also to the interaction between outer texts and the central list aimed at a better access to the prevailing cultural data. In this regard the emphasis will be on the use of synopsis articles in the central list, complemented by alphabetical registers in the back matter to ensure poly-accessible dictionaries which can provide the user with a rapid search route to reach the cultural data.

PARALLEL SESSION (5)

The Treatment of Culture-Specific Lexical Items in Bilingual Dictionaries

Karen HENDRIKS

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

In this paper I intend to undertake an initial exploration of the admission to and treatment of culture-specific lexical items in a bilingual dictionary, which is intended for a multilingual environment, such as the environment that South Africa presents the lexicographer with. It should be emphasized that this study must be understood as nothing more than an introduction, in general terms, of certain issues and questions concerning the role that culture plays in the process of co-coordinating meaning and determining translation equivalents in a bilingual dictionary written for a multilingual environment.

The concept 'culture specific' is in itself problematic. Does the term merely refer to the lexicalization of a semantic value that has its roots in the culture of the speaker? How does one come to an objective, or at least more or less neutral definition of the notion 'culture specific'? According to what and whose rules must the lexicographer identify lexical items that would qualify as culture specific? The issue of defining 'culture specific' becomes even more complex when it's considered within the context of a strong multilingual and culturally diverse society such as South Africa.

The process through which culture-specific lexical items are, and should be, selected for admission to the macrostructure of a bilingual dictionary also presents the lexicographer with certain questions. How does one construct a corpus that is really representative of the culture of the speakers, for instance when the language does not have a representative body of written texts to work with? Should culture-specific lexical items be selected on the grounds of the frequency of use in the database, or are there different measures that ought to apply?

Once the lexicographers have decided upon the admittance of culture-specific items, the problem of providing adequate translational equivalents arises. How does one find or create a translational equivalent in the target language to lexicalize a cultural concept within the source language, which is unfamiliar to the target-language speaker? Furthermore when referring to culture-specific lexical items, one would certainly be referring to multilexical items as well. Multilexical items such as the expressions, proverbs and idioms of a language, which are often

closely related to the culture of the speakers. How does one find a target-language item that lexicalizes the semantic value of the source-language multilexical item, if the context in which for example the idiom is mostly used in the source language is unfamiliar to the target-language user?

Culture should be taken into consideration when the lexicographer attempts to provide translational equivalents for culture-specific items. To what extent culture should have an influence and how much additional explanations could be permitted in the treatment of culture-specific lexical items, remains a question that should be asked and answered in a multilingual environment such as South Africa.

PARALLEL SESSION (6)

Culture and a Dictionary: Evidence from the First European Lexicographical Work in China

Gregory JAMES & Bronson SO Ming Cheung
Language Centre, Hong Kong University of Science and Technology, China

According to his diaries, Matteo Ricci, one of the first Jesuit missionaries to enter China, compiled, with his companion Michele Ruggieri, and possibly one of the first Chinese priests, Father Sebastian, a Portuguese-Chinese glossary for his own and others' use (in c. 1580). The manuscript, once believed lost, was discovered in the Jesuit Archives in Rome in the 1930s, and has recently been published for the first time in a facsimile edition. It comprises a 7,000-entry Portuguese headword list (with some phrases and short sentences, and occasional explicatory synonyms in Latin or Italian), with Chinese character translations and the first known attempts at rendering Chinese into the Latin-script 'phonetic' alphabetic system. At this early stage of European contact with Chinese, there was no conception of the tonal features of the language, and tone is not indicated in the transcriptions. Hitherto, although the Portuguese headword list and the romanisations of Chinese characters have been the focus of some scholarly attention, no work has been undertaken on the semantic content of the dictionary. In this paper, we offer an introductory analysis of some of the cultural features of this, the first dictionary of Chinese known to have been written by a European, and draw tentative conclusions from the evidence of the text as to how the earliest Europeans in China met the challenge of learning Chinese. A significant part of our work has been the translation of all the Portuguese headwords, and the corresponding Chinese translations, into English, and an analysis of the appropriateness of the Chinese as representations of the Portuguese, and the many errors at phonological, syntactic and semantic levels. Indeed, the very selection of the headwords and their translations offers insights into the perceived needs of the missionaries of the period in an unfamiliar cultural context. As might be expected, there are many words concerned with Christianity, but some of these are left untranslated, since at this period, no satisfactory translations for some concepts (e.g. 'grace') had been worked out. Even the word 'God' is missing from the dictionary. 'Saviour' is, however, included, as are 'Confucius' and 'Adam', but not 'Eve'! Indeed, our cross-indexing of the headwords has revealed a general male-centredness throughout the dictionary: while 'man' is described in very positive terms (with accompanying epithets such as 'urbane', 'scrupulous' and 'brave'), 'woman' fares much

less well (described *inter alia* as 'dissolute', 'sinful' and 'immodest'). There is a wealth of examples of names of parts of the body, illnesses and deformities (both natural and inflicted) - the Europeans were very conscious of, and feared greatly, the diseases they could succumb to in an unfamiliar land! Interestingly, there are also many headwords concerned with weapons, torture and punishment, a feature of everyday life in Ming China, especially for missionaries from overseas. In our presentation, all examples will be given via English translations, and we shall demonstrate the features of the innovative web-based relational database and multilingual search facilities we have designed to analyse the dictionary manuscript, and which may be adapted to other similar lexicographical projects.

- *Gregory James is Professor and Director of the Language Centre, Hong Kong University of Science and Technology, and has undertaken the linguistic analysis of the dictionary text.*

- *Bronson So is a graduate in Mathematics and Computing from the Hong Kong University of Science and Technology, and has designed the relational database for this project.*

PARALLEL SESSION (7)

Work in Progress: First Steps in Using a Bilingual Dictionary Framework

Kathryn KAVANAGH
Dictionary Unit for South African English, South Africa &
Philisiwe MANYISA
SiSwati Dictionary Unit, South Africa &
Disebo MOETI
Sesiu sa Sesotho Lexicography Unit, South Africa &
Neo MPALAMI
Sesiu sa Sesotho Lexicography Unit, South Africa

An English framework is being developed by the Dictionary Unit for South African English. It is intended that it should form the basis for a range of bilingual dictionaries for South Africa. The early stages of this project involve making decisions on the nature and level of content and then trialling sample material with lexicographers from other South African lexicography units. The first sample consists of only 100 headwords representing different parts of speech and vocabulary level. This paper reports on the sample created and on the rationale behind some of the decisions taken in setting it up. It describes the approach taken by the lexicographers from the Sesotho and Siswati lexicography units who worked on the sample and includes their comments on the content and style of the framework. Their reactions and ideas will be taken into account when the second, larger sample is created later in 2002.

The framework project aims to make available in a user-friendly form much of the information which will be required to build a bilingual dictionary. It is more than a headword list as it includes information about part of speech, irregular forms, syntax patterns, and register, as well as sense discriminators. For most headwords there are also sentences exemplifying usage. The 100-word sample includes nouns, verbs and function words, general vocabulary, school-curriculum and technical words. The sample text is deliberately not consistent stylistically, because the developers wish to discover which types of expression and

layout are most acceptable to potential users of the framework. Comments are particularly sought on the style and complexity of the sense discriminators. These discriminators are intended mainly as a guide to lexicographers translating the headword or a particular sense, but it is also recognised that they may be included, perhaps in a modified form, in the final dictionary text.

A coding system indicating different levels of vocabulary is being developed for use in the framework. This is intended as a guide to lexicographers selecting from the framework headwords for a particular type of dictionary. Considerable research still has to be done into defining vocabulary level, and in assigning headwords to a particular level.

All lexicographers involved in the project are working in Microsoft Word at this stage, since it is common software. An Access database lies behind the Word front screen. Ultimately the framework is to be set up in database software compatible with the editing software expected to be used by all the South African lexicography units. The layout of the sample material will remain for some time an approximation to the final product.

PARALLEL SESSION (8)

History, Language Contact and Lexical Change: A Lexicography/Terminography Interface in Zimbabwean Ndebele

Langa KHUMALO
ALRI (African Languages Research Institute), Zimbabwe

The paper investigates and analyses linguistic changes and/or developments that the Ndebele language, spoken in Zimbabwe, might have undergone from its earliest attested form to its present-day form and the implications this has for term creation and standardization through lexicography. Language change is oftentimes viewed as obvious and simultaneously mysterious. The Ndebele language of yesteryears is so different from Modern Ndebele. The existence of such differences between early and later variants of the same language raises questions as to how and why languages change over time. This paper will make a very brief historical analysis of the nature and causes of language change in Ndebele. Ndebele has seen a lot of modifications in its lexicon. The paper will, therefore, focus on lexical change in Ndebele.

The paper will first give a brief background of the Ndebele language. It will highlight the movement of the Ndebele people from South Africa where they had linguistic contact with the Zulu, Xhosa, Swati, Sotho, and the Afrikaners among other language groups. Later in Zimbabwe the Ndebele had further contact with the Kalanga, Shona, Venda, Nambya, Tswana, and Tonga whom they subjugated and incorporated into their political system. Even later, they had contact with English and the technological advancement it brought with it. From a linguistic point of view, the above scenario provides a fertile ground for the process of language contact and therefore language change.

The second part of the paper will contrast terminography with lexicography. Lexicography and terminography have much in common: they are both concerned with describing lexical items in a user-friendly format within a dictionary. The specialized nature of the lexical items studied in terminography, however, gives the discipline its own

distinguishing features. The purpose of terminography is to identify and analyse lexical items used in specialized domains of knowledge, such as commerce, medicine, law, computing, etc. In principle, all domain-specific terms are of interest. In practice, however, terminographers are overwhelmingly preoccupied with *new* terms: as domains change and grow, often at a frightening pace, terminographers must document the associated lexical changes. Ndebele lexicographers produced their first corpus-based monolingual dictionary last year. The Ndebele corpus of both spoken and written material demonstrates a large extent of borrowing and also loss of certain lexical items in Ndebele lexical inventory as a result of the reasons stated above. The paper will finally demonstrate that to a large extent Ndebele editors were expected to introduce terms for the purposes of popularising their use and therefore acceptance since some of them can no longer be ignored.

PARALLEL SESSION (9)

On Corpora and the Process of Selecting High Function Words and Their Treatment in Currently Available Dictionaries for Sepedi

Diapo N. LEKGANYANE
Department of Northern Sotho, University of Venda, South Africa

Language politics of South Africa can, simplistically spoken, be divided into two phases. The *first phase* is represented by the constitution of South Africa in the apartheid era and the *second phase* by the post apartheid constitution (1996). The language principles and stipulations in the constitution of the previous government recognised English and Afrikaans as the only two official languages of South Africa, and indigenous languages such as Sepedi were marginalized. The speakers of these languages were made to believe that their languages were inferior to English and Afrikaans, and as a result they developed a negative attitude towards their mother tongues.

The *second political phase* started with the post apartheid constitution of South Africa. It recognised the indigenous languages as well as English and Afrikaans as official languages of South Africa, thereby officially changing their status. The language stipulations in the constitution entail that indigenous languages should be promoted so that they can enjoy the same high functional status as English and Afrikaans. Status-planning for indigenous languages has therefore been accomplished to a certain extent. This kind of language planning must, however, be followed up by government through the promotion and sanctioning of the autochthonous languages as languages of further and higher education.

Efficient status planning makes it possible for acquisition planning to take place without any hindrance. Users may acquire this language (more fully) through speaking and reading textbooks written in African languages. Lexicographers and terminographers can also elevate the status of the language more successfully amongst mother-tongue speakers by compiling African-language monolingual dictionaries, a dictionary type which currently does not exist. The existing African-language bilingual dictionaries can also be improved, not only to assist students and translators, but also to ascertain that these languages are able to take up their place as fully fledged official languages next to a world language such as English. In addition to this

the compilation of bilingual, monolingual and bilingualised learners' dictionaries could be a significant step in making these
languages more accessible to speakers of other languages

In order for these languages to become widely used as high function languages, effective and efficient corpus planning is an imperative. The first step in this process would entail assessment of its functional mobility, i.e. the use of the language across a wide spectrum of social functions, including higher functions.

One way of achieving this goal is to build up a computer corpus of English texts used in higher functions and then use this corpus to determine possible lexical gaps in African languages.

PARALLEL SESSION (10)

Capturing Cultural Glossaries: A Case Study Presentation of N. Sotho Cooking Terms

Matete MADIBA
Technikon Northern Gauteng, South Africa &
Lorna MPHAHLELE
Technikon Northern Gauteng, South Africa &
Matlakala KGANYAGO
Nkoshilo High School, South Africa

This paper is a presentation of a brief cultural glossary on N. Sotho cooking terms. The glossary is mainly composed of names for utensils, ingredients and action words for the processes involved in the preparation of cultural dishes. By means of a case-study approach, the paper seeks to explore ways of capturing cultural glossaries with a view to assist the national dictionary processes. The method that led to the production of this specific glossary starting from a school-based project will be investigated.

There are a number of issues that surfaced in this project that have the potential to serve as a model for the collection of authentic glossaries that can support dictionary-making in African languages. What is considered to be a distinguishing strength in the project is that a meaningful context was used for the collection of this glossary. Contextualisation can then be used as a good organising tool for the collection of other glossaries. The school setting, within which the project is situated, provides a fertile ground for an activity of this nature. The surroundings of the school are dominated by rural settlements, which are even more relevant and useful as authentic resources for cultural embodiments.

Of particular interest is the potential projects of this nature have to capture and put on record cultural words that would otherwise be lost. This work also seeks to investigate how glossaries like these can help to realise and implement innovative methodologies and concepts such as "simultaneous feedback" (De Schryver & Prinsloo 2000) and "hybrid dictionaries", to support lexicographical work in the country. It is also interesting to note that the glossary at hand is a 'secondary' product of the project and not the primary, in the sense that the project had a different aim. This distinctive feature (of being a 'secondary' product) 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 draw conclusions from the process of compiling this brief glossary. It is the exploration of these conclusions which will then be used to propose a possible and authentic model for collection of other glossaries of this nature.

The rationale behind the project is based on the argument that for the formerly marginalized languages it will not be easy to capture cultural terminology from a corpus that is built mainly from written texts. It is therefore argued that a focus on written corpus material has the potential to create gaps in such a way that may exclude cultural terms. The provision of a model for the collection of cultural words and the initiation of similar projects reported in this study will attempt to address these identified gaps.

PARALLEL SESSION (11)

From Corpus Data to the First isiNdebele Dictionaries

P.S. MALEBE
IsiNdebele National Lexicography Unit, South Africa &
Gilles-Maurice DE SCHRYVER
Department of African Languages and Cultures, Ghent University, Belgium &
Department of African Languages, University of Pretoria, South Africa

Although the *Pan South African Language Board* (PanSALB) finalised the establishment of a *National Lexicography Unit* (NLU) for each of South Africa's 9 official African languages in 2000, work at especially the new units has not really come off the ground since then. This is surprising in the light of a series of pioneering articles published in that same year, articles dealing with both metalexicographic issues and practical aspects such as corpus-building and corpus-based lexicography. These publications were specifically written for and aimed at the prospective lexicographers of the 9 African-language dictionary units.

The first task in any lexicographic endeavour is to decide which items are to be treated in the envisaged dictionary, in other words, to draw up the macrostructure. It is widely accepted today that, on the one hand, the actual selection be made with a specific target-user group in mind, and that, on the other hand, the treatment of the items themselves be corpus-based. As several South African languages do not even have a single general-purpose dictionary, the target-user group is in most cases chosen to be as broad as possible, while the corpus is built in such a way that it is of a wide-ranging nature.

Although De Schryver & Prinsloo (2000: 299-302) propose a three-step methodology for the creation of a dictionary's macrostructure, departing from a raw corpus, their approach seems only truly feasible for those African languages for which the degree of conjunctiveness is not too high. In this paper a (four-step) methodology is therefore proposed for the creation of the lemma-sign list of a Nguni-language reference work. The theoretical principles are illustrated throughout with a full-scale case study revolving around isiNdebele.

The suggested methodology departs from a raw corpus and only requires standard, straightforward and widely-available software tools. In Step 1 top-frequency words are extracted from a corpus of running text. This step can be performed with versatile corpus query

software such as *WordSmith Tools*. In Step 2 the dictionary-citation forms are isolated from each of the top-frequency items; in Step 3 the dictionary-citation forms that are equal as well as their corresponding frequencies are brought together; and in Step 4 frequency bands are added to the lemma-sign list. Steps 2 to 4 can easily be performed with spreadsheet software such as Microsoft Excel. The four-step methodology was tested on real data and in real time, and the results indicate that the creation of the macrostructure of a desk-sized dictionary of a conjunctively-written African language need not take more than a month's work.

As case study, we opted for the creation of a macrostructure for isiNdebele – a language badly in need of a scientifically-sound lemma-sign list. Apart from the generic potential of the four-step methodology, the fact that such a list is now available for the very first time for isiNdebele, holds unprecedented promises. Indeed, the availability – right from the early stages of a lexicographic project – of a complete lemma-sign list of a projected reference work, enables the planning of an entire dictionary on a multitude of levels, viz. as regards the number of lemma signs, the number of pages, and the compilation time per alphabetical category. On the management level, the macrostructure can be used as a "ruler" with both *prediction* and *measurement* power. Not only can work be assigned evenly to the various compilers (prediction), but the compilers' performance can now also be computed precisely (measurement).

Finally, we will make suggestions as to how to proceed from here. A transfer of the macrostructure to a database will be suggested, and it will be indicated that one single database can hold various types of dictionaries simultaneously. Populating the database fields with a wide range of microstructural elements will enable any NLU to produce the dictionaries their communities are waiting for. A smooth yet sound methodology to do so has now become available for all African languages, whether these languages are written disjunctively or conjunctively.

PARALLEL SESSION (12)

Divergent Approaches to Corpus Processing: The Need for Standardisation

Esau MANGOYA
ALRI (African Languages Research Institute), Zimbabwe

The proposed paper seeks to focus on the processing of the Shona corpus. Shona is one of the major indigenous languages spoken in Zimbabwe. In corpus linguistics, texts of the written or spoken word is stored and processed on computer for purposes of linguistic research. This allows for research to be done on natural language. The corpus is a body of texts put together in a principled way. It becomes a language bank from which researchers retrieve data for various research purposes. With the corpus, data can be provided to give an authoritative body of linguistic evidence which can support generalisations and against which hypotheses can be tested. The paper would like to make an assessment of the processing of the Shona corpus and discuss how some aspects in the processing impact on the quality of the corpus.

The construction process of the corpus is long and different individuals are involved at different stages. The proposed study seeks to make an analysis of the Shona corpus looking at how different people processing the corpus handled particular aspects of the language at the

different stages of creation and processing. These are critical in determining the quality of the corpus. The paper would like to look at the different stages starting from interviewing, in the case of oral materials, and text writing, in the case of written texts. It will explore how the inputs at that initial stage determine the treatment of particular aspects of the texts in the later stages of the process, an aspect that will also have a bearing on the quality. It seeks to show how the different linguistic backgrounds of the processors affect the appreciation of some vital aspects of the corpus.

The paper will attempt to offer solutions which can be used to avoid or standardise all the efforts put in the processing of the corpus drawing from experiences from other disciplines from which standardisation has been found to be the norm. Team members are supposed to come together and draft manuals that state how issues in which there were divergences and inconsistencies are supposed to be standardised. The paper will try to show how lexicographers in the African Languages Research Institute (ALRI) at the University of Zimbabwe, who rely heavily on corpora have had to standardise certain aspects of their work to come out with standard practices. The current study would like to go a bit back and seek to find and show how the processing of the linguistic resource materials can be for the production of a quality corpus.

PARALLEL SESSION (13)

Prejudice and Reality in a Setswana Monolingual Dictionary – The Systematic and Deliberate Biasing of Cultural Issues

Godfrey Baile MAREME
SETNALEU (Setswana National Lexicography Unit), South Africa

The issue of culture is not as easy as it may seem to be. The culture of a certain group of people is their way of life. A way of life of a people encompasses what they do, but most of all the peculiar way of their language. I totally agree with the *Collins Dictionary of the English Language* (1982) that culture (2nd sense) is "the total range of activities and ideas of a group of people with shared traditions, which are transmitted and reinforced by the members of the group". On this note I will definitely differ with J. Alswang and A. van Rensburg in *An English Usage Dictionary for Southern African Schools* (1987) that culture is "the state of being civilized, having education and good taste." Language is the prime factor that distinguishes people. In most cases their way of speaking affects all other traits with psychological factors inclusive.

How can a dictionary of a people be more representative than another of the same language? A Setswana lexicographer has to take cognisance of the diversity of the dialects involved. Now, how is one dependent on a spoken corpus? This form of corpus will definitely highlight this diversity, which is not represented in the printed materials. Part of this corpus is the one regarded by the Purist and Traditionalist as unsuitable for the consumption of our people.

The culture of Setswana has it that there are certain words and natural acts of life, which are regarded as taboo. Some of these words are common in all the tribes of the Batswana. Some are not well known to others who are also Batswana. Such words are excluded in writing

but may be heard from one speech community to the other. For a Setswana lexicographer, this situation poses a dilemma.

The same Batswana have their own diverse practices amongst themselves. If a Motswana man marries a person from other cultures it can be understandable even if the cultures are to merge or one incorporates the other. But it is unbearable for two (2) Batswana to have some cultural differences. If a Mokgatla negotiates marriage with a Morolong and a Mohurutshe, heads are bound to roll. The practices of paying **lobola** and other issues are different. Hence, the terminology would differ and this needs to be documented. Such terms are worthy of one's corpus i.e. Mokwele, Thobela and Dira. These words are well known in the circles of Barolong and very few groups of Batswana.

If our target is a Setswana Monolingual Learners' Dictionary, can we afford to exclude issues which by tradition were not spoken of? Can it be a convincing book of resolutions? The Batswana keep the information to themselves and hope that one day it will be disseminated to the younger generations. How can our adults remain and die with the information the younger ones need, so that the latter would not opt for other cultures, which are more open and transparent?

As lexicographers, if we were to be judged by transparency issues can we regard ourselves as having done justice to our language? If yes, then we will be producing dictionaries of that era when kids were told that the babies were not born, but being dropped by an "airplane". We can turn out to be catalysts to the non-usage of dictionaries in our communities.

A questionnaire is going to be distributed amongst 100 students & staff wherein they are asked to name their expectations and what they will dislike in a new Setswana Monolingual Learners' Dictionary. Their age range will be between 18-60, fifty (50) of them from the UNW community, the rest from the North West College of Nursing. From that data I will draw my conclusions. My anticipation is that 90% of them will want a dictionary with real information.

At the end I will give examples of certain issues unmentionable in the culture of Setswana and yet in full existence. Not only in dictionaries, but in a whole range of translated works. Thus a new approach should be adopted in as far as transparency is concerned. The NLBs are there to expedite the process of language development and not to deter the progress of the NLUs. Can the NLBs give way to lexicographers recording the language as is?

PARALLEL SESSION (14)

Dictionaries Compiled with French and the Reproduction of the Gabonese Cultures

P.A. MAVOUNGOU, T. AFANE OTSAGA & G.-R. MIHINDOU
Groupe de Recherches en Langues et Cultures Orales (GRELACO), Gabon

The reproduction of culture in dictionaries constitutes one of the fundamental problems confronting lexicographers today. What is the nature of cultural data in dictionaries? To which extent should cultural aspects be transferred from one language to another? How should this transfer take place? This paper attempts to discuss the relevance of the reproduction of Gabonese cultures in the dictionaries compiled with French. One of the main

problems encountered by the compilers of these dictionaries was the transfer and the translation of some cultural aspects.

In order to discuss the nature and extent of cultural information in the existing Gabonese dictionaries, this article will restrict itself to the following focus areas:

Translation of different realities

In some Gabonese dictionaries the French Independence Day **14 juillet**, for instance, has been translated as 'emu awom be<u>n</u>in' ("14th July" in Fang). This translation does not make sense for a Fang speaker, who will not see the relation between the 14th of July and the French Independence Day. The best way to translate this concept is to use the paraphrase of meaning 'emu France a<u>ng</u>a nyo<u>ng</u> fili' ("the day France got freedom"). Numerous examples of the same kind can be found in the existing dictionaries compiled with Gabonese languages.

Role of culture in the change of meaning

Cultural gaps between Gabonese and European languages (French in particular) play an important role in the change of meaning of numerous current words. As far as French is concerned, many words do have another meaning in the Gabonese environment as compared to the meaning they have in the French society. The term **cadeau** for example, means firstly "present" or "gift". But in the Gabonese context, this term also means "free" or "gratis". In the existing Gabonese dictionaries, those cultural specificities have not been taken into account.

Dictionaries and cultural activities

Many dictionaries compiled with Gabonese languages where dialectal differences have been clearly established are biased toward one dialect. This is detrimental to the users of the speech community. When in a given dictionary macrostructural elements are from one dialect, users from the other dialects too often do not recognise themselves in the dictionary in question.

Dictionaries and cultural ethics

It is a well-attested fact that any dictionary should reflect the lexicon of the language being treated. In the existing dictionaries compiled with the Gabonese languages, one finds various terms referring to some cultural taboos (particularly about sex and some parts of the body). Under normal circumstances, Gabonese are extremely decent. The secret parts of the body are taboos and one speaks about it only with metaphors, euphemisms and other rhetorical expressions. It is part of the responsibility of the lexicographer to identify taboo terms and to warn the user against their uncivil nature.

Dictionaries and language registers

It is a common practice in dictionaries to mark, e.g., familiar, popular, and argotic words and expressions. In the majority of dictionaries compiled with Gabonese languages, one hardly finds marks signalling language registers.

Dictionaries and culture revival

Dictionaries try to satisfy the community curiosities. In many dictionaries compiled with the Gabonese languages, a lot of terms referring to old customs and activities are included. Dictionaries should make their users aware of traditional activities and serve as a valuable aid in the culture revival.

Dictionaries and the standardization of culture

For a language with several dialects, dictionaries are usually compiled in one dialect. The compilation of a dictionary in a dialect means that the lifestyle of people from that dialect will be presented in this dictionary. Unconsciously, dictionary users will be influenced by that dialect. By standardizing the language, dictionaries, also standardize the culture. In the Gabonese context, this situation has not been applied yet, either because, in many Gabonese languages, dictionaries do not have a long existence, or because the existing dictionaries are not available for the public at large.

Prior to discussing the above focus areas, a brief description of the dictionaries investigated will be given. After describing how cultural contexts have influenced lexicographers in the choice of macrostructural elements and their treatment, various cultural gaps between source and target languages in the existing dictionaries compiled with Gabonese languages are investigated. The paper concludes with the observation that the majority of existing lexicographical works tend to survey the full vocabulary of the language. Some words are treated in a satisfactory way in the sense that the lexicographical treatment that is offered gives an account of the underlying worldview of the people. For example, the following themes may be found: dietary practices, sexuality, mythology, traditional pharmacopoeia, kinship systems, hospitality, and respect for traditional authority and the elders. However, to be used in the most efficient way, these lexicographical publications need to be revised.

PARALLEL SESSION (15)

Compiling Dictionaries Using Semantic Domains

Ron MOE
Linguistics Consultant, SIL International (formerly Summer Institute of Linguistics)

The text corpus method of compiling dictionaries has much to recommend itself. However for many unwritten minority languages around the world the text corpus method must wait until a sufficiently large corpus is written and keyboarded. Dictionary development in such situations is generally very slow, since words are added to the dictionary one by one as they are encountered in the course of linguistic and lexicographical investigation. An unsystematic approach commonly results in small dictionaries that are uneven in their range and depth of coverage.

To facilitate dictionary development in such situations, the author has been developing a method of collecting a large percentage of the vocabulary of a language. At the heart of the method is a list of 1650 semantic domains. Under each domain the author has included a series of elicitation questions and sample words from English. The questions are based on lexical relations which link the words of the domain. The questions are simply worded, and the sample words are carefully chosen to exemplify the range of lexical items that might belong to the domain in any given language.

The method is effective because the words contained in the mental lexicon are tied together by a variety of lexical relations. These lexical relations tend to cluster around a central or important idea. So a semantic domain can be defined as 'an important idea and the words directly related to it'. Most of these domains have been found to be universal, since they are based on universal human experience. Considerable effort has gone into making the list extensive enough so that it can be used to reasonably classify any word from any language.

In January 2002 the method was used in a two-week workshop for Lunyole, a Bantu language of Uganda. In ten days 30 participants collected over 17,000 lexical items, representing approximately 14,000 unique entries. Since the words were collected domain by domain, the resulting word list was automatically classified. The word list can be efficiently expanded into a dictionary. For instance assigning a part of speech to each entry can be automated to a great extent because of the Bantu prefixal system.

Lexicographers have recommended investigating the semantics and pragmatics of words in lexical sets. This method is made easier since the words of the dictionary are already classified. Writing definitions or example sentences can be done for all the words of a domain at one time, ensuring consistency and revealing insights that would be overlooked if words were dealt with in isolation.

The primary purpose of the list of semantic domains is to collect words. However it can also be used to classify an existing dictionary. The popularity of Roget's Thesaurus in the English-speaking world testifies to the usefulness of publishing semantically organized dictionaries. The major publishing houses have begun publishing a variety of dictionaries organized to some degree by semantic domain.

PARALLEL SESSION (16)

Layered Definitions: A Northern Sotho Case Study

M.P. MOGODI
Sesotho sa Leboa National Lexicography Unit, South Africa

Lexicographers constantly strive to enhance the quality of definitions in monolingual dictionaries to best suit the needs and level of their target users. Landau (2001) clearly states that lexicography is not a theoretical exercise to increase the sum of human knowledge but a practical work to put together text that people can understand. He maintains that the definition must define and not just talk about the word or its usage. It must answer the question, "what is it?," directly and immediately.

The use of simple language by the lexicographer will help the reader to acquire the meaning of the defined word. As a result many dictionaries use a so-called 'defining vocabulary'. The compilers of such dictionaries claim that all defining words in their dictionaries are taken from this restricted vocabulary list. In terms of Zgusta (1971) this is one of the basic principles in lexicography, i.e. not to use words which are 'more difficult' than the word that is being defined. For example, in LDOCE3 it is clearly stated that all the definitions are written in "clear and simple language". LDOCE3 employs the *Longman Defining Vocabulary* of about 2000 common words. Likewise, a major aircraft manufacturer strictly limits definitions in their technical manuals to a much more restricted list of defining words. It

is thus clear that lexicographers should, when defining words, use a simpler language than the word defined. The word must be defined in such a way that the user will get all the answers to the questions that made him or her consult the dictionary. In COBUILD2 it is moreover stated that "[u]sers expect more and more from their dictionaries, and in particular want to gain confidence in using a word". The latter remark also underlines the responsibility of the lexicographer to supply enough encoding information to the user, and even more important, that this information should be on the level of the user.

Landau, while admitting the virtues of Zgusta's rule, i.e. that words which are more difficult than the word defined should not be used, also indicates that it is often impossible to apply this rule. His references to examples violating this rule will be briefly discussed. It will be illustrated that the lexicographer can easily err in his or her compilation of a dictionary in many ways, which results in definitions that are either too difficult or too basic for the target user.

The availability of corpora and the possibility of studying every lemma sign in context prior to the compilation of a definition revolutionised dictionary compilation. Firstly, utilisation of a main or 'general' corpus such as the *Pretoria Sepedi Corpus* (PSC) can help the lexicographer to write definitions for the average layperson, the typical general user of the dictionary. However, in addition to the main corpus, dedicated sub-corpora comprising of a representative sample of reference works used by different target user groups will give a clear indication of the level of compilation for such target users. This means that the different sub-corpora and in particular the corpus lines studied, will reflect the level on which the definition should be compiled right from the start.

The aim of this paper is to experiment with different ways of defining words in Northern Sotho on different levels, depending on the specific target users. The focus will be on three different levels of defining words, namely advanced, intermediate and junior levels. One of the questions to be answered is whether sufficient choice of defining vocabulary exists to present such layered definitions. Another aspect that will be looked in is what the impact on word economy will be for the various levels of defining.

To illustrate all the above, five words will be chosen from the field of **lenyalo** 'traditional wedding'. The behaviour of these words in PSC will be reviewed in terms of frequency-of-use and co-text. It will be attempted to define these words on three different levels, with elaborate motivation for the options selected.

PARALLEL SESSION (17)

The Etymological Aspects of the Idiomatic and Proverbial Expressions in the Lexicographic Development of Sesotho sa Leboa – A Semantic Analysis

V.M. MOJELA

School of Languages and Communication Studies, University of the North, South Africa

The idiomatic and proverbial expressions are important components of the oral tradition of Sesotho sa Leboa and, therefore, a knowledge of the literal meaning of words as they appear in our dictionaries without inclusion of their figurative meaning seems to be a shortcoming. An

idiom, or a proverb, has one basic meaning, viz. the meaning which the idiom is basically meant to refer to, but each proverb or idiom is made up of several lexical items. These lexical items have their own meaning which usually differ from the figurative sense of the proverb/idiom. Even though the meaning of the words in an idiomatic expression seem to differ from the sense of the idiom, there is a relationship to a certain extent. This is the relationship which the lexicographers can explain in their definitions in order to clarify both the literal and the figurative meanings of words in Sesotho sa Leboa.

This research is aimed at stressing the importance of having specialized dictionaries which will give the users detailed lexical explanations concerning the structure of the idiomatic and proverbial expressions as used in Sesotho sa Leboa. This type of dictionary should have definitions covering:

- The basic meaning of the expression.
- The relationship between the literal meaning of the expression and its real meaning.
- The etymological meaning of the individual words in the expression.

Etymology will play a major role in the determination of the relationships between the literal and figurative meaning of the lexical items contained in this special type of dictionary. Etymological data is based on the description of the term 'etymology' as described by scholars such as Svensén:

> "Information about the etymology of words tells us their history: how they were formed and evolved and finally took the shape and meaning they have in the language of today. Etymological facts lie along the time axis, and cut straight across the other information categories, combining elements from all of them: the formal, combinational, and semantic properties of words, and the things and events in the world outside language, all make their contribution." (Svensén 1993: 189)

For instance, the dictionary should answer the following questions concerning the following expression **Go ya ga maotwana hunyela** 'to die':

- What is the meaning of this expression?
- What is the literal meaning of this expression?
- How did this expression get this meaning?
- What is the meaning of the words **hunyela** and **maotwana**?
- How are these words related to 'death'?

If we were to answer the above questions regarding the idiom **Go ya ga maotwana hunyela**, the following would be relevant answers:

- The meaning of this idiomatic expression is 'to die', which is completely different from the literal meaning which is derived from the literal meaning of the individual words constituting this idiom.
- The literal meaning would be 'to go to the place called **maotwana hunyela**'.
- On the etymological level one can say that this expression got its meaning from the traditional method used by the Northern Sotho people for burying the deceased. The corpse sits up in the grave with the knees up against the chest (**go hunyela**), facing West with the entire body covered by the skin of a beast which is slaughtered specifically for this purpose. The beast's meat, called **mogoga**, is eaten with no salt immediately after the burial. This is the background (etymology) underlying the origin of this idiom.
- The word **hunyela** itself means 'to shrink', or in this case 'to bend or to squeeze the body (seated in the grave) with the chin leaning over the knees'. **Maotwana** (literally meaning

'small feet') refers to the feet of a corpse which shrink up in the grave and are covered by soil.

A dictionary of figurative expressions (idiomatic and proverbial) which has intensive definitions like in the abovementioned example will be a valuable tool for the dictionary users:

- As reference for the cultural usage of most of the known words used in figures of speech in Sesotho sa Leboa.
- As a storage of the cultural history of Sesotho sa Leboa in the form of words.

PARALLEL SESSION (18)

Cultural Relativism in Dictionaries: Is this the Right Direction?

L.E. MPHASHA

School of Languages and Communication Studies, University of the North, South Africa

Many Northern Sotho words in various dictionaries are not treated taking into account cultural attributes with which they are associated. Many dictionaries are bilingual and were largely written by the non-native speakers of the language. The verb **tsoga**, for instance, has different connotations. To most non-native speakers of the language it means 'to wake up', to the Christians it means 'to rise from the dead', but culturally it means 'to practice witchcraft'. The other cultural meaning may refer to 'the shopkeeper who was bankrupt but is now again prosperous'.

Ziervogel & Mokgokong's *Comprehensive Northern Sotho Dictionary* (1975) seems to be the only dictionary that contains descriptions of words in various connotations, including their cultural context. It explains meanings and gives good information about them. Most dictionaries, however, list pronunciations, derivations, illustrative quotations, synonyms and antonyms. As eleven African languages in South Africa are official, the necessity of new dictionaries which will explain the words both in general and cultural terms cannot be overemphasized. The cultural context of the words is important because the nature of the interaction between people and their culture is revealed.

Although one may talk of a 'general-purpose' dictionary, it must be realised that every dictionary is compiled with a particular set of users in mind. Many people ask for arbitrary decisions in usage choices, but a reasonable number of linguists feel that, when a dictionary goes beyond its function of recording accurate information on the state of the language, it really becomes a very bad dictionary. Many people, again, encounter dictionaries in the abridged sizes, commonly referred to as 'desk', 'pocket' or 'college-size' dictionaries. For the smaller-sized dictionary, the authors try to select words that are likely to be looked up.

Dictionaries are obliged to give different meanings of words of a language – the 'function words' (those which perform the grammatical functions in a language like pronouns, articles, prepositions, conjunction, etc.); and the 'referential words' (those which show entities outside the language system). Dictionaries have been criticised for not including sufficient cultural information in their explanations.

The social taboos, to a great extent, have also affected dictionaries vis-à-vis adequate interpretation of words. Some of the words which are commonly called obscene have been intentionally omitted, and thus irrational taboos have been strengthened. A perennial problem

in lexicography is the treatment of terms of ethnic insult. When a dictionary is compiled, it must be born in mind that its greatest value is to give access to the full resources of a language and should be seen as a source of information that will enhance free enjoyment of the mother tongue.

PARALLEL SESSION (19)

Cultural Aspects in the Shona Monolingual Dictionary: *Duramazwi Guru reChiShona*

Nomalanga MPOFU
ALRI (African Languages Research Institute), Zimbabwe

The proposed paper seeks to highlight the interrelationship that exits between language, lexicography and culture insofar as a lexicographer has to take into cognisance various cultural aspects, norms and taboos that are embodied in a language in the compilation of a dictionary. The paper will look at how the aspect of culture is interwoven in the practice of dictionary making, that is, how lexicographers handle cultural words and definition styles with the sole purpose of avoiding culturally offending words, yet at the same time aiming to produce as comprehensive and representative a work as possible.

Language is at the core of culture and it is also the major vehicle for the transmission of a people's cultural beliefs and values. Language is also an expression of social structures and attitudes. No culture can exist which does not have at its centre a natural language. A language thus reflects a particular culture. Culture in this paper will be taken to mean whatever a person must know in order to function in a particular society (Wardhaugh 1998: 215).

Two linguists, Edward Sapir (1949) and Benjamin Lee Whorf (1956) wrote extensively on the relationship that exists between language and culture. Their findings came to be referred to as the Sapir-Whorf hypothesis, which postulates that language and culture are inextricably related, so that one cannot understand or appreciate one without knowledge of the other. The paper will use this hypothesis as a point of reference. Though this paper will be informed by this hypothesis, it will not be a sociological study of language, but it will examine the issues from a lexicographic point of view.

The paper intends to look basically at three aspects:
(1) words in a cultural context, that is, the interrelationship between language and culture and its bearing on lexicography;
(2) the lexical entries in *Duramazwi Guru reChiShona* (2001) and how they act as a mirror to the beliefs and social structures of the Shona people (a dictionary, just as a language does, will also be taken as a mirror of the dominant way of thinking of the day);
(3) the definition language used in word categories such as offensive words and names used to refer to certain groups of people (e.g. the handicapped), and how a lexicographer marries cultural observation while still maintaining a balanced dictionary and one that is representative of the language of the day without compromising any information.

These lexical entries will be analysed in terms of cultural meaning, their place in the Shona society and their treatment in Shona lexicography. The paper will not only highlight these aspects but will also look at the problems and challenges that were encountered in handling

cultural aspects in this dictionary. Examples that will form the raw material for this paper will be drawn from the advanced monolingual Shona dictionary, *Duramazwi Guru reChiShona* and the African Languages Lexical (ALLEX) Project and the African Languages Research Institute (ALRI) Shona corpus. Reference will also be made to other Shona dictionaries, both monolingual and bilingual.

PARALLEL SESSION (20)

A Corpus-based Approach to Terminography: An Analysis and Evaluation of the ALRI Corpora

Vezumuzi kaDayisa NDLOVU
ALRI (African Languages Research Institute), Zimbabwe

This paper seeks to examine the efficacy of a relatively new approach to terminography, that is the corpus-based approach to terminographic works being undertaken by the African Languages Research Institute (ALRI). Corpus-based research is now the norm in many language-based fields like linguistics and lexicography. However, the same cannot be said for terminology. Terminology is in itself a relatively young discipline. In most Third World countries where the languages of the colonial masters are dominant, terminography, lexicography and other indigenous language studies are not being undertaken by the government but by independent institutes with the support of outside donors. This means that the language policies of these countries are inclined towards "colonial languages". In the case of Zimbabwe, English is the language of communication at all official levels. This therefore denies the indigenous languages the chance to grow and expand their vocabulary especially in specialised fields. Notwithstanding this policy hurdle, ALRI has taken it upon itself to undertake terminographic projects with a view to develop the indigenous languages terminologically.

This paper will look at the advantages of adopting a corpus-based approach to terminography. Presently the ALRI corpora stand at over 7 million running words. However these corpora have a literary rather than a specialised nature. Most of the corpora are made up of literary novels and oral interviews. This makes the corpora a weak tool in research on specialised subjects. There are only a few technical documents which were translated and inputted into the corpora. This paper will evaluate and analyse the nature of the ALRI corpora, both Shona and Ndebele, and determine whether they are suitable for terminographic purposes. For corpora to be useful to the field of terminography they have to be balanced in terms of the fields they cover, that is, they have to cover most of the spheres of human life including administration, business and commerce, technology, science, law, and various fields of knowledge. The size and representativeness of the corpora should also be suitable for such rigorous research.

Ways of improving the corpora so that they become balanced and useful for all language-based research will be offered. This includes the manner of tagging. When the corpora were collected the primary objective was their use in dictionary research and the tags used were meant to suit that purpose. However, if the corpora are to be utilised for terminographic research the tags have to be altered to make it possible to extract terms using

term extraction techniques. The paper will also look at the advantages, or the lack thereof, of creating parallel corpora for every specialised field instead of using one centralised corpus.

The main objective of the paper is to evaluate the advantages and the practicality of using a corpus-based approach to terminography given the fact that this field is relatively new in Zimbabwe and is mainly driven by the terminographers themselves and not by the users as has become the norm in developed countries.

PARALLEL SESSION (21)

The Lexicographic Treatment of Loan Words in Northern Sotho

Salmina NONG
Department of African Languages, University of Pretoria, South Africa

The aim of this paper is to investigate – from a lexicographic perspective – the preferences of Northern Sotho mother-tongue speakers for loan words *versus* 'traditional' or 'original' words in the language. Results obtained from a survey conducted among 100 speakers from different age and sex groups, backgrounds, places of residence, etc. will be analysed.

The use of loan words *versus* their (more) indigenous counterparts is studied by various disciplines such as science and technology, sociolinguistics, syntax and semantics, and not the least, lexicography. From a lexicographic angle the issue to be investigated links in well with one of the basic approaches in lexicography towards dictionary compilation, namely prescriptiveness *versus* descriptiveness. Within a descriptive approach towards dictionary compilation for an African language, it is imperative to know to what extent loan words in contrast to their 'traditional' or 'original' counterparts are actually and actively used, and to study preferences of the target users of dictionaries in this regard. Not only should the lexicographer strive to lemmatise and lexicographically treat words most likely to be looked for by the target user, (s)he should also be sensitive towards potential changes in preferences regarding the use of loan words *versus* more traditional ones.

Thus within a descriptive approach towards the lemmatisation of loan words in contrast to their 'traditional' or 'original' counterparts it is imperative for the lexicographer to know what the preferences of target users are in this regard.

A total number of 64 words were presented in pairs to the respondents, thus 32 pairs each containing a loan word and the more traditional equivalent, e.g. **radio** *versus* **seyalemoya** 'radio' or **mmotoro** *versus* **sefatanaga** 'car'. Respondents were asked to mark one or both alternatives which they thought should be included into a Northern Sotho dictionary. A third column was added for comments and suggestions of other terms considered to be even better than the two choices offered. Respondents were also invited to report spelling errors or to suggest improvement of spelling or even to motivate why a word should be included or excluded from the dictionary. Finally, an informal conversation was conducted with each respondent.

For each pair, all preferences were carefully calculated and studied especially in terms of the number of respondents who are in favour of using loan words only *versus* those who opted for

using only the original form, compared to respondents who accepted both. The extra information obtained from the additional notes and supplementary conversations were also meticulously documented.

As a next step the frequency count of each of these words, singular as well as plural forms, were taken from the *Pretoria Sepedi Corpus* (PSC) and compared to the respondents' preferences. Lastly, the treatment (or lack thereof) in 9 Northern Sotho dictionaries was investigated.

Analysis of the respondents' comments reveals that there is a general preference towards using original words. A number of respondents feel that loan words should only be used if a reasonable equivalent does not exist. Some even suggest that words should be coined in order to have a Northern Sotho word instead of an adoptive from other languages. Younger respondents tend to accept loan words and original words on a more equal basis. Minor differences exist between the preferences of male *versus* female participants. Frequency counts in PSC reveal an overwhelmingly preference for original words, while loan words and original words are treated on a fairly equal basis in currently available dictionaries.

In order to verify the quality of the feedback a number of carefully selected distracters were built into the questionnaire. Some are both loan words, some do not have the same meaning, and others were added just to find out whether or not the respondents themselves were trustworthy.

PARALLEL SESSION (22)

The Lemmatization of Copulatives in Northern Sotho

D.J. PRINSLOO
Department of African Languages, University of Pretoria, South Africa

For learners of Northern Sotho as a second or even foreign language, the copulative system is probably the most complicated grammatical system to master. The encoding needs of such learners, i.e. to find enough information in dictionaries in order to actively use copulatives in speech and writing, are poorly served in currently available dictionaries. The aim of this paper is to offer solutions to the lemmatization problems regarding copulatives in Northern Sotho and to propose guiding entries for paper and electronic dictionaries which could serve as models for future dictionaries. It will be illustrated that the maximum utilisation of macrostructural and microstructural strategies as well as the mediostructure is called for in order to reach this objective. Prerequisites will be to reconstruct the entire copulative system in a user-friendly way, to abstract the rules governing the use of copulatives and to isolate the appropriate lemmas. The treatment of copulatives in Northern Sotho dictionaries will also be critically evaluated, especially in terms of frequency of use and target users' needs.

It is advisable for the lexicographer (and the compiler of a basic Northern Sotho grammar) to use the user's presumed basic knowledge of the noun class system and the moods, tenses and aspects of common verbs as a point of departure. Learners normally master the nominal and verbal systems first when studying an African language. For dynamic copulatives, **ba(go)**, **be(go)** and **bile(go)** should be lemmatized. For the static copulative, **ké**, **ga se**, a

reduced list of subject concords, for their copulative use, as well as **le(go)**, **se(go)**, and **na(go)** should be lemmatized.

Copulatives in Northern Sotho appear thousands of times in the *Pretoria Sepedi Corpus*. 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.

It will be argued that although entries for copulatives in currently available Northern Sotho dictionaries are technically correct, they only offer limited decoding information and very little encoding information. Lacking in all these dictionaries is treatment of copulatives in the back matter of the dictionary and appropriate use of the mediostructure. The discussion of copulatives in the back matter should fulfil the basic purpose of cross-reference, namely to be the reference address where the user would indeed find more information on copulatives, structured in such a way that it extends the information the user has obtained in consulting the article of the copulative in the central text. For the encoding user it should thus be 'the next logical step' in explaining the correct use of the copulative. Likewise, the back matter should also be the logical step/link to the outside source – thus a comprehensive process from dictionary article to back matter to outside source. In paper dictionaries this does not narrow the gap between dictionary and grammar but at least offers logical steps to the user in the information retrieval process.

In contrast to the paper dictionary, an electronic dictionary can offer the user an exciting new range of data-access routes to the dictionary. The encoding needs of users who look up copulatives in electronic dictionaries for Northern Sotho can, for example, be satisfied by means of pop-up screens.

Compiling user-friendly dictionaries of a high lexicographic standard for African languages poses a great challenge to prospective lexicographers. They are the mediators between complicated grammatical structures and the decoding and encoding needs of their target users. Complicated structures such as nouns, verbs, copulatives, etc. should not be tackled haphazardly as they cross the compiler's way. They should be carefully studied and even researched to obtain a comprehensive overview of the relevant structures. Only then can the lexicographer proceed to planning the macrostructure and microstructure for the lemmatization of a specific construction. 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 and decoding users. The mediostructure should be employed in a sensible way to refer the user to reference addresses where more information can be found. Special attention should be given to references to a well-compiled back matter where cohesion of decontextualised items is restored, thus rendering the 'full picture' to the user.

PARALLEL SESSION (23)

Culture-specific Concepts in Technical Dictionaries

Peter A. SCHMITT

Institut für Angewandte Linguistik und Translatologie, Universität Leipzig, Germany

Common misconceptions in translation studies are the assumptions (1) that "culture" and the idea of "culture-specific concepts" (which are widely accepted for literature translation) are not relevant in the area of technical (in the sense of technological or engineering-related) translation, (2) that technical terms are well-defined or even standardized, (3) that their underlying concepts are interlingually congruent, (4) that translating technical terms is mainly a matter of 1:1 equivalences and mere code-switching, and (5) that making technical dictionaries is, as a consequence, relatively straightforward.

This notion is reflected in the traditional and still mainstream terminological idea (as propagated by Wuester) in which the designation (or term) is language-bound whereas the concept (or significate) is an abstraction independent not only of language but also of culture. This notion also supports the widespread belief that it is possible to produce multilingual term banks. Or to automatically generate several language pairs from two existing pairs. Obviously such multilingual term banks and multilingual technical dictionaries do exist, but this does not prove that they are more reliable and useful than multilingual general dictionaries (which, for good reasons, do hardly exist).

Using examples from renowned technical dictionaries the presentation demonstrates that this somewhat naive approach to technical terminography may give rise to considerable (confusing and potentially costly and even dangerous) misunderstandings. The paper will explain the idea of a *tertium comparationis* which bridges the culture barrier between two semiotic triangles which are embedded in the source and target languages, respectively. In this model, a concept is culture-bound, i.e. its characteristics (features) are not necessarily identical to a corresponding concept in another language (and culture). A simple and striking example is the German term **Hammer** and its obvious English counterpart **hammer**. These terms (or frames) evoke different concepts (or scenes) in the minds of German and English speakers, because the prototypes associated with these words are different - the words are related to concepts with different features. There are, of course, much more complex concepts to be dealt with in technical communication, where culture-specific features are less obvious and much more difficult to explain. It depends on the context whether these differences are relevant or not.

To show the practical application of this approach and a better alternative to many existing technical dictionaries, the paper will comment on culture-sensitive entries in the author's technical dictionaries, including the *PONS Dictionary of Automotive Engineering* and the brand-new two-volume *Langenscheidt's Dictionary Technology and Applied Sciences* (both are large two-volume English-German / German-English dictionaries). The paper will also address the genesis of these dictionaries which are both generated and maintained by means of a custom-designed multi-media terminology management system (www.cats-term.com).

PARALLEL SESSION (24)

The OED as Cultural Icon

Penny SILVA
Oxford University Press, United Kingdom

The first edition of the *OED* was published between 1884 and 1928. The dictionary's 'historical principles' applied the new Darwinian theories of evolution to language, and as well as being a remarkable lexicographical achievement, the text is an icon of Victorian self-confidence, a monument to the British culture of the time. Obviously the bulk of the *OED* defines the central, common vocabulary of English – for example, the words descended from Anglo-Saxon, Danish, and French – in a new and scholarly detail, explaining word origins and semantic development over the centuries and providing quotations from canonical works to illustrate their changing history.

However the *OED* also records the minutiae of British culture – the intriguing local dialect words (sometimes limited to a town or district), the vocabulary of local occupations (both obsolete and extant), and the esoteric slang and other terms limited to the public school and Oxbridge educational systems. Added to these is a comparatively modest sprinkling of 'colonial' terms, reflecting the vocabulary of Empire – and particularly those words which have made their way into general English speech. These items are often treated in a rather less comprehensive manner, compared with the detail presented in British items – for example, the origin of many words from Africa, South America, or Asia is simply given as 'Native language'. The *OED* text was compiled with the educated, British reader's perspective in mind, and while the words descended from European languages were comprehensively recorded and thoroughly etymologised, those from more distant sources were not described in the same detail.

The first edition needs to be seen as a truly remarkable achievement, but also to be understood within the context of its period. The nature of the English-speaking world has changed enormously since 1928. The first ever comprehensive revision of the *OED* began in the early 1990s, and the revised edition is aiming to record English from a more inclusive perspective, as it is spoken across the world. Many more items important to the English used in Australasia, Africa, Asia, and North America are not only being included, but are also receiving treatment as thorough as that formerly given to the specifically British English vocabulary. This process is broadening and deepening the *OED's* documentation of the very varied English vocabulary, and of the cultures of the areas where English has taken root. Where in the first edition the unconscious attitude of 'us' and 'other ' might sometimes be perceived, in the third edition an attempt is being made to describe English as an inter-linked web of varieties rather than as a British 'parent' with many 'children'. In addition, the range of texts used to provide examples of English in use has widened, and now includes sources such as film and television scripts, broadsheets, and popular novels. This paper looks at the cultural change in recreating the *OED* as an icon of modern lexicography, and provides examples of the effects (and challenges) of the new editorial policy as the third edition is gradually compiled and published as *OED Online*.

PARALLEL SESSION (25)

Semi-Automatic Term Extraction for the African Languages, with special reference to Northern Sotho

Elsabé TALJARD
Department of African Languages, University of Pretoria, South Africa &

Gilles-Maurice DE SCHRYVER
Department of African Languages and Cultures, Ghent University, Belgium &
Department of African Languages, University of Pretoria, South Africa

Worldwide, semi-automatically extracting terms from corpora is becoming the norm for the compilation of terminology lists, term banks or dictionaries for special purposes. If African-language terminologists are willing to take their rightful place in the new millennium, they must not only take cognisance of this trend but also be ready to implement the new technology. In this paper it is advocated that the best way to do the latter two at this stage, is to opt for computationally straightforward alternatives (i.e. use 'raw corpora') and to make use of widely available software tools (e.g. *WordSmith Tools*).

The main aim of the paper is therefore to discover whether or not the semi-automatic extraction of terminology from *untagged* and *unmarked* running text by means of *basic* corpus query software is feasible for the African languages. In order to answer this question a full-blown case study revolving around Northern Sotho linguistic texts is discussed in great detail. The computational results are compared throughout with the outcome of a manual excerption, and vice versa. Attention is given to the concepts 'recall' and 'precision'; different approaches are suggested for the treatment of single-word terms *versus* multi-word terms; and the various findings are summarised in a Linguistics Terminology lexicon.

Upon comparison of the manual outcome with the computational results, it will be shown that, for the case study, 74% of the single-word linguistic terms and an astonishing 83% of the multi-word linguistic terms can indeed be extracted semi-automatically. These high figures are obtained with basically just three software tools: WordList, KeyWord and Concord, all part of WordSmith Tools (Scott 1999). Based on this case study one is thus bound to conclude that the semi-automatic extraction of (unlemmatised) terms for the African languages is a *viable* endeavour indeed.

It will also be pointed out that human beings will always remain the final judges in any terminological activity, whether that endeavour be manual or computational. The terms proffered by the software will always need to be scrutinised by the terminologist. Conversely, however, it will be indicated that the research revealed rather surprisingly that the software can *isolate* potential terms and *force* the terminologist to *consider* term status in ways that are less obvious when wading manually through running text. This turns out to be especially valid for multi-word terms, as more than 40% of the multi-word linguistic terms are seemingly missed during manual excerption. Viewed from this angle, the semi-automatic extraction of terms for the African languages is not only viable, but even *crucial* in order to counteract inevitable human errors.

Finally, the various outcomes of the research presented in this paper will be summarised in a tiny special-field *lexicon* in which the terms are listed in their *lemmatised* form. In that lexicon there are 98 terms that were only excerpted manually, 50 that were only extracted computationally, and 187 that were retrieved both manually and computationally. This means that, out of the 335 lemmatised terms, 285 or thus *85% were excerpted manually*, and 237 or thus *71% were extracted computationally*. The difference between the two approaches (14%) is smaller than the number of items not retrieved in either approach. There can thus be no doubt that, when looking at the end product, semi-automatically extracting terminology for and *in* the African languages is indeed a worthwhile venture.

PARALLEL SESSION (26)

A Southern African Lexicographer's Working Definition of the Term Culture

Pieter N. VAN DER WESTHUIZEN
Xhosa Dictionary Project, University of Fort Hare, South Africa

The need for orientation as to one's understanding of culture is addressed. The need becomes more pronounced as one participates in a lexicography conference with the theme "Lexicography and Culture" keeping in mind both how much confusion reigns because of the wide frame of reference for this concept as well as one's desire as a lexicographer to use "culture" in a terminologically effective manner.

A brief survey is done of the definitions of culture as found in dictionaries used in South Africa. This ranges from translation equivalents such as civilization, tradition, custom and art to extended definitions including the sum of human experience as the essence of culture.

An overview is given of the definition offered by the South African specialist dictionary of P.J. & R.D. Coertze, *Verklarende vakwoordeboek vir antropologie en argeologie* (1996). Its salient features are identified and those of importance to lexicography are examined and assessed. The following come to the fore:

The role and nature of the ethnos in its various guises as creator of culture are investigated. The dynamic of the production of culture and the role of the lexicographer in this process is explored. Both the compass and complexity of the phenomenon culture are surveyed and it is shown how language figures in the overall picture as only one of the many aspects in the cultural make-up of an ethnos. On the other hand it is shown that cultures manifest themselves in time, place and extent in widely divergent forms, each the product of a particular ethnos endeavouring to adapt itself to its unique environment (in it broadest sense) so as to survive and effectively sustain itself.

The lexicographer is always part and parcel of this dynamic and this involvement is viewed from the following perspectives. The lexicographer as:

1. Member of an ethnos with an own cultural identity
The following bindings govern and in certain instances restrict the lexicographer as member of an ethnos:

1.1 an own cultural connection or allegiance,
1.2 a particular historical deployment,
1.3 a specific geo-political connection,
1.4 consciousness of an own identity.

These bindings and the limitations they impose need to be recognized by the lexicographer. They are assets that in certain instances provide unique insights and perspectives which enhance performance. Recognition of the limitations on the other hand is often the stimulus for more strenuous lexicographic effort.

2. Broker of culture
The authority ascribed to dictionaries by dictionary users, whether desirable or not, gives the lexicographer inordinate influence as an exegete of or commentator on the culture or cultures

being served by a particular language. Not only does the own culture benefit by enculturation with an addition to the heritage in the form of a dictionary but often acculturation effects the widening of the cultural horizon. In the South African multi-cultural situation the lexicographer is at times expected to act as agent of change to facilitate and accelerate both the processes i.e. enculturation and acculturation.

3. Cultural resource manager

The lexicographer in no small way is called to manage an extremely important resource of the culture of an ethnos or people, i.e. its lexicon. This resource is not only a repository of a large proportion of the artefacts of a culture but is also the instrument which facilitates intercourse with other cultural communities. The lexicographer therefore manages and may at times even manipulate the process of access to the subset of reality which the language brings into focus. This subset is, despite its being unique and also vital to its creators, only a relatively meagre part of reality. The lexicographer manages this cultural resource with the potential to enrich its creators by facilitating and improving the enculturation process as well as contributing to the acculturation effect.

A clear working definition will assist the lexicographer in understanding both the potential and the responsibility of the lexicographic endeavour.

PARALLEL SESSION (27)

The Value of Culture in the Development of Medical Corpora in Zulu

Linda VAN HUYSSTEEN
Department of African Languages, University of South Africa, Pretoria, South Africa

In this paper, the manner in which lexicographical development in Zulu is linked to aspects such as culture is investigated. The 'African Renaissance' can hardly be considered complete without the development of the African languages to their fullest potential. The facilitation of lexicographical projects in these languages is a means of achieving such a goal. In this context of the African Renaissance, it is thus appropriate to discuss some aspects of lexicographical development in one of the continent's most well-known languages, Zulu.

Besides the analytical study of the methods of word-formation in Zulu and the effective use of computerised tools such as the application of frequency counts, concordances, corpus annotation, etc., the practice of lexicographical development in Zulu will not be complete without mentioning the way in which such development is linked to extra-linguistic factors such as culture.

For the purpose of this paper, the term 'culture' is used in its widest general application. However, culture is not to be separated from language because:

> "Language can be studied not only with reference to its formal properties ... but also with regard to its relationship to the lives and thoughts and culture of the people who speak it." (Gregerson 1977: 56)

World view and taboo for instance, are two culture-related aspects which should be taken into account in any type of lexicographical development.

According to the Whorfian hypothesis, a person's mother-tongue offers him/her a framework for his/her perception of the environment or *world view*. Although this hypothesis is yet to be proven, some examples of its application in lexicographical development can clearly be evidenced. However, Hudson (1980:104) offers some perspective to Whorf's views: "We dissect nature by our communicative and cognitive needs rather than by our language." At times these needs have to be fulfilled by developing new terminology within a language. One way to do this is by adjusting existing linguistic items by means of 'semantic expansion'. In this case the existing meaning, of a sometimes culturally-bound word, acquires an expanded or modified meaning in order to name a new, related concept.

Taboo is an umbrella term to refer to terms that are unsuitable for use in a specific identified register or social context. In the African languages of South Africa the application of the concept of taboo is very important for lexicographers who have to devise and eventually lemmatise new terms for sex education, specifically with relation to AIDS. According to Zulu culture, it is taboo to refer to terms with a sexual connotation in a direct fashion. Euphemism is then used to describe the taboo term by an inoffensive expression (an **Inhlonipho**), in order to show respect through avoidance.

In lexicographical practise, by extending the literal meaning of the Zulu word **ucansi** (reeds mat / sleeping mat), for instance, both the concepts of world view and taboo in relation to culture can be captured. A cultural object such as a ' reeds mat / sleeping mat' **ucansi**, clearly reflecting a certain type of world view, is therefore used to indirectly and evasively refer to terms with a sexual (taboo) connotation such as **isifo socansi** (sexually transmitted disease).

It is thus important that culture-related aspects such as world view and taboo deserve prominent mentioning in the form of explanatory notes in the front or back matter of dictionaries in order to complement lemmas and the conditions of their use.

PARALLEL SESSION (28)

Making a Dictionary of an Oral Tradition

Jacques A.J. VAN KEYMEULEN
Woordenboek van de Vlaamse Dialecten (WVD), Ghent University, Belgium

In the sixties and seventies of the last century three major regional dialect dictionaries were started in the southern part of the Dutch language area: the *Dictionary of the Brabant Dialects* (1960-, KUNijmegen / KULeuven, covering the provinces of Northern Brabant in the Netherlands and Antwerp and Flemish Brabant in Belgium), the *Dictionary of the Limburg Dialects* (1961-, KUNijmegen / KULeuven, covering the provinces of Limburg in both the Netherlands and Belgium) and the *Dictionary of the Flemish Dialects* (1972-, RUGent, covering Western and Eastern Flanders in Belgium, Zealand Flanders in the Netherlands and French Flanders in France). The areas of the three dictionaries are geographically complementary. It should be noted that the official language of the Flemings is Dutch; Flemish is used either in a restricted sense for a group of dialects (as is the case in dialectology), or for the 'accent' of Flemings when speaking standard Dutch (as is the case in popular speech and in sociolinguistics).

On the whole, the three projects are set up along the same lines, both with regard to data collection and presentation. The dictionaries are arranged systematically (with alphabetical indices) and have three main sections: I. Agricultural vocabulary; II. Technical vocabularies; III. General Vocabulary. Each editorial board has already published many dictionary fascicles, each one devoted to a specific subject (e.g. 'Ploughing', 'Coopering', 'Birds', etc.).

The dialects of the southern part of the Dutch language area do not have a written tradition, hence the dictionaries cannot be based on a text corpus. The bulk of the data consists of answers to questionnaires filled in by hundreds of volunteering dialect speakers. To this word-collection are added the words taken from other sources (including older dictionaries) since 1880. The whole of the word material is divided in 'lexicographical relevant concepts', which form the basis of the dictionary articles. Thus, every publication consists of a series of concepts relating to a certain conceptual field. For every concept, the heteronymy (= the different lexemes for the same concept) in the different dialects of the area under investigation is presented, including general indications as to frequency and localization. Two types of entry forms are used: Dutchified headwords and so-called lexical variants.

The focus of my lecture will be on the methods of data collection. After a brief summary of metalexicographic considerations (**Why** such a dictionary?), I will briefly discuss the answers given to 5 questions concerning the data collection (**Where**, **Who**, **What**, **How** and **How much**) which are important with regard to macro- and microstructure.

The five questions pertain to geographical matters (Where), the respondents (Who), the selection of the concepts – not of the words! (What), the questionnaire/field work (How) and the representativeness (How much). The emphasis will be on What and How. I will especially dwell on the way the extra-linguistic world is classified in conceptual fields, on the inventory of concepts resulting in onomasiological questionnaires and on the two-phased field work. Of course, I will bear in mind the potential applicability of the 'Flemish' experience for other dictionaries of oral language traditions.

PARALLEL SESSION (29)

**Missionary Influence on Shona Lexicography,
with Special Reference to Father Hannan's Problem of Translation**

Advice VIRIRI
Department of African Languages and Culture, Midlands State University, Zimbabwe

In this paper, I wish to discuss how developments in Shona lexicography during the colonial era have been influenced by the missionaries which later gave birth to the ongoing process of compiling monolingual dictionaries at the African Languages Research Institute (ALRI, formerly the ALLEX project). Various efforts were employed by the missionaries which did not only signal the beginning of an economically exploitative relationship between "the West and the rest of us" but it also had ancillary cultural consequences (Dathorne 1975: 3). Their motive towards the development of African literatures in general and the Zimbabwean lexicographic work in particular were primarily evangelical and not to give impetus to creative writing.

The development of orthographies ushered in an epoch of literary translation that marked the beginning of African-languages literature. A missionary secretary stationed in Rhodesia writes:

> "the books that are in great demand are bibles, hymn books and catechisms. They are regarded by the people as so clearly a part of the necessary apparatus of a Christian that they purchase them without murmur. The Pilgrims Progress enjoys a steady sale in almost every African vernacular into which it has been translated" (Dathorne 1975: 2)

Thus the American Boards of Missions translated and published biblical extracts in Zulu in 1846 while the Berlin Lutheran Missionaries did the same in Northern Sotho. It has been noted that as a result of missionaries settling at different mission stations, a variety of approaches to the making of dictionaries for use by native Zimbabweans were developed. These mission stations represented the different Shona dialects spoken in Zimbabwe, namely ChiZezuru, ChiNdau, ChiManyika and ChiKaranga. The Chikorekore dialect was not represented in the writing by missionaries because the Zambezi Valley where the Korekore people live was Tsetse-fly infested, where there were thus problems of malaria.

Shona lexicography was supposed to play a crucial role in the standardisation of the Shona language. By standard Shona I mean:

> "… a language that has developed a common system of writing or orthography (i.e. spelling, word division and also punctuation) which, when implemented, allows people who speak different varieties of the same languages to write in the same way, while still allowing for stylistic and other variation, as in the choice of vocabulary" (Chimhundu 1992: 87)

The main institution that worked for a common orthography and lexicography was the Southern Rhodesia Missionary Conference which first met in 1903. The main purpose of their meeting was to secure a translation that could be used in all dialects of Mashonaland so as to *"obviate the expense of preparing the Bible in different dialects"* (Fortune, quoted by Magwa 2002: 3).

For the language to develop into a standard one meant serious human efforts were required and that is why among Doke's recommendations was the encouragement to *"unify orthography and pull the vocabularies"* (Doke, quoted by Magwa 2002: 3).

The paper also seeks to evaluate Father Hannan both as a lexicographer and translator in his dictionary titled *Standard Shona Dictionary* and his translation of the Shona Bible. In Father Hannan's dictionary, the inconsistencies and inadequacies are caused by *"the complex language situation in the Shona speaking community"* (Chimhundu 1979: 75). The major problem faced by Father Hannan in his lexicographic and translation work was thus a result of the heterogeneous nature of language and its fluid social situation.

The dictionary-making process among the Shona people came a long way in quest for a monolingual one. This became a dream come true thanks to Dr. Herbert Chimhundu and his team's relentless efforts at ALLEX / ALRI. The monolingual dictionaries prepared by them could not have seen the light if Father Hannan had not published his Standard Shona Dictionary whose main aim, according to the compiler, was:

> "to record Shona words in standard Shona spelling. It has been our aim also to provide by means of the generous number of examples of the use of words, illustrations of the applications of the principles of word division on which standard spelling is based" (Hannan 1959: ix).

Constant changes in Shona orthography further affected Father Hannan's valuable contribution in his dictionary and this was coupled with translation problems caused by the instability of the state of the language. I will conclude with Chimhundu's thoughts succinctly summed up when he says: *"Cultures differ, change and interact, and languages must adapt accordingly 'to suit the occupancy of a new personality'"* (Chimhundu 1979: 8).

[To Table of Contents](#)

Correspondence

›

AFRILEX
African Association for Lexicography
Department of African Languages
University of Pretoria
Pretoria, 0002
Republic of South Africa

Tel.: + 27 12 420 2320
Fax: + 27 12 420 3163
E-mail: prinsloo@postino.up.ac.za (Chairperson:
    Prof. D.J. Prinsloo)
WWW:
    http://www.up.ac.za/academic/libarts/afrilang/homelex.html
    (Home Page AFRILEX)

›

Mariëtta **Alberts**
Systems Development & Research
National Language Service
Department of Arts, Culture Science and
    Technology (DACST)
Private Bag X894
Pretoria, 0001
Republic of South Africa

Tel.: 012 337 8166
Fax: 012 324 2119
E-mail: vt05@dacst5.pwv.gov.za (Dr. Mariëtta
    Alberts)

›

Henning **Bergenholtz**
Centre for Lexicography
The Aarhus School of Business
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

E-mail: hb@asb.dk (Prof. Henning Bergenholtz)
WWW:
    http://www.hha.dk/EOS/LEXC/STAFF/HB_SPROG_FORM.HTM
    (personal)

›

Sonja E. **Bosch**
Department of African Languages
University of South Africa (UNISA)
Pretoria, 0003
Republic of South Africa

E-mail: boschse@unisa.ac.za (Prof. Sonja E.
    Bosch)

›

Emmanuel **Chabata**
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 303298
Fax: +263 4 333674 || +263 4 333407
E-mail: echabata@arts.uz.ac.zw (Mr. Emmanuel
    Chabata)

›

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

E-mail: gillesmaurice.deschryver@rug.ac.be
WWW:
    http://www.up.ac.za/academic/libarts/afrilang/elcforall.htm
    (Electronic Corpora for African-Language
    Linguistics)

›

Rachélle **Gauton**
Department of African Languages
University of Pretoria
Pretoria, 0002
Republic of South Africa

Tel.: 012 420 3715 (W) || 012 361 3355 (H)
Fax: (012) 420 3163
E-mail: rgauton@postino.up.ac.za (Dr. Rachélle
    Gauton)

›

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

Tel.: 021 808 2164
Fax: 021 808 3815
E-mail: rhg@akad.sun.ac.za (Prof. Rufus H.
    Gouws)

›

Karen **Hendriks**
Private Bag X82079
Rustenburg, 0300
Republic of South Africa

Tel.: 014 592 1365 (W) || 014 590 5165 (H)
Cell: 083 355 6404
Fax: 014 592 7647
E-mail: k.hendriks@mweb.co.za (Ms. Karen
    Hendriks)

›

Arvi **Hurskainen**
Box 59, FIN-00014
University of Helsinki
Helsinki
Finland

Tel.: +358 9 191 22677
Fax: +358 9 191 22094
E-mail: arvi.hurskainen@helsinki.fi (Prof. Arvi
    Hurskainen)

›

Gregory **James**
Director, Language Centre
Hong Kong University of Science and Technology
Clear Water Bay
Kowloon
Hong Kong SAR
China

Tel.: +852 2358 7878
Fax: +852 2335 0249
E-mail: lcgjames@ust.hk (Prof. Gregory James)

›

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

Tel./Fax: +27 46 603 8107
E-mail: k.kavanagh@ru.ac.za (Ms. Kathy
    Kavanagh)
WWW: http://www.ru.ac.za/affiliates/dsae/
    (DSAE)

›

Langa **Khumalo**
Head: Ndebele Lexicography Unit
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 333652 || +263 4 303211 ext.
    1780/1788
Fax: +263 4 333674 || +263 4 333407
E-mail: langa@arts.uz.ac.zw ||
    la_nga@yahoo.co.uk (Mr. Langa Khumalo)

›

Diapo N. **Lekganyane**
Department of Northern Sotho
University of Venda
Private Bag X5050
Thohoyandou, Venda
Republic of South Africa

Tel.: 015 962 8578

Cell: 0822022953
Fax: 0159 22045
E-mail: nelsonl@univen.ac.za (Dr. Diapo N.
    Lekganyane)

›

Matete **Madiba**
Academic Development Practitioner
Technikon Northern Gauteng
Private Bag XO7
Pretoria North, 0116
Republic of South Africa

Tel.: 012 799 9293
Fax: 012 799 9167
E-mail: matete@tnt.ac.za (Ms. Matete Madiba)

›

P.S. **Malebe**
IsiNdebele National Lexicography Unit
Department of African Languages
University of Pretoria
Pretoria, 0002
Republic of South Africa

Tel.: 012 420 3944
Fax: 012 420 3163
E-mail: smnguni@postino.up.ac.za (Ms. P.S.
    Malebe, c/o Mr. P. Mnguni)

›

Esau **Mangoya**
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 303298
Cell: +263 11721880
Fax: +263 4 333674 || +263 4 333407
E-mail: emangoya@arts.uz.ac.zw (Mr. Esau
    Mangoya)

›

Mandlenkosi **Maphosa**
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 303298
Fax: +263 4 333674 || +263 4 333407
E-mail: mandlamaphosa@yahoo.com (Mr.
    Mandlenkosi Maphosa)

›

Godfrey Baile **Mareme**
Caretaker-Chief Editor
SETNALEU-PANSALB (U.N.W.)
Private Bag X2046
Mmabatho, 2735

Republic of South Africa

Tel.: 018 389 2343
Cell: 082 200 78 83
Fax: 018 389 2504
E-mail: maremeg@unw001.uniwest.ac.za (Mr. Godfrey Baile Mareme)

›

Kwena **Mashamaite**
School of Languages and Communication Studies
University of the North
Private Bag X1106
Sovenga (Polokwane/Pietersburg), 0727
Republic of South Africa

E-mail: mashamaitek@unin.unorth.ac.za (Mr. Kwena Mashamaite)

›

Webster **Mavhu**
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 303298
Cell: +263 4 023274136
Fax: +263 4 333674 || +263 4 333407
E-mail: webma@arts.uz.ac.zw || vhezh2000@yahoo.com (Mr. Webster Mavhu)

›

P.A. **Mavoungou**
Department of Afrikaans and Dutch
University of Stellenbosch
Private Bag X1
Matieland, 7602
Republic of South Africa

Cell: 083 745 1906
E-mail: 13126733@sun.ac.za (Mr. P.A. Mavoungou)

›

Ronald E. **Moe**
P.O. Box 44456
00100 Nairobi
Kenya

Tel.: +254 2 714 943 (W) || +254 2 719 045 (H)
Cell: 0733 757633
Fax: +254 2 718 220
E-mail: ron_moe@sil.org (Mr. Ronald E. Moe)

›

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

Tel.: + 27 12 420 3076
Fax: + 27 12 420 3163
E-mail: pmogodi@postino.up.ac.za (Ms. M.P.
    Mogodi)

›

V.M. **Mojela**
School of Languages and Communication Studies
University of the North
Private Bag X1106
Sovenga (Polokwane/Pietersburg), 0727
Republic of South Africa

Tel.: 015 268 3108
E-mail: mojelav@unin.unorth.ac.za (Dr. V.M.
    Mojela)

›

L.E. **Mphasha**
School of Languages and Communication Studies
University of the North
Private Bag X1106
Sovenga (Polokwane/Pietersburg), 0727
Republic of South Africa

E-mail: mojelav@unin.unorth.ac.za (Mr. L.E.
    Mphasha, c/o Dr. V.M. Mojela)

›

Nomalanga **Mpofu**
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 303298
Fax: +263 4 333674 || +263 4 333407
E-mail: nmpofu@arts.uz.ac.zw (Ms. Nomalanga
    Mpofu)

›

Vezumuzi kaDayisa **Ndlovu**
ALRI (African Languages Research Institute)
University of Zimbabwe
PO Box MP 167
Mt Pleasant
Harare
Zimbabwe

Tel.: +263 4 303298
Fax: +263 4 333674 || +263 4 333407
E-mail: vezie@arts.uz.ac.zw ||
    vezieask@yahoo.com (Mr. Vezumuzi kaDayisa
    Ndlovu)

›

A.C. **Nkabinde**
PO Box 117
Thornville, 3760
Republic of South Africa

Fax: 033 2510 751 (Attention: Prof. A.C.
    Nkabinde)

›

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

Tel.: + 27 12 420 3076
Fax: + 27 12 420 3163
E-mail: snong@postino.up.ac.za (Ms. Salmina
    Nong)

›

Laurette **Pretorius**
Department of Computer Science and Information
    Systems
University of South Africa (UNISA)
Pretoria, 0003
Republic of South Africa

Tel.: 012 429 6727
E-mail: pretol@unisa.ac.za (Prof. Laurette
    Pretorius)

›

D.J. **Prinsloo**
Department of African Languages
University of Pretoria
Pretoria, 0002
Republic of South Africa

Tel.: + 27 12 420 2320
Fax: + 27 12 420 3163
E-mail: prinsloo@postino.up.ac.za (Prof. D.J.
    Prinsloo)
WWW:
    http://www.up.ac.za/academic/libarts/afrilang/elcforall.htm
    (Electronic Corpora for African-Language
    Linguistics)

›

Daniel **Ridings**
Bräckavägen 17
SE-437 42 Lindome
Sweden

E-mail: daniel.ridings@swipnet.se ||
    daniel_ridings@yahoo.se (Dr. Daniel Ridings)

›

Justus C. **Roux**
Director: Research Unit for Experimental
    Phonology
University of Stellenbosch
Stellenbosch, 7599
Republic of South Africa

Tel.: 021 808 2017
Cell: 083 2888 602
Fax: 021 808 3975
E-mail: jcr@sun.ac.za (Prof. Justus C. Roux)
WWW: http://www.ast.sun.ac.za (African Speech
    Technology) || http://www.sun.ac.za/nefus
    (Research Unit for Experimental Phonology)

›

Peter A. **Schmitt**

Institut für Angewandte Linguistik und
    Translatologie
Universität Leipzig
Augustusplatz 10/11
D-04109 Leipzig
Germany

Tel.: +49 341 973 7600
Fax: +49 341 973 7649
E-mail: schmitt@rz.uni-leipzig.de (W) ||
    pas@paschmitt.de (H) (Univ.-Prof. Dr. Peter A.
    Schmitt)
WWW: www.ialt.de (W) || www.paschmitt.de
    (personal)

›

Penny **Silva**
Director, Oxford English Dictionary
Oxford University Press
Great Clarendon Street
Oxford OX2 6DP
United Kingdom

Tel.: +44 1865 267236
Fax: +44 1865 267811
E-mail: silvap@oup.co.uk (Ms. Penny Silva)

›

Bronson **So** Ming Cheung
Hong Kong University of Science and Technology
Clear Water Bay
Kowloon
Hong Kong SAR
China

Tel.: +852 9612 5512
E-mail: ma_smc@stu.ust.hk (Mr. Bronson So
    Ming Cheung)

›

Elsabé **Taljard**
Department of African Languages
University of Pretoria
Pretoria, 0002
Republic of South Africa

Tel.: 012 420 2494 (W) || 012 332 1357 (H)
Cell: 082 353 6906
Fax.: 012 420 3163
E-mail: e.taljard@freemail.absa.co.za

›

Pieter N. **van der Westhuizen**
PO Box 320
Adelaide, 5760
Republic of South Africa

Tel.: 046 684 0105 || 040 602 2559
Cell: 82 200 3591
Fax: 040 653 2038
E-mail: pwesthuizen@ufh.ac.za (Rev. Pieter N.
    van der Westhuizen)

›

Linda **Van Huyssteen**
Department of African Languages

University of South Africa (UNISA)
PO Box 392
Pretoria, 0003
Republic of South Africa

Tel.: 012 429 8258 (W) || 012 662 0145 (H)
Cell: 07 222 97 303
Fax: 012 429 3221
E-mail: vhuysl@unisa.ac.za (Ms. Linda Van
    Huyssteen)

›

Jacques A.J. **Van Keymeulen**
Emanuel Hielstraat 81
B-9050 Gentbrugge
Belgium

Tel.: +32 9 264 4081(W) || +32 9 231 1364 (H)
Fax: +32 9 264 4170
E-mail: jacques.vankeymeulen@rug.ac.be (Dr.
    Jacques A.J. Van Keymeulen)

›

D.J. **van Schalkwyk**
Editor-in-Chief
Bureau of the *Woordeboek van die Afrikaanse
    Taal* (WAT)
P.O. Box 245
Stellenbosch, 7599
Republic of South Africa

Tel.: 021 887 3113
Fax: 021 883 9492
E-mail: wat@wat.sun.ac.za (Dr. D.J. van
    Schalkwyk)
WWW: http://www.sun.ac.za/wat/index.htm
    (Home Page WAT)

›

Advice **Viriri**
Department of African Languages and Culture
Midlands State University
P. Bag 9055
Gweru
Zimbabwe

Tel.: +263 54 60409
E-mail: adviriri2002@yahoo.co.uk (Mr. Advice
    Viriri)

›

Jill **Wolvaardt**
Associate Editor
Dictionary Unit for South African English (DSAE)
Rhodes University
PO Box 94
Grahamstown, 6140
Republic of South Africa

Tel./Fax: +27 46 603 8107
E-mail: jill@aardvark.ru.ac.za (Ms. Jill Wolvaardt)