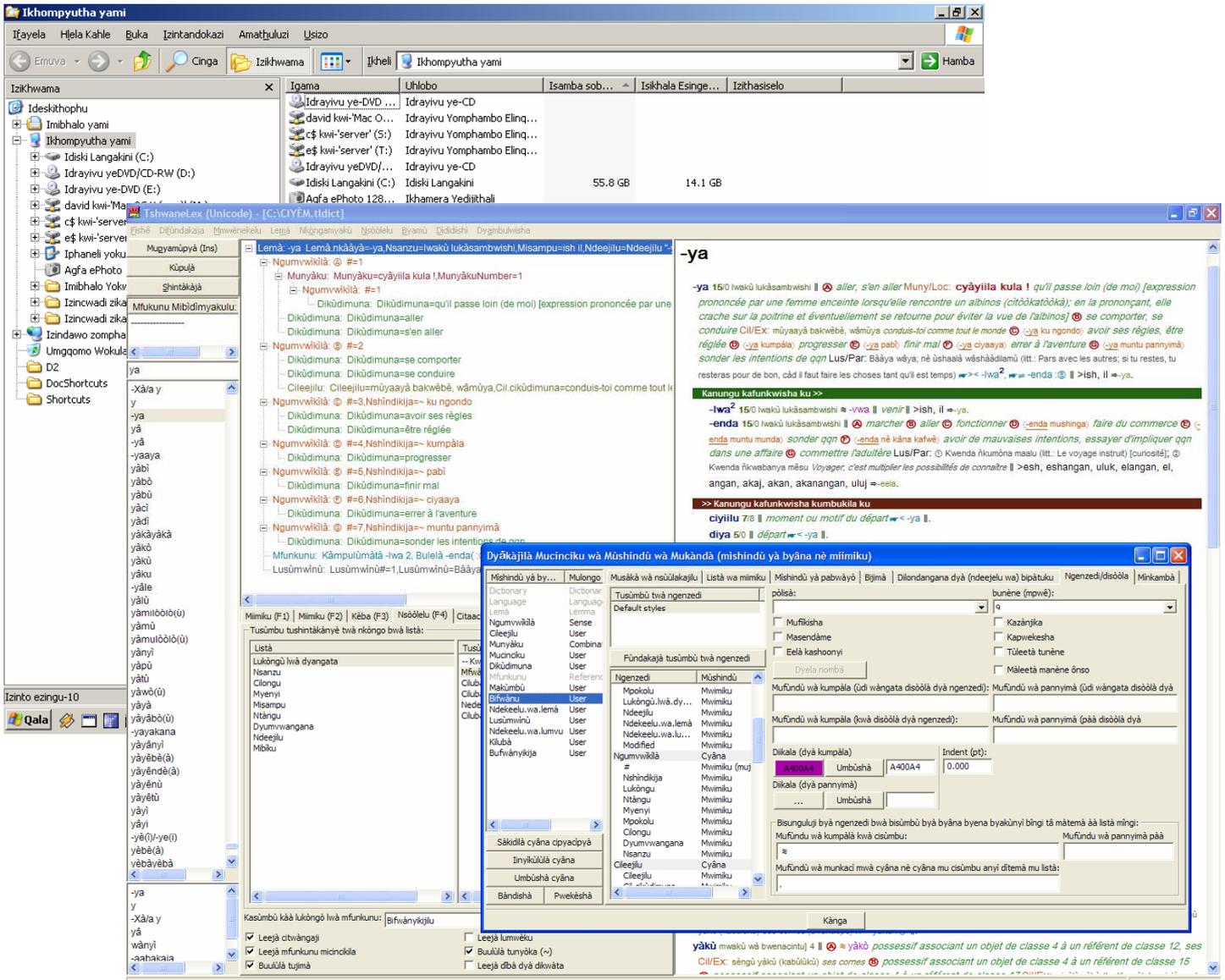


African Association for Lexicography

AFRILEX 2006

The User Perspective in Lexicography



Programme & Abstracts

edited by
Gilles-Maurice de Schryver



11th International Conference of the
African Association for Lexicography

AFRILEX 2006
The User Perspective in Lexicography
Programme & Abstracts

Date: 3-7 July 2006
Organisation: University of Venda for Science and Technology,
Thohoyandou, Republic of South Africa
Local Conference Organiser: Shumani L. Tshikota
Abstract Reviewers: AFRILEX Board
Programme Committee: AFRILEX Board

edited by
Gilles-Maurice de Schryver

Copyright © 2006 by the African Association for Lexicography

ISBN 0-620-36760-1

Pretoria: (SF)² Press

Cover: (1) *Microsoft* LIP (language interface pack) for isiZulu

(2) *TshwaneLex* GUI (graphical user interface) for Cilubà

A FEW WORDS FROM THE PRESIDENT

Afrilex welcomes you to our 11th International Conference which also marks our 11th year of existence. We are proud to be a member of the international *-lex* family and to present you with this Conference Abstracts Booklet, once again meticulously compiled and edited by Gilles-Maurice de Schryver.

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

Afrilex greetings,

Mariëtta Alberts

Table of Contents

Programme	8
Pre-Conference Workshop	12
PRINCIPLES AND PRACTICE OF SOUTH AFRICAN LEXICOGRAPHY	12
Panel Discussions	13
THEME 1: ELECTRONIC DICTIONARIES FOR AFRICAN LANGUAGES	13
THEME 2: PUBLISHERS AND LEXICOGRAPHERS	13
Keynote papers	14
Chimhundu, Herbert ■ Types, Uses and Users of Dictionaries: An African Perspective	14
Sobkowiak, Włodzimierz ■ E-Dictionaries and Phonolexicographic Needs of EFL Users	15
Parallel and Special sessions	16
Alberts, Mariëtta ■ Standardisation, Modernisation and Harmonisation – Users’ Perspectives	16
Bailey, Dwayne ■ Creating a South African Keyboard	17
Beyer, Herman L. ■ Lexicographical Parameters for the Systematic Description of User-friendliness in Dictionaries	18
Bosch, Sonja E., Laurette Pretorius & Jackie Jones ■ Machine-readable Lexicons for Morphological Analysis of the South African Bantu Languages: A Data Model	19
Chikomwe, Justice ■ ALLEX Online Dictionaries: The New Cycle in African Lexicography at ALRI	20
de Schryver, Gilles-Maurice ■ Internationalisation, Localisation and Customisation Aspects of the Dictionary Application TshwaneLex	21
Ekwa Ebanéga, Guy-Modeste & Fatima Tomba Moussavou ■ Survey of Dictionary Use: Case Studies of Gabonese Students at the University of Stellenbosch and the Cape Peninsula University of Technology	25
Gouws, Rufus H. & Henning Bergenholtz ■ Lexicography and Language Policy	26
Klein, Juliane ■ Implementing a Dictionary Culture for South Africa – Attempt at a General Framework for the Official Languages	27
Lepota, Biki ■ The Treatment of Cardinal Number Plurals in Sepedi	28
Mabaso, P. ■ Considerations in the Preparation of a Shangani Dictionary	29
Mafela, Munzhedzi James ■ L.T. Marole: A Forgotten Pioneer in Tshivenda Lexicography	30
Mahlangu, K. Sponono ■ The Lemmatisation of Loan Words in an isiNdebele Bilingual Scholars’ Dictionary	32
Mangoya, Esau ■ A Look at the Historical Influence on Shona Dictionary Products	32
Mheta, Gift ■ The Problems and Challenges of Dealing with Encyclopaedic Information in Shona Lexicography: The Case of <i>Duramazwi reMimhanzi</i>	33
Mihindou, Guy-Roger ■ A Hybrid LSP Dictionary with Microstructural Amalgamation to Enhance User-friendliness	34

Mojela, Maropeng V. ■ Polysemy or Homonymy? A Case Study of the Challenges Pertaining to the Lemmatisation of Polysemous and Homonymous Lexical Items in the Compilation of a Sesotho sa Leboa – English Bilingual Dictionary	35
Motaung, Ntaoleng ■ Microsoft Language Interface Packs (LIPs) for South African Languages	37
Moyo, Nobuhle ■ Some Problems Encountered in the Compilation of the First Specialised Dictionary: The <i>Isichazamazwi Sezomculo</i> (ISM)	38
Mpofu, Nomalanga ■ The User Perspective in Lexicography: Comparison Between a General Purpose Dictionary and a Specialised Terminological Dictionary	39
Nkomo, Dion ■ The Compiler and User Perspectives on the Access Structure of <i>Isichazamazwi Sezomculo</i> (ISM)	40
Nyangone Assam, B. ■ The User Perspective in a Learners' Dictionary	41
Nyathi, Cratilwe & Marble Timbe ■ Steps and Procedures Involved in Data Processing for Lexicographers	42
Nyota, S. and J. Mapara ■ The Impact of Lexicographic Work on Language Use: A Case of Shona Monolingual Dictionaries in Zimbabwe	43
Prinsloo, Danie J. & Elsabé Taljard ■ The Value of Northern Sotho Definitions in LSP Dictionaries for Mathematics and Science: A Preliminary Analysis	45
Roux, Justus C. ■ Human Language Technologies and the Standardisation of Lexical Databases	47
Sibula, Pumlani ■ Furthering the Aims of Multilingualism through Integrated Terminology Development	48
Wolvaardt, Jill ■ Horses for Courses: Researching User Perspectives – Different Approaches for Different Dictionaries	48

Programme

Monday 3 July 2006: Workshop Day 1

PRINCIPLES AND PRACTICE OF SOUTH AFRICAN LEXICOGRAPHY 1/2 Rufus H. Gouws & Danie J. Prinsloo

09:00 – 10:30	Material collection and corpus building The user-perspective
10:30 – 11:00	Tea
11:00 – 13:00	The frame structure of dictionaries Macrostructural aspects
13:00 – 14:00	Lunch
14:00 – 15:30	Microstructural aspects
15:30 – 16:00	Tea

Tuesday 4 July 2006: Workshop Day 2 & Two Panels

PRINCIPLES AND PRACTICE OF SOUTH AFRICAN LEXICOGRAPHY 2/2 Rufus H. Gouws & Danie J. Prinsloo

09:00 – 10:30	Definitions The access structure of dictionaries
10:30 – 11:00	Tea
11:00 – 13:00	Equivalent relations
13:00 – 14:00	Lunch

Panel 1

14:00 – 15:15	ELECTRONIC DICTIONARIES FOR AFRICAN LANGUAGES Convenor: Sonja E. Bosch
---------------	--

- Panellists:
- Justice Chikomwe (Zimbabwe)
 - Gilles-Maurice de Schryver (South Africa & Belgium)
 - Phillip Louw (South Africa)
 - Maropeng V. Mojela (South Africa)
 - Justus Roux (South Africa)
 - Włodzimierz Sobkowiak (Poland)

15:15 – 15:45	Tea
---------------	-----

Panel 2

15:45 – 17:00	PUBLISHERS AND LEXICOGRAPHERS Convenor: Rufus H. Gouws
---------------	--

- Panellists:
- Laetitia Botha (PESA)
 - Philip Louw (OUPSA)
 - Jana Luther (Pharos)
 - Sizakele Ncoko (MML)

Wednesday 5 July 2006: Excursion

05:30 – <i>late</i>	FULL-DAY EXCURSION TO KRUGER NATIONAL PARK
---------------------	--

Thursday 6 July 2006: Conference Day 1

- 08:30 – 09:15 Tea and Registration
09:15 – 09:30 Welcome – President of Afrilex (Mariëtta Alberts)

OFFICIAL OPENING

Acting Vice-Chancellor of the University of Venda for Science and Technology (M.D.R. Ralebipi)

Keynote Address 1

09:30 – 10:25	Włodzimierz Sobkowiak (Poland) E-Dictionaries and Phonolexicographic Needs of EFL Users
---------------	---

- 10:30 – 11:00 Tea
& Demo 1:
Multi Access Dictionary (Włodzimierz Sobkowiak)

Parallel Sessions

11:00 – 11:25	Lexicography and Language Policy Rufus H. Gouws & Henning Bergenholtz	Machine-readable Lexicons for Morphological Analysis of the South African Bantu Languages: A Data Model Sonja E. Bosch, Laurette Pretorius & Jackie Jones
11:30 – 11:55	Survey of Dictionary Use: Case Studies of Gabonese Students at the University of Stellenbosch and the Cape Peninsula University of Technology Guy-Modeste Ekwa Ebanéga & Fatima Tomba Moussavou	Steps and Procedures Involved in Data Processing for Lexicographers Cratilwe Nyathi & Marble Timbe

12:00 – 12:25	A Hybrid LSP Dictionary with Microstructural Amalgamation to Enhance User-friendliness Guy-Roger Mihindou	Human Language Technologies and the Standardisation of Lexical Databases Justus C. Roux
12:30 – 12:55	Lexicographical Parameters for the Systematic Description of User-friendliness in Dictionaries Herman L. Beyer	ALLEX Online Dictionaries: The New Cycle in African Lexicography at ALRI Justice Chikomwe

13:00 – 14:00 Lunch

Parallel Sessions

**Special Session 1:
The user perspective(s)**

14:00 – 14:25	The Value of Northern Sotho Definitions in LSP Dictionaries for Mathematics and Science: A Preliminary Analysis Danie J. Prinsloo & Elsabé Taljard	Horses for Courses: Researching User Perspectives – Different Approaches for Different Dictionaries Jill Wolvaardt
14:30 – 14:55	The Treatment of Cardinal Number Plurals in Sepedi Biki Lepota	The User Perspective in Lexicography: Comparison Between a General Purpose Dictionary and a Specialised Terminological Dictionary Nomalanga Mpfu
15:00 – 15:25	Polysemy or Homonymy? A Case Study of the Challenges Pertaining to the Lemmatisation of Polysemous and Homonymous Lexical Items in the Compilation of a Sesotho sa Leboa – English Bilingual Dictionary Maropeng V. Mojela	The User Perspective in a Learners' Dictionary B. Nyangone Assam
15:30 – 15:55	Furthering the Aims of Multilingualism through Integrated Terminology Development Pumlani Sibula	The Compiler and User Perspectives on the Access Structure of <i>Isichazamazwi Sezomculo</i> (ISM) Dion Nkomo

16:00 – 16:30 Tea

& Demo 2:
Launch of TshwaneLex 2.0 (Gilles-Maurice de Schryver)

16:30 – 17:30 **ANNUAL GENERAL MEETING**

18:30 for 19:00 **Gala Dinner**

Friday 7 July 2006: Conference Day 2

Keynote Address 2

08:30 – 09:25	Herbert Chimhundu (Zimbabwe) Types, Uses and Users of Dictionaries: An African Perspective
---------------	--

Parallel Sessions

Special Session 2:

Localisation

09:30 – 09:55	Some Problems Encountered in the Compilation of the First Specialised Dictionary: The <i>Isichazamazwi Sezomculo</i> (ISM) Nobuhle Moyo	Internationalisation, Localisation and Customisation Aspects of the Dictionary Application TshwaneLex Gilles-Maurice de Schryver
10:00 – 10:25	The Impact of Lexicographic Work on Language Use: A Case of Shona Monolingual Dictionaries in Zimbabwe S. Nyota and J. Mapara	Creating a South African Keyboard Dwayne Bailey

10:30 – 11:00

Tea

& Demo 3:

ALLEX Online Dictionaries (Justice Chikomwe)

Parallel Sessions

11:00 – 11:25	L.T. Marole: A Forgotten Pioneer in Tshivenda Lexicography Munzhedzi James Mafela	Microsoft Language Interface Packs (LIPs) for South African Languages Ntaoleng Motaung
11:30 – 11:55	Considerations in the Preparation of a Shangani Dictionary P. Mabaso	Implementing a Dictionary Culture for South Africa – Attempt at a General Framework for the Official Languages Juliane Klein
12:00 – 12:25	A Look at the Historical Influence on Shona Dictionary Products Esau Mangoya	The Lemmatisation of Loan Words in an isiNdebele Bilingual Scholars' Dictionary K. Sponono Mahlangu
12:30 – 12:55	The Problems and Challenges of Dealing with Encyclopaedic Information in Shona Lexicography: The Case of <i>Duramazwi reMimhanzi</i> Gift Mheta	Standardisation, Modernisation and Harmonisation – Users' Perspectives Mariëtta Alberts

13:00 – 13:10

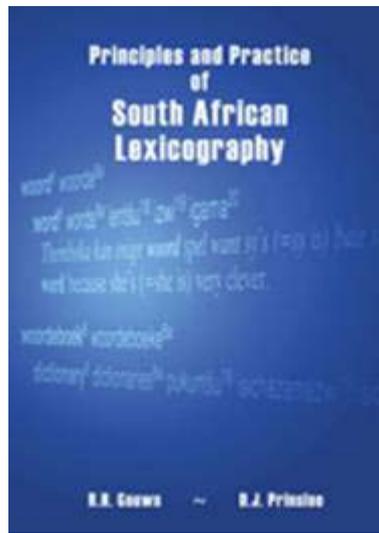
CLOSURE

13:15 – 14:15

Lunch

Pre-Conference Workshop

PRINCIPLES AND PRACTICE OF SOUTH AFRICAN LEXICOGRAPHY



Presenters

- Prof. Rufus Gouws (University of Stellenbosch)
- Prof. Danie Prinsloo (University of Pretoria)

Contents

During this workshop, fundamental issues in South African lexicography are discussed. The discussion is based on the newly-published (2005) book *Principles and Practice of South African Lexicography* by R.H. Gouws and D.J. Prinsloo.

The book *Principles and Practice of South African Lexicography* is directed at experts in the field of practical and theoretical lexicography in South Africa, applying the general theory of lexicography to the South African lexicographic environment.

Themes addressed during the workshop include the following:

- material collection and corpus building,
- the user-perspective,
- the frame structure of dictionaries,
- macrostructural and microstructural aspects,
- definitions,
- equivalent relations, and
- the access structure.

Panel Discussions

THEME 1: ELECTRONIC DICTIONARIES FOR AFRICAN LANGUAGES

Convenor

Prof. Sonja E. Bosch

Contents

When thinking of dictionaries, the first thing that usually comes to mind is paper dictionaries – space constraints, expensive to print, heavy to carry, bulky to store, cumbersome for look-up, especially for newcomers to the field of African languages. But what does the future in the context of human language technologies and digital communication hold for dictionaries in the African languages? This question is the central issue of the panel discussion “Electronic dictionaries for African languages”.

Panellists

Panellists include international experts in the field of electronic dictionaries as well as experts in the field of African languages and lexicography. The following scholars have confirmed their participation in this panel:

- Convenor: Sonja E. Bosch (UNISA – University of South Africa, Pretoria, South Africa)
- Justice Chikomwe (ALRI – African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe)
- Gilles-Maurice de Schryver (TshwaneDJe HLT, Pretoria, South Africa & UGent – Ghent University, Ghent, Belgium)
- Phillip Louw (OUPSA – Oxford University Press Southern Africa, Goodwood, Cape Town, South Africa)
- Maropeng V. Mojela (Sesotho sa Leboa NLU, University of Limpopo, Sovenga, South Africa)
- Justus Roux (SU-CLaST – Stellenbosch University Centre for Language and Speech Technology, Stellenbosch, South Africa)
- Włodzimierz Sobkowiak (Adam Mickiewicz University, Poznan, Poland)

THEME 2: PUBLISHERS AND LEXICOGRAPHERS

Convenor

Prof. Rufus Gouws

Contents

In this panel discussion representatives of different publishing houses have the opportunity to inform us about the lexicographic endeavours and plans of their respective companies.

Panellists

The following representatives have confirmed their participation in this panel:

- Convenor: Rufus H. Gouws (Stellenbosch University, Matieland, Stellenbosch, South Africa)
- Laetitia Botha (PESA – Pearson Education South Africa, Pinelands, Cape Town, South Africa)
- Philip Louw (OUPSA – Oxford University Press Southern Africa, Goodwood, Cape Town, South Africa)
- Jana Luther (Pharos – Pharos Woordeboeke / Dictionaries, Roggebaai, Cape Town, South Africa)
- Sizakele Ncoko (MML – Maskew Millar Longman, Pinelands, Cape Town, South Africa)

Chimhundu, Herbert ■ Types, Uses and Users of Dictionaries: An African Perspective

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

This paper discusses types of dictionaries against presumed vis-à-vis assessed needs and it links lexicography with terminology and language standardization. The presenter questions the relevance of such terms as lexicographic tradition and dictionary culture with reference to marginalized language communities. He argues that, even in the post-colonial era, African languages and cultures remain marginalized. In such a situation, needs assessment and prioritization of types of dictionaries to be produced become a chicken and egg affair because what is made available is what becomes known and appreciated. Products of the fourteen-year old lexicographic experiment known by the acronym ALLEX will be used to show that the views of academic reviewers of new types of dictionaries are sometimes very different from feedback that subsequently filters back from the target users in the speech community. Specific examples will be presented in relation to two general monolingual Shona dictionaries, *Duramazwi reChiShona* (Chimhundu 1996) and *Duramazwi Guru reChiShona* (Chimhundu 2001).

Accordingly, frequent reference will be made to Zimbabwe in order to draw from the experiences of the African Languages Lexical Project (ALLEX Project), which has produced these dictionaries and seven others, since its inception in 1992. The Project, which is now housed at the African Languages Research Institute (ALRI) at the University of Zimbabwe (UZ), has since shifted focus from general language dictionaries to specialized or terminological dictionaries that are intended for use in different subject areas and specific domains. Generalizations will be drawn from the experiences of the ALLEX Project in order to highlight the potential for replication of the pioneering work of ALRI-UZ in other languages of the region.

The main purpose of the paper is to give an assessment of the impact of monolingual lexicography on language raising, with particular reference to standardization and harmonization. It will be argued that monolingual lexicography is not only standard setting but also a form of empowerment for a language community, which must be prioritized for African languages. Further, the presenter will attempt to show that monolingual lexicography has accelerated the process of standardizing the literary written form of ChiShona to the extent that the harmonization its varieties is now a viable proposition. Reference will be made to the work that is now being undertaken jointly by ALRI and CASAS (the Centre for Advanced Studies of African Society) on the harmonization of the Shona varieties that are spoken in Zimbabwe, Mozambique and Botswana. Part of the reason why speakers of all these varieties accept the harmonization program is the success and impact of the ALLEX experiment, which has helped to reverse negative attitudes by Shona speakers towards their own language, and to make them aware of the affinity between the different varieties of the language by simply avoiding dialect labelling in the dictionary entries while also being widely inclusive in headword selection.

All nine of the ALLEX dictionaries that have been compiled so far are prototypes in the sense that, as far as African languages are concerned, they are experimental models for different uses by different mother-tongue users, particularly with reference to education and training. In the absence of an explicit national language policy or an established dictionary culture or a lexicographic tradition established by mother-tongue researchers and compilers, the determination of the types of dictionaries that the Project prioritized and the style manual that was developed for each dictionary were based on presumed rather than researched needs of the target users in the language community. Not surprisingly, when the dictionaries started

to appear, reactions were mixed. The assessment of the impact of the ALLEX experiment on language raising that will be made here will take into account all forms of feedback that the project team has received since 1996, ranging from written reviews of the dictionaries by academics to comments and suggestions made by various users or collected in the field of research from school teachers and college lecturers during workshops that were held with reference groups for the more recent sub-projects focusing on specialized terminologies.

Sobkowiak, Włodzimierz ■ E-Dictionaries and Phonolexicographic Needs of EFL Users

School of English, Adam Mickiewicz University, Poznań, Poland

The phonetic aspect of (EFL) dictionaries is among the most seriously underrated and underdeveloped in (meta)lexicography. Pertinent bibliography is scant and even the best learner dictionaries are found wanting on a number of counts. This contribution is both a summary of my 12-year-long research into (pedagogical) phonolexicography (see bibliography at <http://elex.amu.edu.pl/~sobkow/public.htm>) and a look ahead. I present the current state-of-the-art in phonolexicography with particular attention paid to how the leading pedagogical EFL e-dictionaries relate to the actual and potential phonolexicographic needs of their users, both students and teachers.

The main themes are: (a) the specificity of phonolexicographic needs of EFL users (e.g.: are there EFL dictionaries for teachers?), (b) phonetic representation, both graphic and acoustic, in dictionaries and its problems, (c) phonetic access, i.e. querying the contents of the dictionary via the phonetic code, (d) didactic aspects of phonolexicographic information, i.e. its use in teaching and learning (EFL) pronunciation.

Under (a), a summary of pertinent research is made showing the subordinate nature of pronunciation among self-reported dictionary user needs, but also a stable minority of users eager to run pronunciation queries. I also speculate on the characteristics of the in-spe “(pronunciation) teacher’s dictionary”, as opposed to the learner’s one.

Under (b) I deal, among others, with (i) the long-standing question of the choice and implementation of phonetic transcription in EFL dictionaries, (ii) the representation of phonostylistic variation in headwords, definitions and examples, (iii) the most common errors and inconsistencies in dictionary phonetic representations, e.g. the mismatch of transcription with audio recording, (iv) the L1-sensitivity in the treatment of pronunciation in dictionaries, (v) paraphonetic information, such as difficulty indexes, frequency and style indicators, the phonetic representation of inflections, and the like.

Under (c), my original idea of ‘phonetic access’ conceived of in 1994 (Sobkowiak 1994), and now increasingly implemented in EFL dictionaries, is presented, including the potential structural scope of such access (headwords vs. all dictionary text) and typical available query types (e.g. one-off lookup vs. wordlist generation). Some criticism is voiced of the implementation of phonetic access in the best EFL e-dictionaries currently on the market, and some speculations about its future are made.

Under (d), I briefly look at the existing and potential pedagogical use of phonolexicographic information in EFL e-dictionaries. I also voice an appeal, based on my recent research (Sobkowiak, in press) for a wider use of examples and definitions in monolingual dictionaries as text corpora to feed into a semi-automatic generator of (pronunciation) tasks and exercises for learners and teachers. Some exemplary queries are illustrated.

Finally, an appeal is made for (1) more phonolexicographic research, (2) wider, but reasoned, use of hypermedia (hypertextualised multimedia) in electronic EFL dictionaries, (3) novel standards and methods in electronic pedagogical phonolexicography.

References

- Sobkowiak, W. 1994. Beyond the year 2000: phonetic access dictionaries (with word-frequency information) in EFL. *System* 22.4: 509–523.
- Sobkowiak, W. 1999. *Pronunciation in EFL Machine-Readable Dictionaries*. Poznan: Motivex.
- Sobkowiak, W. (in press). *Phonetics of EFL dictionary definitions*. Poznan: Wydawnictwo Poznanskie.

Alberts, Mariëtta ■ Standardisation, Modernisation and Harmonisation – Users' Perspectives

Pan South African Language Board (PanSALB), Arcadia, South Africa

Language awareness and establishing respect and appreciation for one's mother tongue are regarded to be essential factors in promoting the ideal of exact communication. General communication and language proficiency are indispensable tools in a person's relation with his/her social and economic environment. When it comes to language for special purposes, such as those used in the spheres of education, economics, science and technology standardised terminologies have to be in place with the various domains and labour related fields of interest in order to facilitate successful conceptualisation, knowledge transfer and preparing the nation for the labour market. Terminology facilitates language development and the dissemination of knowledge.

The issue of language in Africa and other developing countries is fraught with complications. The advent of colonialism denied to most people of the African continent the dignity of consideration for their mother-tongue, imposing on them instead a dependency on colonial languages. The role of language in development cannot be over-stressed. Every developmental discourse has taken note of its pragmatic and expressive values. Language is the vehicle and manifestation of culture. Language lies at the heart of culture, and is intrarelationally its most central element in as far as it measures and reflects the thought-world of its user-community (Prah 1998: 2). It is also the vehicle for the transmission of scientific and technical education (Oyetade 1998: 17).

It is a known fact that knowledge is best acquired through the mother tongue or first language. Information is also best presented through the first language of the trainer. The various languages of Africa and other developing countries can be empowered to be potent media of development. *This can only be achieved through a more purposeful system of education which will encourage the teaching and learning of the indigenous languages of Africa, and using them as media of instruction* (Oyetade 1998: 40).

A people's language becomes their symbol of nationality and self-identity (Msimang 1994: 149, 157). Indigenous languages should be developed and modernised. Multilingualism in South Africa and other parts of Africa (and other developing countries) is a sociolinguistic fact to be taken seriously. A large proportion of the indigenous population can only be reached through the means of indigenous languages. Information flow is prevented from being established by factors such as low literacy rates and terminologically poorly developed languages (Alberts 1998: 230, Fourie 1994: 132).

The term *language development* is used by Chumbow and Tamanji (1998: 53) to refer to all language engineering activities undertaken to bring a language from its natural state (the oral form in a pre-literacy state), to a codified and standardised form with some amount of literature and literacy activities as well as vibrant communication activities in terms of the conveyance of knowledge in various aspects of modernism, science and technology.

Modernisation or lexical expansion is the usage of various mechanisms and techniques to provide the language with relevant and adapted terminology to respond to the challenge of communicating knowledge in the arts, science and technology required for self-actualisation and national development (Chumbow & Tamanji 1998: 53).

The aim of language development and modernisation would be the provision of standardised terminology in the domains of mathematics, economics, science, technology and other subject areas and domains to:

- ensure accurate information transfer;
- ensure accurate information acquisition;
- ensure creative thinking in the first language;
- promote technical, scientific and economical communication;
- ensure better education and training results; and
- equip learners, students and workers for utilisation of knowledge.

Terminology is a strategic resource in a multilingual country. It is the medium through which knowledge and information is disseminated. Through the use of correct, standardised terminology, effective scientific and communication skills are developed. Harmonisation could play a vital role in terminology development – especially in a multilingual country.

This paper will address the user's perspective regarding different albeit related issues such as standardisation, modernisation and harmonisation.

Bailey, Dwayne ■ Creating a South African Keyboard

Zuza Software Foundation, Mowbray, South Africa

Translate.org.za is involved in the localisation of Free and Open Source software into the eleven official languages of South Africa. The organisation delivered the first Office productivity suite (wordprocessor, spreadsheet, presentation program) in eleven official languages.

During this process Translate was surprised that Venda translators could not type Venda characters on their computer and that most translators would simply ignore them or go through convoluted processes to insert these characters. This led to the research and development of Translate's South African keyboard.

Although the creation of the keyboard layout was relatively simple, the strategic decisions of the design took much longer. These included:

- Language investigation
- Deciding on one or multiple language keyboards
- Determination of layout

With a keyboard layout created, Venda still has a problem of displaying the characters if you do not have a Venda capable font. To overcome this, the keyboard includes a free font which contains all of the Venda characters.

The keyboard exists and works. We are now working with various partners to get user feedback which we hope will validate the design or give valuable input into improvements.

Translate over and above its software localisation effort continues to undertake work which touches on the broader issues of computers and South African languages. This includes amongst other things the validation of collation sequences and checking of locale information.

Translate has received funding from various organisations including: Department of Communications, CSIR, Shuttleworth Foundation and Hewlett-Packard.

Beyer, Herman L. ■ Lexicographical Parameters for the Systematic Description of User-friendliness in Dictionaries

Department of Germanic and Romance Languages, University of Namibia, Windhoek, Namibia

The aim of this paper is to show how research into user-friendliness from a certain theoretical perspective may approach a point where at least a structurally standardised and coherent metalexicographical answer can be provided for the question: *When is a dictionary user-friendly?*

The notion of the user aspect and user-friendliness represents a central theme in modern lexicographical debate, yet it seems to be difficult to accurately define the specific concept of *user-friendliness*. When the user aspect is involved in planning or evaluating a dictionary, the approach would normally be to determine the extent of the user-friendliness (or non-user-friendliness) of the characteristics of some lexicographical aspect under consideration. The approach outlined in this paper proposes the development of a universal theoretical framework or model according to which the phenomenon of lexicographical user-friendliness can be systematically described without necessarily requiring any specific dictionary as a point of departure, i.e. where instances of user-friendliness in a specific dictionary are not pointed out as much as making use of dictionary data from the complete spectrum of dictionary types to exemplify the exposition of user-friendliness as a lexicographical construct. A model of this nature will draw as much from existing research on the user aspect as it will show the potential to stimulate further research. As such, it cannot represent stagnant and inflexible theory, but must constantly evolve. Invariably, the model will show where and how the (meta)lexicography intersects with other scientific disciplines.

If a dictionary is seen as the carrier of text types, and texts function as media for communication, then a communication situation can be assumed where a dictionary functions. The use of a dictionary may thus be regarded as a communicative act, a “conversation” between the lexicographer and the dictionary user, albeit seemingly a unidirectional one. This assertion offers the opportunity for the lexicographical application of insights gained in communication theory. The general communication model for interpersonal communication is one such insight that shows potential for lexicographical application. With minor adaptations, this model can be used to describe the communicative act that takes place when a dictionary is used. The adapted model presents a functional framework according to which the lexicographical user aspect can be scientifically approached and described.

Consequently the following eleven lexicographical parameters are derived from the model:

- i. the dictionary parameter
- ii. the lexicographer parameter
- iii. the style guide parameter
- iv. the user parameter

- v. the usage parameter
- vi. the user's frame of reference parameter
- vii. the medium parameter
- viii. the channel parameter
- ix. the metatext parameter
- x. the context parameter
- xi. the feedback parameter

Each parameter is defined and described in terms of its comprising variables, and it is shown that the parameters do not function independently, but are highly interwoven. These parameters form the component parts of a *model of lexicographical parameters for user-friendly dictionaries*.

The paper also shows how the model fits into a dictionary plan, and how it focuses a generic definition of the genuine purpose of a dictionary on the user aspect.

Finally, some of the potential research questions that are produced by each parameter of the model are highlighted.

Bosch, Sonja E., Laurette Pretorius & Jackie Jones ■ Machine-readable Lexicons for Morphological Analysis of the South African Bantu Languages: A Data Model

University of South Africa (UNISA), Pretoria, South Africa

Most NLP applications depend in some way on a machine-readable lexicon as a basic resource. For example, machine-readable lexicons are an essential component in the development of morphological analysers. The goal in the development of a machine-readable lexicon is to be as inclusive as possible, thus incorporating all relevant information in the most efficient and economical manner, to be reusable and to conform to suitable and appropriate international standards.

Lexical information for South African Bantu languages is not readily available in the form of machine-readable lexicons. At present the availability of lexical information is restricted to a variety of paper dictionaries. These dictionaries display considerable diversity in the organisation and representation of data. This diversity emanates from factors such as designers' decisions, user needs, intended mode of delivery and economic considerations.

In order to proceed towards the development of reusable and suitably standardised machine-readable lexicons for these languages, a data model for lexical entries becomes a prerequisite.

The purpose of this paper therefore is to address this need in the following way. Firstly, we use a data model proposed by Bell and Bird (2000) as a point of departure and investigate how this model may be applied and modified for the Bantu languages of South Africa. Secondly, we discuss the necessity for modification of the Bell and Bird (2000) model based on the idiosyncrasies of the Bantu languages considering also the recommendations of Weber (2002). We present a data model in the form of an XML (Extensible Markup Language) DTD (Document Type Definition) for the languages in question and demonstrate its use by means of examples. Included in this model are additional particular requirements for complete and appropriate representation of linguistic information as identified in the study of available paper dictionaries.

In this paper we show that previously applied data models (Bell & Bird, 2000) require modification for machine-readable lexicons for the South African Bantu languages. In

particular we question the intuition expressed by Bell and Bird (2000) that “a complete model of dictionaries and lexicons should not need to include recursion of entries”. We instead concur in principle with the notion expressed by Weber (2002:8) that “lexical databases should accommodate (1) derived forms having multiple senses and (2) derived forms ... of the bases from which they are derived”. Indeed, we include our recursion as morpho-syntactic information where from a computational point of view it is available as basic linguistic building blocks for use in, for example, morphological analysis and syntactic analysis. We include the various senses together with their associated derivational forms where they are readily available for use in applications that may require them. We therefore propose an alternative model for machine-readable lexicons, which differs in significant ways from the Bell and Bird model to ensure maximum inclusiveness of all linguistic information. Our model provides flexibility and handles the various representations applicable to Bantu languages in particular and is therefore applicable to diverse uses of machine-readable lexicons.

The collection of data as well as the model we have developed and proposed, is intended to contribute to further discussion and development of a common scheme for storing lexical data not only for the South African Bantu languages, but for the Bantu language family as a whole. We conclude by emphasising that our purpose is mainly aimed at developing machine-readable lexicons as language resources for use in large-scale HLT/NLP applications and the technological development of the South African Bantu languages.

References

- Bell, J. & S. Bird. 2000. A Preliminary Study of the Structure of Lexicon Entries. <http://www ldc.upenn.edu/exploration/expl2000/papers/bell/bell.html>
- Weber, D.J. 2002. Reflections on the Huallaga Quechua dictionary: derived forms as subentries. <http://emeld.org/workshop/2002/presentations/weber/emeld.pdf>

Chikomwe, Justice ■ ALLEX Online Dictionaries: The New Cycle in African Lexicography at ALRI

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

In this paper, the presenter will discuss the need of users to access lexicographic works of the ALLEX Project at ALRI through various media, especially the Internet. Out of the six dictionaries that the Project has published so far, three are about to be available on the Internet, at <http://www.uz.ac.zw/duramazwi/>. These are *Duramazwi reChiShona* (1996), the general Shona dictionary, *Isichamazwi SesiNdebele* (2001), the general Ndebele dictionary, and *Duramazwi Guru reChiShona* (2001), the advanced Shona dictionary. The thrust of the paper is on the usefulness or otherwise of the Internet for the promotion of these works, utilisation of the Net for the collection of new headwords from users of the online versions of the dictionaries, and reviewing definitions of existing entries based on feedback or input from the same users. In this way, a cycle is established from the users to the lexicographer and back to the users again.

All the three ALLEX dictionaries that are online are monolingual. Therefore, they have been availed in Shona and Ndebele respectively. In other words, the presentation of the webpage is (mostly) in the language of the publication. The same applies for the error messages. Traditionally a user who would have encountered a problem with explanations proffered by a given word would proceed to contact the authors by telephone and post mail or

physically visit the authors to seek explanation or clarification and even to make suggestions. These options must remain available, especially since there is limited availability of ICT (Information and Communication Technology) in Zimbabwe and in Africa as a whole. Hence the need to continue with dictionary publication in printed form. However, the online option is more ideal for user input. It has the added advantage that it provides both the lexicographers and the users with fast and powerful searching tools, while it also provides accurate statistical data on usage.

Given the fact that there is always a high probability of a word being left out or not being explained clearly or fully from the user's point of view, the use of the Internet offers an easy solution in that it allows the user to contact the authors quickly and to make a direct contribution to a future revised edition of the dictionary. In the creation of specialised dictionaries this comes in very handy as it enables specialists working as a team on a terminological dictionary to easily coordinate their work online, which also reduces the cost of production.

The discussion will also touch on the role of the IT consultant in all of the above.

de Schryver, Gilles-Maurice ■ Internationalisation, Localisation and Customisation Aspects of the Dictionary Application TshwaneLex

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

Xhosa Department, University of the Western Cape, Bellville, South Africa

TshwaneDJe HLT, Pretoria, South Africa

It should not come as a surprise that in today's global village, the localisation of software is picking up momentum. Just this April for example, the software giant Microsoft launched its isiZulu Language Interface Pack for Windows XP, with the Setswana and Afrikaans versions expected to follow suit. What is true for these generic mass-produced applications has now also become a reality for the highly specialised field of lexicography. Indeed, TshwaneLex – the world's only truly off-the-shelf dictionary compilation software – is now not only fully customisable, but also fully localisable. This thus means that the entire lexicographic process, from initial compilation all the way to final product, may henceforth be conducted in any language of one's choice. The purpose of this paper is to highlight the main aspects that make this possible. The presentation will proceed in reverse, from final product back to the heart of the software environment itself, using a variety of languages.

In Figure 1 an example is shown of a monolingual dictionary, produced and placed online with TshwaneLex, for which both the dictionary data and the entire interface text are in the same language. In this case, a mother-tongue speaker of Sesotho sa Leboa is looking up the word *matšulela* in a digital environment that is entirely in his/her language.

Although seemingly straightforward, what a monolingual dictionary does not reveal is the separation between actual, unique dictionary contents and repetitive metalanguage. If this is done properly, dynamic metalanguage customisation can for example instantly kick in while a user consults an electronic bilingual dictionary. This is illustrated in Figure 2, where a person browsing the dictionary in Kiswahili will see the parts of speech labelled as *kimilikishi nomino* and *kivumishi*, and the cross-reference marker text as *Mzizi*. For a person who browses the same dictionary, but now in English, these same metalanguage strings will automatically be adapted to *possessive pronoun* and *adjective*, respectively *Root*.

Apart from separating dictionary contents from metalanguage, good lexicography software should also clearly separate these data from the structure of the articles. In technical

terms this means that the contents must be ‘governed by’ and thus ‘conform to’ a DTD (Document Type Definition, or dictionary grammar), with this DTD again markedly distinct from the actual formatting (or Style) of the data. This is illustrated in Figure 3, where a Cilubà – French dictionary is being compiled within TshwaneLex. The Tree View on the left is ‘translated’ into one of any number of displays in the Preview Area on the right. What is important to note here is that all the labels for the components of the Tree View are entirely customisable (and have been customised) by the lexicographer: *Ngumvwikilà* stands for ‘Sense’, *Munyàku* for ‘Combination’, *Dikùdimuna* for ‘Translation equivalent’, *Cileejilu* for ‘Example’, etc.

In addition to the above customisation aspects, TshwaneLex has also been internationalised. This internationalisation (often abbreviated as i18n) was achieved through a strict separation between the actual (C++) program code on the one hand, and all the textual data that appears in the GUI (Graphical User Interface) of TshwaneLex on the other hand. This internationalisation, in combination with the fact that Unicode is supported on all levels, means that TshwaneLex can now easily be adapted to other markets. This process is known as localisation (abbreviated as l10n), and at present TshwaneLex has been (or is in the process of being) localised in a variety of languages including Chinese, German, Irish, Sesotho sa Leboa, Setswana, isiXhosa, and Welsh.

Typical tools for software localisation include terminology management systems, translation memory systems, localisation software, and project management tools. In TshwaneLex, however, all of these were brought together into a *single built-in Localisation editor*. All relevant text strings that appear throughout the various menus, dialogue boxes, messages, tabs, buttons, the status bar, etc. are automatically presented in the Localisation editor. This can be seen in Figure 4, where localisation is ongoing for Welsh. Also note that all strings where the English term ‘open’ appears were brought together in this example, which enables the translator to make sure that the translated terminology is consistent. Other powerful features that help ensure consistency and contextual correctness are an indication of where in the GUI (‘Key’ in Figure 4) each text string appears, as well as the possibility to sort on both the original and translated strings, and to sort on the location. Any number of localised versions of the GUI can be prepared and each immediately becomes available within TshwaneLex. Further note that all the text strings are kept in a single file per language, which for example means that all the text of each localised version can easily be spellchecked in one pass.

Figure 5, finally, shows examples of the TshwaneLex GUI being (partially) translated into Welsh. Particularly handy is the fact that the results of the localisation can be seen in real time within TshwaneLex itself. Therefore, while one is using TshwaneLex to localise TshwaneLex, the emerging localised version of TshwaneLex appears in front of the translator’s eyes – arguably the ultimate in localisation.

Acknowledgements

I would firstly like to extend my great appreciation for David Joffe, who not only is the brilliant software engineer behind TshwaneLex, turning the wildest dreams into workable products, but also the chief conceptualiser of the material presented here. I would further like to thank the people who were responsible for the creation of the multilingual data, viz. Maropeng V. Mojela and his team for Sesotho sa Leboa, Sarah Hillewaert and her team for Kiswahili, Ngo S. Kabuta and his team for Cilubà, and Dewi Evans and his team for Welsh.



Figure 1: The first fully monolingual African-language dictionary on the Internet, created and maintained with TshwaneLex (Data © 2004 by Mojela *et al.*)

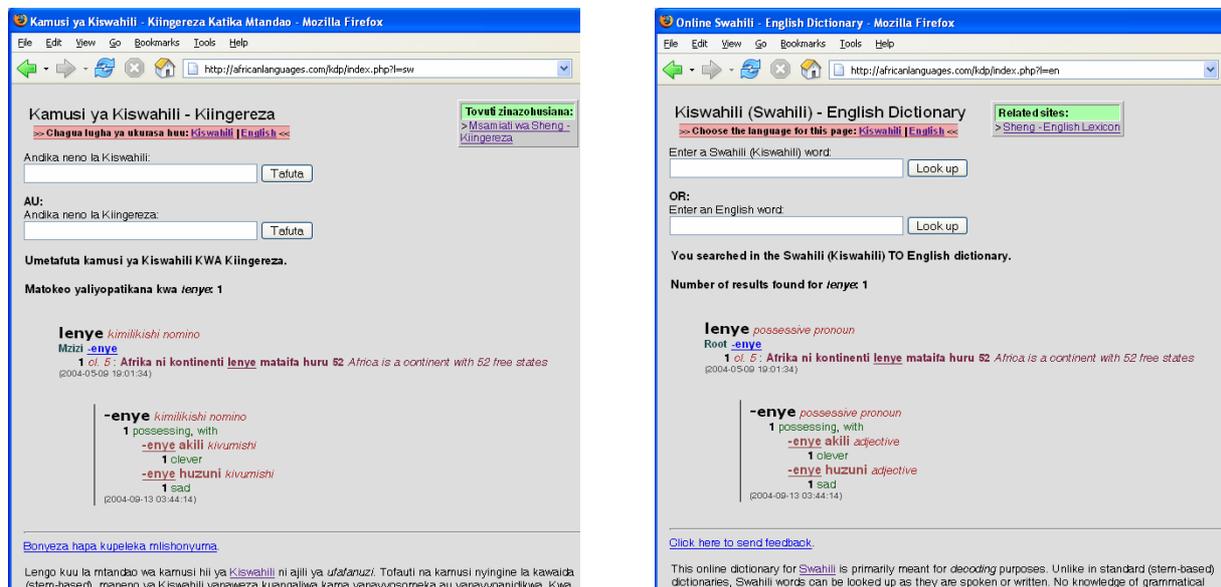


Figure 2: Dynamic metalanguage customisation in real time on the Internet, achieved with TshwaneLex (Data © 2004 by Hillewaert *et al.*)



-ya 15/0 lwakū lukāsambwishi || **aller**, s'en aller Muny/Loc: **cyāyilla kula ! qu'il passe loin (de moi) [expression prononcée par une femme enceinte lorsqu'elle rencontre un albinos (citòòkatòòkà); en la prononçant, elle crache sur la poitrine et éventuellement se retourne pour éviter la vue de l'albinos] se comporter, se conduire CIL/EX: mūyaayà bakwèbè, wāmūya CIL.cikūdimuna=conduis-toi comme tout le monde (-ya ku ngondo) avoir ses règles, être réglée (-ya kumpàla) progresser (-ya pabi) finir mal (-ya ciyaaya) erreur à l'aventure (-ya muntu pannyimà) sonder les intentions de qqn Lus/Par: Bāāya wāya; nè ūshaalā wāshāādillamū (lit.: Pars avec les autres; si tu restes, tu resteras pour de bon, càd il faut faire les choses tant qu'il est temps) =>< -lwa², => -enda :@ || >ish, il =>ya.**

Figure 3: The fully customisable DTD-driven structure for one article (left), and one possible WYSIWYG-output (right), as seen within TshwaneLex (Data © 2006 by Kabuta *et al.*)

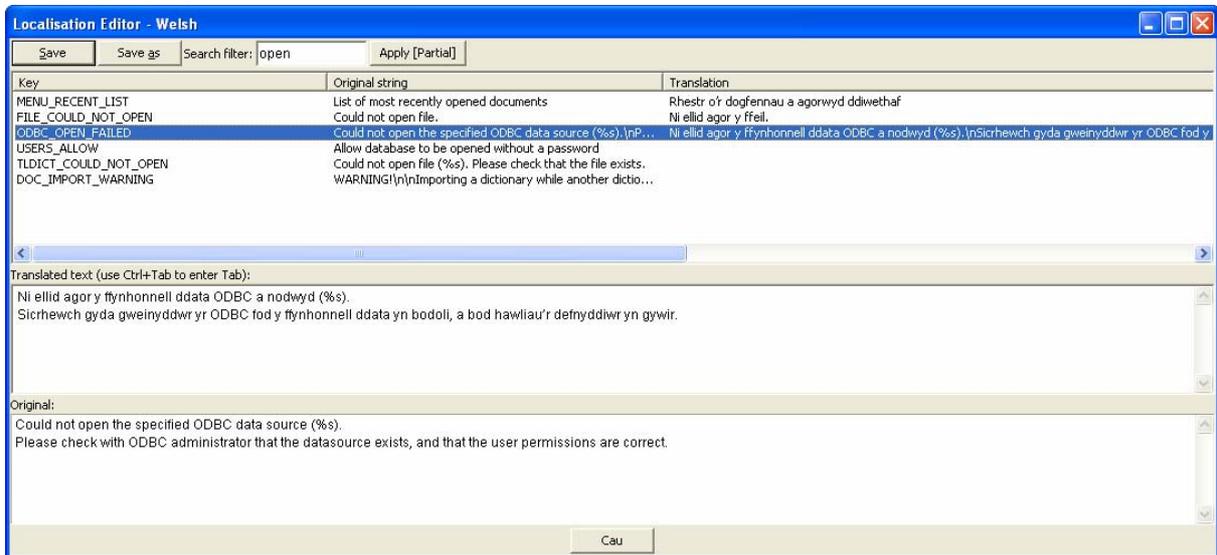


Figure 4: The built-in TshwaneLex Localisation editor (Data © 2006 by Evans *et al.*)

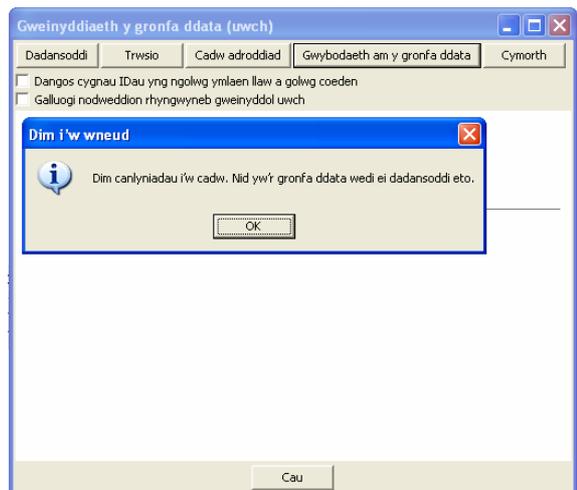
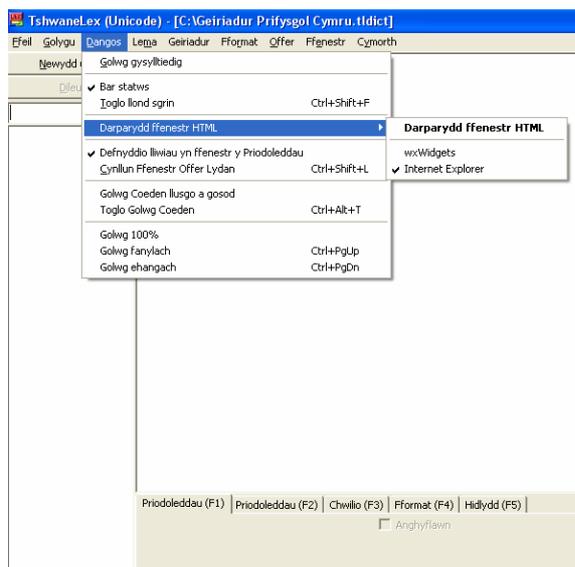


Figure 5: Partially translated GUI sections of TshwaneLex (Data © 2006 by Evans *et al.*)

Ekwa Ebanéga, Guy-Modeste & Fatima Tomba Moussavou ■ Survey of Dictionary Use: Case Studies of Gabonese Students at the University of Stellenbosch and the Cape Peninsula University of Technology

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

Two eras characterize Gabonese Lexicography. The first one is the old era of missionaries and colonial administrators. They are the pioneers in the compilation of dictionaries in Gabonese languages. These dictionaries are all bilingual (or translation) dictionaries biased towards French and aimed at the foreign users, the French speakers. The second area is the modern area of national lexicographers trained in metalexicographical principles to compile dictionaries aimed at the Gabonese user. Up to now, no study or survey has been done on the user profile or the dictionary use with regard to the Gabonese environment.

The aim of this contribution is to investigate the knowledge and the opinions of Gabonese students with regard to dictionary use.

Design and empirical research. The main contribution of this research, which was carried out during the academic year 2005-2006 was to draw a profile of Gabonese students as dictionaries users. Indeed, we want to know their knowledge, opinions, habits of use.

Participants. The survey was conducted among 100 Gabonese who are studying at University of Stellenbosch and Cape Peninsula University of Technology. 36 females and 64 males between 19 and 35 years old took part into this research.

This study was undertaken in the city of Stellenbosch, particularly at University of Stellenbosch and in the city of Cape Town.

Methodology. We collect data through the use of a questionnaire, which includes 33 questions. We ask our participants questions about types of dictionaries, frequency of use, difficulties of use, and instruction on dictionary use, etc. This questionnaire is based on the one developed by Hartmann (1999): “Case study: The Exeter University survey of dictionary use.

Analysis of results. The results obtained from this survey show that:

- More than 50% of the students start using dictionaries in primary school.
- More than 80% own monolingual dictionary.
- More than 70% use most frequently in Gabon a monolingual dictionary.
- Bilingual dictionaries should be the first to be compiled with regard to Gabonese languages.
- Specialised terms are the most difficult type of information to find for more than 70% of the students.
- More than 60% find difficult to use a dictionary when looking up words when it does not provide enough information.
- More than 20% have never been taught how to use a dictionary.
- More than 10% of the students do not read the instruction to the user.
- The teaching of dictionary use is very important for more than 60%, etc.

To summarize the results of this questionnaire, we can point out two general ideas and conclude that:

- Bilingual dictionaries should be the first to be compiled with regard to Gabonese languages.
- Gabonese students need instruction in dictionary skills.

Bearing these conclusions in mind, we hope to eventually move to further research by administering similar questionnaires to a more representative sample of Gabonese students in order to gain the knowledge of their general profile.

Every single lexicographical decision has a language policy relevance and therefore, in the end, a political dimension. This thesis is not new because it is not new to discuss the influence of politics on lexicography. However, it usually applies to politically charged words like *democracy*, *terrorist*, *revolution*, *black*, *racism*, etc. It is new to regard every lexicographical decision as a result of a political decision, more precisely a language policy decision. We do not see the lexicographer as “a harmless drudge that busies himself in tracing the original and detailing the significance of words”, like Dr. Johnson and many lexicographers the last 250 years have underrated themselves. Most lexicographers may be aware of their tremendous influence on language behaviour but fail to realise their influence on cultural and political behaviour.

Loaded political words are the most obvious examples for language policy decisions in lexicography. However, this is only a small part of the problem and the fuller extent will be discussed in this paper. In this regard an important distinction made in this paper is between inter- and intralingual language policy.

Linguistic hegemony, as a special form of interlingual language policy, can influence lexicography in a bi- or multilingual community where all languages do not have the same official status or recognition as preferred language. Dictionaries can be influenced by having to reflect this situation or they can reflect a view independent of the hegemonic approach. This situation is discussed for various lexicographic endeavours in Africa.

A distinction is made between externally-motivated and internally-motivated lexicographic endeavours in Africa. Externally-motivated dictionaries typically are bilingual products, co-ordinating a local African language with a European language, typically the first language of the missionaries or the colonisation officials. The lexicographers of these products are not speakers of the local language but rather of the language of colonisation whereas internally-motivated lexicographic endeavours are the dictionaries compiled from within the given speech community. Externally-motivated lexicographic endeavours often are the results of a prevailing language policy which protects the hegemony of the language of colonisation.

Since the changes of 1994 in South Africa there has been an official change in the favouring of two languages to the detriment of the rest. The establishment of eleven national lexicography units, one for each of the official South African languages, has created the opportunity, at least in theory, for all the official languages to be promoted and to embark on lexicographic projects. In practice, however, the application of the official policy of eleven national languages leaves much to desire and it still has to be seen how this language policy will influence South African lexicography. The strong position of English as language of documentation has already resulted in the choice by a number of national lexicography units to compile bilingual dictionaries with English as partner language. If not managed in a proper way this may lead to a new linguistic and lexicographic hegemony.

Language policy, as a form of intralingual language policy, is based on a particular historical, moral or political conception of proper language. Such a policy resorts primarily but not exclusively on an intralingual level and it should never be interpreted as being predominantly interlingual in nature.

A puristic language policy often is an attempt to keep a language “pure” from either the influence of another language but, especially, from the influence of non-standard varieties, ideologically marked forms or derogatory words and expressions. This is the type of

language policy that can lead to dictionaries eschewing the actual language use of their speech communities. A publisher may adhere to such a language policy and use its dictionaries to promote this approach. The individual lexicographer may adopt such a puristic approach in his/her dictionary, knowing quite well that the intended target user of the dictionary may very well interpret this approach as an official ruling on language. Various aspects of the influence of the puristic urge on lexicography are discussed.

The role of dictionaries in the purist drive is not only to approve or condemn loan words as counterparts to indigenous forms. Dictionaries should also reflect on e.g. derogatory forms. In this regard the function of the dictionary and the target users will determine the possible inclusion of these forms and their treatment. When it comes to a puristic language policy one cannot formulate a single approach for all dictionaries. As with the contents and structures of dictionaries the user and the lexicographic functions of each dictionary should determine the nature and extent of its puristic approach.

The main thesis of the paper is that the lexicographers of concrete dictionaries constantly make decisions that are relevant to language policy. This does not only regard the domain that clearly belongs to party or national politics but also the domain of general decisions regarding language and languages.

Klein, Juliane ■ Implementing a Dictionary Culture for South Africa – Attempt at a General Framework for the Official Languages

University of Leipzig, Leipzig, Germany

Compiling extremely user-friendly dictionaries and teaching people how to use dictionaries efficiently are the two most important tasks for lexicographers in South Africa. A multilingual country as South Africa needs a dictionary culture that includes at least all official languages of the country, i.e. eleven languages in the South African case.

As the South African dictionary users are not a homogeneous group, different approaches are needed in order to establish a dictionary culture. This means that the different user groups - including primary and secondary school pupils, university students, teachers and “ordinary” people - must be taken into account. Another important feature is the distinction between conjunctively written languages, i.e. the Nguni group (isiZulu, isiXhosa, siSwati and isiNdebele) and disjunctively written languages, i.e. the Sotho group (Sepedi, Sesotho, Setswana), Tshivenda and Xitsonga.

The development of a new dictionary culture can roughly be divided into three main phases: first informing people that the dictionaries are available, secondly, teaching the people how to use a dictionary efficiently and thirdly accepting that the use of a dictionary is an everyday activity that children learn from their community.

Informing people about available dictionaries and their benefits is the first step in creating a new dictionary culture. Dictionary awareness campaigns are an essential part in this information phase. Such campaigns must clearly show the people that those new dictionaries are valuable means to empower them in everyday life and not only useless books. The language of such campaigns must be the language the people understand and this implies that it should be carried out in the relevant African languages and not only in English. Another psychological factor is the use of important, famous South Africans who should support such campaigns. A side effect of successful campaigns will be the enhancement of the status of the African languages.

The teaching phase will be quite difficult, as there are different groups to be taught, for example dictionary compilers, students of lexicography, teachers, students – ranging from primary school to university, and adults who would like to consult a dictionary. A second problem is that there are currently only a few people who could do the teaching. Courses for dictionary compilers should consist of three main parts: teaching lexicographic theory including specific grammatical features of the language in question, giving advice on advertising the product (getting the dictionary to the user) and teaching about the dictionary user who wants a user-friendly dictionary. As the dictionary user is the most important factor that decides if a dictionary is successful, compiling user-friendly dictionaries should be the first aim of the lexicographer. Questions to bear in mind concerning the user are: Who are the users? What do they know about the language? What do they know about dictionaries? Do they know how to use a dictionary?

Although a basic knowledge of the theory of lexicography is useful for teachers, a confident use of dictionaries and a profound knowledge about dictionaries are more important. Knowledge about dictionaries includes: dictionary types, the contents of a dictionary and how to extract information. Choosing the right type of dictionary is the first step towards competent and confident use of dictionaries. The next step is to get the correct information out of the dictionary. This means that teachers must know how to find the lemma they are looking for and how to decode the information in the dictionary article. As this can be difficult, especially for conjunctively written languages, the lemmatisation principle of the dictionary must be clear. Dictionary compilers can solve this problem by making it very clear which lemmatisation approach they have used and what they have considered as exceptions and explain it clearly in the front matter (usage guide) of the dictionary. This leads to the next important step, teaching the use of the front and back matters of the dictionary.

Teaching dictionary usage should start as early as possible, that is, in the foundation phase of OBE. Although different approaches are needed for the different age groups, the aim (getting the appropriate information and feel comfortable using a dictionary) is the same for all age groups.

Teaching adults will be more problematic than teaching pupils and will mainly be conducted through dictionary awareness campaigns and the front and back matters of the dictionary itself. This implies that all dictionaries should contain instructions for extracting information.

The last phase will be a stabilization phase in which the use of dictionaries should become a normal activity.

Lepota, Biki ■ The Treatment of Cardinal Number Plurals in Sepedi

Best Books, an imprint of NB Publishers, Hatfield, South Africa

Unit for Academic Literacy, University of Pretoria, Pretoria, South Africa

When treating the cardinal numbers of Sepedi, lexicographers tend to ‘ignore’ the existence of number plurals. Consequently, this category of the noun has never been investigated, although two opposing methods regarding the formation of cardinal number plurals can be distinguished in the existing literature. The aim of this paper is to investigate the proper method of forming cardinal number plurals in Sepedi. As space restrictions do not allow for treatment of all the numerals, the presentation will focus only on the basic numbers, i.e. those from one to ten.

In the same way that cardinal numbers are nouns in English, the same is true for Sepedi cardinal numbers. This implies that each cardinal number has both a singular and a plural form, although a cursory glance at the numbers might lead one to conclude that number plurals do not exist in Sepedi. It is, therefore, not surprising that existing (bilingual) Sepedi dictionaries do not provide the necessary data for a sound treatment of plurals of the cardinal numbers. To put it differently, they do not include class information for every singular number to implicitly tell the user how the plural should be formed. The fact is that in general language plurals of numbers do not appear to be frequently used. At the same time, however, a review of mathematical sources reveals that cardinal number plurals do exist, as those materials do make use of them.

The sources consulted show that there are two conflicting methods used in the formation of cardinal number plurals. As far as the first method is concerned, the prefix *bo-* (of noun class 2a) is added to a number word to denote plurality. In contradiction, the second method affixes the prefix *di-* (of noun class 10) to a number name to pluralise cardinal numbers. There is even a third method that uses the plural prefix *ma-* to change the gender of tens from singular to plural.

From the preceding paragraphs, one can clearly see that the plurals of most numbers end up in *two different classes*. However, notwithstanding the fact that both *bo-* and *di-* are used to express plurality in the Sepedi cardinal numbers of 0 to 9, the type of plurality expressed differs in each case (cf. Van Wyk 1987). The question that arises from this is: which plural prefix should lexicographers use consistently in treating cardinal numbers in Sepedi dictionaries?

This paper will make an attempt to find an answer to the above question in two ways. Firstly, a semantic difference that exists between the plurality expressed by the prefix *bo-* and the one indicated by *di-* will be discussed. Secondly, data obtained from all existing (bilingual) Sepedi dictionaries, existing textbooks, fieldwork results obtained through a questionnaire, corpus queries for the retrieval of frequency information, and the views of Sepedi grammarians will be compared. The interplay between these different types of data will ultimately enable dictionary compilers to provide a sound treatment of cardinal number plurals in Sepedi.

Mabaso, P. ■ Considerations in the Preparation of a Shangani Dictionary

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

African Languages Research Institute (henceforth ALRI) is a non-faculty inter-disciplinary unit at the University of Zimbabwe that is dedicated to research, documentation, development and promotion of Zimbabwe's indigenous languages. It is the first such institute in Zimbabwe that has initiated an agenda in research, documentation and promotion of indigenous languages. It also aims at developing the previously marginalized community languages of Zimbabwe. It is important to note that the term community language is now more commonly used than minority because the speakers of these languages feel that the latter is derogatory. ALRI's major pre-occupation is dictionary compiling in Zimbabwe's indigenous languages, mainly Shona and Ndebele. So far the institute has published four dictionaries in Shona and two in Ndebele. Shona and Ndebele are Zimbabwe's officially recognized national languages. ALRI now wants to facilitate the compilation of dictionaries in Zimbabwe's community languages which are spoken by about ten percent of the total population. Mother tongue

speakers of the community languages are being trained to do the linguistic field work and documentation that is required for the compilation of the dictionaries.

Of all the country's about fourteen community languages, Shangani is the first to embark on a dictionary project. The language is spoken by approximately three percent of the total population. The Shangani speaking people are concentrated in South Eastern Zimbabwe and found in the Southern districts of Masvingo, mainly in Chiredzi South and Mwenezi. They are also dotted around other districts like Zaka and Masvingo South. Not much research has been done on the Shangani language since it has had very little attention from linguists and researchers. It is only maintained in family and community merely as a spoken language. The language had never had a significant role in education during and after the colonial period. Zimbabwe's education policy acted as if community languages like Shangani did not exist (Viriri 2003). It was only in 2004 that Zimbabwe's education act was revised and gave provision for the teaching and examination of Shangani and five other recognized minority languages in schools and tertiary education.

So far ALRI is in the process of building a Shangani corpus which will be used in the compilation of the dictionary in the near future. Since 2001, mother tongue speakers from all the areas Shangani is spoken are being hired to go into the field to collect data for corpus building. They record oral interviews covering a very wide range of different speech situations. These interviews are then transcribed by the mother tongue speakers who have some basic linguistic knowledge. A Shangani mother tongue speaker employed by ALRI then encodes the transcribed data using the diverging orthographies the transcribers use. Unpublished Shangani material is also being collected and added into the corpus. However, there are a lot of issues that need to be considered before the compilation of the dictionary begins.

The paper will discuss some of the general points to consider in the preparation of the dictionary. It will also show how these will impact on the dictionary and how the compilers can go about them. Some of the points to be discussed in the paper are as follows:

- a) The purpose of the dictionary and its prospective users. Is it going to be bilingual or monolingual?
- b) The orthographical problem of the language since it is not yet standardised.
- c) Lack of mother-tongue speakers trained in lexicography.
- d) How to treat the language varieties.
- e) Time factor: how long does it have to take to produce a dictionary that meets lexicography standards?

Mafela, Munzhedzi James ■ L.T. Marole: A Forgotten Pioneer in Tshivenda Lexicography

Department of African Languages, University of South Africa (UNISA), Pretoria, South Africa

Since spoken Tshivenda was reduced to writing by the Berlin Missionaries in the late nineteenth century, very little has been done in the field of lexicography. As Mawela (1999: 251) writes: "Little has been done in the lexicography field in Venda. At present there are only a few dictionaries, all of them translating dictionaries, ...". About a translating dictionary, Zgusta (1971) maintains that its basic aim is to coordinate with lexical units of one language those units of another language which are equivalent in their lexical meaning. In Tshivenda, the major characteristic demand at the time of writing the dictionaries was to

learn Tshivenda on the part of missionaries and other foreign speakers. Translational dictionaries are important in the development of indigenous African languages. As a result of the flourishing increase in international interaction, more and more of people are coming to realize that bilingual dictionaries have become a necessary economic, intellectual and cultural activity (Al-Kasimi 1983). Tshivenda lexicographers who contributed in the Tshivenda lexicography are L T Marole, N J van Warmelo, P J Wentzel, T W Muloiwa, K B Hartshorne and M C Neluvhalani. Even if some of the lexicographers mentioned above were linguists, the majority were not trained lexicographers. They got involved in lexicography out of interest.

L T Marole is a pioneer in Tshivenda lexicography. He is the first Muvenda to publish a lexicography work in Tshivenda. Up to 1955, four lexicography books were published by Marole, i.e. *Phrase Book for English and Venda* (1932), *Phindulano: English – Venda Phrase Book* (?), *English – Venda Vocabulary* (1954), and *Afrikaans – Venda: Vocabulary and Phrase Book* (1955). The lexicography books are small in size but very valuable to foreign speakers who intend to learn Tshivenda. Although Marole is a pioneer in Tshivenda lexicography, he is hardly mentioned by Tshivenda lexicographers when they discuss lexicography matters. Dictionary users hardly use his works in their daily references to dictionaries. All this is because they do not know their (dictionaries) existence because Marole is not accorded the recognition he deserves.

Marole was influenced by the Berlin Missionaries to write the *Phrase Book for English and Venda*. This lexicography work is meant for foreign language speakers, especially missionaries who wanted to learn Tshivenda at the time. Entries in the translational dictionary are arranged according to the parts of the body, wild animals, domestic animals, prepositions, seasons, relationships, sickness, house work and interrogative pronouns. *Phindulano: English – Venda Phrase-Book* consists of two parts, i.e. translation of phrases and bilingual glossaries. Lexical entries are arranged according to themes like wild animals, birds, and domestic animals. *English Venda Vocabulary* is the best dictionary Marole has written. It is even better than some of the dictionaries that have been produced recently. The author has included parts of speech in his definition of the entries. In some instances, multiple equivalents have been provided for one entry. Unlike in the works mentioned above, the entries have been arranged according to the letters of alphabet. Marole did not only cater for English speaking people who want to learn Tshivenda, Afrikaans speakers have also been considered. In 1955 he produced an *Afrikaans Venda: Vocabulary and Phrase Book*. Like the other lexicography works mentioned above, this contribution is made up of two parts, i.e. translation of phrases and bilingual glossaries.

Besides that Marole is a pioneer in Tshivenda lexicography, his contribution is an asset to the Vhavenda. He got involved in lexicography writing without a financial support from the government or any other private organisation. He established his own publishing company to have his works read by the society. Yet, his lexicography works receive little recognition from both present lexicographers and dictionary users. The aim of this paper is to highlight Marole's contribution towards the development of Tshivenda lexicography.

References

- Al-Kasimi, A.M. 1983. *Linguistics and the Bilingual Dictionary*. Leiden: E.J. Brill.
Marole, L.T. 1932. *Phrase Book for English and Venda*. Sibasa: Marole Book Depot.
Marole, L.T. 1954. *English – Venda Vocabulary*. Sibasa: Marole Book Depot.
Marole, L.T. 1955. *Afrikaans – Venda: Vocabulary and Phrase Book*. Sibasa: Marole Book Depot.
Mawela, A. 1990. The state of Tshivenda Lexicography. *Lexikos* 9.
Zgusta, L. 1971. *Manual of Lexicography*. Prague: Academia.

Mahlangu, K. Sponono ■ The Lemmatisation of Loan Words in an isiNdebele Bilingual Scholars' Dictionary

IsiHlathululi-mezwi SesiNdebele, Department of African Languages, University of Pretoria, Pretoria, South Africa

IsiNdebele is a standard language that was recognized in the early 1980s and introduced into schools in 1985. Its 1st Orthography and Spelling rules came to light in 1983 and was revised in 1995 and 2005 respectively (Skhosana 2003). Its serious engagement in the dictionary making process, first by means of corpus building, started in 1999. The bilingual scholars' dictionary compilation project started in 2001, when the Unit had a corpus of 1.2 million. This project was completed in late 2005 and officially launched in March 2006.

Researchers in the fields of linguistics and lexicography such as Nkondo and Jafta (in Van Huyssteen: 115) and Kamwangamalu (2000: 89) are of the opinion that no language is lexically self-sufficient. This notion has been argued and proven correct especially in African languages, where the technical terms and vocabulary for social interaction has been mainly drawn from foreign languages such as English and Afrikaans. Borrowing is said to empower these official languages in terms of economic, educational and political status.

Lexicographers such as Al-Kasimi, Wells and Gove are also in agreement that the responsibility of a dictionary is to record the language and not to set its style. From a lexicographic point of view, this puts a lexicographer in a dilemma especially when a lexicographer has to enter and define words as they are used in everyday speech. Traditionally, a lexicographer was seen as the judge or ruler of the language. The lexicographer was the one who pointed out words that should be included and those that are unacceptable in the dictionary with the sole aim of using a dictionary to rectify and purify the language.

Challenges with regard to the treatment of borrowed words in African-language dictionaries have been dealt with in other African languages by scholars such as Nong, De Schryver & Prinsloo (2002) in Sepedi, Drame (2001) in isiXhosa, Ncube (2005) in Zimbabwean Ndebele, and Mpofu and Mangoya in Shona. In isiNdebele, no attempt has ever been made to investigate the challenges in the lemmatisation of borrowed words. De Schryver (2004) investigates and discusses the creation of a macrostructure of a Nguni Dictionary, and applies this to isiNdebele. Skhosana (1999) discusses the history of the *isiHlathululi-mezwi sesiNdebele* Dictionary Unit before 1997.

Therefore, in this paper the lemmatisation and selection of loan words into the isiNdebele Bilingual Scholars Dictionary is investigated and discussed. In conclusion, the paper will illustrate (a) challenges regarding lemmatisation of borrowed words, (b) whether a lexicographer could be the final judge of which borrowed terms to include in the dictionary compilation or not, (c) whether language users at grassroots level are the ones who should decide for the lexicographer or not.

Mangoya, Esau ■ A Look at the Historical Influence on Shona Dictionary Products

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

It is said that most of the artistic works tend to be influenced by the historical environment in which they are produced. They tend to reflect the general feeling and mood of the time that they are produced. The proposed paper would like to make an analysis of Shona lexicography

and try and look at aspects that could have influenced some levels of headword selection, defining formats and the inclusion of and handling of some grammatical aspects in the dictionaries. It will also make an analysis of the defining language used in the definitions. The paper would like to discuss how such environments would also impact on the overall dictionary product. The discussion will be from the background that missionaries played a great role in the development of the indigenous languages of this country particularly the setting up of the writing systems of the languages and the early grammatical analysis of the languages. It has to be noted that they even were the pioneers in the compilation of dictionaries in the indigenous languages of this country.

However lexicographic work has continued to be done and other people without this missionary background have also been involved as well. The paper would like to make comparisons of the dictionaries that have been compiled in the country checking on the historical environment they have been produced and how that has influenced the content of the products. In other words it is an analysis of how other dictionary products compare to ones produced by the missionaries.

It will also look at the various use of the dictionary and how the users would view the dictionaries in terms of the lay down of the information in the dictionary. That is to say it will assess how the historical background leads to particular trends in dictionary formation and how that will impact on the user-friendliness of the dictionary. The study will try to show the trends and patterns in dictionaries emanating out of the historical environments in which the dictionaries are produced.

Mheta, Gift ■ The Problems and Challenges of Dealing with Encyclopaedic Information in Shona Lexicography: The Case of Duramazwi reMimhanzi

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

The paper will presents the problems and challenges of dealing with encyclopaedic information in Shona lexicography. It will focus on how encyclopaedic information is collected, selected, entered and defined. Emphasis will be on identifying the problems and challenges experienced by lexicographers at the four main stages of dictionary making. It also presents solutions to the impediments in question. It brings out discernible patterns in the treatment of encyclopaedic information. The analysis, from the collection stage up to the definition stage is done vis-à-vis the relatively long history of compiling linguistic dictionaries in the Shona language. Emphasis will also be on how the lexicographers deal with challenges with the objective of satisfying the needs of the targeted end users of the dictionary.

The paper will highlight the experiences and challenges faced by the African Languages Research Institute in the compilation of an encyclopaedic dictionary of music, henceforth *Duramazwi reMimhanzi*. The paper will be based on my experiences and the observations that I have made as the Chief Editor of the dictionary. The dictionary project is the first of its kind in the lexicographic history of Zimbabwe. It was published in March 2006. *Duramazwi reMimhanzi* is trilingual in nature. It has three languages namely, Shona, Ndebele and English, the national official languages in Zimbabwe. However, the dictionary has one alphabetizing language i.e. Shona, an indigenous Zimbabwean language that is made up of five dialects and is spoken by seventy-five percent of the country's population. In *Duramazwi reMimhanzi* Shona headwords are followed by Ndebele and English equivalent terms and then followed by definitions, which are all in the Shona language. At the end of the

dictionary section there is a section on illustrations that is followed by two indices, the first one comprising Ndebele and English equivalent terms and the other one made up of Ndebele headwords and Shona glosses. The dictionary is targeted at tertiary institutions and some secondary schools that offer music as a course or subject.

Before presenting problems and challenges of dealing with encyclopaedic information in Shona lexicography it is essential to briefly outline the history of Shona lexicography in general as it has a bearing on some of the challenges that are currently being faced in the compilation of Shona dictionaries that are encyclopaedic in nature.

As background, the Shona language has a lexicographic history of nearly one and half centuries. It can be traced to as early as 1856 when W.H.I. Bleek published, *The Languages of Mosambique*, the first bilingual dictionary in Shona and English. Ever since, the trend has been that of compiling linguistic, synchronic dictionaries mainly those that are bilingual in nature. The most popular of these dictionaries is Father Hannan's *Standard Shona Dictionary* that was published in 1959 and subsequently revised in 1984. The trend of compiling bilingual dictionaries continued until 1996 when the African Languages Lexical (ALLEX) Project, then housed in the Department of African Languages and Literature produced the first ever monolingual Shona dictionary. This was followed by the publication of *Duramazwi Guru reChiShona*, the advanced Shona monolingual dictionary by the African Languages Research Institute in 2001. Ever since the publication of this dictionary, there has been a paradigm shift in Shona lexicography, from the production of general linguistic dictionaries to the compilation of specialized dictionaries as evidenced by the publication of a medical terms dictionary, *Duramazwi reUrapi Neukoti* in 2004, *Duramazwi reMimhanzi* 2006, on which this paper is based and the currently running projects on dictionaries in specialized fields such as the *Shona Linguistic and Literary Terms Dictionary* and the *Shona Children's Dictionary*. It is against this strong background of producing general linguistic dictionaries that the present researcher will discuss the problems and challenges encountered in the compilation of a specialized dictionary of musical terminology, *Duramazwi reMimhanzi*.

References

- Bleek, W.H.I. 1856. *The Languages of Mosambique*. London: Harrison.
Chimhundu, H. 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
Chimhundu, H. 1996. *Duramazwi reChiShona*. Harare: College Press.
Hadebe, S. 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
Hannan, M.S.J. 1959. *Standard Shona Dictionary*. Harare: Literature Bureau.
Hannan, M.S.J. 1984. *Standard Shona Dictionary*. Harare: College Press.
Mheta, G. 2006. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
Mpfu N. 2004. *Duramazwi Reukoti Neurapi*. Gweru: Mambo Press.
Zgusta, L. 1971. *Manual of Lexicography*. Prague: Academia.

Mihindou, Guy-Roger ■ A Hybrid LSP Dictionary with Microstructural Amalgamation to Enhance User-friendliness

University of Stellenbosch, Stellenbosch, South Africa

The concept of *user-friendliness* in lexicography advocates for a better treatment of data in such a manner that the presentation of data in the dictionary can facilitate both the access of data, and the retrieval of information the user is looking for. Thus, the lexicographer has to devise innovative ways to achieve this user-friendliness in order to satisfy users.

Consequently, the typology of the dictionary may play a role by specifying the type of user the dictionary intends addressing his or her needs.

In the case of a Language for Specific Purpose Dictionary (LSPD), for instance, the target users are primarily specialists of a given field who would like to further their knowledge in a particular area of expertise. It will be expected from such a dictionary that structures like macrostructure and microstructure as well as the distribution of data within the dictionary, will slightly differ from those in a Language for General Purpose Dictionary (LGPD). Concerning microstructure, a monolingual LSP dictionary for example will display a comprehensive paraphrase of meaning of the lemma, whereas a bilingual or multilingual LSPD will present a very condensed treatment with translation equivalents and some illustrations.

The hybrid LSPD as proposed in this paper will take into account the monolingual and the multilingual dictionary features. The accepted lexicographical view among scholars is that the classic macrostructure has to be presented alphabetically or thematically. However, this paper argues in favour of a possible combination of the alphabetical and thematic ordering of lexical items. Furthermore, it is possible to have three different languages, e.g. French, Yipunu, and English treated equally in the same dictionary article. The paper refers to this microstructural treatment as “microstructural amalgamation”. The implication of this amalgamation leads to questions such as which language should be considered as a language of lemmatisation and based on which criteria.

The consequence of this model is that the central list has to be structured in a hybrid manner to facilitate the dictionary accessibility. This type of dictionary will present both the paraphrase of meaning and translation equivalent with French as a language of lemmatisation. It will also be argued that the outer text (viz. front matter and back matter texts) should play an important role to facilitate rapid accessibility. The target users in this contribution are experts (e.g. medical doctors), semi-experts (e.g. students of a particular field), and lastly laypeople (i.e. any person interested on the particular field for which the dictionary is compiled).

Finally, attention is also drawn to the nature and the status given to the three distinct articles amalgamated into one. Some terminological propositions (e.g. ‘parallel article’) are suggested in order to avoid confusion with the main article, which includes amalgamated dictionary articles or parallel articles.

Mojela, Maropeng V. ■ Polysemy or Homonymy? A Case Study of the Challenges Pertaining to the Lemmatisation of Polysemous and Homonymous Lexical Items in the Compilation of a Sesotho sa Leboa – English Bilingual Dictionary

Sesotho sa Leboa National Lexicography Unit, University of Limpopo, Turfloop Campus, Polokwane, South Africa

This research investigates the challenges regarding the lemmatisation of lexical items which are either polysemous or homonymous in the compilation of bilingual dictionaries in Sesotho sa Leboa and English. These sense relations are properly defined by Lyons (1977: 550), Macdonald (1977: 625), Leech (1983: 228) and many other respected semantic scholars. Leech (1983: 228) defines homonymy as:

“—roughly two or more words having the same pronunciation and/or spelling”

According to Macdonald (1977: 625), homonymy pertains to instances where a word has the same sound and perhaps the same spelling as another, but differs in meaning and origin, while polysemy is defined by Hurford and Heasley as follows:

“a case of polysemy is one where a word has several very closely related senses”

As such, the problem pertaining to the distinction between the lexical items which fall within these two sense relations emanates from the lack of sufficient knowledge regarding the etymological aspects of the lexical items. This insufficient etymological knowledge in the distinction between these two sense relations results in inaccuracy in the classification of the lexical items, i.e. either classifying lexical items as homonyms while these are in fact polysemous, and vice versa.

These problems are usually encountered with words or lexical items where the lexicographers have little, or no etymological background knowledge, which result in improper entry in the dictionary. Sometimes, even the mother tongue speakers of the language who are not conversant with the etymological history of the lexical items do have a problem with the lemmatisation of these words. These words include lexical items such as the following:

- *tshela* ‘six’
- *tshela* ‘jump’

- *tshela* ‘pay tax’
- *tshela* ‘pour’

- *šupa* ‘point’
- *šupa* ‘seven’

- *lekarapa* ‘helmet’
- *lekarapa* ‘person coming for holiday from the cities (mines)’

In order for the lexicographer to be able to give a clear unambiguous distinction between the lexical items which are either homonymous or polysemous the following criteria, which are suggested by Lyons (1977: 550), need to be considered:

- The *relatedness/unrelatedness criterion*. In accordance with this criterion, the lexicographer will need to determine the extent to which the lexical items are related before entering them in the dictionary.
- The *etymological criterion*. This criterion will help the lexicographer to determine whether the lexical items are related or not, basing the relationship on the historical aspect of the words. As such, knowledge of the etymological aspect of the lexical items will be a prerequisite to be able to give correct and accurate lexical entries in the monolingual and bilingual dictionaries.

Motaung, Ntaoleng ■ Microsoft Language Interface Packs (LIPs) for South African Languages

Microsoft South Africa

In partnership with organisations such as the Pan South African Language Board, Microsoft hopes to bring technology to the doorstep of every South African, irrespective of what their mother tongue is, through our Local Language Program.

Despite incredible advances in innovation and the proliferation of personal computers over the past decade, ninety percent of the world's population still lacks access to the technology tools they need to compete in today's knowledge economy.

To address this challenge, Microsoft is actively partnering with nongovernmental organisations (NGOs), governments, and others to find innovative product and community investment strategies that can increase technology access and software availability for individuals and communities around the world.

We were inspired by an experience that Andy Abbar, director International Strategic Initiatives for Microsoft, had during a visit with the ministry of communications in Vietnam some time ago. He noticed that a receptionist at the ministry had covered her computer screen with little Post-it notes. They had writing on them in Vietnamese.

Andy asked a colleague to please translate what were all these notes sitting on the computer screen up and down. And he discovered that down in the lower left the Post-it note on the screen translated the word "Start." In the upper left notes translated "File, Edit" and so on. When the receptionist left the office so that another receptionist could sit down at the computer, the next receptionist could benefit from the same Post-it notes. So they were finding ways to work around the language barriers, not very elegant, but simple and affordable.

As a result, we designed local-language interfaces of our English versions of Windows and Office, making words like File, Print, Save, etc. accessible in local languages.

Launched in March 2004, the Local Language Program is a global Microsoft initiative to build partnerships with governments, universities, and local language authorities to develop language localisation resources for Windows XP Home, Windows XP Professional, and Office 2003 Standard Edition.

Using Web tools provided by Microsoft, sponsoring groups and volunteer translators can develop community glossaries of standardised technical terminology. These glossaries in turn provide the foundation for Language Interface Packs (LIPs), which provide computer users the ability to adapt their copies of Windows and Office to display many commonly used features in their local language.

When complete, each new Language Interface Pack is available as a free download from the Microsoft Download Centre and may also be distributed by participating third parties such as governments, local language authorities, and universities. Through this collaboration, many people around the world will be able to work with PCs – some for the first time – in their native languages.

Besides making technology and its benefits more readily available to people in many different languages, the Local Language Program is also helping to preserve many native and indigenous languages and cultures by making it easier for people to continue working and learning in their native languages.

From the start, we recognised that these interfaces really belong not to Microsoft but to the linguistic communities in the countries themselves, and that we should involve those people very much in the development of these translations and localisations. We worked with governments and universities to let local volunteers suggest the right terms for their language.

These glossaries are all in the public domain so that anyone, a competitor of ours, a local software developer in the country, can use them to create additional software that will benefit the linguistic community, and we believe that this Local Language Program should help stimulate the growth of local software enterprises, help the communities keep their languages alive and help extend the benefits of technology to more people.

Through our work, we identified a localisation pyramid that is used to define the five levels of this Local Language Program. It separates those languages that we view are strategic in nature – these amount to more than 65 languages throughout the world, while 37 have been earmarked as commercial.

- Enablement only
- Community glossary
- LIP localisation
- Partial localisation
- Full localisation

In South Africa, we are pleased to confirm that the first of our eleven official languages, which have been earmarked as strategic to our local business, are LIP localisations. To date, we have announced the availability of isiZulu and Setswana LIP for Windows XP, and are getting ready to announce details for the Afrikaans LIP for both Windows XP and Office 2003.

We're very excited about this programme and look forward to a continued partnership with PanSALB and the greater linguistic community to deliver additional languages.

Moyo, Nobuhle ■ Some Problems Encountered in the Compilation of the First Specialised Dictionary: The *Isichazamazwi Sezomculo* (ISM)

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

The compilation of a specialised dictionary involves several important aspects within the field of specialised lexicography. These basically involve the transfer of scientific and technological knowledge from one linguistic community to the other. It thus becomes a problem to achieve highest translation equivalence for specialised dictionaries within a culture-dependent subject field. Five years after the publication of the *Isichazamazwi SesiNdebele*, a monolingual dictionary, the Ndebele language has seen another compilation of the first specialised dictionary in the area of music, the Ndebele Musical Terms Dictionary termed *Isichazamazwi Sezomculo*, henceforth ISM. It is in the area of music because this is a well established discipline in Zimbabwe which centres on different cultures. It is important as it represents the people's way of life hence it transmits its culture. The ISM comes as a pioneering work in the area of specialised lexicography. The dictionary is trilingual in nature with Ndebele headwords carrying the definitions in Ndebele whilst English and Shona carry the headword equivalents. The ISM comprises of both modern and traditional musical terms and is mainly targeted at tertiary institutions where musical education is done. It is hoped that the documentation of specialised terms will help in the teaching of music in indigenous languages hence facilitating their standardisation. It being the first of its kind the compilers experienced some problems in the various stages of its production. First of all, the very fact that it is a specialised dictionary required special knowledge of the subject field which the compilers did not have. Therefore subject field experts in the area of music had to be

consulted in order to produce a quality product. This required a capacity for planning and co-operation with these specialists. The coining of new words from the Western culture was one of the main challenges that the compilers experienced in the compilation of the ISM. This had to be done in a manner where the created term would relate directly to the concept being referred to and express it clearly. Though, this is one of the strategies that can be implemented in term creation, it posed problems to the compilers. Translation equivalence was also not wholly achievable owing to the absence of cultural relevance of some Shona and Western musical terms in the Ndebele musical scene. It should also be born in mind that following the Ndebele orthographic rules, borrowing or adopting from the source language was problematic as this had to be done within the morphological and phonological structures of the Ndebele language so as to avoid violating the natural word formation techniques of the linguistic community. Some of the challenging areas included the alphabetisation of borrowed terms bearing in mind that a user-friendly dictionary was to be produced at the end. Dealing with synonyms was also problematic as the usual norm in general lexicography is defining the lesser synonyms with the major synonyms; this poses challenges in specialised lexicography as some vital information may be lost in the quest for championing space economy. This article will discuss these problems and give suggestions on how these challenges were handled in the dictionary.

Mpofu, Nomalanga ■ The User Perspective in Lexicography: Comparison Between a General Purpose Dictionary and a Specialised Terminological Dictionary

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

Lexicography in Zimbabwe spans almost a century. Tracing the trends of dictionaries in Zimbabwe shows that their focus and user perspective has shifted through the years. The onset of lexicography in Zimbabwe (then Rhodesia) was through missionary efforts. The bilingual translation dictionaries that were compiled by missionaries served a complementary purpose to the reference works that were available at the time. The aim of these dictionaries was to provide, by means of a number of examples, the use of words in sentences and phrases and illustrations of the application of the principles of word-division on which Standard Spelling is based (Hannan 1959: vii). Lexicographic work can thus be traced back to the colonial days when missionaries were attempting to construct a common orthography for Shona. This common orthography was to be used to construct vocabularies in Shona which would be translations of religious texts into English and vice versa. This was to eventually lead to the translation of the Bible using a common orthography.

The proposed paper thus focuses on the shift of the user perspective in lexicography in Zimbabwe. The focus of the dictionaries that were compiled by the African Languages Research Institute (ALRI) was until recently, on general purpose monolingual Shona and Ndebele dictionaries. The ALRI research agenda is now on specialised terminological dictionaries in specialist subject areas. To date, two specialised terminological dictionaries have been published in Shona, a biomedical terms dictionary, *Duramazwi reUrapu neUtano* (2004) and *Duramazwi reMimhanzi* (2006, with two more in press). The specialist terminological dictionaries are to cater for specialist areas such as medicine, music/ethnomusicology and linguistic and literary terms. The paper will focus on the Shona musical terms dictionary, *Duramazwi reMimhanzi* and it carry out a comparative analysis of selected words in the general purpose monolingual dictionary, *Duramazwi Guru reChiShona*

(2001) and in the specialised musical terms dictionary, *Duramazwi reMimhanzi* (2006) with the aim of highlighting the differences in the needs assessment between the two dictionaries vis-à-vis the intended users of the dictionaries. In other words, defining specialist terminology requires a different needs assessment from a general purpose dictionary for the benefit of both the specialist user and the non-specialist user.

The words selected as illustrative examples are found as entries in both dictionaries. By looking at words such as *ngoma* ‘drum’, *mbende* ‘a kind of dance’, *mbira* ‘mbira’, *chinamwari* ‘initiation ceremony for girls’, *chigubhu* ‘plastic container/double-headed drum’, one can immediately see the differences in depth of analysis between a definition in a general purpose dictionary and that in a specialist terminological dictionary. The comparison between the two dictionaries will be in terms of:

1. the microstructural aspects
2. the macrostructural aspects

By analyzing the difference in scope between the two dictionaries the paper will highlight the difference between lexicographic definitions and encyclopaedic definitions. It will also show that when terminology moves from a general to a specialist area the user-driven approach and lexicographic function govern the overall structure of the dictionary.

Nkomo, Dion ■ The Compiler and User Perspectives on the Access Structure of *Isichazamazwi Sezomculo* (ISM)

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

“The aim of any lexicographic project should be the production of dictionaries to satisfy the needs of the target users of a specific language community” (Gouws 1996: 100) and “The ultimate reason for the existence of dictionaries is the users’ need for information” (Hartmann 1996: 102). As such “all dictionaries are motivated by and judged against the lexical needs of the language user whom they serve” (Hartmann 1983: 5). Svensen goes further to note that:

Not merely information needs of different categories of users are important. An equally important factor is their ability to find and make use of the information given (Svensen 1993: 10).

This paper looks at the access structure of *Isichazamazwi SezoMculo*, a Ndebele Musical Terms Dictionary, henceforth the ISM from both the compiler and the user perspective. The access structure of any dictionary is a set of guide structures that provides a framework within which the availability and accessibility of different information types in that particular dictionary can be evaluated (Louw 1999). The compiler perspective looks at a dictionary or any aspect of a lexicographic work from the point of view of the dictionary-maker, usually guided by the principles of dictionary-making. On the other hand, the user perspective evaluates a dictionary, or any aspect of a lexicographic work from the point of view the user. It considers the reference needs, reference skills and even dictionary culture of the user. As such, this paper therefore presents the challenges faced by the editors of the ISM, of which this presenter is part, in creating an access structure based on the principles of lexicography on one hand and the consideration of the target users on the other. The main thrust was to make all the provided information accessible without violating the principles of lexicography,

a task that was not so easy considering the type of the dictionary, the rules of the language and the users of the dictionary among other factors.

The ISM is the first terminological dictionary in Ndebele, coming five years after the publication of *Isichazamazwi SesiNdebele*, the first monolingual Ndebele dictionary, henceforth the ISN. Unlike the ISN, the ISM is targeted at a restricted group of users who are mainly music students, lecturers, teachers and other practitioners who may want to interpret and understand music genres, instruments and other concepts in Ndebele. Such a group of users generally share a low level of dictionary culture with the entire Ndebele community and not much in terms of the user perspective, except the shared background of reference needs and reference skills which could not be easily relied upon to guide the editors. Of course the compilers benefited from the expertise of a Reference Working Group (RWG) which comprised of mainly Ndebele music lecturers and individuals who are involved in music teaching and research one way or the other.

The dictionary is trilingualised in the sense that while it alphabetises and defines in Ndebele, it also gives Shona and English equivalents of the headwords. About 1 250 entries constitute its macrostructure. The headword's first letter largely determines its location in the macrostructure, although there are special cases of verbs and nouns with vowel-commencing stems. Definitions are encyclopaedic in order to capture clearly the concepts represented by the terms. In some cases the microstructure includes variants, etymology and synonyms. Illustrations of mainly musical instruments and costume types are presented after the main text followed by a selection of Ndebele songs presented using musical notes and symbols in order to illustrate their use in musical notation. Shona and English reverse indices form the last part of the dictionary.

It would be noted that in order to present and make all the provided information accessible, traditional lexicographic conventions and innovations were employed. A guide to dictionary use is included in the front matter and a network of referencing, both implicit and explicit, has been created between the main text and other texts, e.g. from an entry in the main text to an illustration in the back matter text. Such conventions and innovations are an integral part of the access structure of the ISM and this paper highlights the challenges encountered by the editors in creating it in a manner that conforms to the principles of dictionary making while at the same time facilitating users' access to all the information. Through the paper, the presenter hopes to provide insights with regard to the access structure of the dictionary thereby facilitating the consultation and use of the dictionary. The presenter also hopes to provide more insights with regards to African languages lexicography, especially the dictionary structure, the compiler vis-à-vis the user perspective.

Nyangone Assam, B. ■ The User Perspective in a Learners' Dictionary

School of Languages and Literature, University of Cape Town, Cape Town, South Africa

The user of a dictionary is without question the central figure in lexicography. The dictionary is compiled with the needs of the intended users in mind. Therefore the target users should determine the purpose of a dictionary, the character, and the material presented in it. This paper aims to present ways in which this can be achieved in the planning phase of a lexicographic endeavour.

The planning, the design of any dictionary should evolve around the goal of making a dictionary that will satisfy the needs of its target users. This paper analyses some of the aspects that can help to attain this goal, namely the users' needs, their reference skills, their

expectations, their usage situation. The stress is on the importance of planning every step of the compilation process with the needs of the users in mind.

The dictionary planning should be done according to the needs of its identified users and its genuine purpose. In present-day lexicography, the users' needs have a definite influence on the design, structure and contents of any dictionary. The utility of the dictionary has never been in doubt, but the question is whether the dictionary can satisfy or solve the users' linguistic needs and problems.

There are many reasons why users consult a dictionary, one of the main reasons being that they experience linguistic difficulties with the meaning, spelling, or translation equivalent of a lexical item. Lexicographers must determine in which situations the user will use the dictionary and for what purpose, whether it be communication-orientated or knowledge-orientated or even both. The usage situation and user situation have great influence on the structures of the dictionary. The user-profile will determine the functions of the dictionary.

The task of the lexicographic team will be made easier by adding a proper need assessment to the process of identifying the target users. The user perspective necessarily plays an important role in the conception of new learner's dictionaries. The conception of future dictionaries will have to adhere to a lexicographical-pragmatic approach orientated to at least three parameters: the user, the situation in which the dictionary will be used, and the nature and extent of the data treated in the dictionary.

This paper appeals to any lexicographic endeavour to turn from the tradition of focussing only on the transfer of data, to concentrate on the importance of the user in modern-day lexicography. This paper will analyse the above avenues with special reference to learner's dictionary and to stress that dictionaries should be designed, planned with a specific set of users in mind and for their specific needs. It is about formulating in the planning stage the genuine purpose of the intended dictionary.

According to Gouws (2004) the genuine purpose implies that a dictionary is produced so that the target user who uses the dictionary in a typical usage context will have an instrument to assist him/her in achieving a successful dictionary consultation procedure by reaching the goals that motivated the search. The genuine purpose of a dictionary should be to ensure successful dictionary consultation procedures. A successful dictionary consultation procedure depends on the way in which the needed linguistic information can be retrieved. Therefore the user should be the central focus at the starting point, as well as the end of a lexicographic project the final product the dictionary will be evaluated by its capacity to satisfy the user search. Lexicographers have to find ways in the presentation of the data to facilitate the user's quest and aim to his/her success in the consultation process.

Nyathi, Cratilwe & Marble Timbe ■ Steps and Procedures Involved in Data Processing for Lexicographers

African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe

The co-presenters will give a comprehensive description and demonstration of the steps and procedures followed by the ALRI Data Entry Operators (DEOs) to process and store both written and oral materials collected during field work by teams of Shona and Ndebele lexicographers in the ALLEX Project. These materials are used to build electronic corpora and sound databases that are matched with the text files. The presenters will give a practical demonstration of the computational tasks that are performed by the DEOs as part of the

support team for the lexicographers. The procedures range from the collection of materials in the field to the manipulation and preservation of the various materials using quite a number of storage methods. There are two viable units namely Shona Language Unit (SLU) and the Ndebele Language Unit (NLU). These units have built a library and archive with several audio tapes from field research, transcription of the same interviews in long hand, books, technical reports, electronic files, presentations and various other reports. These materials are used in building electronic corpora for the two major national languages of Zimbabwe namely Shona and Ndebele although the same procedures can also be applied to other minority languages in future. The presentation aims at describing and demonstrating computational processes used in the processing and storage of lexicography work collected during fieldwork by researchers and then processed and stored by Data Entry Operators in the ALLEX Project since 1992. The process ranges from data collection, processing, storage, to sending text to the corpus. The paper will highlight major procedures involved such as Scanning (converting from a hard copy into an electronic copy), Encoding (keying in or typing transcribed materials into a computer), Digitalisation (transferring sound from audio tapes to the computer), CD Burning (creating sound saved from audio tapes into the computer through digitalisation), Proof reading, Tagging (assigning linguistic codes or rules to the encoded electronic files), Parsing (a programme specifically designed to identify grammatical errors in the tagged file). These processes assist lexicographers in building corpora of the indigenous languages. The corpora acts as a databank that students, teaching staff, scholars, researchers and the general public can consult. The paper will highlight methods used for the preservation and conservation of the originals for use by generations to come. The paper shall also highlight how ALRI has used the database to process and store data. The software used for the database is File Maker Pro Versions three, seven and eight. Using this program an application was created by Professor Christian Emil Ore which is specifically designed to store information using different fields for future reference. It includes the Interview Registry (record of encoded books), Tape Registry (tape information) and the CD Registry (CD information). These databases are the major output of the processing and the DEOs' routine tasks include creation, expansion and maintenance of the three registries. As part of the description of their work, the DEOs will show how data that has been collected since the inception of the ALLEX Project in 1992 has processed and stored at ALRI. These steps and procedures are now standardised and can be replicated for any other African language. The main focus of this presentation therefore, is to show how data collected for the purposes of making dictionaries and other lexical tools has been stored and processed using information technology tools since 1992.

Nyota, S. and J. Mapara ■ The Impact of Lexicographic Work on Language Use: A Case of Shona Monolingual Dictionaries in Zimbabwe

Masvingo State University, Department of Curriculum Studies & Department of Languages, Literature and Music, Masvingo, Zimbabwe

This paper examines the impact of the current lexicography work in Zimbabwe on some sectors of Shona language use. It looks specifically at monolingual lexicography work in the Shona language being carried out by the African Languages Research Institute (ALRI) and the African Languages Lexical Project (ALLEX). The paper analyses how users of the Shona language in different sectors are responding to different lexicographical products being published by ALRI. Noted sectors are education (from primary to tertiary level), the

electronic and print media, advertising and the medical field. The paper maintains that Shona monolingual lexicography has resulted in language raising and awareness. It has also resulted in term creation and standardisation of the Shona language. The language has also been given the abstractive power it needs to explain it and other concepts. All these have enabled diglossia leakage from Shona Low (L) to Shona High (H) in some areas of the Shona language usage. The overall effect is that Shona is now used in some formal sectors previously reserved for the English language in Zimbabwe.

The paper asserts that the production of monolingual Shona dictionaries, beginning with the first one in 1996 have had a tremendous impact on all the sectors mentioned above. It has been observed that the publishing of general dictionaries *Duramazwi reChiShona* (1996) and *Duramazwi Guru reChiShona* (2001) resulted in some sectors requesting for specialised Shona dictionaries in their areas. The publication of *Duramazwi reUrapi neUtano* (2004) was in response to a request by the medical sector for a specialised medical Shona dictionary. Following *Duramazwi reUrapi neUtano* the same sector has even asked for a bigger Shona medical dictionary with more entries and depth.

Likewise the country's schools examinations body, the Zimbabwe School Examination Council (ZIMSEC) has also responded by asking ALRI and other stakeholders in education to help in standardising Shona linguistic and literary terms and is now closely working with ALRI and their publisher Mambo Press to produce the glossary that was requested by ZIMSEC. Another response is a result of the observation of wide spread code-switching in the teaching of mathematics in the primary schools made by some lecturers in the Curriculum Studies Department at Masvingo State University. This has resulted in them embarking on a specialised Junior ChiShona Mathematics Dictionary project and this Junior ChiShona Mathematical dictionary is to be published by end of 2007. The Shona Language and Culture Association (SLCA) has also welcomed the publication of these dictionaries and is encouraging its members to write Shona text and reference books using the resultant standardised language.

The paper asserts that the publication of these Shona monolingual dictionaries opened floodgates for the publication of Shona grammars and other Shona language texts like *Manyorerwo eChiShona* (1999) and *Dudziramutauro reChiShona* (1999). All these were published three years after the publication of the first Shona monolingual dictionary in 1996. Before these publications, the texts for Shona grammar were in English and the language itself (Shona) was taught in English. Previously it was hitherto thought that the Shona had no power to explain grammatical ideas and concepts. The same is true regarding rules and regulations that related to Shona orthography. It was also observed that some tertiary institutions have responded by using ChiShona as a medium of teaching the Shona language. Such institutions are Masvingo State University, the University of Zimbabwe's Faculty of Education and all teacher training colleges in the country.

The general dictionaries have also favourably impacted on both the print and electronic media. What was once a bilingual weekly newspaper was unbundled for wider coverage and depth after standardisation of Shona and Ndebele through lexicography into two separate newspapers namely *Kwayedza* (Shona) and *Umthunywa* (Ndebele). Another observation is that standardisation of the Shona language resulted in the increase of the time slotted for news in Shona on both radio and television from five (5) to fifteen (15) minutes and in addition the news has been neatly divided into sections to cater for local, regional, arts and culture and sports news.

Another notable impact since the publication of these monolingual dictionaries is in advertising. There has been an increasing interest in advertising in Shona by both the informal business sector and the corporate world. For the corporate world English adverts were the norm but the standardisation of Shona is one of the reasons why there has been a

shift from English to Shona in advertising. One cannot escape noticing advertisements that are in Shona whether reading the newspaper, shopping, watching television or travelling.

The paper concludes that there has been a positive response in lexicography from users of the Shona language.

Prinsloo, Danie J. & Elsabé Taljard ■ The Value of Northern Sotho Definitions in LSP Dictionaries for Mathematics and Science: A Preliminary Analysis

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

As pointed out by Gouws and Prinsloo (2005: 42), the majority of African language speakers find themselves in a pre-dictionary culture environment, characterized by poor dictionary using skills on the side of the users and limited access to high quality dictionaries. This is even truer with regard to the use of LSP dictionaries, since there is an almost total absence of LSP dictionaries for these languages. A notable exception in this regard is the Multilingual Explanatory Science and Math Dictionaries, published in 2005 by Clever Books as part of the Teacher Mentorship Programme, which is run by the University of Pretoria. The aim of this programme is to assist teachers and learners to understand the concepts and questions in the Natural Sciences and Physical Sciences curricula of grades 8 to 12. In these dictionaries, terms are defined first in English, followed by translations of the English definitions into isiZulu, Setswana, Sepedi and Tshivenda.

The research upon which this paper is based forms part of an evaluation of the efficacy of these dictionaries in the multilingual educational environment. The focus of our investigation is to ascertain (a) whether the target user possesses the necessary dictionary using skills to effectively make use of the dictionary, and (b) whether the benefit of exposure to definitions of terms in the home language is significant in the decoding of the meaning of science and math terms. Data were collected by means of a questionnaire that was filled in by members of the intended target user group, a number of teachers and selected students at the University of Pretoria. The focus of the eventual analysis will be on the performance of learners in Grades 8-12. Presented here will be the results for the total number of respondents, i.e. 53.

The questionnaire consists of three sections that were administered separately and consecutively. The first section is answered without the aid of the dictionary itself. The questions in this section are aimed mostly at general dictionary using skills; testing whether learners are, for example, familiar with the alphabetical ordering system on both single and multiword levels. Also in this section, multiple-choice questions relevant to the natural sciences curriculum are put to learners. In the second section, the same multiple-choice questions are asked, but this time learners are provided with the English definitions of the relevant terms. Only when answering the third section of the questionnaire, do learners have access to the dictionaries. The multiple-choice questions are posed for a third time, but this time learners have the definitions of terms in both English and their home language at their disposal. This is specifically aimed at measuring the impact of having the definitions in the home language on the decoding process. This section furthermore contains practical exercises on dictionary use, where learners are instructed to answer questions using the dictionary. Learners are also provided with the opportunity to express their needs and preferences in this section.

The presumed low dictionary culture was confirmed by the survey since the overwhelming majority of respondents indicated that they rarely use any dictionary. It was

however ascertained that respondents do possess basic dictionary skills in the sense of understanding the access structure of the dictionary and an ability to use it effectively in the look-up procedure. Respondents also indicated a positive attitude towards dictionary use, especially in the final section of the questionnaire where they expressed their opinions in respect of the value of English and Northern Sotho / Sesotho sa Leboa definitions to understand the meaning of a randomly selected set of Mathematical and Scientific terms. As for the main aim of the investigation, i.e. whether consultation of the Northern Sotho / Sesotho sa Leboa definitions over and above the English definitions would benefit the users, the results indicate an increase in the decoding of the meaning of the Mathematical and Science terms. The following table indicates the increase in correct percentages of answers for 11 multiple choice questions on the basis of (1) without the assistance of a dictionary, (2) access to an English definition and (3) access also to paraphrase of meaning in Sesotho sa Leboa.

Table 1: Percentages of correct answers: 1 = no dictionary → 2 = with English definition → 3 = adding Sesotho sa Leboa definitions

1	41.67	58	64	48	74.51	69.39	67.31	46.81	61.22	53.06	31.11
2	76	83.67	64.71	56	76.47	75	92.31	73.47	81.63	86.54	52.94
3	87.76	85.42	68.63	56.25	78.85	77.08	94.23	81.25	87.23	96	54.9

Focusing on the added value obtained by the Northern Sotho / Sesotho sa Leboa definitions, the arrows in the following graph indicate an increase in the number of correct answers given in the multiple choice sections A versus B versus C of eight randomly selected Mathematical and Science terms. The first two bars in each set of six indicate the split between correct and incorrect answers without the benefit of a dictionary, the third and fourth bar reflect the split between correct and incorrect answers using an English definition and the final pair of bars in each block the split when Northern Sotho / Sesotho sa Leboa definitions were consulted.

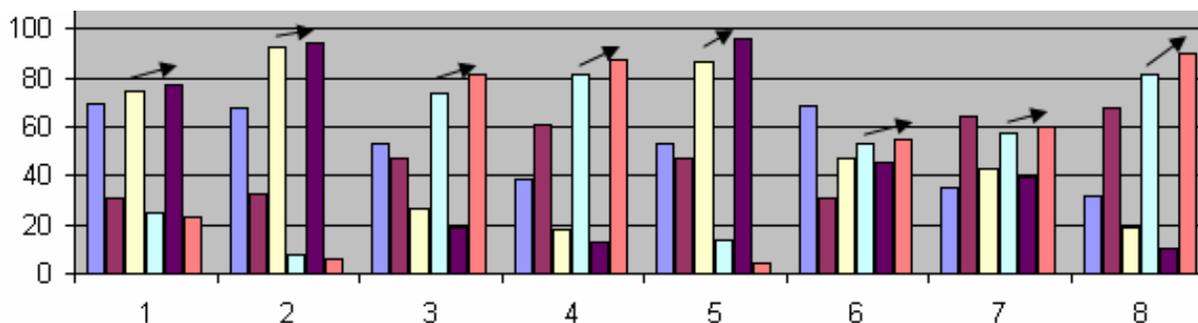


Figure 1: Added value obtained by the Northern Sotho / Sesotho sa Leboa definitions

These preliminary results therefore clearly indicate that the target-user of the Multilingual Explanatory Science and Math Dictionaries does benefit from having access to definitions of basic chemistry and math terms in their home language.

Roux, Justus C. ■ Human Language Technologies and the Standardisation of Lexical Databases

Stellenbosch University Centre for Language and Speech Technology (SU-CLaST), Stellenbosch, South Africa

Human Language Technologies (HLT), as enabling technologies, are regarded as core technologies of the twenty first century. The growth in research and development in this field has been extra-ordinary over the last two decades, and as such, hold great promise for deployment of language based products in a multilingual environment such as South Africa.

The aim of this presentation is to exemplify the nature of different types of HLT applications, focusing on the vital role of the lexicon in the process, and the need for standardisation of lexical databases.

HLT research and development focus on the creation of language and speech based products for an extremely wide range of applications. These products can either be speech and/or text based and invariably comprise of a linguistic as well as a technological component. An appropriate lexicon normally underpins the development of the linguistic component where natural language processing (NLP) and / or natural language understanding (NLU) needs to take place. The nature and format of the lexicon is of prime importance, not only for the functioning of the system itself, but also for the system potentially interacting with other systems.

HLT systems find their applications in machine translation, information extraction and/or retrieval, natural language generation, multilingual speech based communication systems, etc. within environments such as e-learning, e-business, e-government etc. At this point the relevance of classical dictionaries and terminologies for special purposes will be considered against the backdrop of a need for the development of interoperable domain specific lexicons.

The development of interoperable domain specific lexicons is highly dependent on standardised formats of presentation, hence the ISO TC37/SC4 committee embarked on the development of a Lexical Markup Framework (LMF) which is regarded as a metamodel to provide a standardised framework for the construction of computational lexicons. "One of the crucial aspects impacting human language technologies (HLT) in general and natural language processing (NLP) in particular, with special consideration of human-orientated translation technologies, is the need to optimize the production, maintenance, and extension of lexical resources. ISO 24631:2004(E), p.1.

Some of the main features of the LMF will be presented and discussed against the backdrop of the findings of a workshop, *LMF: working to reach consensual ISO standards on lexicons*, that will be attended by the author at the LREC2006 Conference that is to take place in Genoa at the end of May 2006.

The paper will conclude with some remarks on the challenges this model will pose for lexicon development and standardisation in the local languages of South Africa, as well as the role of the local StanSA TC37 committee in this regard.

Sibula, Pumlani ■ Furthering the Aims of Multilingualism through Integrated Terminology Development

Stellenbosch University Language Centre, Stellenbosch, South Africa

The Stellenbosch University, being an Afrikaans institution in its tuition, is faced with the challenge of catering other students, especially the black students who have no previous knowledge of Afrikaans language. In responding to the Constitution of South Africa, which provides for multilingualism and development of linguistic heritage, Stellenbosch University Language Centre has recently established a Unit for isiXhosa for a primary purpose of promoting isiXhosa to a status of academic language, as one of the three official languages of the Western Cape Province, through a project of terminology development. With this project, Stellenbosch University wishes to make its own contribution to the development of the indigenous language of South Africa. IsiXhosa, respectively, being the most important indigenous language in the region.

It is the primary objective of this terminological project to assist isiXhosa students to understand Afrikaans textbooks being compiled in their own language, which will result in significant effects in terms of increasing their academic performance given that cognition and conceptualization has taken place through their home languages. This terminological project is done in an integrated way, as it is done in close collaboration with the different academic departments and faculties to ensure that the terminology of the day in each of these institutions is accommodated. In compiling these terminologies both the subject and language specialists are fully consulted through physical and virtual contacts. These terminology lists are compiled from English as a source language (SL) and are defined within the needs the specific subject fields, such that each term is given to one concept. After that, the result is officially presented to the concerned departments in printed form, e.g. a booklet.

Secondly, this abstract explores different methods employed in term formation of isiXhosa as a scientific and technical terminology, e.g. loan words, transliteration, etc.

In conclusion, the Unit for isiXhosa is not only engaged in promoting isiXhosa through terminology project, but also by creating opportunities to further use of isiXhosa through offering accredited language acquisition courses to both staff members and students who have no previous knowledge of isiXhosa.

Wolvaardt, Jill ■ Horses for Courses: Researching User Perspectives – Different Approaches for Different Dictionaries

Dictionary Unit for South African English (DSAE), Rhodes University, Grahamstown, South Africa

The Dictionary Unit for South African English (DSAE) has two commercially published dictionaries currently on the market. The *Illustrated School Dictionary for Southern Africa* (Francolin, 1997; Maskew Miller Longman, 2001) is a monolingual primary school dictionary aimed at learners who speak English as an additional language. The *South African Concise Oxford Dictionary* (Oxford, 2002) is a monolingual English desktop dictionary, with the user assumed to be fluent in English. A key feature each of these dictionaries has in common is that they are based on original dictionaries which were not designed with the South African English speaker – whether first, or additional language – in mind. The DSAE's task was to adapt the original material to make it 'user-friendly' to South Africans.

This presentation will discuss the methodology of the user research conducted for each dictionary. The approaches were quite different: in one case, sample materials were widely trialled in the field by expert practitioners; in the other small-scale localised research was conducted by the lexicographers themselves. Reference will also be made to the research conducted by Oxford University Press (OUP) and the University of Waterloo, Canada, in preparation for the online edition of the *Oxford English Dictionary*.

In preparing the *Illustrated School Dictionary* the publisher convened a panel which included representatives of two of South Africa's leading literacy non-governmental organisations (NGOs), the Molteno Project, and the READ Educational Trust, as well as the DSAE. The brief was to take an existing database of entries from a dictionary for English second language (ESL) primary school learners in India, and to tailor it for South African schools. The dictionary was to include everyday English words, curriculum vocabulary, and a small core of words considered to be entrenched in South African English. The DSAE prepared sample text, which was then printed in a trial material booklet. Molteno drew up a set of exercises for testing the material in schools. Between them, the Molteno Project and the READ Educational Trust trialled the sample materials in 46 schools across six provinces. The results of these tests and their effect on the final product will be discussed.

Research for the *South African Concise Oxford Dictionary* (SACOD) was conducted differently, largely because the target user could be assumed to share many of the characteristics of the user of the *Concise Oxford Dictionary* (COD), on which SACOD was to be based. However, the DSAE editors felt it important to check this assumption, and to discover if there were any dictionary-use conventions which were unfamiliar or could be improved to enhance accessibility. As the DSAE is based in Grahamstown, a university city with a large English-speaking population engaged in a spread of occupations, it was felt that a sufficient cross-section of prospective users could be found to enable the research to be conducted locally. The team drew up a questionnaire in which respondents were asked to consider different elements of a spread of entries as they appeared in the COD. A significant constraint in drawing up the questionnaire was that there was little room for manoeuvre in changing the design of the existing dictionary, so any adaptations which might appear useful as a result of the research would have to fit largely within the already established format. Once again, the research results and the impact on the finished dictionary will be discussed.

The session will conclude by considering the rationale behind the respective user-research, and examining whether there are any general principles which might be applied for future dictionaries.