



**African Association  
for Lexicography**

---

**AFRILEX 2008**  
**Thirteenth International Conference**

*Printed and Electronic Dictionaries  
and the Future*

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

30 June – 3 July 2008

## A FEW WORDS FROM THE PRESIDENT

On behalf of AFRILEX, I wish to welcome you to Stellenbosch for our Thirteenth International Conference. We hope you will enjoy the programme, which contains 38 papers by our colleagues, featuring amongst them the keynote addresses of Proff. Patrick Hanks, Ngo Semzara Kabuta and Piet van Sterkenburg, as well as the contact with fellow lexicographers, the hospitality of the Bureau of the WAT, the University and the town of Stellenbosch, previously referred to by Prof. Van Sterkenburg as a *dictionopolis*.

We wish to thank you for attending the pre-conference workshop and the conference and for your loyal support of our association and for lexicography in Africa.

In AFRILEX, we have already become accustomed to the high technical quality of our journal *Lexikos* and the meticulous care with which Dr. Johan du Plessis and Mrs. Riette Ruthven prepare this journal. The same winning team volunteered to produce this conference brochure — and the quality shows! On behalf of AFRILEX, thank you ever so much.

Rufus H. Gouws  
President AFRILEX

Abstract Reviewers: AFRILEX Board  
Programme Committee: AFRILEX Board  
Tel.: +27-12-420-2494  
Fax: +27-12-420-3163  
E-mail: [elsabe.taljard@up.ac.za](mailto:elsabe.taljard@up.ac.za)  
Internet: <http://afrilex.africanlanguages.com/>  
Conference Venue: Letteregebou / Arts Building  
Cnr Merriman & Andringa Streets  
Stellenbosch University Campus

## CONTENTS

<b><i>Pre-Conference Workshop</i></b>	1
Practical Lexical Analysis Using Corpus Tools	
<b><i>Conference Programme</i></b>	2
Printed and Electronic Dictionaries and the Future	
<b><i>Keynote Addresses</i></b>	
<i>Patrick Hanks (Masaryk University, Brno, Czech Republic)</i> ▪ The Structure, Typology, and Function of Similes	
<i>Ngo Semzara Kabuta (Ghent University, Ghent, Belgium)</i> ▪ Lexicography in Central Africa, with Special Focus on Cilubà	
<i>Piet van Sterkenburg (Institute for Dutch Lexicology, Leyden, The Netherlands)</i> ▪ Lexicography and Magic Spells	
<b><i>Parallel and Special Sessions</i></b>	
<i>M. Alberts</i> ▪ Translation-oriented Terminography in the Electronic Age	6
<i>H.L. Beyer</i> ▪ A Functionalistic Approach to the Lexicographic Treatment of Idioms	7
<i>A. Cichocka and K. Brits</i> ▪ Dictionaries as One of the Intercultural Communication Tools: South African Cultural Terms on Polish Soil	8
<i>A.E. Cloete, B.P.D. Gabriels, F.J. Lombard and G.J. van Wyk</i> ▪ Comprehensive Dictionaries in the Electronic Era: Challenges and Solutions	10
<i>G. De Pauw and G.-M. de Schryver</i> ▪ Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes	11
<i>G.-M. de Schryver</i> ▪ An Analysis of the <i>Oxford Bilingual School Dictionary: Northern Sotho and English</i>	14
<i>V. de Tier</i> ▪ The <i>Zeeuwse Klapbank</i> , a Multimedia Exhibition about Zeeland Dialects	17
<i>H. Fourie</i> ▪ The Lexicography of Sign Language: Its Past, Present and Future in South Africa	19
<i>R.H. Gouws and D.J. Prinsloo</i> ▪ Surrogate Equivalence in Bilingual Dictionaries with Special Reference to Dictionaries Bridging English and African Languages	20
<i>S. Hadebe</i> ▪ The Uncertain Future of Dictionary-making in Zimbabwe	22
<i>U. Heid and D.J. Prinsloo</i> ▪ A Bilingual Dictionary for a Specific User Group: Supporting Setswana Speakers in the Production and Reception of English	24
<i>M. Kadenge and G. Mheta</i> ▪ <i>Duramazwi reDudziramutuario neUvaranomwe</i> : A Milestone in the Development of Shona Lexicography	26
<i>L.M. Kganyago</i> ▪ An Analysis of the Interface between Meanings and Translation: A Case Study of the Northern Sotho Lord's Prayer 'Tatawešo wa Magodimong'	27
<i>J. Klein</i> ▪ South Africa's New African Language Dictionaries and their Use for the African Language Speech Communities	28

<i>M.J. Mafela</i> ▪ The Component of Etymology and Bilingual Dictionaries: The Case in African Languages	29
<i>W.M. Mojapelo</i> ▪ Pronunciation and Spelling of Foreign Lexical Items: A Case Study of the Writing of Foreign Lemmas as Adoptives in Sesotho sa Leboa	31
<i>V.M. Mojela</i> ▪ Borrowing and Loan-words: A Convenient and Accurate Method of Lemmatizing Newly Acquired Lexical Items in Sesotho sa Leboa	32
<i>D.M. Mphahlele</i> ▪ Lemmatization Problems of Neologisms in Northern Sotho-English Dictionaries: A Lexicographical Perspective on Media Influence on Northern Sotho Spoken Language	33
<i>M. Nabirye</i> ▪ Dictionary Testing in Un-/Less Documented Languages: A Case Study of Lusoga	35
<i>H.S. Ndinga-Koumba-Binza and J.C. Roux</i> ▪ Vowel-length Representation in Civil Orthography: Implications for Dictionaries	36
<i>E. Ndllovu</i> ▪ Sense Relations in the Treatment of Meaning in the Ndebele Monolingual Dictionary <i>Isichazamazwi SesiNdebele</i>	38
<i>D. Nkomo</i> ▪ An Analysis of Subject-field Representation in Special-field Dictionaries: In Pursuit of a Theory of Lemma Selection for Special-field Lexicography	39
<i>B. Nyangone Assam and H.S. Ndinga-Koumba-Binza</i> ▪ Lexicography and the French Language in Gabon	42
<i>H.J. Ojwang</i> ▪ A Study of Luo Ethnobotanical Dictionaries	44
<i>C.L. Pienaar</i> ▪ Towards a Corpus of South African English: Corraling the Subvarieties	45
<i>T.Z. Ramaliba</i> ▪ The Significance of Noun Classes in the Lemmatization of Nouns in a Tshivenda Dictionary	46
<i>L.N. Sakwa</i> ▪ An Analysis of the Usage Labelling Practices in English Lexicography	47
<i>M.R. Selokela</i> ▪ The Relevance of <i>Pukuntšuthaloši ya Sesotho sa Leboa</i> in a Classroom Situation	48
<i>P. Swanepoel</i> ▪ Dictionary Criticism: The Lame Leg of Lexicographic Theory and Practice	49
<i>K.H. Tan</i> ▪ Media Specific Features of Electronic Dictionaries: Dictionary Users' Perception	50
<i>S. Tarp</i> ▪ A Lexicographic Approach to Language Policy	51
<i>K. Thangalimodzi</i> ▪ The Role of an Electronic and Online English-Chichewa Dictionary in Technology Pedagogy and E-Learning	52
<i>S.L. Tshikota</i> ▪ An Investigation of the Treatment of Nouns in Tshivenda Dictionaries: A Morphosyntactic Approach	53
<i>M. van der Merwe</i> ▪ The Pedagogical Purpose of Dictionaries in South Africa: A Perspective on Dictionary Education	54
<i>J. Zondo</i> ▪ Affirming Verb Lexemes in <i>Isichazamazwi SesiNdebele</i> and in <i>A Practical Ndebele Dictionary</i> : The Case of <b>-wa</b> Verbs	55

## PROGRAMME

**Monday 30 June 2008: PRE-CONFERENCE WORKSHOP**

**Theme:** *Practical Lexical Analysis Using Corpus Tools*

**Presenter:** *Patrick Hanks*, Masaryk University, Brno, Czech Republic

**Time:** 08:30–16:30

**Place:** Room 693, Sixth Floor

Letteregebou (Arts Building)

Cnr Merriman & Andringa Streets

Stellenbosch University Campus

08:30–09:00: Registration: Foyer of the Arts Building

09:00–10:30: **Session 1**

10:30–11:00: Tea

11:00–12:30: **Session 2**

12:30–13:30: Lunch

13:30–15:00: **Session 3**

15:00–15:30: Tea

15:30–16:30: **Session 4**

**Tuesday 1 July 2008: CONFERENCE DAY 1**

**Theme:** *Printed and Electronic Dictionaries and the Future*

**Place:** Letteregebou (Arts Building)

Cnr Merriman & Andringa Streets  
Stellenbosch University Campus

08:00–08:30: Registration: Foyer of the Arts Building

<p><b>Official Opening: Venue A: Room 229</b> 08:30–09:00: Word from the Vice-President of AFRILEX <i>M.J. Motsapi</i> Welcome by the Rector and Vice-Chancellor of the University of Stellenbosch <i>Prof. H.R. Botman</i></p>
---

<p><b>Keynote Address 1: Venue A: Room 229</b> 09:00–09:55: The Structure, Typology, and Function of Similes <i>Patrick Hanks</i> (Masaryk University, Brno, Czech Republic)</p>
--

10:00–10:25: Tea

**Parallel Sessions**

	<b>Venue A: Room 229</b>	<b>Venue B: Room 227</b>
10:30–10:55:	The Component of Etymology and Bilingual Dictionaries: The Case in African Languages <i>M.J. Mafela</i>	A Lexicographic Approach to Language Policy <i>S. Tarp</i>
11:00–11:25:	Dictionary Testing in Un-/Less Documented Languages: A Case Study of Lusoga <i>M. Nabirye</i>	Dictionary Criticism: The Lame Leg of Lexicographic Theory and Practice <i>P. Swanepoel</i>
11:30–11:55:	A Functionalistic Approach to the Lexicographic Treatment of Idioms <i>H.L. Beyer</i>	Sense Relations in the Treatment of Meaning in the Ndebele Monolingual Dictionary <i>Isichazamazwi SesiNdebele</i> <i>E. Ndlovu</i>
12:00–12:25:	Surrogate Equivalence in Bilingual Dictionaries with Special Reference to Dictionaries Bridging English and African Languages <i>R.H. Gouws and D.J. Prinsloo</i>	Lemma-tization Problems of Neologisms in Northern Sotho–English Dictionaries: A Lexicographical Perspective on Media Influence on Northern Sotho Spoken Language <i>D.M. Mphahlele</i>

12:30–13:25: Lunch

## Parallel Sessions

### Venue A: Room 229

### Venue B: Room 227

13:30– 13:55:	An Analysis of Subject-field Representation in Special-field Dictionaries: In Pursuit of a Theory of Lemma Selection for Special-field Lexicography <i>D. Nkomo</i>	Translation-oriented Terminography in the Electronic Age  <i>M. Alberts</i>
14:00– 14:25:	An Analysis of the <i>Oxford Bilingual School Dictionary: Northern Sotho and English</i> <i>G.-M. de Schryver</i>	The Uncertain Future of Dictionary-making in Zimbabwe  <i>S. Hadebe</i>
14:30– 14:55:	An Analysis of the Interface between Meanings and Translation: A Case Study of the Northern Sotho Lord's Prayer 'Tatawešo wa Magodimong' <i>L.M. Kganyago</i>	Media Specific Features of Electronic Dictionaries: Dictionary Users' Perception  <i>K.H. Tan</i>
15:00– 15:25:	The Relevance of <i>Pukuntšuthaloši ya Sesotho sa Leboa</i> in a Classroom Situation  <i>M.R. Selokela</i>	South Africa's New African Language Dictionaries and their Use for the African Language Speech Communities  <i>J. Klein</i>
15:30– 15:55:	The Significance of Noun Classes in the Lemmatization of Nouns in a Tshivenḡa Dictionary <i>T.Z. Ramaliba</i>	Towards a Corpus of South African English: Corraling the Sub-varieties  <i>C.L. Pienaar</i>

16:00–16:25: Tea

16:30–18:00: **GENERAL MEETING: Venue A: Room 229**

19:00: **Conference Dinner**

## Wednesday 2 July 2008: CONFERENCE DAY 2

### Keynote Address 2: Venue A: Room 229

08:30–09:25: Lexicography in Central Africa, with Special Focus on  
Cilubà

*Ngo Semzara Kabuta* (Ghent University, Ghent, Belgium)

### Parallel Sessions

#### Venue A: Room 229

#### Venue B: Room 227

09:30– 09:55:	Vowel-length Representation in Civili Orthography: Implications for Dictionaries <i>H.S. Ndinga-Koumba-Binza and J.C. Roux</i>	The <i>Zeeuwse Klapbank</i> , a Mul- timedia Exhibition about Zeeland Dialects <i>V. de Tier</i>
------------------	---	---

10:00–10:25: Tea

### Parallel Sessions

#### Venue A: Room 229

#### Venue B: Room 227

10:30– 10:55:	<i>Duramazwi reDudziramutaro neUvaranomwe</i> : A Milestone in the Development of Shona Lexicography <i>M. Kadenge and G. Mheta</i>	The Pedagogical Purpose of Dictionaries in South Africa: A Perspective on Dictionary Edu- cation <i>M. van der Merwe</i>
11:00– 11:25:	Publishers' Session	<i>WikiOmega Presentation</i>
11:30– 11:55:	<b>Venue: Bureau of the WAT, 115 Banghoek Road</b> Comprehensive Dictionaries in the Electronic Era: Challenges and Solutions <i>A.E. Cloete, B.P.D. Gabriels, F.J. Lombard and G.J. van Wyk</i>	Pronunciation and Spelling of Foreign Lexical Items: A Case Study of the Writing of Foreign Lemmas as Adoptives in Sesotho sa Leboa <i>W.M. Mojapelo</i>

12:00–18:00: **Conference Excursion**



**Thursday 3 July 2008: CONFERENCE DAY 3**

**Keynote Address 3: Venue A: Room 229**

08:30–09:25: Lexicography and Magic Spells

*Piet van Sterkenburg* (Institute for Dutch Lexicology,  
Leyden, The Netherlands)

**Parallel Sessions**

**Venue A: Room 229**

**Venue B: Room 227**

09:30– 09:55:	Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes <i>G. De Pauw and G.-M. de Schryver</i>	An Investigation of the Treatment of Nouns in Tshivenda Dictionaries: A Morphosyntactic Approach <i>S.L. Tshikota</i>
10:00– 10:25:	The Role of an Electronic and Online English–Chichewa Dictionary in Technology Pedagogy and E-Learning <i>K. Thangalimodzi</i>	A Bilingual Dictionary for a Specific User Group: Supporting Setswana Speakers in the Production and Reception of English <i>U. Heid and D.J. Prinsloo</i>
10:30– 10:55:	Borrowing and Loan-words: A Convenient and Accurate Method of Lemmatizing Newly Acquired Lexical Items in Sesotho sa Leboa <i>V.M. Mojela</i>	An Analysis of the Usage Labelling Practices in English Lexicography <i>L.N. Sakwa</i>

11:00–11:25: Tea

**Parallel Sessions**

**Venue A: Room 229**

**Venue B: Room 227**

11:30– 11:55:	Dictionaries as One of the Intercultural Communication Tools: South African Cultural Terms on Polish Soil <i>A. Cichocka and K. Brits</i>	Affirming Verb Lexemes in <i>Isichazamazwi SesiNdebele</i> and in <i>A Practical Ndebele Dictionary</i> : The Case of <b>-wa</b> Verbs <i>J. Zondo</i>
12:00– 12:25:	Lexicography and the French Language in Gabon <i>B. Nyangone Assam and H.S. Ndinga-Koumba-Binza</i>	The Lexicography of Sign Language: Its Past, Present and Future in South Africa <i>H. Fourie</i>
12:30– 12:55:	(To Be Announced)	A Study of Luo Ethnobotanical Dictionaries <i>H.J. Ojwang</i>

13:00–13:10: **CLOSURE: Venue A: Room 229**

13:15–14:00: Lunch

## ABSTRACTS

### Keynote Addresses

---

**Hanks, Patrick ▪ The Structure, Typology, and Function of Similes**

*Masaryk University, Brno, Czech Republic (hanks@fi.muni.cz)*

**Kabuta, Ngo Semzara ▪ Lexicography in Central Africa, with Special Focus on Cilubà**

*Ghent University, Ghent, Belgium (ngosemzara.kabuta@ugent.be)*

**Van Sterkenburg, Piet ▪ Lexicography and Magic Spells**

*Instituut voor Nederlandse Lexicologie, Leyden, The Netherlands*

*(pvansterkenburg@planet.nl)*

---

### Parallel Sessions

---

**Alberts, Mariëtta ▪ Translation-oriented Terminography in the Electronic Age**

*Pan South African Language Board (PanSALB), Pretoria, Republic of South Africa (marietta@pansalb.org.za)*

---

Information is distributed and knowledge is acquired through terminology. The terminology of each subject field or domain is increasing with every new development or invention. The supplying of appropriate scientific, technical, educational and economic terms should be a national priority, especially in a multilingual dispensation.

Should the terminology of the minority/marginalised/developing or standardised languages of the country be developed into functional terminologies, South Africans would be equipped with effective communication tools. Terminology development also appreciates the innovative skills of the language and subject related communities within the country. In this sense, subject areas such as science, technology and economy can play a role in the development of languages as they have done with English, French, German, Chinese, Japanese, and Afrikaans to mention just a few. Languages can develop into functional languages through terminology development efforts undertaken by terminology offices, private initiatives, publishers and developers of electronic lexicographical supporting devices.

Terminology plays a pivotal role in language development and the promotion of multilingualism. The availability of (multilingual) polythematic terminology is an indicator of development since specialised communication has a central axle or hub in terminology. Standardised terminology contributes to the quality of translations, editing, interpreting services, dictionary compilation and specialised or subject related communication. Streamlined translation and interpreting services provide competitive advantages.

It is important to develop official languages into functional languages in all spheres of life. Information transfer, assimilation and retrieval should be through the first language (L1) or mother tongue. It is a given that information is best acquired (decoding process) and conveyed (encoding process)

through the L1. Standardised terminology leads to exact communication and misinterpretation or misunderstanding is avoided.

Terminology, therefore, is a strategic resource and has an important role in a country regarding the functional development of languages and their users. Effective economic, scientific and technological transfer and assimilation of knowledge and skills amongst subject specialists, laypeople and language practitioners are developed through the use of correct terminology, improving the communication abilities of the citizens of a country.

In South Africa, terminology is harvested in terms of three basic terminographical practices: subject-oriented terminography, translation-oriented terminography and linguistic community-oriented terminography.

The official national terminology office, the Terminology Coordination Section (TCS), of the National Language service (NLS), functions mainly according to traditional subject-oriented principles. The TCS works in a specific conceptual domain (subject area) and uses a specific documentation language to define the concepts. English is usually the documentation or source language (SL). Translation equivalents have to be supplied in the other ten official South African languages (target languages (TLs)).

The translators employed by the NLS and other language bureaux need to partake in translation-oriented terminography (TOT). Terms in SL texts and term equivalents in TL translations should be aligned and harvested. These harvested SL and TL term lists should form the basis for translation activities.

The third practice is community-oriented terminography where terminology is harvested from rural and urban speech communities and documented.

This paper discusses the value of each of the terminography principles with special emphasis given to translation-oriented terminography. With the use of electronic devices, TOT could form the basis for electronic special purpose dictionaries.

---

### **Beyer, Herman L. ■ A Functionalistic Approach to the Lexicographic Treatment of Idioms**

*Department of Germanic and Romance Languages, University of Namibia, Windhoek, Namibia (hbeyer@unam.na)*

---

It is well documented that idioms should be regarded as lexical items in their own right and should as such be regarded as lemma candidates when any standard linguistic dictionary is planned. Idioms that are selected as treatment units should therefore be lemmatised and represented in the macrostructure of the dictionary, be it a descriptive or translation dictionary.

The exact procedures of and mechanisms for the lemmatisation of idioms are not universally identical, however. The variety of lemma signs that represent idioms in different dictionaries are the product of various approaches to and perceptions of idioms as lexical items, determining the place where they could be accommodated in macrostructures that are inevitably *word-centred* and not *lexical item* based. Some dictionaries address this issue

by allocating certain search areas in dictionary articles to idioms that contain the word equivalents of the lemma signs of those articles, based on one linguistic criterion or other that is usually explained in the user's guide. Other dictionaries use inserted inner texts to deal with idioms, but the answer to the question as to the exact article, in which such an inner text should be placed more often than not depends on the exact article, in which a search area for a specific idiom should be accommodated, so that the different approaches yield the same result in terms of macro- and microstructural placement, if not treatment. The target user (and lexicographic layperson) might well argue that this state of affairs does not represent a matter of theory, but only one of presentation. As lexicographers, however, we would argue that theory underlies presentation. In fact, a number of scientific papers deal specifically with this question and pose viable solutions. A more appropriate question would be whether the theory is adequate. More specifically, do these solutions and their manifestations in dictionaries fulfil the requirement of lexicographic functionality? A review of recent literature reveals that the issue is tackled from a linguistic rather than a lexicographic point of view. The question, for which (good) answers are proposed, is more or less an equivalent of 'Where can we fit idioms into a certain dictionary based on their lexical characteristics?'

Recent developments in metalexigraphy have seen the focus shift away from a strictly linguistic approach to dictionaries towards an approach that accounts for dictionary structures and lexicographic functions. While these developments have taken place, it would seem that the issue of the lexicographic treatment of idioms has been left as solved within a linguistic approach. A functionalistic perspective on this issue, however, begs a somewhat differently formulated question: 'How can we treat idioms in a certain dictionary so as to contribute to the fulfilment of that dictionary's functions?' This paper attempts to present this perspective and explore the potential that a functionalistic approach to the lexicographic treatment of idioms might offer the lexicographer and ultimately the dictionary user. The role of the target user is placed centrally in (a) a brief review of current literature on the topic, (b) a review of the treatment of idioms in a number of dictionaries that claim to be polyfunctional dictionaries, and (c) the presentation and evaluation of suggested treatment strategies for idioms from a functionalistic perspective.

---

**Cichocka, Aleksandra and Karien Brits • Dictionaries as One of the Intercultural Communication Tools: South African Cultural Terms on Polish Soil**

*School of English, Adam Mickiewicz University, Poznań, Poland (olaluk@ifja.amu.edu.pl); and (karien@ifja.amu.edu.pl)*

---

As scholars working in cultural studies, we want to analyse dictionaries as documents of culture and as pedagogical tools. We doubt whether the existing material in the Polish medium covering South African cultural concepts

is sufficient and adequate. We have discovered that certain definitions do not provide objective and/or accurate information. Linguistic and cultural contact between Poland and South Africa has been limited, which might explain the misrepresentation of South African cultural concepts in Polish dictionaries. We aim to propose a possible solution to fill the existing gap, drawing from our experiences as teachers of South African cultural and linguistic comparative studies.

Firstly, we want to discuss the presence of lexical items, which constitute South African loanwords in Polish. For this purpose the major Polish monolingual dictionaries will be examined. The next step will be to compare the definitions of these loanwords with the entries in the major dictionaries of English and South African English. The reason for using British dictionaries and English as a reference point is twofold. Firstly, it is important from a linguistic perspective, as English is one of the most eminent languages of communication in South Africa and also a medium of instruction in our courses. Secondly, it is valuable from a cultural point of view, as the cultural contact between South Africa and Britain has been far more extensive than that between South Africa and Poland. This initial stage would serve to establish that the image a dictionary user might form of South African culture on the basis of existing dictionaries is incomplete.

We hope that compiling a comprehensive dictionary of cultural terms will provide a more objective and informed image of South Africa. This dictionary could then be used not only for pedagogical purposes but also serve as a valuable source of information for those interested in other cultures. The scope of interest will be limited to mainly Afrikaans terms. It does not mean that references to other South African languages will be omitted; however, they will not form the focus of the current discussion. Several examples illustrating different sections of the prospective dictionary will be considered. The four main sections are: (a) words of common origin, (b) false friends, (c) common cultural concepts, and (d) cultural concepts specific to South Africa.

The first part would acquaint speakers of Polish with Afrikaans words of common origin (e.g. Latin, French, Greek and English); we expect that learning words similar in form might reduce the fear of 'the other' by making the language more familiar. Although traditional dictionaries do not treat false friends separately, awareness thereof is crucial in acquiring competence in intercultural communication — thus the second section would focus dictionary users' attention on this matter, while at the same time offering a comprehensive list of problematic terms. During our classroom interaction, we have discovered a number of examples of cultural terms present in Polish and Afrikaans but without counterparts in English. The third section would accordingly be devoted to these concepts. The last section would form the main part of the dictionary and consist of specifically South African terms with neither Polish nor English counterparts. Here we want to present the keywords crucial in interpreting and understanding South African culture and reality.

It is an ambitious project and we hope that the paper will stimulate discussion, resulting in feedback vital to our further research.

---

**Cloete, Alet E., Basil P.D. Gabriels, Frikkie J. Lombard, Gerhard J. van Wyk • Comprehensive Dictionaries in the Electronic Era: Challenges and Solutions**

*Bureau of the WAT, Stellenbosch, Republic of South Africa (wat@sun.ac.za)*

---

Never has there been a better suited time than the present to compile a comprehensive dictionary. Not only did the past decades experience an explosion of knowledge and new terminologies, which must all be incorporated into dictionaries, but the greater accessibility and refinement of electronics and computers make it possible to employ ways to expedite the unlocking of information which was hitherto a tedious process. New technology, however, poses some challenges to lexicographers, and solutions to these must be found.

In an earlier age, dictionary-making involved a limited number of index cards which had to be sorted manually before they were also manually processed. Today, the electronic corpora which are utilized, are much bigger and more dynamic. They are constantly updated and expanded, and they can be processed by a computer. On the one hand, it is therefore much easier and quicker today to work through a great deal of material and to distinguish nuances of meaning, but, on the other hand, the danger of being overwhelmed and 'derailed' by the sheer mass of information is also present.

Lexicographers of today have the same amount of time at their disposal as those of the past, but the modern lexicographer must be able to process more data quicker and in a more cost-effective way. Because time equals money, it is vitally important to finish a dictionary volume within a certain time-frame, and therefore the lexicographer must use electronic aids wisely.

Computerization of the dictionary-making process may initially have a positive effect on production, but unlimited access to electronic corpora and sophisticated search engines may have a retarding effect when not used in a disciplined and purposeful way.

Availability of dictionary text on a CD-ROM has long since been a necessity, rather than an ideal. The benefits of such a CD are obvious: it is compact, it is easily handled and information can be accessed much quicker. But the makers of comprehensive dictionaries are also faced with a particular number of problems. Because comprehensive dictionaries which are still under compilation, are published in discrete volumes, information from A to Z is not simultaneously available. This means that a CD-version must be updated constantly, and it is problematic to market an unfinished product. Because the editorial system used in older volumes differs significantly from the present system, it is difficult to achieve a uniform character on a CD. Cross-references to lemmas in later volumes do not necessarily add up and lemmas which have become archaic or obsolete still figure prominently in

previous volumes. Because it is easier and more cost-effective to make changes to a CD, a discrepancy can and does develop between the printed and electronic media.

To utilize modern electronic aids, lexicographers must find mechanisms to help them cope with the challenges new aids invariably bring. It is necessary to have clear guidelines for the efficient and economic use of corpora as well as a refining of the processes used to convey information in dictionary format. The ideal of a finished comprehensive dictionary in Afrikaans can only be achieved by using electronic aids judiciously.

---

**De Pauw, Guy and Gilles-Maurice de Schryver ■ Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes**

*Language Technology Group, University of Antwerp, Belgium; and School of Computing and Informatics, University of Nairobi, Nairobi, Kenya (guy.depauw@ua.ac.be); and Department of African Languages and Cultures, Ghent University, Ghent, Belgium; Xhosa Department, University of the Western Cape, Bellville, Republic of South Africa; and TshwaneDJe HLT, Pretoria, Republic of South Africa (gillesmaurice.deschryver@ugent.be)*

---

Computational morphological analysis is an important first step in the automatic treatment of natural language. The internal constitution of a word can hold a lot of syntactic and semantic information that is useful for other natural language processing tasks as well, such as part-of-speech (POS) tagging and syntactic analysis. While the success of the established two-level morphology paradigm has resulted in a great number of word analysers for Indo-European languages, not many computational tools for the morphological analysis of Bantu languages exist. This is partly due to the morphological richness of these languages, which makes the construction of an adequate rule set cumbersome, but also because other useful resources, such as digital dictionaries and morphologically annotated lexicons are often lacking.

Recent machine learning approaches to computational morphological analysis have provided a viable alternative to the rule-based paradigm, typically factoring out the expensive human element in the design of the system. Supervised machine learning approaches do this by employing only minimally annotated data to bootstrap morphological analysis (Van den Bosch and Daelemans 1999). Unsupervised approaches go one step further by only requiring raw word lists and performing advanced pattern-matching operations to unearth morphological patterns in the language (Linden 2008).

In De Schryver and De Pauw (2007), it was shown how, with a minimum of human effort, machine learning techniques can successfully be employed to obtain POS-tagged corpora that can be used for lexicographic purposes. The experiments were run for Northern Sotho, and it was indicated how the results were seamlessly integrated with the dictionary compilation and terminology management software *TshwaneLex* (Joffe et al. 2008). A

demo version of the Northern Sotho POS-tagger itself may be found at <http://aflat.org/?q=node/177>. In this talk, the computational morphological analysis for Bantu languages is taken to a next level, and exemplified for Swahili.

We present experiments with a supervised memory-based learning method for morphological analysis. The training material for this task is extracted from the annotated *Helsinki Corpus of Swahili* (HCS) (Hurskainen 2004), which for many words provides the associated lemma. In the preprocessing phase, the word form is matched to this lemma, which results into a prefix (P) and suffix (S) part and a (possibly bound) root morpheme (R). For instance, the word form *aadilishwe* is lemmatized in the HCS as *adilishwa*. Preprocessing combines this information into a surface representation  $a@P+adilishw@R+e@S$  and an underlying representation  $a@P+adilishwa@R+e@S$ .

In the first classification phase, we perform morphological segmentation and labelling of the respective morphemes using a single classifier. The features are defined on the level of the grapheme. We use a sliding window approach to extract instances, with a left and right context of 5 graphemes and a focus grapheme. This instance is associated with a morphological class, indicating the presence and nature of a morpheme boundary — see Table 1. This data is then used as training material for a memory-based classifier that can consequently process new data.

L	L	L	L	L	F	R	R	R	R	R	class
#	#	#	#	#	a	d	i	l	i	s	P
#	#	#	#	a	d	i	l	i	s	h	0
#	#	#	a	d	i	l	i	s	h	w	0
#	#	a	d	i	l	i	s	s	h	w	0
#	a	d	i	l	i	s	h	w	e	#	0
a	d	i	l	i	s	h	w	e	#	#	0
d	i	l	i	s	h	e	#	#	#	#	0
i	l	i	s	h	w	#	#	#	#	#	R
l	i	s	h	w	e	#	#	#	#	#	S

**Table 1:** Feature extraction for the morphological segmentation task

In the next processing phase, we 'repair' the root morpheme, which after the first phase may still be bound. For this task, we built a second classifier that likewise works on the level of the grapheme, but uses rewrites as classes — see Table 2.

L	L	L	L	L	F	R	R	R	R	R	class
#	#	#	#	#	a	d	i	l	i	s	a
#	#	#	#	a	d	i	l	i	s	h	d
#	#	#	a	d	i	l	i	s	h	w	i
#	#	a	d	i	l	i	s	s	h	w	l
#	a	d	i	l	i	s	h	w	e	#	i



a	d	i	l	i	s	h	w	e	#	#	s
d	i	l	i	s	h	e	#	#	#	#	h
i	l	i	s	h	w	#	#	#	#	#	wa
l	i	s	h	w	e	#	#	#	#	#	e

**Table 2:** Feature extraction for the morpheme restoration task

Despite the Indo-European bias of the methodology and the intricate morphological richness of Swahili, our 2-step memory-based learning approach is able to correctly segment words in a held-out and manually verified test set with an accuracy of 87%. This is a significant improvement over previously reported results using a semi-supervised finite-state technique (Linden 2008). It is also able to improve on the actual stemmer used to annotate the *Helsinki Corpus of Swahili*, apparently smoothing over inconsistencies and remaining ambiguity in the annotation.

Since this morphological analyser also *de facto* constitutes a stemmer, finding base forms for (often highly inflected) word forms is an additional useful tool for corpus-based dictionary compilation, as it is able to quickly and accurately peruse large corpora, in search of different instantiations of the same lemma, as well as discover previously unrecorded ones.

### References

- De Schryver, G.-M. and G. De Pauw.** 2007. Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*. *Lexikos* 17: 226–246.
- Hurskainen, A.** 2004. *HCS 2004 – Helsinki Corpus of Swahili*. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC.
- Joffe, D. et al.** 2008. *TshwaneLex Suite – Dictionary Compilation and Terminology Management Software*. Available from: [http://tshwanelex/](http://tshwanelex.com/tshwanelex/).
- Linden, K.** 2008. A Probabilistic Model for Guessing Base Forms of New Words by Analogy. *Computational Linguistics and Intelligent Text Processing, Ninth International Conference, CICLing 2008, Haifa, Israel, February 17-23, 2008, Proceedings*: 106-116. Berlin: Springer.
- Van den Bosch, A. and W. Daelemans.** 1999. Memory-based Morphological Analysis. *Proceedings of the Thirty-seventh Annual Meeting of the Association for Computational Linguistics*: 285-292. University of Maryland.

---

**De Schryver, Gilles-Maurice ■ An Analysis of the *Oxford Bilingual School Dictionary: Northern Sotho and English***

*Department of African Languages and Cultures, Ghent University, Ghent, Belgium; Xhosa Department, University of the Western Cape, Bellville, Republic of South Africa; and TshwaneDJe HLT, Pretoria, Republic of South Africa*  
*(gillesmaurice.deschryver@ugent.be)*

---

John Sinclair's impact on the lexical analysis of the English language as well as his pioneering work in corpus-driven linguistics and lexicography is well-known and well-documented. What is less widely known is his impact on lexical studies and dictionary making for languages other than English. In this talk it is shown how Sinclair's revolutionary insights have been and are being adopted, reanalysed and taken further during the production of various bilingual dictionaries for Bantu languages. This work has proceeded from a modest attempt ten years ago to compile a Cilubà–Dutch learner's dictionary (De Schryver and Kabuta 1998), to an online Swahili–English reference work (Hillewaert, Joffe and De Schryver 2004) and finally, and also the core of this talk, a Northern Sotho–English school dictionary published by Oxford University Press Southern Africa (De Schryver 2007).

Although the latter product is Sinclairian in many aspects, it also displays features that transcend the monolingual level, as two corpus-based analyses had to be mapped onto one another. Interesting new questions arose as a result, which revolve around the need to reach idiomatic (or at least idiomatic-sounding) bilingual example sentences. A frequency-based approach to lexical and grammatical gaps was adopted, and there was a need to produce a seamlessly integrated 'corpus-based dictionary mini-grammar'.

The presentation of the Northern Sotho dictionary will proceed from a discussion of the dictionary planning phase, to an analysis of macrostructural and then microstructural aspects. On the macrostructural level, special attention will go to the so-called star-rating which was adopted to mark the top-frequent lemmas, while 'real-text examples' (Fox 1987: 144) will be one of the core foci on the microstructural level.

Figures 1 and 2 show two sample articles, one for each side of the dictionary.

<p><b>itše</b> *** /itšê/ verb + perfect (le) ☞ RE' ■ said</p> <ul style="list-style-type: none"><li>◆ Polelong ya gagwe o itše, ga go yo a swanetšego go hlorišwa. <i>In her speech, she said no one must be abused.</i></li></ul> <p>◊ e itše = EITŠE</p> <p>► <b>itšego</b> /itšêgo/, <b>itšeng</b> /itšêng/ ■ who/which said</p> <p>► <b>[DEM + SC +]</b> itšego /itšêgo/ verbal relative construction ■ (a) certain; (a) specific</p> <ul style="list-style-type: none"><li>◆ Diphefo tše di tliša pula ditikologong tše di itšego. <i>These winds bring rain to certain regions.</i></li></ul>
--

**Figure 1:** 'Grammatical constructions' in the Northern Sotho → English side (Note that DEM and SC stand for a demonstrative and subject concord respectively)

**edge** \*\* noun (pl. edges) 1 ■ nthla ♦ Don't walk too close to the **edge** – you will fall! O se ke wa sepela kgauswi kudu le ntheng – o tla wa! 2 (Maths) ■ morumo ● the place where two surfaces of an object meet or the outline of a shape lefelo leo go lona mafelelo a mabedi a selo a kopanago goba magomo a sebopego ♦ Does the shape have a straight **edge** or a curved **edge**? A sebopego se na le morumo wo o lego thwii goba morumo wo o kgopamego?  
 ◇ be on **edge** ■ tshwenyega ♦ He has been on **edge** the whole day. O hwele a tshwenyegile letšatši ka moka.

**Figure 2:** A 'curriculum sense' in the English → Northern Sotho side (Note the accompanying label and definitions)

Following this presentation, an in-depth analysis will be performed against some Sinclairian research findings. More in particular, attention will go to:

- (a) normal realizations and patterns of co-selection,
- (b) the 'open-choice / idiom' principle,
- (c) textual well-formedness, and
- (d) semantic prosody.

This analysis will be supported by ample corpus data, and will include numerous detailed summaries such as those shown in Tables 1 and 2.

**Table 1:** Collocates of the Northern Sotho verb *akga* (with N = 92; *f* = frequency; OC = object concord)

<i>seatla</i> 'hand' → <i>f</i> = 0	<i>letsogo</i> 'arm' → <i>f</i> = 6 + 1 (OC) + 1 (– prefix) = 8
<i>diatla</i> 'hands' → <i>f</i> = 33 + 2 (+ le) = 35	<i>matsogo</i> 'arms' → <i>f</i> = 9 + 1 (+ { le}) = 10 <i>mabogo</i> 'arms' → <i>f</i> = 1
<i>lenao</i> 'foot' → <i>f</i> = 0 { <i>dinao</i> 'feet' → <i>f</i> = 28 + 1 (– prefix) = 29 <i>manao</i> 'feet' → <i>f</i> = 0	<i>leoto</i> 'leg; foot' → <i>f</i> = 0 <i>maoto</i> 'legs; feet' → <i>f</i> = 6
<i>dinkatana</i> 'underwear' → <i>f</i> = 1; <i>molala</i> 'neck' → <i>f</i> = 1; <i>themo</i> 'chisel' → <i>f</i> = 1	

**Table 2:** Collocates of the Northern Sotho verb *fahla* (raw frequencies; top 20)

<i>N</i>	<i>Word</i>	<i>Total</i>	<i>Left</i>	<i>Right</i>	<i>L3</i>	<i>L2</i>	<i>L1</i>	*	<i>R1</i>	<i>R2</i>	<i>R3</i>
1	<i>fahla</i>	101	0	0	0	0	0	101	0	0	0
2	<i>go</i>	53	43	10	2	9	32	0	0	2	8
3	<i>a</i>	40	25	15	8	5	12	0	2	5	8
4	<i>ka</i>	34	10	24	3	2	5	0	6	16	2
5	<i>le</i>	27	14	13	6	8	0	0	5	6	2
6	<i>magotlo</i>	27	0	27	0	0	0	0	24	3	0
7	<i>e</i>	19	16	3	5	2	9	0	0	0	3
8	<i>o</i>	17	13	4	3	4	6	0	0	1	3

9	<i>ya</i>	16	9	7	5	3	1	0	0	4	3
10	<i>re</i>	14	12	2	3	1	8	0	0	0	2
11	<i>ge</i>	11	3	8	2	1	0	0	0	3	5
12	<i>sa</i>	11	7	4	3	2	2	0	0	1	3
13	<i>ba</i>	10	7	3	2	3	2	0	1	1	1
14	<i>ke</i>	9	7	2	4	2	1	0	0	2	0
15	<i>wa</i>	9	3	6	0	2	1	0	2	2	2
16	<i>mo</i>	8	7	1	0	0	7	0	0	1	0
17	<i>mahlo</i>	7	0	7	0	0	0	0	6	1	0
18	<i>tša</i>	7	3	4	1	2	0	0	0	1	3
19	<i>se</i>	6	3	3	2	0	1	0	0	2	1
20	<i>yo</i>	6	6	0	0	0	6	0	0	0	0

In conclusion it will be pointed out that corpora, even small ones, allow to fast-track the production of reference works, act as arbiters on what to include in and what to omit from both the macro- and microstructures, enable mother-tongue speakers to map meanings onto use, while all along the raw material may be lifted out of the corpora and dropped straight into the dictionary articles that are being compiled. Idiomaticity and various corpus statistics drive this entire process. Following a decade of building, analysing and extracting data from corpora for lexicographic purposes of the Bantu languages, it is hard to imagine how one can even attempt compiling dictionaries without them. The production of this Northern Sotho dictionary once more proved that.

### References

- De Schryver, G.-M. and N. S. Kabuta.** 1998. *Beknopt woordenboek Cilubà–Nederlands & Kalombodi-mfundilu kàà Cilubà (Spellingsgids Cilubà). Een op gebruiks-frequentie gebaseerd vertalend aanleerders-lexicon met decodeerfunctie bestaande uit circa 3.000 strikt alfabetisch geordende lemma's & Mfundilu wa myakù idì itàmba kumwèneca (De orthografie van de meest gangbare woorden)*. Second Edition; First Edition 1997. Recall Linguistics Series 12. Ghent: Recall.
- De Schryver, G.-M.** 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English/Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.
- Fox, G.** 1987. The Case for Examples. Sinclair, J.M. (Ed.). *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London: Collins ELT.
- Hillewaert, S., P. Joffe and G.-M. de Schryver.** 2004. *Online Swahili–English Dictionary/Kamusi ya Kiswahili–Kiingereza Katika Mtandao*. Available from: <http://africanlanguages.com/swahili/>.

---

**De Tier, Veronique • The *Zeeuwse Klapbank*, a Multimedia Exhibition about Zeeland Dialects**

*Department of Dutch Linguistics, Ghent University, Ghent, Belgium  
(veronique.detier@ugent.be)*

---

**Introduction**

Dialects are disappearing quickly: the numbers of dialect speakers diminish, dialect vocabulary has become less extensive and specific phonetic characteristics are fading. That is why the Stichting Cultureel Erfgoed Zeeland (SCEZ), Foundation Cultural Heritage Zeeland, Middelburg, the Netherlands, has decided to stage a multimedia exhibition about Zeeland dialects. *De Zeeuwse Klapbank* is a travelling exhibition about Zeeland dialects with dialect audio extracts, stories in dialect and a few songs. The name of the exhibition refers to a central place, mostly a bench ('een bank') in a village or in a harbour where people gathered to talk about daily life. Within the word *klapbank* is the old verb *klappen*, meaning 'to talk'.

**The origins of the exhibition**

The idea of an exhibition with dialect audio extracts is not new. There is a permanent exhibition in the 'Language Chamber' in Het huis van Alijn, Ghent, about dialects in Dutch-speaking Belgium. The main difference is that the Flemish exhibition resides permanently in one museum whereas the *Zeeuwse Klapbank* is a travelling exhibition that is set up in libraries and museums. These two projects have sparked off other initiatives in other Dutch regions like Twente.

**Construction of the exhibition**

The exhibition consists of three main parts: the listening booths, the cinema seats, and a computer set-up.

**1. The listening booths**

This subdivision is the central part of the exhibition. In a tubular structure, five cells or listening booths have been built. There are five DVD-players in these booths on which can be heard dialect recordings from diverse Zeeland regions, equally divided, as far as possible, among the islands. A Dutch 'translation' text can be seen on a screen so that non-dialect speakers or people from other regions can also follow easily.

The recordings used in the exhibition were made between 1961 and 1980 by the Meertens Instituut in Amsterdam and by the Department of Dutch Philology at the University of Ghent. From conversations lasting more than fifty hours a selection was made of thirty four- to five-minute recordings about three different topics. In this way the exhibition attempts to give a realistic picture through dialect of the sixties and seventies of the last century. There are ten conversations about the 1953 flooding, ten about agricultural life and ten about everyday life. On the homepage can be found data about the recording: the name of the speaker, date of the recording, etc.

In the listening booths further information is given on seven vinyl screens about the exhibition and about the Zeeland dialects (vocabulary, etymology, pronunciation, etc.).

## **2. The cinema seats**

The second part of the exhibition consists of three groups of four cinema seats where visitors can listen to stories and to some songs from north to south Zeeland. The Dutch versions can be found in a book in which there is also an introduction on the specific dialect of that region. For each story information is given about dialect peculiarities, comments about the vocabulary used, dialect maps, etymology and grammatical information.

## **3. The computer**

The third part is a computer centre. Here visitors will have access to further information such as the digital *Zeeland Dictionary*, speaking dialect maps, references to the *Zeeland Etymological Dictionary*, a link to the digital *Woordenboek der Nederlandse Taal* (WNT), dialect maps from the language atlas, excerpts from the Flemish exhibition, etc.

## **4. Newspaper and DVD**

Finally, each visitor is given a 'dialect newspaper' with some more extensive information about the Zeeland language. This newspaper, accompanied by a DVD with ten recordings, could, for example, also be used in schools.

## **5. The future**

A possible spin-off could be an educational project about language variation. Starting from the dialect recordings of this exhibition, more information could be given to students about lexicological aspects as well as about syntax and phonology. With each of the recordings as basis, specific phonological characteristics or syntactical peculiarities of a given dialect could be highlighted. By comparing the recordings, differences and similarities will become clear.

As far as lexicology is concerned, interesting dialect words used in the recordings (many of which have already disappeared) can be linked to the *Zeeland Dialect Dictionary* and with the I-WNT, which in turn will be linked to the three big regional dialect dictionaries for the whole of the southern language region (*Zeeuws-Vlaanderen* is part of the *Dictionary of Flemish Dialects*, *Woordenboek van de Vlaamse Dialecten* (WVD)) and in this way also linked to the Dutch and Afrikaans etymological dictionary. Through links with the regional dictionaries of the southern Netherlands, it also becomes possible to call up general maps of the entire southern region, to which also audio excerpts can be linked.

## **Conclusion**

Exhibitions about dialects are relatively new because the digitalisation of old dialect recordings has only now made it possible to have easier access to this material. Zeeland is proud of its dialects, as is shown in the attention paid to them on provincial level. In the past, it was the first province in the Netherlands that had a scientific dialect dictionary and an etymological dialect dictionary; it was also the first province to have its dictionary digitalised and it is now the first province to have an exhibition about its dialects. After years of scientific dialect studies (lexicological, syntactic and phonological) and thanks to digitalisation, it is now possible to bring interested laypersons, who want to know more about their own dialect into contact with a small part of

their intangible cultural heritage. Because of digitalisation this can be done in a simple and pleasant, yet scientifically correct way.

---

**Fourie, Hanelle ▪ The Lexicography of Sign Language: Its Past, Present and Future in South Africa**

*Bureau of the WAT, Stellenbosch, Republic of South Africa (hfourie@sun.ac.za)*

---

The *Dictionary of American Sign Language on Linguistic Principles* by William Stokoe and some of his colleagues at Gallaudet University rang in the era of linguistic research into Sign Language in America in 1965. With this revolutionary dictionary it was possible to look up the meaning of a sign with the aid of a notation system (similar to the phonetic alphabet) that was used to group the signs. To date there is no similar lexicon of South African Sign Language (SASL) that has been researched and compiled according to the principles of lexicography.

The research that has already been done in the domain of the lexicography of SASL is not of a truly lexicographical nature, even though a few dictionaries do exist, i.e. *Talk to the Deaf* (Nieder-Heitmann 1980) and *Dictionary of South African Signs* (Penn 1992). Most of the researchers so far have been audiologists, speech therapists or educationists, and have approached their research from a *pathological* point of view. Only a few, like Debra Aarons, Philemon Akach, Ernest Kleinschmidt and Helen Morgans (the latter two both deaf themselves) have made a *linguistic* contribution to the research.

On the level of both theoretical and practical lexicography, little has been done for SASL. In this regard it is important that the development that has taken place in theoretical lexicography over the past few years, amongst others the establishment of a comprehensive general theory of lexicography, results in the design of a lexicon of SASL. The general theory of lexicography, as developed by H.E. Wiegand and others, is an important basis for the theoretical model of a sign language lexicon. Especially the different structural components of dictionaries will be highly relevant here. Given the distinctive needs of the target users of such a dictionary, both the nature of the external and internal access structures and the choice of an appropriate frame structure must be applied to the model of a sign language lexicon.

A lexicon of signs is ideally represented in a multimedia format, which makes it much easier for the user to look up words with the press of a few buttons rather than having to master a complex notational system.

The ideal sign language lexicon will be a bidirectional system, which not only contains the sign for a given concept, but which also makes it possible to look up the definition of a sign. Because the existing dictionaries of South African signs do not make provision for such a function, these dictionaries receive sharp criticism from the deaf community.

The speaker is currently a doctoral candidate investigating the aspects mentioned. The study hopes to contribute to both the study of sign language

and the metalexigraphy, and to eventually create a basis for new developments in lexicography. The theoretical model proposed for a sign language lexicon could lead to the compilation of a user-oriented dictionary that recognises not only the needs of a clearly identified target user but also the reference skills of this target user.

An important impact of this study will therefore be the creation of a theoretical foundation for further study of sign language in South Africa, and the dictionary model will make an essential contribution to extending metalexicographical literature and the scientific determination of typological models. The dictionary which could flow from this will be meaningful to both the deaf community and any hearing persons who have close contact with this community (e.g. interpreters, hearing educators at deaf schools or in inclusive education schools, medical practitioners, etc.).

### **References**

- Nieder-Heitmann, N.** 1980. *Talking to the Deaf: A Visual Manual of Standardized Signs for the Deaf in South Africa*. Department of Education and Training, Republic of South Africa, and South African National Council for the Deaf.
- Penn, C.** 1992. *Dictionary of Southern African Signs for Communicating with the Deaf*. Pretoria: Human Sciences Research Council.

---

### **Gouws, Rufus H. and Danie J. Prinsloo ▪ Surrogate Equivalence in Bilingual Dictionaries with Special Reference to Dictionaries Bridging English and African Languages**

*Department of Afrikaans and Dutch, University of Stellenbosch, Republic of South Africa (rhg@sun.ac.za); and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (danie.prinsloo@up.ac.za)*

---

### **Introduction**

Three major types of equivalent relations prevail, i.e. full equivalence, partial equivalence and zero equivalence. These different relations of equivalence confront lexicographers with different challenges to ensure that users will be able to achieve an optimal retrieval of information from a given dictionary article. Ideally, suitable translation equivalents in the target language would be available for each source language item. Instances where a suitable translation equivalent is not available occur in any given language pair to be treated as source and target language in a bilingual dictionary. This is referred to as zero equivalence and can be regarded as the most complex type of equivalence to be dealt with in a bilingual dictionary.

This paper will focus on various ways in which lexicographers of different dictionaries need to deal with the lack of equivalence and the subsequent use of surrogate equivalents. A number of strategies are available to the lexicographer when treating instances of zero equivalence, e.g. the use of glosses, paraphrases of meaning, pictorial illustrations and even text boxes with lexicographic comments. Depending on the nature, importance to the target user and the complexity of such instances, various cross-referencing



options should also be utilized to refer users to dictionary internal reference addresses, e.g. other articles and/or the back matter, and/or dictionary external reference addresses where they could find more information.

A linguistic gap is identified where the speakers of both languages are familiar with a certain concept but where the one language does not have a word to refer to it, whereas the other language does have such a word. A referential gap can be postulated where a lexical item from language A has no translation equivalent in language B because the speakers of language B do not know the referent of the lexical item from language A. Different users will need different types of surrogate equivalents to ensure their optimal retrieval of information from a given dictionary. In an Afrikaans–English bilingual dictionary compiled for use in South Africa, the word *padkos* should be entered as an Afrikaans lemma. No exact translation equivalent exists and the lexicographer will have to use a surrogate. For the average member of the Afrikaans and English speech communities in South Africa, a brief description like 'food for the road' will suffice. In a dictionary directed at primary school learners, a slightly more comprehensive explanation might be needed. In the choice of surrogate equivalents, cognizance should also be taken of the cultural differences prevailing between the members of the speech communities of the language pair presented in the dictionary. A failure to do this often results in surrogate equivalents that do not achieve either semantic or communicative equivalence between source and target language. In a dictionary with English and one of the Nguni languages as treated language pair, the Nguni word *lobola* will have to be included as a lemma. English has a lexical gap in this regard and the lexicographer will have to use a surrogate equivalent in the treatment of the lemma *lobola*. English does have a partial equivalent that can be offered in the treatment of this lemma, i.e. the equivalent *bride's price*. However, this English equivalent does not convey the cultural value of the word *lobola* and the lexicographer should preferably give some additional entries, representing a more comprehensive treatment, as surrogate equivalents to complement the partial equivalent and to ensure a proper understanding of the meaning of the word *lobola*. If this dictionary has been compiled for use in South Africa, where the English-speaking user of a dictionary may have some exposure to the concept of *lobola* the surrogate equivalent may be less comprehensive than in a dictionary compiled for use in, say, Great Britain, where the typical user will not have been exposed to this concept. Yet again, it is necessary that a generic approach to the treatment of zero equivalence is supported by dictionary-specific approaches, where the user needs of the specific dictionary determine the nature and extent of the surrogate equivalents.

#### **Surrogate equivalents displaying a cognitive function**

The treatment of, for example, culture-bound lexical items confronts lexicographers with unique challenges to achieve communicative equivalence; cf. the following example from a bilingual dictionary with English and isiXhosa as treated language pair with regard to some aspects of the use of surrogate equivalents.

**indyifolo** bead ornament worn by men round the neck; it is broad and flat with a number of pendant strings which are crossed and thrown back over the shoulders where they are again crossed.

The comprehensive target language entry is primarily due to the culturally bound nature of the isiXhosa lexical item. Both linguistic and encyclopaedic data are needed in this surrogate equivalent. The inclusion of encyclopaedic data is in accordance with the approach that the distinction between linguistic and encyclopaedic data may be of interest to the linguist but is irrelevant to the lexicographer. The lexicographer is only interested in conveying enough data to ensure an unambiguous retrieval of the information needed by the intended target user.

#### **A hierarchy of surrogate equivalents**

Acknowledging different degrees of complexity in the relation of surrogate equivalence leads to a tiered view of surrogate equivalents. The first level in the hierarchy provides for linguistic gaps where a mere gloss or brief paraphrase of meaning will suffice. More complicated are gaps where the surrogate equivalent also has to provide grammatical guidance. Referential gaps demand a more comprehensive description characterized by encyclopaedic entries. The top tier in the hierarchy provides for referential gaps where taboo, culture-specific or sensitive values have to be expressed. This hierarchy will be illustrated with detailed examples.

---

### **Hadebe, Samukele ▪ The Uncertain Future of Dictionary-making in Zimbabwe**

*Department of African Languages and Literature, University of Zimbabwe, Harare, Zimbabwe (shadebe@arts.uz.ac.zw)*

---

The last decade and a half saw an upsurge in lexicographic work in Zimbabwe resulting in a number of monolingual dictionaries in Shona and Ndebele. The case of Zimbabwean lexicographic works is well documented and at surface level it is apparently a success story. While the present situation looks good, the future is not promising and the high hopes that were brought by the current achievements may soon lead to disappointment. In this paper, it is argued that monolingual lexicography has reached its peak in Zimbabwe and that there are concrete signs of its future decline.

The following reasons are put forward to explain the possible decline and bleak future: lack of recognition of dictionary-making as academic work, reliance on donor funding, the project orientation of dictionary-making and the language policies that favour English only as medium of instruction and language of codification. A cursory look into the history of lexicography will show that recognition and crediting of the work of lexicographers were not achieved without a protracted struggle. While most universities now have lexicography units/departments and dictionary work is duly acknowledged as an academic enterprise, in Zimbabwe, especially at the University of Zimbabwe, this is unfortunately not the case. Although only

the University of Zimbabwe has a lexicography unit, the African Languages Research Institute (ALRI), it does not acknowledge dictionary work as an academic enterprise. Academics therefore can only continue making dictionaries either for monetary gain or out of pure interest. None of these reasons can sustain the rigour of dictionary-making since so far no dictionary has brought profit for its editors. Similarly, pure interest in the work cannot be sustained without compensation for the hard work and time expended. In short, there is no motivation or security for professional lexicographers, and future dictionary projects would either not commence or may never be completed because of lack of practitioners.

Secondly, the dictionaries produced so far were mainly, if not wholly, funded by external donors. Generally donor-driven projects are results-oriented and time-tabled. Apart from the generous funding and technical support, donor projects usually have publicity and some prestige associated with them. These factors usually diminish once the donor period lapses. Of course, part of the initial agreements before such projects commence is the promise of sustainability and continuity long after the lapse of the donor funding. Findings elsewhere have shown that there is no control to ensure sustainability after the donor period has lapsed as most workers and consultants usually leave the project. The resources to carry on with the projects are often inadequate and there is no support to motivate the remaining staff, most of whom would be juniors and inexperienced. It is unlikely that any new dictionaries will be produced, especially for the other African languages of Zimbabwe other than Shona and Ndebele. The project orientation of dictionary-making has meant that almost five dictionaries have so far been produced for Shona, two for Ndebele and none for Tonga, Nambya, Shangani, Venda or Kalanga. The personnel for the Shona and Ndebele projects are of little use to the dictionary projects of other languages.

Finally, the paper shows how the continued policy of the domination of English in the education system and in all official spheres hinders the growth of lexicography in the African languages. In a situation where the African languages are only taught as optional subjects and have no other domain of use than the home, there is therefore a limited use for monolingual African dictionaries. Actually, monolingual African dictionaries are bound to decline and bilingual dictionaries to increase in demand. The high cost of dictionary production cannot be justifiable under the current language conditions where mother-tongue education has been relegated to infant schooling and is not even practised at that lowest level. Unless the African languages play more functional roles in society, the future of African language lexicography is bleak. While the focus is on Zimbabwe, the future of African language dictionaries and their compilers might be generalized for other countries that more or less resemble the Zimbabwean situation.

---

## **Heid, Uli and Danie J. Prinsloo • A Bilingual Dictionary for a Specific User Group: Supporting Setswana Speakers in the Production and Reception of English**

*Institut für maschinelle Sprachverarbeitung – Computerlinguistik, University of Stuttgart, Stuttgart, Germany (uli@ims.uni-stuttgart.de); and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (danie.prinsloo@up.ac.za)*

---

### **1. Introduction**

The aim of this presentation is to focus on the design of a bilingual dictionary for languages A and B where the needs of a narrowly defined group of target users are limited to (a) text and speech *production* in language B, in the section of the dictionary where B is the target language and (b) text *reception* from language B in the section of the dictionary where B is the source language. The design of an English–Setswana dictionary is taken as a case in point. It will be shown how available treatment options could be maximally utilized in order to compile a dictionary with a limited macrostructure but more extended microstructure in the A → B component versus a large macrostructure with limited microstructure in the component B → A. We expect this layout to be clearly adapted to the specific needs of the target user group, mother-tongue speakers of Setswana, to be able to identify the data they are looking for in order to retrieve the necessary information and to utilize it in a productive (encoding) and receptive (decoding) way.

### **2. Defining the target user**

The target users are narrowly defined as Setswana-speaking school children in two categories, Upper Primary (10–11 years old) and Junior Secondary (13–15 years old), seeking empowerment in English — thus the principle idea of a productive Setswana → English component or section of the dictionary and a receptive English → Setswana side. The design envisages an equal number of pages for these two components, but the A → B (Setswana → English) component will contain fewer lemmas with more exhaustive treatment and the B → A (English → Setswana) component a larger quantity of lemmas but with shorter articles.

### **3. A → B (Setswana → English): a rich microstructure for production / encoding**

#### **3.1 Macrostructural selection criteria**

Departing from a restriction of 400 pages for this section of the dictionary and the average length calculated from a number of test articles, we determined that a maximum of 10 000 articles could be accommodated. To ensure a proper balance between alphabetical stretches, the Setswana Ruler and Block System (Prinsloo 2004) was used.

#### **3.2 Microstructure: indications for English**

To assist users to produce English utterances or texts, a detailed microstructural programme is required. It will have a rather prescriptive character to ensure the production of correct standard English. Setswana lemmata are followed by English translation equivalents, example sentences in both languages, and indications of the following properties of the English equiva-

lents: pronunciation, morphological or morphosyntactic specificities (e.g. plural form and use), valency, preposition use, as well as collocations and idioms where appropriate. The examples, given in abundance, should reflect typical situations of usage in Setswana with a view to correctly express them in English. Translating the examples in English, fully and even in more than one way where applicable, should be a priority.

### **3.3 Microstructure: indications for Setswana**

Setswana lemmata include simple terms, but also semantically autonomous derived forms. Verbs can have several hundreds of derivations, which may or may not be morphologically regular and semantically transparent or opaque. We lemmatize all Setswana verb stems which are either frequent or morphologically irregular or opaque, and we cross-reference derived forms and their basic stems.

Contrary to existing English–Setswana, Setswana–English dictionaries, comments on form of the Setswana lemmata should be limited and preferable presented in a smaller font to create an *easy-to-jump-over* effect and/or be placed at the end of the article.

### **4. B → A (English → Setswana): a large macrostructure for reception / decoding**

The lexicographic approach in this section of the dictionary contrasts with the rich microstructure programme advocated for the Setswana → English section where being productive in English was the overriding criterion resulting in elaborate microstructures. To the contrary, in the English → Setswana section, the focus is on English reception and a larger macrostructure resulting in a slimmer microstructure. In this section of the dictionary, the microstructure is thus characterized by only reception-related entries on a minimalistic basis.

#### **4.1 Selection of lemmata and balance of alphabetical stretches**

Aspects underpinning the selection of English lemmata are frequency of use in the general language with special accommodation for vocabulary used in the prescribed books of the target users. For the compilation of the English lemma list, the design of De Schryver and Prinsloo (2003) will be followed according to which a general English corpus is compared to what they term a *dedicated* corpus consisting of prescribed literature at the level of the target users. In terms of this approach, words having a low or even zero frequency in the general corpus are lemmatized and included by virtue of a high occurrence in the dedicated corpus, e.g. curriculum words. This in itself constitutes a 'motivated imbalance' that contributes positively to the needs of the target users in finding the words they would most likely look up. As in the case of the Setswana → English side of the dictionary, a Ruler and Block System for English are employed on frequency lists culled from both corpora to ensure balance in terms of alphabetical stretches, dictionary compilation schedule, length of articles, etc.

#### **4.2 Microstructure**

Translation equivalent paradigms should be given for all, or at least the most frequent, meanings of the lemma, as well as short examples in both

English and Setswana. Special attention should be given to those polysemous values which constitute lexical gaps. In contrast to the Setswana → English section, a more descriptive approach is followed in the English → Setswana section.

### References

- De Schryver, G.-M. and D.J. Prinsloo.** 2003. Compiling a Lemma-sign List for a Specific Target User Group: The *Junior Dictionary* as a Case in Point. *Dictionaries* 24: 28-58.
- Prinsloo, D.J.** 2004. Revising Matumo's *Setswana-English-Setswana Dictionary*. *Lexikos* 14: 158-172.

---

### **Kadenge, Maxwell and Gift Mheta ▪ *Duramazwi reDudziramutauro neUvaranomwe*: A Milestone in the Development of Shona**

#### **Lexicography**

*Department of Linguistics (kadengezw@yahoo.com); and African Languages Research Institute (ALRI) (giftmheta@yahoo.com), University of Zimbabwe, Harare, Zimbabwe*

---

This paper critically looks at and analyses *Duramazwi reDudziramutauro neUvaranomwe* (DRU), a 2007 product of the University of Zimbabwe's African Languages Research Institute (ALRI). In this regard, it evaluates DRU, the first terminological dictionary of Shona literary and linguistic terms. Furthermore, the paper traces the compilation of the dictionary from the collection, selection, entry, and definition to the final editing stage. It also examines both the microstructure and the macrostructure of DRU. Furthermore it highlights how the dictionary has been received by its target users, i.e. both students and teachers in high schools and tertiary institutions offering Shona as a subject or course.

As background information, the paper gives a brief history of Shona lexicography. It goes back to the period preceding the arrival of the Pioneer Column and the creation of Southern Rhodesia by the British South Africa Company. It discusses writings on the Shona language such as Bleek's *The Languages of Mosambique* (1856). It shows that ever since the trend has been that of compiling linguistic bilingual synchronic dictionaries. This trend continued until 1996 when the African Languages Lexical (ALLEX) Project, then housed in the Department of African Languages and Literature at the University of Zimbabwe, produced the first Shona monolingual dictionary, *Duramazwi reChiShona*, which was followed by the advanced Shona monolingual dictionary, *Duramazwi Guru reChiShona*. 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 which include DRU analysed in this paper.

In the analysis, the collection of the data for DRU is assessed. The major sources of data discussed are:

- (a) written sources,
- (b) the Shona corpus,
- (c) tape-recorded Shona lectures, and
- (d) interviews with reference working groups.

Each source is evaluated according to its positive contribution to the compilation of DRU. The disadvantages of each source are also discussed.

The selecting, entering, defining and editing stages of headwords for DRU are examined in terms of how they have posed challenges to lexicographers and how such challenges were overcome. Focus is also on how these stages are carried out computationally. The paper shows how modern information technology has revolutionised Shona lexicography by lessening the burden of the lexicographer and how the end-users of the dictionaries, in particular DRU, are now reaping the benefits of technological advancement.

Both the microstructure and the macrostructure of the dictionary are discussed, highlighting how the target users of DRU are likely to benefit from the structure of the dictionary at both micro and macro levels. Statistical data from a survey consisting of the feed-back from users of DRU is also presented. Principles of lexicography as articulated by Hartmann and James, Zgusta and other metalexicographers are applied throughout this discussion of DRU.

---

**Kganyago, Linkie M. ▪ An Analysis of the Interface between Meanings and Translation: A Case Study of the Northern Sotho Lord's Prayer 'Tatawešo wa Magodimong'**

*Language Services Section, Parliament of RSA, Cape Town, Republic of South Africa  
([lkanyago@parliament.gov.za](mailto:lkanyago@parliament.gov.za))*

---

Different translators from various languages have translated the Bible into Northern Sotho but the Northern Sotho translation and revisions do not show much improvement on the original translation. The reason for this is that the modern version of 'Tatawešo wa Magodimo' consists of expressions whose meanings do not correspond to the English version. Most of the language used in *Bibele* (2000) is still the same as the language used in *Bibele* (1951). The only difference existing between these two Bibles is the orthography, for example the opening phrase 'Tata-wešo wa Maxodimong' was replaced by 'Tatawešo wa Magodimong'. The latter version of course adheres to the current orthography of Northern Sotho.

In translation, misrepresentation of meaning can cause an irreversible communication blunder. The language used in the Northern Sotho Lord's Prayer need to be revisited because some items seem to be incomprehensible while some are no longer applicable to modern standards. For example, the already mentioned opening phrase *Tatawešo* was not adequately treated to match the contextual meaning presented by its source form 'Our Father'. The possessive construction *Tatawešo* only embraces those to whom one is related, unlike 'Our Father' where everybody is included in the prayer. This misrepresentation of meaning may even cause disputes among races because it could be argued that this was done deliberately by the colonial and apartheid regimes to manipulate the black community as most Bible translations were done prior to 1994.

It is important for translators to establish the kind of readership that is expected to use the translation versions of the Bible. As the Northern Sotho Lord's Prayer is meant for everyone, educated or illiterate, it should be approached with caution because it is one of the most important texts in the Bible. If translators in this regard experience zero-equivalence, they should use surrogate equivalents as a strategy or establish the kind of lexical gap that prevails in the target language to resolve such issues. For example, in the opening phrase of the Northern Sotho Lord's Prayer, the adverb of place *Magodimong* (Heaven) should have been located as a referential gap because the concept of Heaven is unknown to the Northern Sotho speech community. If concepts like these are not reconsidered, the Northern Sotho speech community will be denied understanding these concepts the way the source audience understands them. It is therefore the duty of translators to give the receptor readers an insight into the source culture.

Therefore, this paper analyses the interface between meaning and translation by making a textual analysis of the Lord's Prayer ('Tatawešo wa Magodimong') presented in the *Good News Bible* (1995) (English as source language) and *Bibele Taba e Botse* (2000) (Northern Sotho as target language). The focus will be on methods and strategies, which had an impact on the translation of this text and the meaning it acquires. The aim of this paper is to verify whether the intentions of the original were adequately captured and if not, recommendations will be made on how to approach such texts in future.

---

**Klein, Juliane ▪ South Africa's New African Language Dictionaries and their Use for the African Language Speech Communities**

*Institut für Afrikanistik, Leipzig University, Leipzig, Germany*  
(klein.juliane@freenet.de)

---

South Africa is a multilingual country with eleven official languages, of which nine are African languages. These nine languages still lack the status and standardization level English and Afrikaans have already reached. Although all official languages have at least some bilingual dictionaries, most South Africans still live in a pre-dictionary-culture environment. This implies that dictionaries are unfamiliar to many people and that these people are unfamiliar with the possibilities dictionaries hold for them. As dictionaries are valuable means to enable communication, they are important for empowering people in all domains of life where any form of communication is needed. Such fields include, among others, the whole education system, everyday communication between members from different speech communities, doctor–patient communication, or the law sector.

General dictionaries, ranging from smallish learner's, desk-top and college to large multivolume comprehensive dictionaries, have three main functions for the speech community. Firstly, they are compiled to cover the general vocabulary of the language, i.e. they document the written (standard) language. Their second function is psychological, as such a dictionary can be



seen as a status symbol for the language. Having a general dictionary shows the speech community of languages lacking a high prestige that their language has enough words and is 'good and important' enough to be recorded in a dictionary. Such a dictionary helps the user to communicate on a broad range of everyday topics and is essential for decoding and encoding language. This is especially important in multilingual societies where people must communicate with people from other speech communities.

Technical dictionaries also have three main functions. Firstly, they help to standardize the formerly underprivileged and underdeveloped languages by recording technical terminologies. Secondly, they allow people to communicate in their first language as well as in other African languages in more domains of life, i.e. they do not need to use English but can express themselves in their own language. The last function is psychological, as technical dictionaries can also be seen as status symbols for the language proving that the African languages can indeed be used in fields as, for example, medicine or law.

Online dictionaries empower people because they are means which enable communication on a broad range of topics, i.e. they give people the necessary words to speak and write and allow them to decode and encode any text on any topic. They are ideal for inexperienced users as they can be made more user-friendly than paper dictionaries. As they do not have space restrictions, this implies that more lemmas and derivational word forms can be included, allowing users to consult the dictionary successfully without having to master different lemmatization principles which are necessary for paper dictionaries. Other advantages are that they are easy to access via cell phones and computers and are affordable because of their low production costs.

---

**Mafela, Munzhedzi James • The Component of Etymology and Bilingual Dictionaries: The Case in African Languages**

*Department of African Languages, University of South Africa (UNISA), Pretoria, Republic of South Africa (mafelmj@unisa.ac.za)*

---

The main reasons for consulting dictionaries are to find the meaning of words and check their spelling. Jackson (2002: 126) says that dictionaries have a double function: as a record of the vocabulary of a language, i.e. a lexical description, and as a reference work to meet the needs of users for information about words and their usage. A complete definition of a lexical item involves the provision of word category, pronunciation, spelling, structure, illustrative examples and etymology. A full description of a lexical item (involving all the elements mentioned) can only be achieved in large dictionaries, especially monolingual dictionaries. Small and bilingual dictionaries in particular are mainly concerned with the meaning and spelling of lexical items. Elements such as word category, pronunciation, structure and illustrative examples are rarely included in small and bilingual dictionaries.

African languages are developing languages. Lexicographers of these languages concentrate more on bilingual dictionaries than monolingual dictionaries. In most cases, the element of etymology is excluded from the definition of a lexical item in bilingual dictionaries. This is imputable to a lack of space. In many instances, information regarding the etymology of a word is given in larger general purpose dictionaries. Etymology is defined as the branch of linguistics dealing with word origins. Kirkness (1984: 109) says: 'Etymology may be defined for present purposes as the description of the origin and history of both form and content of a lexical item with regard to related items (i.e. to word families and semantic fields) and to equivalent or cognate items in related languages.'

African languages, especially South African indigenous languages, are characterized by the use of archaic words and borrowings. Much of their terminologies, especially in the field of science, mathematics, economics and health, depend on those of languages such as Afrikaans and English. Sometimes even terminologies of the developed languages such as Afrikaans and English are derived from Greek and Latin. Borrowing is not only restricted to developed languages such as Afrikaans and English, words can also be borrowed from other African languages. Archaic and borrowed words are not only reflected as lexical items in larger dictionaries, but are also included in small and bilingual dictionaries. Therefore, the element of etymology must also form a part of the definition of lexical items in small and bilingual dictionaries, especially for archaic and borrowed words. If the etymology of archaic and borrowed words is not included, a dictionary user will find it difficult to understand their meaning and spelling. Dictionaries are not only consulted for the purposes of meaning, but also for the history of the lexical entry. Landau (2001: 127) argues that although meaning is certainly most often sought by dictionary users, it does not have the cachet of etymology, which combines knowledge of other languages. According to Drysdale (1979: 47), three reasons for including etymology in dictionaries are: to satisfy the need of the scholar or the student of the history of the language, to increase one's understanding of one's language and stimulate interest in it, and to provide clues to the history of the culture and its relationships to other words. The source of words in a language are diverse, a small proportion is original, many are borrowed from a range of other languages, therefore there should be some justification for providing information at least about the immediate origin of a word (Jackson 2002). The aim of this paper is to highlight the importance of the inclusion of etymology in the definition of a lexical item, with special reference to Tshivenda dictionaries.

### References

- Drysdale, P.** 1979. Dictionary Etymologies: What? Why? and for Whom? *Papers of the Dictionary Society of North America* 1979: 39-50.
- Jackson, H.** 2002. *Lexicography: An Introduction*. London: Routledge.
- Kirkness, A.** 1984. The Etymology of Europeanisms, or: Lexicographers' Difficulties with 'Lexicographer'. Hartmann, R.R.K. (Ed.). *LEXeter '83*

*Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 109-116. Tübingen: Max Niemeyer.

**Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York: The Scribner Press.

---

**Mojapelo, Welhelmina M. ▪ Pronunciation and Spelling of Foreign Lexical Items: A Case Study of the Writing of Foreign Lemmas as Adoptives in Sesotho sa Leboa**

*Sesotho sa Leboa National Lexicography Unit, University of Limpopo, Turfloop Campus, Polokwane, Republic of South Africa (mojapelo@up.ac.za)*

---

The main purpose of this paper is to analyse the imbalances associated with the pronunciation of adoptives used as lemmas in Sesotho sa Leboa or Northern Sotho. This pertains to lexical items which are morphologically written in the same form as in the source language, even though their pronunciation warrant a change of spelling different from the original lexical items. These adoptives include lexical items such as the following taken from the *Cambridge Advanced Learner's Dictionary* (2003):

**celsius** [not gradable] of a measurement of temperature on a standard in which 0° is the temperature at which water freezes, and 100° the temperature at which it boils

**data** information, especially. facts or numbers, collected for examination and consideration and used to help decision-making, or information in an electronic form that can be stored and processed by a computer

**delta** an area of low flat land, sometimes shaped approximately like a triangle, where a river divides into several smaller rivers before flowing into the sea

**menu** a list of choices that can be made to appear on a computer screen

These lexical items are recorded in the following publications: *Multilingual Mathematics Dictionary* (2003), *Natural Science and Technology Terminology Worklist* (2004) and *Northern Sotho Terminology and Orthography No. 4* (1988). The Sesotho sa Leboa National Lexicography Unit used the same spelling for these lexical items in the monolingual dictionary *Pukuntšuthaloši ya Sesotho sa Leboa*. The spelling for these lexical items in the source language (English) and Northern Sotho are as follows:

<b>celsius</b> /sel-si-əs/	<b>celsius</b> , instead of <i>sēlēšiasē</i>
<b>data</b> /dei-tə/	<b>data</b> , instead of <i>deitha</i>
<b>delta</b> /del-tə/	<b>delta</b> , instead of <i>dēltha</i>
<b>menu</b> /men·ju/	<b>menu</b> , instead of <i>menyu</i>

In Northern Sotho the lemmas are written exactly as they appear in the source language, English. The pronunciation of these English lexical items is more or less the same as in *Pukuntšuthaloši ya Sesotho sa Leboa*. The fol-

lowing lemmas, i.e. **data**, **delta** and **menu**, have indigenous Northern Sotho equivalents such as the following:

**data**    *tshedimošo / dintlhakgoboketšwa / tsebišo / dintlha tše di filwego*  
**delta**    *manwaneng*  
**menu**    *lenaneotirišo*

---

**Mojela, Victor Maropeng. • Borrowing and Loan-words: A Convenient and Accurate Method of Lemmatizing Newly Acquired Lexical Items in Sesotho sa Leboa**

*Sesotho sa Leboa National Lexicography Unit, University of Limpopo, Turfloop Campus, Polokwane, Republic of South Africa (mojelav@ul.ac.za)*

---

The influence of foreign languages on Sesotho sa Leboa or Northern Sotho leads to borrowing, or foreign acquisition, which ultimately leads to the increase in the vocabulary of this language. The languages which influenced developments in the vocabulary of Northern Sotho include, inter alia, the following: English, Afrikaans, Xitsonga, Tshivenda and the Nguni languages. The borrowing of lexical items from these languages is effected either through direct borrowing or indirect borrowing. Direct borrowing leads to the formation of loan-words, while indirect borrowing occurs through the coinage system, where only the meaning or the sense of the lexical item is taken and this meaning is associated with the indigenous Northern Sotho lexical item, ultimately leading to extensions in the meaning of this indigenous lexical item.

The main objective of this paper is to give a critical analysis of the adoption of foreign lexical items in the lemmatization of vocabulary in Northern Sotho, especially with regard to the purist attitude of most academics and intellectuals. Most Northern Sotho academics and intellectuals are reluctant to adopt loan-words as a means of developing the vocabulary of Northern Sotho even when confronted with heavy economic and scientific pressures for foreign linguistic acquisition. When faced with these pressures for foreign language acquisition, the purists usually prefer the use of coinage rather than a direct conversion of the foreign lexical item into a Northern Sotho loan-word. The critical corpus analysis of this purist policy reveals the following discrepancies:

- (1) The development of the vocabulary of the language is disadvantaged, since there is no increase of lexical items in the language, but only the meanings are added to the same number of words.
- (2) The vocabulary of the language is not balanced since there will be excessive use of polysemous lexical items originating from the excessive use of the coinage system.
- (3) In most cases, previously adopted loan-words are mistaken for original indigenous lexical items and are, therefore, given preference to newly acquired lexical items, which have direct and accurate bearing on the

meaning of the newly discovered or acquired concepts or objects. The following are examples in this regard:

<b>Previously adopted</b>	<b>New loan-words</b>	
<i>mohlagase</i>	<i>elektirisithi</i>	electricity
<i>keiti</i>	<i>heke</i>	gate
<i>modulasetulo</i>	<i>tšhemene</i>	chairman
<i>sekero</i>	<i>tlilipara</i>	hair clipper
<i>garafo</i>	<i>sepeiti</i>	spade
<i>dipala</i>	<i>diphoulo</i>	pole
<i>malekere</i>	<i>maswiti</i>	sweets
<i>sekipha</i>	<i>T shete</i>	T-shirt
<i>pere</i>	<i>hose</i>	horse
<i>sepoko</i>	<i>koust</i>	ghost

- (4) The purist policy encourages ambiguity in the language by the excessive use of coined lexical items.

The origin of these loan-words much depends on which source language was the first to interact or to come into contact with the indigenous Basotho ba Leboa community.

---

**Mphahlele, Daniel Matlape • Lemmatization Problems of Neologisms in Northern Sotho–English Dictionaries: A Lexicographical Perspective on Media Influence on Northern Sotho Spoken Language**

*School of Languages and Communication, University of Limpopo, Turfloop Campus, Polokwane, Republic of South Africa (matlapedaniel@kayafmmail.co.za)*

---

This paper investigates the lemmatization of neologisms in Northern Sotho–English bilingual dictionaries. In this investigation, works of sociolinguists such as McLaughlin (1978), Newmark (1988), Baker (1992) and Webb and Kembo-Sure (2000) will be referred to.

According to Prinsloo et al. (2000: 232), coinage is witnessed in most African languages. This is due to a constant contact with different language and dialect users. Thus, as Prinsloo et al. suggest, borrowed words from other languages and dialects are not neologisms, because they existed or exist in those languages.

Naudé (2005: 43) defines neologisms as newly created lexical items that take the place of specific words in the lexicon. New research (www.yahoo.com) further expatiates on this definition by adding 'which do not appear in dictionaries'. Therefore, according to these explanations, neologisms are new terms which, even if they are of use or in use expanding the lexicon of a given language, are not immediately included in the dictionaries of that language.

Newly created words can expand and elaborate the host language or impair the vocabulary and terminology in the existing approved thesauruses and other word lists generated by the government-appointed language bodies such as PanSALB, which, according to language policies, are to be regularly revised. Some neologisms in a given language may become obsolete without being widely recognized.

This paper also investigates the way new lexical units are adapted, the reasons why neologisms are in use, and the importance of these recently created linguistic units (their origin and functions) for the Northern Sotho lexicon.

The conclusion is that they are 'new', because they do not originate as derivatives from existing lexical units in other languages. They are different terms taking different connotations or near similar functions (polysemous senses) in the language. They create in themselves new semantic and pragmatic senses in the lexicon. It is for this reason that Prinsloo et al. classify them as newly created words, which are spread in a very short time in the speech community by language users and through media coverage, particularly Radio Thobela.

Original lexical units in the standardized language lose their basic meanings because of these new words. Words are key or fundamental units of a language conveying the intended and meaningful messages in discourse. As long as a language is an inseparable aspect of culture, hyponyms and other word categories are culture specific and carry with them specific meanings. Neologisms are not lexicalized and, according to Baker (1992: 21), such linguistic units create a problem in translation practice.

Through the big influence of the broadcasting media, specifically Radio Thobela, the Northern Sotho spoken language is today making much use of newly coined words. This paper examines how spoken Northern Sotho forged its expansion with unapproved vocabulary, thus affecting language users, because such terms are irretrievable. This results in the devaluation of the original words in both written and spoken language. However, words should gain popularity of usage before inclusion into the Northern Sotho lexicon. Areas mostly affected are trade, technology and commerce. Languages evolve as much as cultural traits and practices. This is to accommodate functional purposes as well as the descriptive functions of entities as they emerge in time compared to their indigenous connotations. Reasons for these changes are largely ascribed to radio broadcasting and not to the revised thesauruses, grammar books, and manuals by government-appointed bodies such as PanSALB. For this reason, broadcasters are to be trained and alerted that listeners are quick to emulate the language used on the radio. This will help much in forcing broadcasters to prepare themselves thoroughly by checking word lists and dictionaries before going onto the air.

#### **References**

- Baker, M.** 1992. *In Other Words*. London: Routledge.
- McLaughlin, B.** 1978. *Second Language Acquisition in Childhood*. Hillsdale, N.J.: Erlbaum.
- Newmark, P.** 1988. *A Textbook of Translation*. London: Pergamon.
- Naudé, J.A.** 2005. *Course Book Manual for TPV644/744*. Bloemfontein: UFS Press.
- Prinsloo, D., A.R. Chuwa and E. Taljard.** 2000. Webb, V.N. and E. Kemba-Sure (Eds.). 2000: 220-244.
- Webb, V.N. and E. Kembo-Sure (Eds.).** 2000. *African Voices: An Introduction to the Languages and Linguistics of Africa*. Cape Town: Oxford University Press.

---

**Nabirye, Minah ▪ Dictionary Testing in Un-/Less Documented Languages: A Case Study of Lusoga**

*Institute of Languages, Makerere University, Kampala, Uganda (mnabirye@yahoo.com)*

---

The process of compiling a dictionary consists of three major stages, namely word collection, the compiling process and testing. Testing of dictionaries, however, has received relatively little attention compared to the other two stages. Hartmann (2001) gives guidelines for testing dictionaries, but these guidelines are yet to be fully appreciated and implemented by lexicographers. Traditionally dictionaries were tested using different methods and the same still applies. For example, at the Makerere Institute of Languages, no specific testing procedure for dictionaries has been used, and all the dictionary testing methods were conducted in an ad hoc manner. Testing of the whole dictionary has also proved to be limited in obtaining credible criticism for all the information within dictionaries, especially in un-/less documented languages.

Lexicographers continue to assume that target users are knowledgeable in many fields such as linguistics, semantics, etc. Testing dictionaries in un-/less documented languages, however, requires a lexicographer to make proposals about the nature of the language being documented. In most cases, there are no official bodies responsible for the acceptability of new developments in these languages, yet the proposals need to be vetted and passed. Therefore, finding critics for such a task raises a challenge. Identification and grouping of dictionary information is required as specifications for the critics. Therefore a review of the testing procedure is necessary.

Testing would be more beneficial if it focused on specific areas of the dictionary to be tested by specific interest groups. Interest groups do not necessarily have to be the intended users of the dictionary. Target users are those whose language is being specified and documented. Interest groups are those who are knowledgeable about the different information categories in the dictionary and can therefore criticize the information without necessarily speaking the language.

Documented languages already have Language Boards responsible for vetting and passing new changes and most grammatical aspects, e.g. the orthography, parts of speech, etc. Most words included in the dictionary already exist in corpus banks and are readily available for reference unlike in un-/less documented languages. These language instruments have to be created to cater for the critical appraisal of dictionary information in un-/less documented languages. For example: Language Boards can be used to vet and pass changes in the orthography and the alphabet. Native speakers can be used to check for missing words in the dictionary and the natural forms of words, while linguists can be asked to criticize the citation forms used. Regarding the phonetic form, native speakers can be asked to pronounce words and the phonetician can assign tone marks on entries as pronounced. Therefore there is a need for devising methods of testing individual parts or areas of the dictionary by specific audiences.

The study gives a critique of the testing methods used for the monolingual dictionaries of Luganda and Runyankore-Rukiga at the Makerere Institute of Languages in comparison with the testing of the *Eiwanika ly'Olusoga* (a monolingual Lusoga dictionary) and suggests guidelines for testing specific information in dictionaries by specific audiences.

#### Reference

**Hartmann, Reinhard R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.

---

#### **Ndinga-Koumba-Binza, H. Steve and Justus C. Roux • Vowel-length Representation in Civili Orthography: Implications for Dictionaries**

*Stellenbosch University Centre for Language and Speech Technology, Stellenbosch, Republic of South Africa (13751719@sun.ac.za); and (jcr@sun.ac.za)*

---

The paper deals with vowel-length representation in the orthography of Civili, a developing language spoken in Gabon and a few neighbouring countries. It provides an assessment of existing orthography proposals in these languages and offers materials for a new proposal that takes account of the issue of vowel length particularly in Civili. Thereafter, implications for dictionaries are presented since some of the existing proposals have been in use for a number of reference works such as dictionaries and religious textbooks.

Often, orthographic systems of various languages are built upon alphabets of these languages, and these alphabets are based on the phonetic and/or phonological systems of the languages.

The first record of an orthographic alphabet available for Civili is the one that appeared in Marichelle's dictionaries. Since then, a number of proposals have been made for a unified orthography of Gabonese languages.

First, Blanchon (1990) presented not only an outline of the Civili phonology but also a proposal for the alphabet and orthography of Civili. His work consists of a brief overview of the phonological system of Civili: he lists consonants, vowels and tones, and then illustrates his proposed alphabet and orthography with a short text written in Civili. However, Blanchon does not explain how he analyzed the significance of vowel length mentioned in his proposed Civili orthography. He uses a succession of two identical segments in each context where it is possible to hear vowel lengthening.

Unlike Blanchon, the works of Ndamba (1977), Mabika Mbokou (1999), and Ndinga-Koumba-Binza (2000) were not intended to contribute directly to the establishment of a standardized alphabet and orthography for Civili. These authors do not agree on vowel duration. For instance, from the Civili texts in Ndamba's appendices, it appears that he did not intend to indicate vowel length orthographically: the following Civili words have phonetically audible vowel length; however, orthographically Ndamba and Blanchon will write them differently.

<b>Ndamba</b>	<b>Blanchon</b>	
mama	maama	<i>mum</i>



tata	taata	<i>dad</i>
saku	saaku	<i>bag</i>

Blanchon's approach might lead to a long vowel being regarded as a sequence of two distinct syllabic nuclei, i.e. as two syllables. This could pose some theoretical problems as well as practical difficulties during language description.

The current orthographic practice seems facultative according to the user, despite the requirements of the Orthography of Gabonese Languages of 1999. In fact, the revised orthographic system for all Gabonese languages requires doubling of the vowel for any vowel duration. However, the writing system of Civili, as for many Gabonese languages, has experienced a number of problems with the implementation of this requirement.

Two problems need to be mentioned here:

- (a) Since 1999, there has not been any campaign to inform the population of the proposed Civili writing system. In fact, the Government has been slow to endorse and implement the new orthography.
- (b) The practice of different users may be based on that of different views among researchers, as demonstrated above. A number of textbooks and publications reflect the non-standardization or non-harmony of the orthography with regard to vowel duration in Civili. Because different writing systems are modelled for users by the different researchers, it is advisable that the practices of the various researchers are reconciled.

In order to bring about such a reconciliation, Ndinga-Koumba-Binza (2007) has recently proposed that vowels are doubled, even in predictable contexts, in order to make Civili learners aware of vowel duration. However, this proposal is not based on any evident phonetic fact.

Answers to the questions of the physical nature of vowel duration and of the difference between length and sequence of two adjacent identical vowels might help the process of standardizing the orthographic system of Civili.

Moreover, an assessment of existing proposals for the orthography of Civili, as well as for all Gabonese languages, revealed inconsistencies and inapplicability for popular use. This is due to the fact that most of these proposals did not comply with certain methodological and socially acceptable principles for orthographic conventions. The situation with regard to an acceptable and standardized orthography for Civili is indeed very confusing. In order to set a standard and an effective orthographic system for Civili, this paper suggests considering functions and characteristics of an efficient orthography as presented by Capo (2002: 8-9).

### References

- Blanchon, J.A.** 1990. Civili. *Revue Gabonaise des Sciences de l'Homme* 2: 141-142.
- Capo, H.B.C.** 2002. The Pan-dialectal Approach to Orthographic Conventions: The Case of the Gbe Languages of West Africa. Prah, K.K. (Ed.). *Writing African. The Harmonization of Orthographic Conventions in*

- African Languages*: 5-25. Cape Town: Centre for Advanced Studies in African Society (CASAS).
- Mabika Mbokou, L.** 1999. *Les phénomènes analogiques en civil: étude phonologique et morphologique*. Unpublished M.A. Thesis. Libreville: Omar Bongo University.
- Marichelle, C.** 1902. *Dictionnaire vili-français*. Loango: Imprimerie de la Mission.
- Marichelle, C.** 1912. *Dictionnaire français-vili*. Loango: Imprimerie de la Mission.
- Ndamba, J.** 1977. *Syntagme nominal et groupe nominal en vili (H12): langue bantoue du Congo*. Unpublished Doctoral Dissertation. Paris: Université Sorbonne Nouvelle (Paris 3).
- Ndinga-Koumba-Binza, H.S.** 2000. *Phonologie du civil de Mayumba: langue bantu du Gabon (H12a)*. Unpublished M.A. Thesis. Libreville: Omar Bongo University.
- Ndinga-Koumba-Binza, H.S.** 2007. Propositions pour l'orthographe du civil. Hubert, J. and P.A. Mavoungou (Eds.). *Écriture et standardisation des langues gabonaises*. Stellenbosch: SUN PRESS.

---

**Ndlovu, Eventhough ▪ Sense Relations in the Treatment of Meaning in the Ndebele Monolingual Dictionary *Isichazamazwi SesiNdebele***

*Department of African Languages and Literature, University of Zimbabwe, Harare, Zimbabwe (endlovu@arts.uz.ac.zw)*

---

This paper evaluates how meaning has been treated in *Isichazamazwi SesiNdebele* (ISN) (2001) through the application of the sense relations of hyponymy, meronymy, and synonymy. It attempts to establish the adequacy or inadequacy of meaning in definitions given to headwords. It assesses how the application of sense relations contributes to the user-friendliness of definitions, and the accessibility and semantic analysis of lexemes in lexicography. This will hopefully give insights that will help refine and improve future dictionary definitions, especially for the proposed *Advanced Ndebele Dictionary*. These insights will help lexicographers to compile user-friendly and accessible dictionary definitions meeting users' needs and expectations. It will provide an approach to lexicography that accounts for more than a selection of semantic facts of lexemes and phrases.

The paper shows that hyponymy and meronymy facilitate the formulation of more user-friendly and accessible dictionary definitions. The two sense relations form the basis of the analytic, classical, intentional and Aristotelian definition which is considered the most accessible and user-friendly in lexicography (Jackson 1988). Hyponymy and meronymy prove to be the most productive and illuminating ways of defining lexemes in lexicography. As possible defining formats, they can assist lexicographers to adhere to the essential principles of defining terms/headwords. On the other hand, synonym definitions reduce the user-friendliness and accessibility of the defini-

tions in dictionary types other than synonym dictionaries. Synonym definitions waste space, since all words used in the definition must be entered and defined. Unless synonym definitions are purely used as semantic commentaries in the onomasiological part of a dictionary when they are listed under an explicit cross-reference, they are not space-saving.

The paper recommends that in monolingual and other dictionary types which are not synonym dictionaries, synonyms should as much as possible be avoided in the definitions. They should be included after the definition proper as cross-references. It is recommended that hyponymy and meronymy should be used effectively and extensively in defining semantic sets and part-whole relationships. Another view is that the use of synonym definitions compromises the essential principles of defining, that of defining the entry, addressing the question 'What is it?' directly and immediately, avoiding circularity and improving accuracy. This study shows that the use of synonym definitions appear to be the source of *ukukiliza* (the ungrammatical or erroneous use of words), yet a dictionary must be a trusted and respected depository of facts about a language and culture in addition to being an instrument that protects the language from unacceptable and corrupting influences. Finally it is argued that synonym definitions in dictionaries like ISN which are not synonym dictionaries, compromise the value of the dictionary as a resource of what is considered to be the 'best' or 'correct' language or as a tool that represents some form of final authority in matters of lexical meaning and use.

#### References

- Hadebe, S. et al. (Eds.).** 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Jackson, H.** 1988. *Words and their Meaning*. Harlow: Pearson Education.

---

#### **Nkomo, Dion ■ An Analysis of Subject-field Representation in Special-field Dictionaries: In Pursuit of a Theory of Lemma Selection for Special-field Lexicography**

*Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, Republic of South Africa (deouf@yahoo.co.uk)*

---

The notion of lexicographic functions is singled out as one theme in the historical development of lexicography which emerged once in 1940, disappeared and reappeared to become of paramount importance in modern-day lexicography (Gouws and Prinsloo 2005: 7). Metalexigraphers distinguish between knowledge-oriented and communication-oriented (linguistic) functions (Tarp 1995, 2000 and 2002, and Gouws and Prinsloo 2005). In spite of this distinction, there seems to exist a strong link between the two types of functions. Concentrating on subject-field dictionaries, which are the focus of this paper, all communication-oriented functions are based on the knowledge-oriented function of subject-field dictionaries, which is primarily to provide 'special data about the subject field' (Gouws and Prinsloo 2005: 14).

The provision of special data reflects on the knowledge of the subject field either in the users' native or foreign language and therefore fulfils the following specialised dictionary communicative functions listed and discussed in detail by Tarp (1995: 22-24, 2000 and 2002):

- native language text production
- native language text reception
- foreign language text production
- foreign language text reception
- translation from the native language to the foreign language
- translation from the foreign language to the native language

By facilitating these communication-oriented functions, the provision of special data about the special field becomes central to special-field lexicography. The success of the special-field dictionary in fulfilling the above listed communicative functions primarily rests on the extent to which it represents the subject field in question. A principled lemma selection is a prerequisite for a systematic representation of the subject field, for which a subject-field dictionary is being compiled. Lemma selection in special-field lexicography would be systematic if it is preceded by a systematic classification of the subject field in question (Pedersen 1995: 83-90). It is more than a decade since Bergenholtz and Tarp (1995: 98) expressed concern that lemma selection, together with equivalent selection for special-field dictionaries, had received little attention from metalexigraphy. An analysis of recently published special-field dictionaries reveals that while a fair share of attention has been given to lemma selection by metalexigraphers, some of whom are also practising lexicographers, practical lexicography continues to suffer from a lack of adequate theoretical guidance. This is regrettable considering that lemma selection primarily distinguishes between special-field lexicography and general lexicography (Mihindou 2004). General dictionaries basically deal with language for general purposes (LGP), while specialised dictionaries treat the various special fields of the lexicon (Gouws and Prinsloo 2005: 6). Besides this, special-field lexicography, like any other part of lexicography, needs first and foremost to be guided by strong theoretical principles of lexicography.

This paper investigates subject-field representation mainly in three subject-field dictionaries produced by the ALLEX project in Zimbabwe, namely *Isichazamazwi SezoMculo* (ISM), the Ndebele dictionary of musical terms, its Shona counterpart *Duramazwi reMimhanzi* (DRM) and *Duramazwi reDudziramutauro neUvaranomwe* (DDU), the Shona dictionary of linguistic and literary terms. Attention is paid to the nature of the subject fields (domains) in question, their LSPs, dictionary scope, dictionary size, the lexical status of its lemmata, and the treatment of synonyms and variants. The paper thus brings to the fore the challenges of lemma selection encountered and the principles employed by their editors during lemma selection. A closer look at the macrostructural elements of the dictionaries takes this paper into the domain of dictionary criticism. Admittedly, this is a challenge

to this writer who does not only write as a former ALLEX member but also as co-editor of one of the dictionaries, the ISM. The inclusion of the other two dictionaries, which were compiled by other team members for this discussion is meant to counter the paradox of self-criticism. However, it presents another challenge of having Shona dictionaries criticized by a non-Shona metalexigrapher or native speaker. In the case of this writer, little Shona language competence is supported by the inclusion of English reverse indices in both Shona dictionaries and also a Ndebele reverse index in the DRM. Encyclopedic competence of this writer is supported by having been a member of the Musical Terms Dictionary project, which led to the production of both the ISM and the DRM, and also the writer's background as a linguistic and African languages student in the Department of African Languages and Literature at the University of Zimbabwe. All these facts, as well as being a researcher in special-field lexicography, should enable this writer to arrive at metalexigraphically sound conclusions in pursuit of a theory for lemma selection for special-field lexicography.

### References

- Bergenholtz, H. and S. Tarp.** 1995. Selection. Bergenholtz, H. and S. Tarp (Eds.). 1995: 98.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries.* Amsterdam: John Benjamins.
- Chimhundu, H. and E. Chabata (Eds.).** 2006. *Duramazwi reDudziramutauro neUvaranomwe.* Gweru: Mambo Press.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography.* Stellenbosch: SUN PReSS.
- Mheta, G. (Ed.).** 2005. *Duramazwi reMimhanzi.* Gweru: Mambo Press.
- Mihindou, G.-R.** 2004. Some Features of Monolingual LSP Dictionaries. *Lexikos* 14: 118-136.
- Mpofu, N. et al. (Eds.).** 2004. *Duramazwi reUrapu neUtano.* Gweru: Mambo Press.
- Nkomo, D. and N. Moyo (Eds.).** 2006. *Isichazamazwi SezoMculo.* Gweru: Mambo Press.
- Pedersen, J.** 1995. Systematic Classification. Bergenholtz, H. and S. Tarp (Eds.). 1995: 83-93.
- Tarp, S.** 1995. Specialised Dictionary Functions. Bergenholtz, H. and S. Tarp (Eds.). 1995: 20-28.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2002. Basic Elements of Lexicographic Theory. Emejulu, J.D. (Ed.). 2002. *Éléments de lexicographie gabonaise. Tome II:* 7-20. New York: Jimacs-Hillman.

---

**Nyangone Assam, Blanche and H. Steve Ndinga-Koumba-Binza ■**  
**Lexicography and the French Language in Gabon**

*French Matters, Stellenbosch, Republic of South Africa (bnyangone@hotmail.com);  
and Stellenbosch University Centre for Language and Speech Technology,  
Stellenbosch, Republic of South Africa (13751719@sun.ac.za)*

---

This paper focuses on lexicography and the French language in Gabon. It suggests reconsidering the place of French within the emerging Gabonese lexicography. French is the sole official language in Gabon, a country where language studies are experiencing the fast-growing and emerging discipline of lexicography. Though it has its origins as a foreign language, French has been granted the rank of national language through socio-political, economic and contextual means (Ndinga-Koumba-Binza 2004 and 2005a, Pambou 1998). This leads to the question whether Gabon should discontinue importing French dictionaries in order to compile its own (Ndinga-Koumba-Binza 2005b). This question gives rise to another: What is the place of French within the emerging Gabonese lexicography? Apart from introducing French Gabonese lexicography and its contribution to education, reconsidering the place of the French language within Gabonese lexicography would cover the following topics:

**1. A review of the various positions of French in Gabon**

A number of studies have been conducted on the situation of French in Gabon (Couvert 1982, Moussirou Mouyama 1984 and 1986, Ogden 1984, Blanchon 1994, Pambou 1998, Tomba Moussavou 2001, Mouloungui Nguimbyt 2002 and Ndinga-Koumba-Binza 2004 among others). Yet, many have tried to circumscribe the status of French in Gabon (Moussirou Mouyama 1984, Blanchon 1994, Pambou 1998 and Ndinga-Koumba-Binza 2004).

The most comprehensive and focused on its current status is that of Pambou (1998). He has identified five positions taken by French in Gabon, namely as:

- (a) second language,
- (b) foreign language,
- (c) mother-tongue, first language or initial language,
- (d) real initial language, and
- (e) official language and language of education.

According to our understanding of the situation of French in Gabon, this order of presentation and the designation of each status should be reorganized in terms of foreign language, official language, second language, mother-tongue and initial language, local and national language, and appropriated language.

**2. The current diversity within the language itself**

The idea of a unified French language is purely fictional. In reality, several aspects diversify French: chronological, geographical, functional, etc. The focus will be on the geographical diversities. French is a worldwide language diffused over all continents and various countries. This geographical dispersal brings the language into daily contact with new usages in Europe

(Switzerland, Belgium, Luxemburg), North America (Canada, Louisiana), Africa (Senegal, Morocco, Gabon, the Congo, etc.), Asia (Cambodia, Vietnam) and the Middle East (Lebanon). This also reflects different cultures where French has been adopted and usages have diversified from standard French, as recognised by the Académie Française.

It is important that, in the above-mentioned places, when it comes to learning and teaching French as foreign or second language or even for any other purpose among native speakers, standard French is the norm. Nevertheless, many of these geographical diversities have claimed the right to be recognised as different forms of French (e.g. by the Académie Royale de la Langue Française in Belgium, the Académie Canadienne-Française, now the Académie des Lettres du Québec and the Office de la Langue Française au Québec, in Canada, and the Délégation de la Langue Française in Switzerland). One of the ways to bring this about is through the compilation of dictionaries presenting these differences.

### References

- Blanchon, J.A.** 1994. Gabon: Language Situation. Asher, R.E. (Ed.). 1994. *The Encyclopedia of Language and Linguistics. Volume 3.* New York/Oxford: Pergamon Press.
- Couvert, C.** 1982. *Langue française en République gabonaise.* Paris: Haut Comité de la Langue Française.
- Mouloungui Nguimbyt, F.V.** 2002. *De la variation dialectale en français au Gabon.* Unpublished M.A. Thesis. Libreville: Omar Bongo University.
- Moussirou Mouyama, A.** 1984. *La langue française au Gabon: Contribution sociolinguistique.* Doctoral Dissertation. Paris: University René Descartes.
- Moussirou Mouyama, A.** 1986. L'introduction de la langue française au Gabon. *Annales de L'Université Omar Bongo du Gabon: Série Lettres, Droit, Sciences et Médecine:* 55-63.
- Ndinga-Koumba-Binza, H.S.** 2004. *Le statut socio-politique du français et la promotion langues locales au Gabon.* Paper presented at the Eighteenth International Conference of the Association of French Studies in Southern Africa (AFSSA), Stellenbosch, University of Stellenbosch, 8–11 September 2004.
- Ndinga-Koumba-Binza, H.S.** 2005a. Politique linguistique et éducation au Gabon: un état des lieux. *Journal of Education* 4(1): 65-78. Réduit: Mauritius Institute of Education.
- Ndinga-Koumba-Binza, H.S.** 2005b. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ogden, J.** 1984. French in Gabon. *Contemporary French Civilization* 8: 339-348.
- Pambou, J.-A.** 1998. Le français au Gabon: une langue à multiples statuts. *Iboogha* 2: 127-149.
- Tomba Moussavou, F.** 2001. *Les réalisations des séquences des consonnes en français de Libreville.* Unpublished M.A. Thesis. Libreville: Omar Bongo University.

---

**Ojwang, Humphrey J. • A Study of Luo Ethnobotanical Dictionaries**

*Institute of African Studies, University of Nairobi, Nairobi, Kenya*  
([hjojwang@yahoo.com](mailto:hjojwang@yahoo.com))

---

Traditionally, people have used plants for different purposes: food, building, furniture, medicine and even decoration or beauty. In Africa, people have developed an intimate knowledge of plants since time immemorial. In order to use plants as a valuable resource, communities have evolved systematic taxonomies of plants enabling them to identify and name different plants and their uses.

This paper focuses on Luo ethnobotanical dictionaries, specifically on the printed dictionaries by Prof. John O. Kokwaro of the Department of Botany, University of Nairobi, Kenya. Kokwaro has published two important dictionaries, namely the *Luo–English Botanical Dictionary* and the *Luo Biological Dictionary* (the latter covering both the names of plants and animals).

In the *Luo Biological Dictionary*, Kokwaro and Johns (1998) argue that the use of plants and animals, as well as their subsequent products, reflects their way of life wherever they are found in East Africa. These lexicographers have given detailed descriptions of biological species in the Luo country. Most of the ethnobotanical materials in the biological dictionary come from Kokwaro's original work published in 1972, namely the *Luo–English Botanical Dictionary of Plant Names and Uses*.

This paper seeks to analyze Luo ethnobotanical dictionaries according to a futuristic perspective. An attempt is made to address the gaps identified in the two ethnobotanical dictionaries, namely:

- Translating Luo plant names into English
- Interpreting the meanings of Luo plant names
- Paraphrasing any culture-specific bits of information that may be gleaned from Luo oral folklore

Future Luo ethnobotanical dictionaries will certainly be influenced by these two seminal works. There will be need to include some derivational features of plant names in the Luo ethnobotanical dictionaries, namely:

- Meanings derived from characteristics of plants
- Meanings derived from usages of plants
- Meanings derived from human attributes of plants
- Meanings derived from animal attributes of plants
- Meanings derived from folklore
- Meanings derived from borrowed or loan words from other languages

An ethnolinguistic analysis of plant names in a speech community relies on what ethnobotanists call *data*, i.e. 'the broad range of information they collect on how local people interact with the natural environment' (Martin 1995: 10). In ethnobotanical dictionaries, recorded data will necessarily include the plant names and their meanings in a given language. In this paper, information about plants from the printed dictionaries of Kokwaro constitutes the data to be analyzed. A futuristic approach to printed dictionaries of plant



names is adopted, given the fact that Kokwaro in his seminal works does not give the derivation of the meaning(s) of plant names, but simply gives local Luo names and matches them with their scientific equivalents. The analytical approach adopted in this paper seeks to retrieve the meaning(s) of plant names from the cultural and linguistic reality of Luo speakers.

The study derives from the culture-specific concerns of ethnobotanical terminology and the relevance of enhancing Luo lexicographic practice. It also identifies the patterns followed in naming plants in the Luo speech community. The paper concludes by stating that any study of Luo ethnobotanical terminology in a culture-specific configuration must necessarily have lexicographic implications for dictionary making. This approach should enable lexicographic and terminological practices to be deployed in the botanical sciences and related disciplines in culturally relevant contexts.

### References

- Kokwaro, J.O.** 1972. *Luo-English Botanical Dictionary of Plant Names and Uses*. Nairobi: East African Publishing House.
- Kokwaro, J.O. and T. Johns.** 1998. *Luo Biological Dictionary*. Nairobi: East African Educational Publishers.
- Martin, Gary J.** 1995. *Ethnobotany*. London: Chapman and Hall.

---

### **Pienaar, C. Leela ▪ Towards a Corpus of South African English: Corraling the Subvarieties**

*Dictionary Unit for South African English, Rhodes University, Grahamstown, Republic of South Africa (l.penaar@ru.ac.za)*

---

Within the last twenty years, the use of a corpus for language research has become the *sine qua non* in many areas of linguistic enquiry. This trend is particularly evident in the discipline of lexicography, which has become increasingly and overtly 'corpus-driven'.

This paper draws on research from a Master's project which involved the collection of a small corpus of Indian South African English (ISAE), an acknowledged component or subvariety of South African English (SAE). The discussion highlights the importance of aiming for a balanced representation of the known subvarieties of a language when compiling corpora for lexicographic and linguistic investigation.

Since ISAE is primarily an oral dialect, specific focus is given to the methodological challenges involved in compiling a spoken corpus. Methodological insights from local as well as international corpus research were used to guide and inform the process. These include the Xhosa English Corpus, the New Zealand Corpus of Spoken English and the Hong Kong Corpus of Conversational English. The various stages in the research process are described, together with explanations of how problems such as the issues of corpus design, the selection of corpus contributors, the data-collection process and developing guidelines for consistency during the corpus compilation were addressed.

The paper provides a keyhole view of the main lexical and syntactic features of ISAE exemplified in the corpus and juxtaposes these against the backdrop of general SAE and trends in World English. It concludes with a motivation for the collection of parallel corpora of other subvarieties of SAE which will provide an objectively compiled repository of language in use to enable researchers to discern the linguistic features at the core and periphery of SAE. It is argued that the establishment of corpora of the various known subvarieties of SAE could constitute an important step towards the creation of a truly representative large corpus of SAE and ultimately towards a better definition and understanding of SAE.

---

**Ramaliba, T. Zacharia • The Significance of Noun Classes in the Lemmatization of Nouns in a Tshivenda Dictionary**

*Tshivenda National Lexicography Unit, Venda University of Science and Technology, Thohoyandou, Republic of South Africa (zacharia.ramaliba@univen.ac.za)*

---

This paper looks at the significance and role of noun classes in the compilation of Tshivenda dictionaries, especially bilingual dictionaries where Tshivenda and English are involved. It needs no emphasis to state that nouns on their own are fully-fledged lexical items in the lexicon. Nouns as linguistic expressions are regarded as independent, which is why no dictionary can afford not to lemmatize them. The problem though is that currently Tshivenda dictionaries, especially bilingual ones, do not include much information about the classes to which the various nouns in Tshivenda belong.

It is important to realize that all nouns in Tshivenda belong to a specific noun class and these noun classes are distinguished by different class prefixes. This is, however, not applicable to English and it will therefore not be important to classify English nouns in the same way as Tshivenda nouns. The paper endeavours to indicate that without including details regarding a variety of noun classes in Tshivenda dictionaries, it will be very difficult, if not impossible, to compile a proper and user-friendly dictionary in Tshivenda. It must be borne in mind that nouns in Tshivenda consist of 18 classes, i.e. **mu-vha** [+1-2], **mu-mi** [+3-4], **li-ma** [+5-6], **tshi-zwi** [+7-8], **n-dzin** [+9-10], **lu** [+11], **vhu** [+14], **u** [+15], **fha** [+16], **ku** [+17], **mu** [+18], **ku** [+20] and **di** [+21].

It cannot be overemphasized that each of these noun classes can only accommodate specific nouns and meanings (especially referential and affective meanings). It therefore follows that should a lexicographer ignore the impact of nouns when compiling a dictionary, a user will find it very difficult to comprehend what is being described, as Tshivenda noun classes identify and classify nouns according to their content such as human beings, time, estimation of time and things, rivers, plants, parts of the body, natural phenomena, animals and birds, domestic utensils, fluids, and languages, to mention but a few examples. In other words, it is not only the macrostructural entries that should receive attention from lexicographers, but the micro-

structural features of nouns must also be considered as equally important when it comes to the lemmatization of data. In addition, in order to compile a quality dictionary in Tshivenda, the paper will deal with semantic issues associated with these noun classes. The paper examines this issue as lexicographers often do not regard the semantics of noun classes as independent multi-lexical items, therefore not lemmatizing them.

Semantic issues associated with noun classes that receive attention in this paper, pertain to both referential (as already alluded to above) and affective meaning. The latter type of meaning refers to aspects such as feelings, attitudes and opinions. There are noun classes used to express regular meaning, augmentative meaning and derogative meaning. Should a lexicographer not be aware of these meanings and thus exclude them from the dictionary, the user may end up confused. For example, the noun **muthu** (person) belongs to class [+], but it can acquire a derogative meaning if prefixes of other noun classes are attached, as shown in the following examples: **dithu** [+21] (a big, ugly, unattractive person), **luthu** [+11] (a very lean, insignificant person), and **kuthu** [+20] (a very small, short person). This is just some of the issues that receive attention in this paper.

From this brief explanation, it is evident that noun classes play a major role as far as the lemmatization of nouns in Tshivenda is concerned. For illustrative purposes, the paper bases its argumentation on dictionaries already available in Tshivenda such as those by Van Warmelo (1989), and Wentzel and Muloiwa (1982), and the terminology lists compiled by the Departments of Arts and Culture (National and Provincial).

#### References

- Van Warmelo, N.J.** 1989. *Venda Dictionary Tshivenda–English*. Pretoria: J.L. van Schaik.
- Wentzel, P.J. and T.W. Muloiwa.** 1982. *Trilingual Elementary Dictionary Venda–Afrikaans–English*. Pretoria: University of South Africa.

---

#### **Sakwa, Lydia Namatende • An Analysis of the Usage Labelling Practices in English Lexicography**

*Faculty of Education, Uganda Martyr's University, Nkozi, Uganda (lsakwa@umu.ac.ug)*

This paper analyses the usage labelling practices in English lexicography. Landau (2001: 217) stipulates that *usage* refers to 'any or all uses of language'. It is the study of good, correct, or standard uses of language as distinguished from bad, incorrect, and nonstandard uses of language. Usage may also include the study of any limitations on use, whether geographic, social or temporal. It basically alerts the user that certain terms should not be naively employed in communication. Usage is included in dictionaries as a guide to speakers of a language in making their decisions on how to use words appropriately. To help learners with these difficulties, a number of labels are introduced in the dictionary to denote the stylistic values of words or the technical fields in which they are used; these are called usage labels.

Although some users may interpret all dictionary data as authoritative on how to spell, pronounce, interpret meaning or check usage, the question surrounding the appropriateness of usage labelling as found in most dictionaries is one that has raised much concern among many linguists.

This paper intends to expose the dilemma surrounding usage labels, which makes them confusing and sometimes misleading from the point of view of the second language learner. The focus of this study is on exposing the problems of the labels assigned to words from a second language speaker's perspective. The study is therefore useful in an attempt to guide future lexicographers to make more informed usage labelling decisions having been exposed to inadequacies attributed to usage labelling in English lexicography.

The analysis is based on the labelling practices in six monolingual English dictionaries namely: the *Oxford Advanced Learner's Dictionary*, the *Macmillan English Dictionary*, the *Longman Dictionary of Contemporary English*, the *Cambridge International Dictionary of English*, the *World Book Dictionary* and the *New Shorter Oxford English Dictionary*.

Discrepancies in the contextual usage labelling in the dictionaries are inclusive of the inconsistencies that surround usage labelling in English lexicography (with the same word bearing different labels in different dictionaries), unreliable front and/or back matter information, inadequate definitions of the usage labels, ambiguity, questionable selection and classification practices, as well as the tendency to adopt a prescriptive and/or proscriptive tone, as opposed to the recommended descriptive tone in usage labelling. All these discrepancies seem to emanate from the absence of an agreed-upon criterion for making usage decisions in English lexicography; this is the reason for the inconsistencies which characterize usage labelling in English dictionaries, and the problems which emerge from such inconsistencies.

#### **Reference**

**Landau, S.I.** 2001. *Dictionaries. The Art and Craft of Lexicography*. Second Edition. Cambridge: Cambridge University Press.

---

#### **Selokela, Mokgadi R. ▪ The Relevance of *Pukuntšuthaloši ya Sesotho sa Leboa* in a Classroom Situation**

*Sesotho sa Leboa National Lexicography Unit, University of Limpopo, Turfloop Campus, Polokwane, Republic of South Africa (mokgadis@ul.ac.za)*

---

The aim of this paper is to investigate the relevance of the newly published Northern Sotho Monolingual Dictionary, *Pukuntšuthaloši ya Sesotho sa Leboa* (2007), in a classroom situation. It is designed to be used by learners as reference material as they use their text books in a classroom situation.

Dictionaries are mostly used by people, especially students, to look up meanings. To further comprehend the text or subject matter with which they are dealing, students usually consult dictionaries and it is through this function that dictionaries play a role in the classroom teaching situation. Like any

dictionary, *Pukuntšuthaloši ya Sesotho sa Leboa* is primarily compiled to be consulted by any indigenous speaker of Northern Sotho to look up meanings. This paper tries to investigate its extended function: its role in a classroom learning situation.

Undoubtedly many dictionaries fulfil a pedagogical function, be it at home, in the library, in the office or in a classroom. *Pukuntšuthaloši ya Sesotho sa Leboa* would be no exception in fulfilling this function. It will follow other dictionaries as a pedagogical publication. But the interest of this paper is on whether this dictionary can, in the meantime, play a role as school dictionary since there are no available monolingual school dictionaries in Northern Sotho at present.

In investigating its function in classroom activities, the study tries to answer the following questions: whether *Pukuntšuthaloši ya Sesotho sa Leboa* could satisfy the needs of the learners in a classroom, whether the structure of the dictionary, i.e. macrostructural and microstructural elements, could make it easier for information retrieval by the learners, and whether learning Northern Sotho, i.e. its grammar, phonetics, phonology, syntax, morphology, etc., could be facilitated by the use of this dictionary.

Indigenous knowledge is very important in the pedagogical development of a child. Since this dictionary defines Northern Sotho culture, the research shows how this knowledge can be imported through the use of this dictionary into the classroom learning situation.

---

### **Swanepoel, Piet ▪ Dictionary Criticism: The Lame Leg of Lexicographic Theory and Practice**

*Department of Afrikaans and Theory of Literature, University of South Africa (UNISA), Pretoria, Republic of South Africa (swaneph@unisa.ac.za)*

---

Although dictionary criticism (Wörterbuchkritik; cf. Wiegand 1993, 1998) has been indicated as one of the major components of a theory of lexicography, very little research has in fact been forthcoming in this field, and, above all, very little thought has gone into questions as to exactly what the main goals of such research should be, and what data, theories and methodologies would be needed to achieve these goals.

In this paper a distinction is made between dictionary criticism as praxis and the scientific study of dictionary criticism. Problems currently experienced in both these areas of lexicography are discussed at some length. This is followed by a proposal for relevant goals, data, theories and methodologies for both of these fields. Special attention is given to the most important aspects of dictionary evaluation criteria, viz. their content, format, justification and application. A number of suggestions are put forward for research in both these components of dictionary criticism.

#### **References**

**Wiegand, H.E.** 1993. Zur Einführung (of the Thematic Part *Wörterbuchkritik*). *Lexicographica* 9: 1-7.

**Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Berlin/New York: Walter de Gruyter.

---

**Tan, Kim Hua ▪ Media Specific Features of Electronic Dictionaries: Dictionary Users' Perception**

*School of Language Studies and Linguistics, Kebangsaan Malaysia University, Bangi Selangor, Malaysia (kimtjinlyn@yahoo.com)*

---

A corollary of the electronic era in the area of electronic lexicography is the development and use of electronic dictionaries in language learning. This paper is concerned with the identification and examination of dictionary features perceived as helpful by ESL learners in assisting them in accessing information. The data were obtained from the analysis of literal comments of a hundred ESL learners after the subjects have completed a sense differentiation task with the help of an electronic dictionary. The participants were university undergraduates majoring in English Language Studies at a local public university in Malaysia and all of them did not have any systematic training in dictionary use. The set sense differentiation task required the participants to identify and circle the correct sense of a particular word in a sentence from the given range of senses of that word. Participants were then asked to write down dictionary features that particularly helped them in the task and also dictionary features they would prefer in an electronic dictionary. The CD-ROM version of the *COBUILD English Dictionary for Advanced Learners* was used as reference in the task completion exercise.

Anecdotal data in the form of literal comments of the participants or dictionary users were analyzed through a process of data reduction, data display and conclusion drawing and verification (Miles and Huberman 1994). At the data reduction stage, the literal comments were categorized into media-related features and generic-related features. Media-related features are dictionary features specific to the electronic medium and not the print medium. Generic-related features refer to dictionary features related to dictionary content regardless of the delivery medium. Both categories of dictionary features identified were further classified into macrocategories, microcategories and microtypes and accorded 'helpful' and 'unhelpful' status based on the literal comments. Qualitative coding categorization was used to classify both media-related and generic-related dictionary features at three levels: the macrostructural level, the microstructural level, and the mediostructural level. Macrostructure in e-dictionaries refers to the procedural structure of how the user goes about accessing entries. Access and retrieval of information may not be constrained by the alphabetic organization of the dictionary. Microstructure refers to the structure of an entry. It is the way the content of each entry is organized. It is the internal design of a reference unit. Mediostructure denotes the way in which a lexicon's structure integrates itself into resources external to the lexicon (Burke 1998). Dictionary

features existing within these definitions are classified accordingly and grouped in their respective levels.

Percentage based on frequency count was tabulated at the data display stage and conclusions drawn from the analyzed data. Results generally indicated that preferred dictionary features mostly occurred at the microstructural level and that generic-related features were a priority as compared to media-specific features. Findings in this study pointed towards the fact that dictionary users appreciated generic-related features such as definitions, example sentences and annotational features more than other dictionary features. For media-specific features, among the occurrences at the three levels, features most frequently quoted were at the microstructural level, followed by macrostructural and mediostructural levels. Search procedure (search speed, search type), navigation within (hyperlinks), typography (bold print, font type and size, colour), layout (screen layout, text layout), sound (pronunciation) and navigation without (external links) were most frequently quoted. The implication here being, the structure of each dictionary entry is deemed more important than linking features within the dictionary, or out of the dictionary.

#### **References**

- Burke, S.M.** 1998. *The Design of Online Lexicons*. Unpublished M.A. Thesis. Evanston, IL: Northwestern University.
- Miles, M.N. and A.M. Huberman.** 1994. *Qualitative Data Analysis*. Second Edition. Newbury Park, CA: Sage Publications.

---

#### **Tarp, Sven ■ A Lexicographic Approach to Language Policy**

*Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (st@asb.dk)*

---

Whereas there is a long tradition of discussing the concepts of description and prescription in terms of lexicographic theory and practice, very few contributions have dealt with the complex relation between general language policy and the conception of lexicographic works. This paper will show that even if the lexicographers accept the indications given by language policy at a macro level, they cannot transfer them blindly to the lexicographic micro level without undermining the general language policy. On the one hand, as Wiegand (1986) argued, a 'descriptive' dictionary will always have a prescriptive value or force. On the other hand, a completely prescriptive dictionary will in many cases not allow the users to access the needed data. On this basis, the paper suggests a proscriptive approach in terms of dictionary making.

#### **Reference**

- Wiegand, H.E.** 1986. Von der Normativität deskriptiver Wörterbücher. Zugleich ein Versuch zur Unterscheidung von Normen und Regeln.
- Henne, H. 1986. *Sprachnormen in der Diskussion. Beiträge vorgelegt von Sprachfreunden: 72-101*. Berlin/New York: De Gruyter.

---

**Thangalimodzi, Kondwani ▪ The Role of an Electronic and Online English–Chichewa Dictionary in Technology Pedagogy and E-Learning**  
*Mzuzu University, Mzuzu, Malawi (thangalimodzi@yahoo.com)*

---

The online English–Chichewa dictionary is a tool for promoting e-learning in Africa. Chichewa (also known as Nyanja) is a language of the Bantu language family, spoken in parts of East, Central, and Southern Africa. It is used in Malawi where, from 1968 until recently, it has served as the national language. It is also spoken in Mozambique, especially in the provinces of Tete and Niassa, in Zambia (especially in the Eastern Province), as well as in Zimbabwe where, according to some estimates, it ranks after Shona and Ndebele as the third most widely used local language. The countries of Malawi, Zambia, and Mozambique constitute the central location of Chichewa. It is also spoken in the southern part of Tanzania. About 60% of the general Malawian population is considered illiterate and this makes it difficult for them to learn technologies in foreign languages. Most of these people considered illiterate have a functional literacy or an active command of Chichewa. Therefore this software is of great help in promoting e-learning technology in Africa.

The online English–Chichewa dictionary software project is an Internet dictionary for English and Chichewa. The dictionary system was developed using PHP scripting, MySQL and Linux. It can be hosted on an Internet server using LAMP/WAMP technology. Most Western cultures have greatly advanced technologically because their languages have been used in science and technology. The best way therefore of improving technology in Africa is by the implementation of local languages in science and technology. This can ease understanding among the less educated majority in Africa. With the help of the English–Chichewa dictionary software online, e-learning technology can be utilized by local Africans residing in some parts of Sub-Saharan Africa who do not have a full understanding of English. The other part of the dictionary will have an online translating tool.

With the increasing use of ICT in Africa on the one hand, and worldwide exploration of the multilingual potential of the technology on the other, there is an opportunity to develop tools that will help promote e-learning technologies. As Chichewa is a cross-border language, it clearly indicates that several countries are benefiting from initiating the project. The advantage of using an online dictionary is that it can be dynamically modified at runtime, eliminating the need to struggle with printing as is the case with hard copy dictionaries. With the introduction of new technologies, terminology development needs to be a continuing process and this can be facilitated with the use of online dictionaries like the English–Chichewa dictionary software.



---

**Tshikota, Shumani L. ▪ An Investigation of the Treatment of Nouns in Tshivenda Dictionaries: A Morphosyntactic Approach**

*Tshivenda National Lexicography Unit, Venda University of Science and Technology, Thohoyandou, Republic of South Africa (shumanileonard.tshikota@univen.ac.za)*

---

This study aims to investigate what a morphosyntactic noun looks like in a Tshivenda dictionary. The question with regard to the dictionary is what information should be included with a morphosyntactic noun. The study will trace grammatical information that one would expect to find in a dictionary entry, and how this information is provided, either explicitly or implicitly.

Attention is on how the syntactic operation of the noun in relation to other words could be indicated in the dictionary. For example, verbs are traditionally marked as 'transitive' or 'intransitive' but this is insufficient information to enable the dictionary user to form acceptable clauses. The dictionary entry needs to specify in detail into which clause patterns a verb may enter, what complementation a verb may take, and which items are obligatory or optional or deletable. Additionally, the dictionary needs to specify whether a verb cannot undergo a particular grammar rule, i.e. whether it is able to form 'progressive' tenses or to enter into the 'passive' construction. Verbs determine to a large extent which other elements may be present in their clauses.

What is true for the specification of the syntactic performance of verbs is also true, to a lesser extent, for nouns. Verbs are marked by the heads of their phrasal structures. Which labels are used to mark nouns, where is the syntactic operation of the noun indicated, and, if given, is it implicitly or explicitly?

The paper also explores the role of lexicographers in making nouns satisfy their syntactic performances, and issues of morphology which represent an inherent feature of nouns. The paper discusses the tradition used by Tshivenda dictionaries in treating nouns and whether the practice has been consistently followed, i.e. whether the question regarding the morphological and syntactic category noun can be answered in a principled way and whether there is a big difference between the morphological and syntactic category noun. It thus is necessary to look for the answer within these two components which deal with nouns, and whether the answer can be apparent from the tradition of dictionaries in the African languages. The other critical role is to look at the treatment of the noun and how the treatment makes provision for user-friendliness in the dictionary.

The paper explores why the major lexical items in Tshivenda always appear with prefixes and sometimes suffixes, why the formal category adjective in Tshivenda has to appear within a syntax with at least a noun class prefix, although it may also appear with a suffix such as **-ana**, and how the category adjective is treated in Tshivenda dictionaries and what information is apparent with the category adjective which is not needed.

The category verb also appears with a number of prefixes and suffixes, both inflectional and derivational. The verb is never entered into a dictionary with its inflectional morphemes such as agreement, tense, mood, negative

and aspect. The verb is always entered as a root or stem with at least a suffix [-a] which does not express any grammatical function. Most dictionaries, however, do include derivation affixes such as the applicative and the causative. In the same way, it is considered how nouns appear with the inflectional morphemes in Tshivenda.

The paper aims to establish whether the exact entry in a dictionary has to be considered from the viewpoint of the syntactic and morphological category only. Furthermore it has to be consistent with all the dictionaries in the specific language to guarantee the user-friendliness of the dictionary. With noun class as an inflectional morpheme, it is apparent that all nouns have to appear with some reference to a class prefix because noun class will always determine agreement on the verb and the nominal modifier.

However, with lexical derivation, it will be necessary to establish whether the various affixes may occur freely with all nouns or whether they have a very limited distribution. The answer to this issue will determine whether these derivations need to be entered into a dictionary.

It seems as if only specific nouns may appear with gender affixes and as if the locative and diminutive affixes may appear with almost all the nouns. The meaning expressed by the nouns derived with certain noun class prefixes is apparent in the original noun. Augmentative and diminutive nouns or nouns with a derogatory sense are derivations which can be realised with all nouns. What therefore determines their inclusion in the dictionary? Lastly the paper investigates the treatment of infinitives in Tshivenda dictionaries.

---

### **Van der Merwe, Michele ▪ The Pedagogical Purpose of Dictionaries in South Africa: A Perspective on Dictionary Education**

*Faculty of Education and Social Sciences, Cape Peninsula University of Technology, Wellington Campus, Wellington, Republic of South Africa (vandermerwem@cput.ac.za)*

---

Issues such as the following are raised:

- What is dictionary education?
- What is the status of dictionary education in South Africa?
- What are the benefits of dictionary education for a country?
- What is the perception of dictionary education among teacher-students?
- What is the role of dictionary education in establishing a dictionary culture?

Very little research has been done on dictionary education in schools, be that in Africa or in Europe. Carstens (1995) refers to research done in Europe and compares that to the situation in South Africa. In Europe it is mentioned in school curriculums that learners should be able to use reference works, but nothing is mentioned of dictionary education or the teaching of reference skills. A similar situation in South Africa exists, except for the fact that a few outcomes of dictionary use are mentioned in the school curriculum.

The focus of the paper is on the teaching of dictionary use at primary school level. Prerequisites for teaching dictionary use are an awareness of

dictionary typology, a realization of the importance of dictionary reference skills and knowledge of practical dictionary situations in the classroom. But do teachers have the know-how and learning materials to teach dictionary use? Textbooks do not furnish information on dictionary use and do not mention strategies to incorporate dictionary use with the teaching of linguistics. This can lead to the perception among teachers that dictionary use is not as important as language acquisition and the linguistics programme. Suggestions will be made for practical, outcomes-based ways to teach dictionary use, as well as some examples of dictionaries to be used.

Are teachers aware of the importance of dictionary education and the role it has to fulfil in the teaching of a language, especially linguistics? In the South African context of multilingualism, the teaching of linguistics plays a very important role to ensure successful communication between people. If a learner is taught well in his mother language, a good foundation for dictionary use is being laid. If dictionary use is being taught in a classroom, a good foundation for the acquisition of a new language is being created. Many teachers in South Africa are not adequately qualified to teach linguistics and most of them have not been trained in the very basics of dictionary typology and use. It would be understandable if they do not see dictionary education as a necessity in language teaching.

But what are the perceptions of a next generation of language teachers? The perceptions of teacher-students towards dictionary education were tested at the Cape Peninsula University of Technology by means of a questionnaire. The questionnaire dealt with issues such as their own use of dictionaries, their dictionary education and their perception of dictionary education in the primary school. An analysis of the questionnaire is given.

---

**Zondo, Jerry ▪ Affirming Verb Lexemes in *Isichazamazwi SesiNdebele* and in *A Practical Ndebele Dictionary: The Case of -wa Verbs***

*Department of African Languages and Literature, University of Zimbabwe, Harare, Zimbabwe (jerryzondo@arts.uz.ac.zw)*

---

Hadebe in *Isichazamazwi sesiNdebele* (ISN) (2001) and Pelling in *A Practical Ndebele Dictionary* (PND) have made entries of a number of verbs ending with **-wa**. Pelling has 18 verbs while Hadebe has 35. The user of both dictionaries could easily take the **-wa** verbs for the **-iw/-w-** word forms of the stylized verb extension, passive, as both lexicographers let slip into their dictionaries extension forms rather than typical lexemes. Lexicographers can therefore easily enter as headwords the passivized verb, a word form, while mistaking it for the original **-wa** ending verbs. Consequently other verbs ending with verbal extension morphemes like **-el-**, **-ek-**, **-is-**, **-isis-** and **-an-** can also be entered as headwords in dictionaries with lexicographers believing that they are dealing with original verbs.

Both Hadebe and Pelling have taken pains to establish original **-wa** verbs, treating them adequately and accurately, and selecting them carefully

as headwords. The dictionary user will notice, however, that both lexicographers have not subjected their verbs to a sufficiently selective and discriminating process that can beyond doubt differentiate the **-wa** original ending from the **-iw/-w-** extension.

A major observation is that the **-wa** ending verbs should/can be inflected by **-w/-iw-**, **-is-**, **-an-**, **-el-**, **-ek-** and **-isis-** in the usual manner. Incidentally, the process of inflection involves the principle of radical intactness despite the morphophonemic changes that might occur, like palatalization under [w] influence. Furthermore the same verbs can be used as bases for derivation of nouns or ideophones.

The paper clearly distinguishes between a lexeme and a word form. It then emphasizes the need for lexicographers to use lexemes as headwords rather than word forms. Both ISN and PND use word forms instead of lexemes, giving rise to inappropriate entries. All the **-wa** verbs will be lexemes while the **-iw/-w-** verbs are actually word forms and will not readily qualify as headwords if the appropriate tests were applied.

The paper calls for an appropriate front matter for Ndebele dictionaries which will avail the user with adequate information on inflection and derivation. Once word forms, in the case of verbs, have been explained in the front matter, all stylized verbal extensions will easily be accommodated by the dictionary user's 'ability' to inflect and use such forms. The dictionary user will be aware of the possibilities of inflecting and will hopefully do so automatically. There will be positive consequences on space utilization as opposed to space wastage, since space is considered as key in dictionary making.

While Pelling will not necessarily 'test' verbs to establish the use or otherwise of an extension, Hadebe obviously would, since he has had much linguistic experience. It does look, however, as if some of his entries have escaped his careful scrutiny prior to selection as headwords. The tests emphasized in the paper will easily separate lexemes from word forms in verbs.

Future lexicographic attempts could possibly use the tests not only for **-wa** ending verbs but also for verbs that end in **-eka**, **-ela**, **-isa**, **-isisa** and **-ana**, when all such forms are not verbal extensions but merely original endings. When the extension is clearly distinguished from an original verb ending, then the verb headwords can with absolute certainty be identifiable, and the process of headword selection perfected. If Ndebele lexicography should move towards the *Advanced Ndebele Dictionary*, specialized Ndebele dictionaries, Ndebele terminology dictionaries, and other types of dictionaries, a more consistent and reliable testing mechanism will help in arriving at more accurate headword selections which save space and avoid unnecessary entries.

While the accent in this paper is on the **-wa** ending verbs, similar research can be done on other endings (which could easily be taken for stylized verbal extension forms) to establish how both dictionaries have treated such verbs.

## AFRILEX MEMBERSHIP

<b>Annual Membership fee for:</b>	<b>Africa</b>	<b>Rest of the world</b>
Individual Members	ZAR 150	US \$50
Corporate Members	ZAR 600	US \$200

Membership is R150 (US \$50) per year and includes a **free** copy of *Lexikos*. Corporate membership is R600 (Africa) and US \$200 (outside Africa) and includes four copies of *Lexikos*. Published annually since 1991, *Lexikos* is a prestige international journal on lexicography. Contributions published in *Lexikos* are indexed by the *ISI Web of Science Citation Index*.

Members from abroad who are not from the USA should pay the equivalent of US \$50 in their own currency, e.g. Pounds drawn on a UK Bank, Euros in the Euro zone, etc. Due to the huge amounts charged for bank commission by the banks of members as well as our bankers, members are advised to pay for two or three years in advance. Members from Africa can also pay for up to three years in advance if they so wish.

Payments can be deposited directly into the AFRILEX Savings Account, giving easily recognizable identification, e.g. surname and initials:

ABSA Bank    Branch: Hatfield Pretoria    Branch Code: 008675  
Savings Account no.: 9054460560

The Proof of Payment, as well as the following completed AFRILEX Membership Form, must then be faxed, emailed or mailed to the Treasurer, Prof. E. Taljard:

**Fax:** + 27 (0)12 420 3163

**Email:** elsabe.taljard@up.ac.za

**Address:**

Prof. E. Taljard  
Department of African Languages  
University of Pretoria  
PRETORIA 0002  
Republic of South Africa

The AFRILEX Membership Form is also available on our website, at <http://afrilex.africanlanguages.com/membership.htm>



**AFRILEX MEMBERSHIP FORM**

Title, Initials and Surname: .....

.....

Work Address: .....

.....

.....

.....

Postal Address (if different from work address):

.....

.....

.....

Tel.:(W).....

(H) .....

Cell:.....

Fax: .....

E-mail: .....

I would like to become a member of AFRILEX. Included please find:

A. My cheque / postal order for

.....

(Please state currency and amount. Please make cheques / postal orders payable to: AFRILEX)

OR

B. My proof of payment for

.....

(Please state amount and date of payment)

