



# **AFRILEX**

## **African Association for Lexicography**

### **Programme & Abstracts**

**17th Annual International Conference**  
**University of Pretoria**  
**2<sup>nd</sup> – 5<sup>th</sup> July 2012**

*Hosted by:* Department of African Languages, Faculty of  
Humanities, University of Pretoria, South Africa

*Conference Organiser:* Prof. E Taljard and Prof. DJ Prinsloo

*Abstract reviewers:* Dr M Alberts, Prof G-M de Schryver, Dr PA Louw, Prof.  
RH Gouws, Dr T Otlogetswe, Dr VM Mojela, Dr HS  
Ndinga-Koumba-Binza, Prof. DJ Prinsloo, Prof. E Taljard

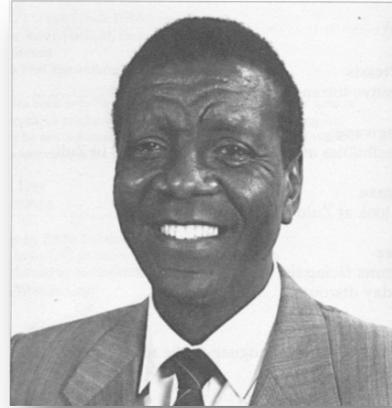
*Abstract booklet editor:* Ms J Wolvaardt

© 2012 AFRILEX, African Association for Lexicography  
ISBN 978-0-620-53469-7

## AFRILEX HONORARY MEMBERS



Professor R. H. Gouws



Professor A.C. Nkabinde

## AFRILEX Board 2011 – 2013

<b>President:</b>	Prof. G-M de Schryver
<b>Vice-President:</b>	Dr VM Mojela
<b>Secretary:</b>	Dr M Alberts
<b>Treasurer:</b>	Prof. E Taljard
<b>Members:</b>	Dr HL Beyer
	Dr PA Louw
	Dr HS Ndinga-Koumba-Binza
	Prof. DJ Prinsloo
<b>Lexikos Editors:</b>	Prof. E Taljard (2011)
	Prof. DJ Prinsloo (2012)

## **Afrilex 2012 – A Few Words from the President**

Dear Conference Participant,

Welcome to the 17<sup>th</sup> *International Conference of the African Association for Lexicography: AFRILEX 2012*. Let me begin with what you're holding, this conference booklet, once again meticulously compiled and edited by Jill Wolvaardt, the Executive Director and Editor-in-Chief at the *Dictionary Unit for South African English (DSAE)*. Thanks are due to her, as well as to her team of adjudicators, which included all Board Members of AFRILEX, as well as external specialists.

As the programme and abstracts in this conference booklet clearly indicate, AFRILEX 2012 promises to be yet another successful gathering, bringing together not only all lovers of African-language dictionaries, but also all those of us who help develop lexicography as an independent discipline. Members from academia, research institutions, civil servants and the trade will once again exchange their latest views and products, with no fewer than 40 scheduled presentations.

We especially welcome our keynote speakers: From Wellington (New Zealand), Rachel McKee will speak on "*Sign-Language Lexicography*", while her husband, David McKee, will simultaneously sign her presentation — a first at an AFRILEX conference. From Libreville (Gabon), Paul A. Mavoungou will entertain us on "*Gabonese French Dictionaries: Survey and Perspectives*". Also from overseas, from Valladolid (Spain), we welcome Pedro A. Fuertes-Olivera, who accepted our invitation to present a pre-conference workshop on "*LSP Lexicography*".

A final work of thanks goes to the local organizing team, headed by Elsabé Taljard and D. J. Prinsloo, both from the Department of African Languages at the University of Pretoria. We are happy to thank our sponsors: the Faculty of Humanities of the University of Pretoria for flying over Rachel McKee, as well as Pearson and Pharos (NB Publishers) for supplying office materials.

Welcome to Pretoria!

Gilles-Maurice de Schryver

President: AFRILEX

# CONTENTS

<b>Programme</b>		<b>7</b>
<b>Keynote Addresses</b>		<b>12</b>
<b>I.</b> New Zealand Sign Language Online: Insights from the making of a modern, multimedia dictionary <b>Rachel Locker MCKEE and David MCKEE</b>		12
<b>II.</b> Gabonese French Dictionaries: Survey and Perspectives <b>Paul A. MAVOUNGOU</b>		13
<b>Parallel Sessions</b>		<b>13</b>
<i>Somali Spelling Correction for Native Speakers of English</i>	<b>Nikki ADAMS</b>	13
<i>Terminology as an aid in mother-tongue education</i>	<b>Mariëtta ALBERTS</b>	16
<i>Access to data in lexicographic tools from the perspectives of lexicographic and information seeking concepts</i>	<b>Henning BERGENHOLTZ &amp; Theo BOTHMA</b>	17
<i>What is a dictionary?</i>	<b>Henning BERGENHOLTZ</b>	19
<i>The Potential of a Diagnostic Text Evaluation Model for Dictionary Criticism</i>	<b>Herman BEYER</b>	20
<i>Lexicography and the relevance criterion</i>	<b>Theo BOTHMA &amp; Sven TARP</b>	22
<i>A corpus of South African English for diachronic lexicographical research: what, why and how</i>	<b>Richard BOWKER</b>	24
<i>Bantu Forensic Lexicography: a Kikongo case study</i>	<b>Jasper DE KIND, Gilles-Maurice DE SCHRYVER &amp; Koen BOSTOEN</b>	25
<i>Establishing a dictionary culture inside the classroom for better dictionary use outside the classroom</i>	<b>Rufus GOUWS &amp; Danie J. PRINSLOO</b>	27

<i>Who is a lexicographer?</i>	<b>Rufus GOUWS</b>	28
<i>Expectation levels in dictionary consultation and compilation</i>	<b>Inge KOSCH</b>	29
<i>The role of sense relations in constructing definitional phrases for synsets in the Tswana WordNet</i>	<b>Jurie LE ROUX</b>	31
<i>From Subject-Based to User-Based Dictionaries: Rejecting the Cultural Opposition Thesis in Specialized Lexicography</i>	<b>Patrick LEROYER</b>	32
<i>The lemmatization of qualificatives in Setswana and Silozi</i>	<b>John LUBINDA</b>	34
<i>Treatment of taboo words in Gabonese lexicography: a case study of five reference works</i>	<b>Ludwine MABIKA MBOKOU</b>	35
<i>A critical analysis of the use of Xitsonga terms in Public Services – electronic and otherwise in South Africa</i>	<b>Pearl MAGAMANA</b>	36
<i>A comprehensive and extra-lexicographic process in support of teaching local languages in Gabon</i>	<b>Edgard MAILLARD ELLA</b>	37
<i>‘Balanced’ and ‘representative’ Corpus; A case study of the imbalance and representativeness of the Northern Sotho corpus</i>	<b>Victor MOJELA</b>	38
<i>A Macrostructure for an English-French Dictionary of Phonetic Sciences</i>	<b>Steve NDINGA-KOUMBA-BINZA</b>	40
<i>Standardisation of the principles and methods on terminology development in South Africa: A case of Statistics South Africa.</i>	<b>Lufuno NETSHITOMBONI</b>	42
<i>An exploration of Setswana cattle colour terms and their lexicographic treatment</i>	<b>Thapelo OTLOGETSWE</b>	43
<i>A critical evaluation of dictionaries with amalgamated lemma lists</i>	<b>Danie J. PRINSLOO</b>	44

<i>On-screen data presentation of collocations within a framework of lexicographic and usability requirements</i>	<b>Hilma A. REINHARDT</b>	46
<i>Old wisdom: The surprising and highly relevant lexicographical knowledge we can get from a specialized dictionary published in 1774</i>	<b>Sven TARP</b>	48
<i>Developing a mathematics and science school dictionary in isiXhosa: some challenges</i>	<b>Zola WABABA &amp; Julius DANTILE</b>	49
<b>Special Session on SABS TC37</b>		<b>50</b>
<i>An overview of SABS TC37 – a South African mirror committee of the international ISO/TC37</i>	<b>Mariëtta ALBERTS</b>	50
<i>An overview of SABS TC37/SC3 - Systems to manage terminology, knowledge and content</i>	<b>Karen CALTEAUX</b>	52
<i>The standardisation of the application of principles and methods in terminology, terminography, lexicography, translation and interpreting work – the role of TC37/SC2</i>	<b>Hatu MACHABA and Mapula GAFFANE</b>	54
<i>Ontologies and standardisation in ISO/TC37</i>	<b>Laurette PRETORIUS</b>	55
<i>Standards for Language Resource Management (ISO/TC37/SC4)</i>	<b>Justus ROUX &amp; Sonja BOSCH</b>	57
<b>ADDENDUM: Abstracts omitted from published conference booklet</b>		<b>58</b>
<i>Luxdico: A Lëtzebuergesch - German online dictionary: what could it learn from South African online dictionaries?</i>	<b>Juliane KLEIN</b>	58
<i>Double Pluralism of class 6 (ama-) and 10 (iiN-) nouns in isiNdebele: A case study in the isiNdebele Bilingual Dictionary</i>	<b>K. Sponono MAHLANGU</b>	60

# AFRILEX 2012: PROGRAMME

Monday 2 July 2012

**Pre-conference workshop: 09:00 – 16:00**

**Conference Centre: SRC Chamber**

LSP lexicography

**Pedro A. FUERTES-OLIVERA** in co-operation with **Mariëtta ALBERTS**

**19:00 – Cocktail Party**

**Adlers restaurant: Huis en Haard Building**

Tuesday 3 July 2012

08:30 – 9:10 **Registration**  
**Conference Centre: SRC Chamber**

09:15 – 9:30 **Official Opening**  
**Conference Centre: SRC Chamber**

Word from the President of Afrilex

**Gilles-Maurice de SCHRYVER**

A word of welcome on behalf of the University of Pretoria

**Keynote Address I: Conference Centre: SRC Chamber**

09:30 – 10:25 Sign-Language Lexicography  
**Rachel MCKEE & David MCKEE**

10:30 – 10:55 Tea

**Parallel Sessions**

	<b>Conference Centre: SRC Chamber</b>	<b>Graduate Centre L-64</b>
11:00 – 11:25	Access to data in lexicographic tools from the perspectives of lexicographic and information seeking concepts <b>Henning BERGENHOLTZ &amp; Theo BOTHMA</b>	A Macrostructure for an English-French Dictionary of Phonetic Sciences <b>Steve NDINGA-KOUMBA-BINZA</b>

11:30 – 11:55	Old wisdom: The surprising and highly relevant lexicographical knowledge we can get from a specialized dictionary published in 1774 <b>Sven TARP</b>	Somali Spelling Correction for Native Speakers of English <b>Nikki ADAMS</b>
12:00 – 12:25	What is a dictionary? <b>Henning BERGENHOLTZ</b>	Developing a mathematics and science school dictionary in isiXhosa: some challenges <b>Julius DANTILE &amp; Zola WABABA</b>
12:30 – 12:55	Who is a lexicographer? <b>Rufus GOUWS</b>	Standardisation of the principles and methods on terminology development in South Africa: A case of Statistics South Africa. <b>Lufuno NETSHITOMBONI</b>

13:00 – 13:55 Lunch  
**Adlers restaurant: Huis en Haard Building**

#### **Parallel Sessions**

	<b>Conference Centre: SRC Chamber</b>	<b>Graduate Centre L-64</b>
14:00 – 14:25	Treatment of taboo words in Gabonese lexicography: a case study of five reference works <b>Ludwine MABIKA MBOKOU</b>	A corpus of South African English for diachronic lexicographical research: what, why and how <b>Richard BOWKER</b>
14:30 – 14:55	An exploration of Setswana cattle colour terms and their lexicographic treatment <b>Thapelo OTLOGETSWE</b>	A critical analysis of the use of Xitsonga terms in Public Services – Electronic and otherwise in South Africa <b>Pearl MAGAMANA</b>

15:00 – 15:25 Tea

<b>15:30 – 16:30 Publishers' Session: Conference Centre: SRC Chamber</b>
Pharos: Spreading one's wings in a fierce local market <b>Wanda SMITH</b> MEDI Publisher (Botswana): Design and compilation of <i>Tlhalosi ya Medi ya Setswana</i> <b>Thapelo OTLOGETSWE</b> Oxford University Press: New horizons <b>Phillip LOUW</b> Pearson Education: t.b.a. <b>Jana LUTHER</b>

**16:30 – 18:00 Software Session: Conference Centre: SRC Chamber**

CText: The Autshumato Terminology Management System & Pharos'  
Multilingual Illustrated Document

**Wildrich FOURIE**

HLT @ CSIR: Tools, systems and applications

**Nic DE VRIES, Georg SCHLUNZ, Mpho KGAMPE**

TshwaneDJe HLT: Ten years of TshwaneLex

**David JOFFE**

18:00 – 19:00 **Annual General Meeting**  
**Conference Centre: SRC Chamber**

19:00 **Conference Dinner**  
**Adlers restaurant: Huis en Haard Building**

Wednesday 4 July 2012

**Keynote Address II: Conference Centre: SRC Chamber**

09:00 – 09:55	Gabonese French Dictionaries: Survey and Perspectives <b>Paul A. MAVOUNGOU</b>
---------------	---

10:00 – 10:25 Tea

**Parallel Sessions**

	<b>Conference Centre: SRC Chamber</b>	<b>Graduate Centre L-64</b>
		<b>Special Session on SABS TC37</b>
10:30 – 10:55	From Subject-Based to User-Based Dictionaries: Rejecting the Cultural Opposition Thesis in Specialized Lexicography <b>Patrick LEROYER</b>	An overview of SABS TC37 – a South African mirror committee of the international ISO/TC37 <b>Mariëtta ALBERTS</b>
11:00 – 11:25	Lexicography and the relevance criterion <b>Theo BOTHMA &amp; Sven TARP</b>	An overview of SABS TC37/SC3 - Systems to manage terminology, knowledge and content <b>Karen CALTEAUX</b>
11:30 – 11:55	Establishing a dictionary culture inside the classroom for better dictionary use outside the classroom <b>Rufus GOUWS &amp; Danie J. PRINSLOO</b>	The standardisation of the application of principles and methods in terminology, terminography, lexicography, translation and interpreting work – the role of TC37/SC2

		<b>Hatu MACHABA and Mapula GAFFANE</b>
12:00 – 12:25	Expectation levels in dictionary consultation and compilation <b>I.M. KOSCH</b>	Ontologies and standardisation in ISO/TC37 <b>Laurette PRETORIUS</b>
12:30 – 12:55	The Potential of a Diagnostic Text Evaluation Model for Dictionary Criticism <b>Herman BEYER</b>	Standards for Language Resource Management (ISO/TC37/SC4) <b>Justus ROUX &amp; Sonja BOSCH</b>

13:00 – 13:55 Lunch  
**Adlers restaurant: Huis en Haard Building**

### Parallel Sessions

	<b>Conference Centre: SRC Chamber</b>	<b>Graduate Centre L-64</b>
14:00 – 14:25	'Balanced' and 'representative' Corpus; A case study of the imbalance and representativeness of the Northern Sotho corpus <b>Victor MOJELA</b>	Bantu Forensic Lexicography: a Kikongo case study <b>Jasper DE KIND, Gilles-Maurice DE SCHRYVER &amp; Koen BOSTOEN</b>
14:30 – 14:55	The lemmatization of qualificatives in Setswana and Silozi <b>John LUBINDA</b>	Terminology as an aid in mother-tongue education <b>Mariëtta ALBERTS</b>

15:00 – 15:25 Tea

### Parallel Sessions

	<b>Conference Centre: SRC Chamber</b>	<b>Graduate Centre L-64</b>
15:30 – 15:55	Double Pluralism of class 6 and 10 in isiNdebele: A case in the isiNdebele Bilingual Dictionary <b>Sponono MAHLANGU</b>	On-screen data presentation of collocations within a framework of lexicographic and usability requirements <b>Hilma A. REINHARDT</b>
16:00 – 16:25	Luxdico: A Lëtzebuergesch - German online dictionary: what could it learn from South African online dictionaries? <b>Juliane KLEIN</b>	The role of sense relations in constructing definitional phrases for synsets in the Tswana WordNet <b>Jurie LE ROUX</b>

16:30 – 16:55	A critical evaluation of dictionaries with amalgamated lemma lists <b>Danie J. PRINSLOO</b>	A comprehensive and extra-lexicographic process in support of teaching local languages in Gabon <b>Edgard MAILLARD ELLA</b>
---------------	--	--

17:00 – 17:10    **Closure**

Thursday 5 July 2011

**Post-conference excursion to Freedom Park**



The President's tree, Freedom Park  
<http://www.freedompark.co.za>

## KEYNOTE PRESENTATION I

### **New Zealand Sign Language Online: Insights from the making of a modern, multimedia dictionary**

**Dr Rachel Locker MCKEE and Dr David MCKEE**

Victoria University of Wellington, New Zealand

Capturing the lexis of a relatively young signed language presents some challenges that differ in some respects from making a dictionary of a spoken language that has an historical existence in written form. Challenges for signed language lexicographers relate to modality, organization, analysis of signs as lexemes, and issues of sociolinguistic variation and authenticity. However the online medium affords liberating possibilities for representing the actual form of a signed language and organizing content in ways that offer versatility in the ways that first and second language users can interact with it. The Online Dictionary of New Zealand Sign Language (ODNZSL), launched in 2011, is an example of a contemporary sign language dictionary that leverages the technical advantages of a digital medium and an existing body of descriptive research, including a small electronic corpus of NZSL and a linguistically informed print dictionary (Kennedy et al 1997). These 21st century conditions have enabled the production of a bi-directional, multimedia dictionary that offers an enriched experience for users (compared to print dictionaries), boosts the public accessibility of NZSL, and serves as a dynamic repository of lexical and cultural reference material about the NZSL community.

Our presentation will introduce the principles that underlie design of entries and search features of the ODNZSL and other modern sign language dictionaries, and discuss some challenges in its making. Although the ODNZSL database rests on a previous dictionary, our editorial team encountered lexicographical challenges in revising, migrating and expanding dictionary content to its new online version. Some of these challenges stem from the modality and usage conditions of signed languages, while others will be familiar to those working in languages that are non-written, and/or of limited diffusion. Decisions were required about the treatment of sociolinguistic variants, and, in a dictionary with video content, variation issues extended to managing the embodied representation of language forms by identifiable members of the NZSL community. Like previous sign language lexicographers, we also grappled with determining lexemes, citation forms and word class in a language that is prone to synthesising localised meanings that are not strictly lexicalised. Keeping in mind two main user groups (NZSL 'natives', and native English speakers learning NZSL), decisions had to be made about approximating equivalence in English glosses, and deriving authentic, yet comprehensible, usage examples from a small corpus.

User feedback on ODNZSL to date indicates that it is technically user-friendly, appealing in terms of production values and informational content, and is having a social impact by making NZSL accessible to user groups (such as families, workmates and other associates of Deaf people) who previously would not have found an accessible reference tool. Although language learning requires far more than dictionaries, expanding public access to NZSL in this form obviously contributes directly to Deaf people's opportunity to communicate with others in society, and to the recognition of their language identity.

## KEYNOTE PRESENTATION II

### Gabonese French Dictionaries: Survey and Perspectives

**Prof P.A. MAVOUNGOU**

Department of Language Sciences, Omar Bongo University, Gabon

This paper is a study of lexicographic activities involving the French language as it is spoken in Gabon, a francophone country in central Africa. The modern era of Gabonese lexicography has witnessed the production of various dictionaries focusing on this African variety of French. Before 1999, due to the status of French as the sole official language of Gabon, French dictionaries used there were mostly French products. Monolingual French dictionaries produced in Gabon by Gabonese scholars are very limited, and despite the fact that they were compiled for use by Gabonese, they are what is known as ‘differential dictionaries’ as they tend to include references to Parisian standard French. A differential dictionary is a dictionary focusing on differences between two language varieties or two similar languages, and all current Gabonese French dictionaries are written this way. It is this situation, and the ways it can be corrected that this paper will focus on.

The idea is to produce a complete dictionary of Gabonese French. Such a dictionary will reflect a clear emancipation from the French language spoken in France (i.e. Parisian standard French). For reasons of national sovereignty, I plead for the production of a complete dictionary of Gabonese French: Gabonese lexicographers and linguists do not need to define the lexicon of their language differentially.

At present, four reference works have been published on the Gabonese French variety; Boucher (1999), Dodo-Bounguendza (2008), Ditougou (2009) and Moussounda Ibouanga (2011). The central list of Boucher’s work comprises a set of lexical items as used by the younger generation of Libreville inhabitants, aged from 15 to 30. Fair attention is given to different style or normative levels, that is, standard French, official French (acrolectal level), common French (mesolectal level) and popular French (basilectal level). Dodo-Bounguendza’s dictionary encompasses macrostructural elements taken from common French (mesolectal level) with attention being paid to Gabonisms, that is, Gabonese-specific lexical creations. The same lexical creations comprise the central list in Ditougou (2009) with particular pictorial and contextual illustrations. As was the case for Boucher (1999), Moussounda Ibouanga’s dictionary is based on fieldwork conducted among the youth of Libreville in order to determine different sociolinguistic styles or normative levels of French. It is an update of Boucher’s work.

This paper argues that Gabonese French bears the features of a national variety like other national varieties such as Belgian French, Canadian French, Ivorian French, Haitian French and Senegalese French. In fact, the “Gabonisation” of the French language has reached a significant point and it is no longer a foreign language per se. The paper subsequently reviews the existing dictionaries in this “Gabonese language”, which by any reckoning appears to be the national language of Gabon and has acquired a significant place in the Gabonese language landscape (Ndinga-Koumba-Binza 2011, 2007 & 2005). Next, attention is given to concrete projects that are currently being undertaken in order to improve the development of the variety of the French language spoken in Gabon. Two projects have to be mentioned here: (i) the compilation of a dictionary of collocations and idiomatic phrases (Mavoungou & Moussounda Ibouanga, forthcoming) and (ii) the production of a general monolingual dictionary of Gabonese French

(Mavoungou 2008-2009). The latter project will require input from all Gabonese trained lexicographers as well as the expertise of linguists involved in lexicographic activities. This paves the way to the recognition of an emerging branch within modern Gabonese lexicography, that is, Gabonese French lexicography.

## PARALLEL SESSIONS

### Somali Spelling Correction for Native Speakers of English

**Nikki ADAMS**

University of Maryland-College Park, USA

It is well-known that foreign language learners have difficulty accurately perceiving sounds in the languages they learn. This is as true for English-speaking learners of Somali as it is for any other language pair. Phonemic distinctions made in Somali but not in English include vowel length, a phonemic glottal stop, and a series of voiced and voiceless velar, uvular, and pharyngeal stops and fricatives. This has obvious implications for the ability of students of Somali to find words in dictionaries. The aim of this project is to apply to an electronic dictionary a system of spelling correction that is based on the types of errors English speakers are likely to make. This spelling correction system is that which was developed by and discussed in Rytting et al (2011) for English-speaking learners of Arabic, here modified for Somali. We aim to show that such a system suggests the correct spelling of a misspelled word at higher accuracy than a simple Levenshtein-based weight system (where all insertions, deletions, and substitutions are equally likely).

For online Somali dictionaries, if a word is misspelled, either no results are returned or results for which the entered word is a substring may be returned. As an example of the latter, if a user misspells *walaal* “sibling” as *walal*, it will return *walalac* “twinkling”, but not *walaal*. As with the Arabic version, however, the Somali specific system not only returns *walaal*, but also produces the desirable result of listing it as a more likely correction than, for example, *walqal* “christening.” Both require exactly one substitution, but this spell checker recognizes that ‘aa’ is a more likely substitution for ‘a’ than ‘q’.

### Methodology

A confusion matrix based on the hypothesized likelihood of one letter or digraph being mistaken for another was created based on feedback from learners of Somali, and linguistic knowledge of the general closeness of any two sounds. It therefore takes into account both confusions based on sounds English speakers have trouble discerning in Somali, for example the voiceless uvular fricative versus the voiceless pharyngeal fricative, and sounds that English differentiates but which are, nevertheless, phonetically close, for example a bilabial nasal and an alveolar nasal. Beyond these likely confusions, any substitution was allowed for at a cost higher than any of the other specified confusions, for example, ‘b’ can be inserted or deleted anywhere at a high cost. Another confusion matrix was created where all substitutions and deletions were given equal weight, the Levenshtein-based system against which ours was compared.

Dictionary Form. Lower Weight = More Likely Confusion														
User Input	Weight 1			Weight 2			Weight 3			Weight 4				
a	aa			e	y		i	o	u	ee	w	oo	ii	uu
aa	a			ee			ii	oo	uu	e	o	i	u	
b	bb						m							
bb	b						mm	m						
c	<>	<>	x	kh							h			
d	dd	dh		t	r		l							

The results reported here are from an initial trial, with expansion planned in the upcoming months. Three beginner students of Somali participated in a dictation exercise where they listened to audio recordings of approximately 25 Somali sentences a maximum of 4 times and typed the sentences. Example errors to test were taken from this dictation exercise. A small Somali wordlist (roughly 3,500 unique words, to be expanded shortly) was collected from a variety of Somali language web pages. This represented the lexicon in which words were searched.

A finite state transducer was created from the confusion matrix as described above. The mistakes were run against our Somali-specific spell corrector and the Levenshtein-based one, with the top 3 results returned from each. As these students' Somali knowledge was just at a beginning level, and they did not have much vocabulary, many of their errors involved incorrect word boundaries. For the purpose of demonstrating the system, their word boundaries were corrected, but no examples were used where errors may have been caused by the lack of knowledge of the boundary, that is, those errors which occurred at word boundaries and which were not replicated frequently throughout their dictation exercise. Some example return results are shown below; more detailed results, including the percentages for which our spell-checker versus the Levenshtein-based one return the correct word in the top n results, will be prepared shortly.

**Examples of Preliminary Results from Somali-Specific Spelling Corrector (SSSC) and Levenshtein-based Spelling Corrector (LBSC)<sup>1</sup> (with total costs in parentheses)**

Correct form <i>meelood</i> mistyped as <i>meelot</i>		Correct form <i>joog</i> mistyped as <i>juq</i>		Correct form <i>ukun</i> mistyped as <i>oqun</i>	
SSSC	LBSC	SSSC	LBSC	SSSC	LBSC
meelood (3)	meelo (1)	joog (4)	deeq (2)	ukun (2)	aqoon (2)
meelo (5)	meel (2)	suuq (5)	dhaq (2)	ogaan (5)	kun (2)
meel (9.5)	meelood (2)	uk (6)	. . . j<oo>g (2)	kun (5.5)	. . .ukun (2)

<sup>1</sup>The LBSC returns, on average, more results of the same weight for shorter words. For example, for *ukun* misspelled as *oqun*, there were 6 results returned as the same cost. When this happens, I list the results in alphabetical order, but skip where needed to show the correct result if it was returned in the top 3 weights.

In sum, we (i) describe the confusions English-speaking learners of Somali are likely to make and the rationale for assigning the various weights to these confusions, (ii) describe briefly how the language-specific spell corrector works, (iii) show the results of our Somali-specific spell corrector compared to a Levenshtein-based one, and (iv) show that ours produces more accurate and informative results.

**Reference:**

C. Anton Rytting, David M. Zajic, Paul Rodrigues, Sarah C. Wayland, Christian Hettick, Tim Buckwalter, Charles C. Blake. (2011) "Spelling Correction for Dialectal Arabic Dictionary Lookup." Association for Computing Machinery Transactions on Asian Information Processing. 10(1). March 2011.

## Terminology as an aid in mother-tongue education

Dr Mariëtta ALBERTS

Northwest University, South Africa

This paper deals with the value of mother-tongue/first language education and the role of terminology in the teaching of subject matter.

There are several arguments underpinning the value of mother-tongue education. It is, however, probably not possible, currently, to provide mother-tongue education in South Africa from the primary phase through to the tertiary phase of education. There are several reasons for this situation, but this paper will focus on the lack of terminology in the various subject-related matters.

Afrikaans is an example of a language that was initially regarded as a “kitchen language” and which developed into a functional language in *inter alia* science, technology, economy and education. There is therefore no reason why South African indigenous official languages could not develop into languages of higher learning. The hegemony of English results in the indigenous languages often struggling with terminology development. Most terms denoting new concepts reach South Africa through the medium of English and special efforts need to be made to translate these terms into the various official languages. The terms that are available in these languages also need to be standardized to ensure exact communication by subject specialists, language practitioners, the media and lay people. It frequently happens that the same concept is denoted by means of more than one term equivalent and this leads to communication problems. One could therefore argue that, although mother-tongue education is preferable, it is not necessarily achievable.

This paper will focus on the terminology of mathematics since many learners battle with the subject. It is also evident that teachers have difficulty coping with this subject area, cf. “*Teachers battle with simple mathematics...*” (SAPA 2011-04-04; *News24* 2011-04-04). It is true that mathematics contains difficult subject matter to be explained and comprehended. What makes this issue even worse is that this subject is often taught in many learners’ second or third language.

It is widely recognised that concepts are best understood in one’s first language or mother tongue. It is also easier to convey information (e.g. teach or train) through the medium of the first language of both the teacher and the learner. It is therefore essential to internalise information in one’s first language before switching to an additional language. As a result, education should be in the mother tongue in the primary, secondary and tertiary educational phases. It is easier to switch to another language (e.g. English) later in life, for example in the work environment.

Terminology is a tool for communication in languages for special purposes, for example subject areas (mathematics, physics) and domains (art, music, sport). Terminology allows for communication between subject specialists and between subject specialists and lay people. Specialised information is encoded to be conveyed from a sender (communication source) to a receiver (recipient of information) who will decode the message and then react on the stimulus received.

Terminology uses the grammar of the standard language. Terms are created according to the word-forming principles of the standard language. Term creation adheres to the spelling and

orthography rules of the standard variety of the language. Terms are documented and systematised in order to compile technical dictionaries, terminology lists, or electronic term banks. The primary aim is to promote communication in scientific and technological environments.

Specialised terms are needed that:

- provide access to texts in other languages, e.g. source language to target language
- have exact meaning, with no emotional connotations
- function within a conceptual system
- are standardised
- are harmonised in related (inter)national languages
- are harmonised in related South African languages

Terminology development adheres to the language policy of the country. In the previous dispensation, with a bilingual policy, terms were supplied in English and Afrikaans. Since 1994 South Africa has had a multilingual policy and therefore terminology should be available in the 11 official languages. Terms should also be available in South African Sign Language – in this case using signs to denote the underlying concepts. Concepts should be defined in the source language and terms to denote the concept should be created in that language (process: primary term creation). Only then term equivalents could be supplied in target language(s) (process: secondary term creation).

Terminology should be standardised to allow exact communication. The standardisation process entails that the majority of subject specialists should agree to denote a definite concept with a specific term through:

- the limiting of meaning by means of a proper definition – one concept one term
- systematisation and usage of language rules e.g. word-forming principles, spelling and orthography rules
- standardisation of terms through the approval by National Language Bodies, and the dissemination, usage and penetration of terms.

This paper will discuss various options to resolve problematic issues regarding terminology development of subject-related matter which could benefit mother-tongue education.

### **Access to data in lexicographic tools from the perspectives of lexicographic and information seeking concepts**

**Henning BERGENHOLTZ**

Centre for Lexicography, Aarhus School of Business, Denmark & Dept of Information Science, University of Pretoria, South Africa

**Theo JD BOTHMA**

Department of Information Science, University of Pretoria, South Africa

Information is abundantly and ubiquitously available in the information society. Because of databases and electronic information tools, it is much easier for someone to find relevant data when they are seeking information. In lexicography we speak about the *access process* (Bergenholtz/Gouws 2010). It can be shown that this process begins with the *origin of the*

*problem* and then leads to an *information source usage situation*; the whole access process contains at least 19 different parts, and each part can contain different steps. The most important terms here are *access route* and *access time*, but not *access structure*. The latter term is not only ill defined – it is less relevant, because each user has different access routes, and also the same user makes use of different routes in different cases. But besides that, it is clear that lexicography has the main concepts of information need ⇒ access process ⇒ searching. The whole list of different parts of the access process, given here, could eventually be expanded, but the table below lists the parts currently described in lexicographic terms in the sequence of the access process:

1. Origin of the problem	11. Combined search strings
2. Information source usage situations	12. Access by means of an alphabetical macrostructure
3. Choice of information source	13. Access by means of a systematic macrostructure
4. Choice of the component of an information source	14. Index access
5. Consultation of an information source	15. Search in a part of a component
6. Search string	16. Search route
7. Search option	17. Search step
8. Situation-oriented access	18. Search speed
9. User type-oriented access	19. Search time
10. Accuracy of the access and the data presentation	

**Table 1: The different steps in the access process, as listed in Bergenholtz/Gouws 2010**

Some parts of the access process listed above are equally relevant to paper-based information sources (e.g. index access) but some are unique to the electronic environment (e.g. *combined search strings*). This paper, however, focuses mainly on the electronic environment.

In information science the information seeking process is embedded in the broader term *information behaviour*. This process is in essence the same as in lexicography, but there are smaller or larger differences in terminology. Some of the terms are the same, for example *information need*, *information source*, and *search string*. The term ‘search’ in lexicography is, however, used differently in information science and it includes the information science concepts of *searching*, *browsing* and *navigating*. Searching, in information science, implies that the user has to define a search string (which could consist of a single word, a phrase, or a number of words/phrases combined by means of Boolean operators, that is, it requires a specific intellectual input from the user to identify the specific words/phrases and to combine them in the correct way (either in a command line interface, or through drop-down menus). Navigating and browsing are also distinguished; both imply following predefined links provided to the user (either by the system or by the original author), but with the difference of intentionally following a path to find specific information, in contrast to serendipitously following links without necessarily knowing where this will lead. Finding the required relevant information often requires, in information science terminology, a combination of searching and navigating, but each aspect of the process is clearly defined.

The aim of this paper is to compare and contrast the access process terminology of lexicography and information science and to show how these disciplines complement one another by providing clearer definitions to describe the different access processes. The definitions will be illustrated with examples from e-dictionaries, other information tools and general web environments.

## Reference:

Bergenholtz, H. & Gouws, R. H. 2010. A New Perspective on the Access Process. *Hermes. Journal of Language and Business Communication* 44: 103–127.

## What is a dictionary?

**Henning BERGENHOLTZ**

Centre for Lexicography, Aarhus University, Denmark

Some words have a clear and unambiguous meaning. No user with a given language as mother tongue will therefore ever have reception problems and will therefore not need to use a dictionary when finding this word in a text. This situation was the reason for an earlier edition of a Danish dictionary not even to give any definition for such words. One of those words was *ordbog* (*dictionary*). Now this dictionary in a new edition has changed the approach and gives definitions for all lemmas, also the lemma *dictionary*, where it gives a traditional definition with reference to some items necessary for being a dictionary:

*dictionary* a book with words in a systematical order and with information about the spelling, word class, inflexion and meaning of the words or their translation to another language

This definition and those in most other dictionaries, handbooks and theoretical contributions can easily be criticised. You do not find word class items in every dictionary, you do not find meaning items in every dictionary, and so on. The only kind of information you mostly find is implicit spelling information, giving the lemma in at least one of its spelling variants. But we have dictionaries without lemmas, see Bergenholtz (2011). Therefore, no item type is to be found in every kind of dictionary. Many definitions claim – but not the quotation from the Danish dictionary – that a dictionary has an alphabetical macrostructure. It is of course not the case; they are not ordered systematically at all in many cases, at least not in a way that the user can see. Many electronic dictionaries have no macrostructure at all. If you search with a certain search string you often get exactly one dictionary article. In some electronic dictionaries the alphabetical order of neighbouring words is listed, but that is not what we normally understand as macrostructure, that is an access structure. Another problem in recent definitions is the attempt to clarify borderlines between dictionary, lexicon, encyclopaedia, word list, thesaurus, etc. It can be done in theory. The naming of concrete lexicographical tools does not follow clear rules; a lexicographical tool is often called a dictionary or lexicon or another name if the dictionary company believes that such a name will sell better. In addition it must be mentioned that dictionary, lexicon, encyclopaedia, thesaurus etc. are used differently in different languages.

In some cases we find almost empty discussions about classifications or exact definitions. If we select this or another classification or definition it does not influence practice or decisions anyway. That is not the case for the definition of *dictionary*. We have a bibliography problem in the case that we have one database, but various dictionaries extracted from this database, for example in the following case: in a list of dictionaries you find the dictionary name "Musikordbogen" (the music dictionary), but if you go to this "dictionary" you learn that it is not one, but four dictionaries:

Inger Bergenholtz i samarbejde Henning Bergenholtz.

Database: Richard Almind og Martin Gyde Poulsen: Betydning af musikudtryk.

Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2555-1)

Inger Bergenholtz i samarbejde Henning Bergenholtz.  
Database: Richard Almind og Martin Gyde Poulsen: Viden om musikudtryk.  
Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2299-4)

Inger Bergenholtz i samarbejde Henning Bergenholtz.  
Database: Richard Almind og Martin Gyde Poulsen: Find et musikudtryk.  
Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2556-8)

Inger Bergenholtz i samarbejde Henning Bergenholtz.  
Database: Richard Almind og Martin Gyde Poulsen: Fremmedordbog med musikudtryk.  
Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2557-5)

How can we then quote all four dictionaries in one literature item? Or can't we? But this practical question implies decisions to be made for the list of research results at individual universities. If the government pays a certain amount of money for each new book, would and should they if not only four, but 23 dictionaries can be extracted out of one and the same database? We think they should, because a database is not a dictionary.

We use in our proposal *dictionary* as the common name for all kinds of lexicographic tools. We make proposals too for parts of a typology of dictionaries as mentioned above. But the most important problem in recent definitions is that the word *dictionary* has not just one meaning, but at least two:

1. for a single lexicographical tool
2. for a collection of lexicographical tools

We will, in this contribution, give a critical overview of types of definitions found in different information tools and theoretical contributions, and connect this discussion to new definitions in a way that the same definition can be applied equally to paper and to electronic dictionaries.

## **The Potential of a Diagnostic Text Evaluation Model for Dictionary Criticism**

**Herman BEYER**

Dept. of Language and Literature Studies, University of Namibia

A study of dictionary reviews reveals that there seems to be no systematic or universal approach to evaluating dictionaries. In one particularly weak review that was studied, the reviewer all but condemns the dictionary based virtually on a singular and not so critical deficiency, and a close reading suggests that the reviewer indeed seems to have a personal axe to grind with the lexicographer. Such a review cannot be regarded as very scientific. In this case it can be argued that another reviewer might have reached a vastly different conclusion about the dictionary under review, even (and especially) if the approach had also been primarily subjective.

A communicative metalexigraphy benefits from existing communication theory and, as an extension, from existing theory of textual communication. When it comes to dictionary criticism, the dictionary is approached as a product of textual communication within the broad framework of a lexicographic communication model. This approach supports the notion that dictionaries can be regarded as carriers of text types, as proposed by Wiegand's theory of lexicographic texts (cf., e.g. Wiegand 1996).

In the theory of textual communication a model for text evaluation has been developed. This model, of which the applicability extends to practical editing, was developed by the Dutch text scientist Prof. Jan Renkema and named the *CCC model*. The three Cs stand for *Correspondence*, *Consistency* and *Correctness* as the main evaluation criteria, and the model is widely utilised for evaluating text quality (cf. Carstens & Van de Poel 2010).

This paper investigates the extent to which the CCC model can be adapted to become a model for the evaluation of dictionaries and perhaps even for the evaluation of dictionary evaluations, that is, for dictionary criticism. It is probably impossible that dictionary criticism would not contain any subjective judgements whatsoever. A systematic model of criticism might however contribute to evaluations based to a lesser extent on subjective (sometimes even idiosyncratic) criteria and more on a framework of reasoning that is perhaps more scientifically accountable.

The CCC model's criteria of Correspondence, Consistency and Correctness are each aligned to five characteristics of texts, namely, text type, content, structure, formulation and presentation, as in Table 1 below:

**Table 1: The CCC Model for Text Evaluation (cf. Carstens & Van de Poel 2010: 63)**

	Correspondence	Consistency	Correctness
Text type	1	2	3
Content	4	5	6
Structure	7	8	9
Formulation	10	11	12
Presentation	13	14	15

The representation of the CCC model in the grid in Table 1 illustrates a resulting 15 measures of text quality. For each measure a number of questions can be posed that can collectively form a type of checklist according to which the quality of a text can be judged. It is characterised as a *diagnostic* model, as it seeks to identify and contribute to the correction of text deficiencies.

The CCC model is used for evaluating business letters, brochures, reports, etc., but its potential value for the discipline of dictionary criticism has not been explored, although Carstens & Van de Poel (2010: 68) lists dictionaries as a type of informative referential text that could be evaluated by using the model. This paper represents a first attempt at interpreting the model for application to the evaluation of dictionaries. This is done as part of a research project currently underway at and funded by the University of Namibia that aims to identify evaluation criteria for school dictionaries for Namibian schools.

The exposition will show how the main evaluation criteria (Correspondence, Consistency and Correctness) could be interpreted if the text characteristics pertain to dictionaries, and which questions could be posed for each of the 15 measures.

**References:**

Carstens, W.A.M. & K. van de Poel. 2010. *Teksredaksie*. Stellenbosch: SUN Press.  
 Wiegand, H.E. 1996. A theory of lexicographic texts: an overview. *South African Journal of Linguistics* 14(4): 134-149.

## Lexicography and the relevance criterion

**Theo JD BOTHMA**

Dept of Information Science, University of Pretoria, South Africa

**Sven TARP**

Centre for Lexicography, Aarhus University, Denmark

Until recently, lexicography and information science could rightly be considered two disciplines which had developed along parallel lines but with no or very little formal relation between them. Although the two disciplines developed in almost completely isolation from each other, during the last few years it has nevertheless become increasingly clear that they have a lot in common. This trend began within lexicography which started viewing lexicographical works as a special kind of tools designed to be consulted in order to obtain information. Upon this basis, it has been suggested that lexicography should be considered a part of information science and, hence, integrated into it (cf. Tarp 2007, 2009, 2010, Leroyer 2011). It is evident that this integration of two hitherto independent disciplines with long traditions of their own is not something to be solved overnight and neither can it be a unilateral process. Tarp (2011: 56) formulates the challenge as follows:

[...] lexicography, on the one hand, *has a lot to contribute* to other theories dealing with punctual consultation tools and to information science in general, and on the other hand, *has a lot to learn* from these theories and this science.

In this paper we will discuss the concept of relevance which is relevant to both disciplines. This will be done within the framework of the function theory of lexicography, as discussed in the many works of Tarp and Bergenholtz (e.g. Tarp and Bergenholtz 2002) and others, and relevance theory in information science as defined by Saracevic (1975, 1996), Cosijn and Ingwersen (2000) and others.

According to the function theory of lexicography, the types of information needs relevant to lexicography should never be considered abstract needs, but specific and even concrete needs which are determined by the types of potential users of lexicography's practical works as well as the types of extra-lexicographical situations where lexicographically relevant information needs may occur. This also means that user needs should always be understood in their intimate relation to a specific type of situation (i.e. the context of the information need) and that they can never be defined only based upon the characteristics of a potential user however relevant these characteristics may be. At this stage, the function theory works with four fundamental types of lexicographically relevant situations (cf. Tarp 2008a), viz. communicative, cognitive, operative and interpretive situations.

What the function theory claims is that the relevant user characteristics depend on the situations where the relevant needs occur. To this end, the function theory has elaborated a list of lexicographically relevant user characteristics – an open list to which new characteristics can be added – from which the characteristics that are relevant to each type of extra-lexicographical situation can be selected. In this respect, the function theory works with a set of variables that have to be taken into account when determining the specific type of information needed in each case (cf. Tarp 2008b).

According to the function theory, the user's information need is met by a corresponding set of lexicographical data which have been selected, elaborated and prepared (by the lexicographer)

and made accessible for consultation. The selection, elaboration and preparation of the data may be performed following various techniques and methods but in order to produce high-quality lexicographical works it is necessary that these techniques and methods in the final analysis are built upon the criteria of relevance determined by the specific type of information need in question.

When an individual person experiences an information need, he or she may then access the data and retrieve the needed information from these data through a complex cognitive process. Finally, the information retrieved may be used in various ways in order to satisfy the user's specific type of need, e.g. to solve a communicative or cognitive problem, to store it as knowledge, to perform a task etc.

In all these cases, relevance is centered on the information needs and their satisfaction and combines extra- and intra-lexicographical elements. However, it must be admitted that nowhere in the lexicographical literature known to the authors of this contribution, it is possible to find a definition or a more extended discussion of the criterion of relevance in terms of lexicography.

In information science research, Saracevic has, already in 1996, stated that “[n]obody has to explain to users of IR [Information Retrieval] systems what relevance is, even if they struggle (sometimes in vain) to find relevant stuff. People understand relevance intuitively” (Saracevic 1996). However, Saracevic and others (e.g. Borlund (2003), Cosijn (2003), Cosijn and Bothma (2005)) agree that there are many dimensions to relevance. Two basic classes of relevance are distinguished, viz. objective relevance which is system-based (system relevance) and subjective relevance which is user-based (which may then be subdivided into, amongst others, cognitive and situational relevance).

Even from this brief discussion it is evident that there are points of contact and difference between lexicography and information science. This paper will explore the concept of relevance in both disciplines in more detail and show, at the hand of examples from lexicographical tools, how the theoretical frameworks of both disciplines can complement one another.

## References:

- Bergenholtz, Henning and Sven Tarp.** 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 18: 253-263.
- Borlund, Pia.** 2003. The concept of Relevance in IR. *Journal of the American Society for Information Science and Technology* 54(10):913-925.
- Cosijn, Erica.** 2003. *Relevance Judgements in Information retrieval*. Unpublished DPhil thesis. Pretoria: University of Pretoria.
- Cosijn, Erica and Theo Bothma.** 2005. Contexts of Relevance for Information Retrieval System Design. Fabio Crestani & Ian Ruthven (Eds.): *Context: Nature, Impact, and Role. 5<sup>th</sup> International Conference on Conceptions of Library and Information Sciences, CoLIS 2005 Glasgow, UK, June 4-8, 2005*: 47-58. Berlin & Heidelberg: Springer.
- Cosijn, Erica and Peter Ingwersen.** 2000. Dimension of Relevance. *Information Processing and Management* 36: 533-550.
- Leroyer, Patrick.** 2011. Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographical Information Tools. Pedro A. Fuertes-Olivera & Henning Bergenholtz (Eds.): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 121-140. London & New York: Continuum.
- Saracevic, Tefko.** 1975. Relevance: A Review of and Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science and Technology* 26(6): 321-343.
- Saracevic, Tefko.** 1996. Relevance reconsidered '96. Peter Ingwersen and Niels Ole Pors (Eds.): *Integration in Perspective*. Proceedings of CoLIS2, Second International Conference on Conceptions of Library and Information Science, Copenhagen, 13-16 October 1996: 201-208.

- Tarp, Sven.** 2007. Lexicography in the Information Age. *Lexikos* 17, 170-179.
- Tarp, Sven.** 2008a: Revival of a Dusty Old Profession. *Hermes. Journal of Linguistics* 41: 175-188.
- Tarp, Sven.** 2008b. *Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner's lexicography.* Tübingen: Niemeyer.
- Tarp, Sven.** 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Henning Bergenholtz, Sandro Nielsen & Sven Tarp (Eds.): *Lexicography at a Crossroads: Dictionaries and Encyclopedias today, Lexicographical Tools tomorrow:* 17-32. Bern: Peter Lang.
- Tarp, Sven.** 2010: Reflections on the Academic Status of Lexicography. *Lexikos* 20: 450-465.
- Tarp, Sven.** 2011: Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Pedro A. Fuertes-Olivera & Henning Bergenholtz (Eds.): *e-Lexicography: The Internet, Digital Initiatives and Lexicography:* 54-70. London & New York: Continuum.

## **A corpus of South African English for diachronic lexicographical research**

**Richard BOWKER**

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

This paper considers a number of key issues in the rationale for, and the design, construction and utility of a corpus of South African English (SAE) for diachronic lexicographical research at the Dictionary Unit for South African English (DSAE).

In terms of rationale, while a large amount of material exists on the Internet and is accessible and searchable for lexicographical purposes (e.g. researching new lexical items and building senses for these, and antedating existing headwords and researching changes in senses of these) even with the use of tools such as Google Books and Google Scholar, the pages of the worldwide web function as an archive rather than a corpus – there remains much ‘noise’ that detracts from accurate representations of the frequency and collocations of relevant lexical items. As importantly, there remains much material not accessible online that is as yet unresearched by the DSAE as it works toward the second edition of *A Dictionary of South African English on Historical Principles*. With this in mind, the compiling of the corpus will initially focus on digitising and processing material that is not yet available electronically.

Furthermore, while small corpora of sub-varieties of SAE (e.g. Indian South African English, Xhosa South African English) have been compiled in the last decade or so, lexicography generally requires very large corpora in order to analyse language adequately – and no large corpus has yet been constructed to investigate broad patterns in the use of SAE (and its sub-varieties). A large monitor corpus of this nature is particularly relevant considering the low-frequency of SAE in the English used in South Africa, and the complex, predominantly multilingual environment in which SAE is found.

In terms of the design of the corpus, the paper will discuss the use of established corpus design criteria (e.g. Atkins, Clear & Ostler 1992) as well as the possibility of using more recent developments such as agile corpus creation (cf. Voormann & Gut 2008) and the relative contributions these can make to the compilation of a corpus of SAE. The kinds of annotations required to optimally query the corpus will be outlined. These are annotations of both linguistic and extra-linguistic criteria – examples of the latter include, at the level of the text producer, the biographical information available on the authors and whether or not English is their first or subsequent language; and, at the level of the document, its source, date of publication, and so on. It is argued that with a requisite degree of sophistication in the annotation system, the need for corpora of sub-varieties of SAE can be obviated as these can all be researched within a unitary

corpus structure which can foreground, among other things, the differences and relations between them. At later stages in its development the corpus would need to be flexible enough to incorporate different kinds of material, such as transcripts of oral material where there is more than one speaker.

Regarding the construction of the corpus, an outline is given of a pilot project that is being conducted by the DSAE and Rhodes' Department of Linguistics in the course of 2012 and how the knowledge gained for this smaller project will help inform the construction of the broader monitor corpus of SAE. A sketch of the kinds of material already available for insertion into the corpus is given, as are strategies for the ongoing accessing of further material. How long-standing issues in corpus creation (e.g. balance and representativeness) will be dealt with is described. In addition, points of potentially useful intersection between the DSAE and other departments at Rhodes University (and further afield) – such as software development, use of the corpus (or sub-corpora thereof) across research projects and multidisciplinary projects– are outlined. That is, it is envisaged that the ongoing project will provide datasets for research by a number of disciplines beyond lexicography.

Lastly, what we expect to be able to do with a corpus of SAE is delineated. Among the functions envisaged are, for example, searches and analysis of the data by chronology, genre, region, sub-variety, domain and gender.

#### **References:**

- Atkins, S., Clear, J. & Ostler, N. (1992) 'Corpus design criteria' in *Literary and Linguistic Computing* 7(1)  
Voormann, H. & Gut, U. (2008) 'Agile corpus creation' in *Corpus Linguistics and Linguistic Theory* 4(2)

#### **Bantu Forensic Lexicography: a Kikongo case study**

**Jasper DE KIND, Gilles-Maurice DE SCHRYVER & Koen BOSTOEN**

KongoKing Research Group, Department of Languages and Cultures, Ghent University

In this paper, the oldest Bantu dictionary hitherto known is explored, that is the *Vocabularium Latinum, Hispanicum, e Congense*, handed down to us through a manuscript from 1652 by the Flemish Capuchin friar Joris van Gheel, missionary in the Kongo (present-day north-western Angola and the southern part of the Lower Congo Province of the DRC). The date of this dictionary implies that Bantu lexicography is much older than has been generally accepted. Benson (1964) stated that it was only a century old, and nowadays, Bantu lexicography still seems to be considered only 150 years old. The *Vocabularium Congense* pushes this date back to 360 years ago.

The dictionary was compiled at the Capuchin Mission in northern Angola in the 17<sup>th</sup> century, and although the handwriting is clearly Van Gheel's, it is most likely that the Flemish priest copied it from another Capuchin. Moreover, it is very probable that the vocabulary list was compiled with the aid of a native speaker of Kikongo, namely Manuel Roboredo, a Kongolese Capuchin. The precise dialect that is described in the dictionary cannot yet be pinpointed with sufficient linguistic evidence, but as we show in this paper, it is very likely the Kisikongo variant spoken in the vicinity of Mbanza Kongo, the former capital of the Kongo Kingdom. Since this handwritten manuscript is hard to read, Joseph Van Wing and Constant Penders, two Belgian Jesuits, reworked the text in 1928, producing a decidedly more user-friendly version. However, considerable changes were executed with respect to the original manuscript. First, they changed

the translation direction, from Latin/Spanish ⇔ Kikongo, to Kikongo ⇔ French/Flemish, replacing the Latin/Spanish with French/Flemish. Second, they adjusted the Kikongo orthography used in the original to the modern (i.e. early 20<sup>th</sup> century) orthography. As a consequence, the phonological changes Kikongo underwent from the 17<sup>th</sup> to the 20<sup>th</sup> century were completely obscured. Moreover, they suppressed all diacritic marks, of which the precise meaning has not yet been retraced. In other words, this paper argues that the dictionary Van Wing and Penders compiled does not reflect the Kikongo variant spoken in the 17<sup>th</sup> century, but should be considered a new dictionary, reflecting the early 20<sup>th</sup> century dialect.

The original manuscript is highly interesting in a variety of ways. Through a comparison with later sources, it allows for the study of the phonological and morphological changes the Kikongo dialect underwent from the 17<sup>th</sup> century onwards, and it reveals how the Capuchin missionaries compiled their vocabulary lists. The *Vocabularium Congense* — and the vocabulary list(s) it was copied from — are likely based on a Latin ⇔ Spanish dictionary of Antonio de Nebrija. This inevitably implies that Kikongo equivalents were sought for each Latin lemma, regardless of the absence of these concepts in the Kongo language. As a result, the need for neologisms was high. Various techniques were used, such as extending meanings of existing Kikongo concepts, the use of paraphrases and also, unsurprisingly, the frequent use of loanwords. On the other hand, misnamings also occur, as European concepts were wrongly attributed to existing Kikongo terms. For instance, the European concept of ‘wolf’ was designated with the Kikongo term for ‘hyena’. This requires one to tread with caution when using the manuscript for yet another interesting linguistic/historical discipline, namely the ‘Words-and-Things Method’, by which a population’s history is reconstructed through their language (and more specifically their lexicon).

The lexicographical structures of both dictionaries – Van Gheel’s (1652) manuscript, and Van Wing and Penders’ (1928) reworking of it – have been studied in detail. When contemplating the re-edition of 1928, one notices an advanced lexicographical structure, complete with combinatory aspects, example slots, as well as grammatical notes. The Kikongo lemmata are thus often provided with extra information on their use and meaning. This material cannot always be traced back to the original manuscript of 1652, where the structure is also more linear. Occasionally the Capuchins did provide extra information (in Latin), over and above mere translation equivalents, but overall it seems as if Van Wing and Penders truly added material. A comparison of the two works also reveals that both convergence as well as divergence were employed in reversing out the data, in addition to ‘nonvergence’. These features will be amply illustrated during the presentation.

Currently, both the original manuscript and the re-edition are being digitized, linked and added to an interlingual and multimedia database that revolves around Kikongo and the early history of the Kongo kingdom. As such, it will be used (amongst others) to study Kikongo language change and the history of its speakers.

### References:

- Benson, T.G. 1964. A Century of Bantu Lexicography. *African Language Studies* 5: 64-91.
- Van Gheel, J. 1652. *Vocabularium Latinum, Hispanicum, e Congense*. Manuscript kept in the National Central Library of Rome: Fondo Minori 1896, MS Varia 274.
- Van Wing, J. and C. Penders. 1928. *Le plus ancien dictionnaire bantu / Het oudste Bantu-woordenboek. Vocabularium P. Georgii Gelensis*. Louvain: J. Kuyt-Otto.

## **Establishing a dictionary culture inside the classroom for better dictionary use outside the classroom**

**Rufus H. GOUWS**

Department of Afrikaans and Dutch, University of Stellenbosch

**Danie J. PRINSLOO**

Department of African Languages, University of Pretoria.

The speech communities of many of the languages of South Africa and also of Africa have had a number of lost generations with regard to knowledgeable dictionary use. This is due to the lack of a prevailing dictionary culture within these speech communities. In theoretical lexicography a valid distinction is made between a societal and an individual dictionary culture. One often finds traces of individual dictionary culture within a given speech community because some speakers are familiar with dictionaries and the use of dictionaries whereas the majority of the members of that speech community are not familiar with dictionary use. This constitutes the lack of a societal dictionary culture.

Within both an emerging and an established standard language, a societal dictionary culture is of extreme importance to ensure access to data in a variety of reference sources and the successful retrieval of information from these sources. A dictionary culture is part of a more comprehensive reference culture that assists users to access and use different reference sources. For the average member of a speech community a dictionary is the first reference source they need to consult. Therefore an early introduction to a dictionary culture is necessary within the classroom. From the first phases of acquiring this culture it should be clear that it is not an abstract or only theoretical endeavour – young members of a speech community should be exposed to dictionaries; by becoming familiar with these dictionaries they participate in strengthening not only their individual but also their societal dictionary culture.

The successful use of a broad typological spectrum of dictionaries in accordance with the genuine purpose of each dictionary demands dictionary using skills. These skills need to be acquired and the sooner this can be achieved the better are one's chances of embarking on a successful encounter with dictionaries and other reference tools. Learners' possession of dictionary skills should not be assumed and teachers should embark actively on a process of teaching these skills so that a dictionary culture can be established. Only when this is done is an optimal use of dictionaries possible.

This paper focuses on ways and means to establish dictionary culture at an early age and to continue its strengthening throughout the school years. This includes knowledge of dictionary typology, knowledge of the structure, contents and functions of a given dictionary and knowledge of access to data and a successful retrieval of information from a given dictionary. This study utilises some of the findings of existing comprehensive user studies.

Even at the first introductory level a sound theoretical foundation is needed to ensure the appropriate participation in active dictionary use. Various approaches to the teaching of dictionary skills are introduced with an interactive relation between theory and practice. Learners need to be constantly exposed to dictionaries compiled as practical instruments that respond to the specific needs of specific users in specific situations of dictionary use. In order to achieve this, a gradual introduction to a variety of dictionaries is suggested where students become familiar with both language for general purposes (LGP) and language for special purposes (LSP)

dictionaries. By becoming familiar with LSP dictionaries learners should also realise that dictionary use should not only be regarded as an activity within a language curriculum. The use of general and specialised dictionaries within all the different subject fields at school should be the target of a wide-ranging dictionary culture. Having acquired dictionary using skills the students are in a position to utilise both dictionaries and other reference tools as part of a process of life-long learning. Outside the classroom they will know how and when to resort to which dictionaries when requiring assistance.

The required input inside the classroom versus dictionaries used outside the classroom varies. In a first world environment good dictionaries are normally available and most users grow up in a dictionary culture so have a reasonable amount of dictionary skills when they enter the classroom for the first time. The teacher builds on this platform which in turn stimulates dictionary use outside the classroom. In contrast to this stands a third world setup, for example in Africa, where the majority of dictionary users do not grow up with a dictionary culture and experience dictionary use for the first time inside the classroom: the task of the teacher in this case is to introduce them to dictionaries and to guide them to successful dictionary use. This paper aims to show that dictionary using skills need to be acquired within the classroom. The skills should include using LSP and LGP dictionaries but also other reference sources. These skills acquired within the classroom will help to enhance the practical use of dictionaries outside the classroom and will lead to the establishment of a societal dictionary culture. A societal dictionary culture enables a speech community to access data and to become active members of an information and knowledge environment.

### **Who is a lexicographer?**

**Rufus H. GOUWS**

Department of Afrikaans and Dutch, University of Stellenbosch.

Samuel Johnson defined the word *lexicographer* as "a harmless drudge, a maker of dictionaries". Although dictionaries are still compiled by lexicographers it can be argued that the scope of the word *lexicographer* goes far beyond the explanation supplied by Johnson. This paper looks at different definitions of the word *lexicographer*, as found in present-day dictionaries and publications on lexicography. These definitions are then compared with the reality of both theoretical lexicography and lexicographic practice in order to suggest a revised definition of the word *lexicographer*.

Lexicography is a field consisting of two subfields, that is the theoretical domain and the practical domain. Participants in both these domains typically are known as lexicographers – practical and theoretical lexicographers. The term *metalexicographer* has been introduced to refer to theoretical lexicographers. However, a clearer explanation is needed. With reference to theoretical lexicography the paper takes a closer look at the activities of those persons referred to as *metalexicographers*. Is a *metalexicographer* someone with a formal university qualification in lexicography or any person talking in a scientific way about dictionaries? Will someone working within the South African context in a PanSALB subcommittee on lexicography or someone serving on the Board of Directors of a national lexicography unit be called a (meta)lexicographer? Instead of merely arguing that *metalexicographers* are "theoretical lexicographers" a new distinction is suggested to make provision for different types of *metalexicographers*, for example, those contributing to the formulation of lexicographic theories

without applying their theories to the practice and those metalexigraphers conceptualising theoretically-based dictionary plans and models but also going further to apply these models to the lexicographic practice and by doing so becoming involved in practical dictionary-making. It is suggested that the term *metalexigrapher* is polysemous and that it refers to different categories of lexicographers. These categories will be discussed.

On the level of the lexicographic practice the typical approach today is that the making of a dictionary is regarded as a team effort and not that of a lone ranger. Within such a team each member has a very specific assignment. The question, however, is whether all the members of such a team can be regarded as lexicographers. No one will doubt that every member of the team that actually writes a dictionary article can be called a lexicographer. The leader of the team, that is, the editor-in-chief, does not necessarily write any articles but oversees the project. His/her status as lexicographer will not be questioned. But what about the computer expert devising the program or the person responsible for the layout? It is a common fact that the layout of dictionaries plays an important role in determining their success as practical tools. Does this make the person responsible for the dictionary-specific layout, that is, a form of layout not utilised in any other products but dictionaries, a lexicographer? What are the minimum criteria to apply in order to determine whether someone qualifies to be called a lexicographer? Do the same criteria apply with regard to printed and electronic dictionaries? Is the person compiling the corpus for a given dictionary or the person seeing to the technical aspects of the data bank a lexicographer? Is a lexicographer only involved in making products known as dictionaries? What about the compiler of an encyclopedia or a lexicon or a telephone directory? Is someone compiling a dictionary dealing with languages for special purposes a terminographer or a lexicographer?

This paper will not attempt to give answers to all these questions. But an attempt will be made to give a clearer and less ambiguous identification of that member of the species *Homo sapiens* that can be defined as a lexicographer. The need for more clarity is determined by the need for a more exact and scientific description of lexicography as an independent discipline – and not as a part of linguistics.

### **Expectation levels in dictionary consultation and compilation**

**Inge KOSCH**

Department of African Languages, University of South Africa

Dictionary consultation and compilation is a two-way engagement between two parties, that is, a dictionary user and a lexicographer. How well users cope with looking up words in a Bantu language dictionary and to what extent their expectations are met, depends on their consultation skills, their knowledge of the structure of a Bantu language and the dictionary design. From the user's perspective there are certain expectations, such as to be able to establish the meaning of words (this being the highest ranked type of information that a dictionary is consulted for according to Svensén, 1993:14), spelling, pronunciation, grammar, application in context, synonyms and etymology. Users of Bantu language dictionaries, especially novices, are invariably guided by their intuition and will expect to find the information sought in the first place they look. More often than not, this expectation is not met due to the agglutinative structure of the Bantu languages, which requires a special approach to lemmatization, different from the European languages, for example.

Conversely, there are also demands that the lexicographer places on the user. These demands are of varying degrees, ranging from low-level expectations, where basic look-up skills according to the letters of the alphabet are assumed to have been mastered already (as needed when using a word-based dictionary), to medium-level expectations (such as being able to look up words in a stem-based dictionary), to high-level demands (such as being able to look up words which have undergone phonological changes). In a paper dictionary space-saving considerations are obviously important. To avoid duplication of verb stems Snyman, Shole and le Roux (1990), for example, have generally omitted verb stems with the reflexive prefix *-i-* in the central list. However, for the user to be able to find the meaning of reflexive verb stems, the compilers have provided the necessary support in the form of a table of sound changes in the front matter which can be understood at one glance, even by a novice. Kriel (1982) includes a number of reflexive forms of verb stems under the letter ‘I’, but the list is by no means exhaustive. This constitutes a problem, as the user is given no instructions in the front matter on how to look up the meaning of non-listed reflexive forms.

The necessity to complete paradigms may be experienced by a user as a medium- to high-level challenge, depending on the clarity of the directives. A user-friendly dictionary would enable a user to derive ‘missing’ forms in a paradigm with the maximum amount of support and the minimum amount of effort. As far as deverbatives are concerned, Endemann (1911:36) explains that users can create nouns on their own according to a set pattern. He gives one example of a regular formation and one where sound strengthening is involved in the initial sound of a verb stem as in the case of prefix-less classes (as when *thato* ‘love’ in class 9 is formed from *-rata*). This would probably count as a high-level activity as there is no further support at that point and no cross-reference to page 15 where sound strengthening is explained. Ziervogel and Mokgokong (1975:117-123), on the other hand, offer a more comprehensive overview of the sound changes in their front matter. A fine balance must be struck between the instructions given and the presumed dictionary consultation skill of the target user.

Compilers should consider that when compiling dictionaries, scientific needs should not be served above the practical needs of users. The level of expectation should be pitched at the level of the target user. In this vein Gouws and Prinsloo (2005:39) endorse that “Modern dictionaries are judged by the success with which the *user* is able to retrieve the sought information and not in the first instance by ‘linguistic achievement.’”. Thus the ultimate test of a good dictionary is how well it can be used. The compiler can often only gauge whether the demands placed on the user were realistic once the dictionary is in active use. Lessons learnt can then be implemented in the revision of the dictionary.

This paper endeavours to look into various levels of expectation in a number of Bantu language dictionaries, both from the perspective of the user and the compiler.

### References:

- Endemann, K. 1911. *Wörterbuch der Sothosprache*. Abhandlungen des Hamburgischen Kolonialinstituts, Band VII. (Reihe B. Völkerkunde, Kulturgeschichte und Sprachen Band 4). Hamburg: L. Friedrichsen & Co.
- Gouws, R.H & Prinsloo, D.J. 2005. *Principles and practice of South African lexicography*. Stellenbosch: Sun Press.
- Kriel, T.J. 1982. *Pukuntšū woordeboek: Noord-Sotho – Afrikaans, Afrikaans – Noord-Sotho*. (3<sup>rd</sup> ed.). Goodwood, Kaap: Nasionale Boekdrukkery.
- Snyman, J.W., Shole, J.S. & le Roux, J.C. 1990. *Dikšinare ya Setswana English Afrikaans: Dictionary – Woordeboek*. Pretoria: Via Afrika Limited.
- Svensén, B. 1993. *Practical Lexicography: Principles and methods of dictionary-making*. Oxford, New York: Oxford University Press.

## **The role of sense relations in constructing definitional phrases for synsets in the Tswana WordNet**

**Jurie C. LE ROUX**

Department of African Languages, University of South Africa

The Tswana WordNet is a semantic dictionary that is designed as a semantic network by presenting words and concepts as an interrelated system based on specific sense relations. Synonymy is taken as the basic sense relation since sets of synonyms (synsets) are the basic units of the WordNet. These synsets are again linked by means of a number of semantic-conceptual relations, including hyponymy and meronymy. Semantic opposition as a lexical relation, functions as a linking device on word level.

This paper deals with the role the sense relations play in constructing the definitional phrase for each synset. With synonymy as the basic sense relation underlying the semantic network of the WordNet, the formation of synsets in the Tswana WordNet is scrutinised. It should be understood that the notion of synonymy is not used in terms of interchangeability in all contexts in WordNet. The viewpoint is rather that synonyms are interchangeable in some contexts and that one should speak of synonymy relative to a context. This criterion is presupposed rather than asserted and it affects the role of synonymy when used as a device to construct definitional phrases in the Tswana WordNet. It is pointed out that the mere translation of a synset from English to Tswana does not always give an acceptable result. Different rules apply for members of a synset to be synonymous or not in Tswana. It inevitably leads to a revision of the synonymous relationship within the Tswana synsets. The synonymous differences between the English and Tswana counterparts are also reflected in the construction of the definitional phrases when other sense relations also start to play a role when the definitional phrase for each synset is constructed.

As this paper focuses on the capacity which the definitional phrases of synsets in the Tswana WordNet have to convey the sense of synsets by virtue of the associative ties they have in relation to other synsets, one inevitably needs to relate to the field of semantics in terms of the specific approaches to it. Aspects of structural semantics as well as semantic field theory are therefore applied. In terms of structural semantics the application of semantic relations is investigated and explained when definitional phrases are constructed. Where sense relations do not directly affect the construction of definitional phrases, attention is given to the organisation of synsets into semantic fields within which synsets interrelate and define each other in various ways. Examples from conventional Tswana monolingual dictionaries as well as the Tswana WordNet are related to this discussion.

As for polysemy, polysemous words are usually indicated by separate entries in conventional dictionaries. In WordNet a synset is not equal to an ordinary dictionary entry and a polysemous word that will have separate different explanatory glosses in a conventional dictionary will have a single gloss in WordNet. When constructing a definitional phrase in WordNet this difference needs to be taken into consideration where polysemy applies. This is illustrated and explained with examples from conventional Tswana monolingual dictionaries as well as the Tswana WordNet.

Experience gained from formulating definitional phrases for synsets in the Tswana WordNet shows the unique role played by hyponymic and meronymic relations when dealing with definitional phrases for agglutinating languages. Antonymy may also be functional when formulating definitional phrases. The role of this semantic relation is therefore also scrutinised with reference to examples from conventional Tswana monolingual dictionaries as well as the Tswana WordNet.

In terms of semantic relations it is not only the paradigmatic semantic relations that play a role when definitional phrases are constructed. The syntagmatic relationships of collocation may also play a role in constructing definitional phrases for synsets. Collocational relations are therefore also examined as a supportive tool to construct definitional phrases.

There are, however, some instances where semantic relations are not directly usable when definitional phrases are formulated, for example for colour terms. In such instances the organisation of synsets into semantic fields does help to assist in constructing definitional phrases. The meaning of these synsets can only be understood by placing them in relation to the other synsets which occur with them in the semantic field by demarcating the characteristics which constitute the semantic field. Some examples will be discussed and examined in this regard.

In conclusion it is argued that because the main objective of WordNet is to make the semantic relations between word senses more explicit and easier to use, sense relations play a major role when constructing definitional phrases for synsets.

### **From Subject-Based to User-Based Dictionaries: Rejecting the Cultural Opposition Thesis in Specialized Lexicography**

**Patrick LEROYER**

Centre for lexicography, Aarhus University, Denmark

Leroyer and Bergenholtz (2012) state that the treatment of cultural items in dictionaries is determined by the views held by lexicographers on culture and by the methodology applied in dictionary making. They argue that the treatment of cultural items should be governed by the functions of the dictionary. The treatment of culture has been one of the most salient topics in language for general purposes (LGP) lexicography, particularly in relation to bi- or multilingual dictionaries, but specialized dictionaries are also subject to culture and to the widely accepted opposition between culture-dependent domains (e.g. politics, law) and culture-independent domains (e.g. physics, engineering). The central assumption of the opposition thesis is that certain conceptual systems modelling human knowledge (i.e., ontologies) are shaped by cultural constraints while others are held to be universal truths. Yet this is an epistemological issue, and the opposition thesis is fostered by a positivistic philosophy of science. This paper argues that the cultural opposition is the result of a terminological approach in which ontology-based linguistics take up the lion's share. The subject-based opposition has been harmful to theory making in specialized lexicography and has impaired the innovation process; its most pernicious effect has been the capsizing of any lexicographic good sense, as lexicographers have been working from terms to users rather than from users to terms, and it must be rejected. We defend here a user-based view on specialized lexicography in which focus is on the genuine purpose of the dictionary, namely the satisfaction of information needs in specific social contexts. The

dictionary functions are determined by answers to such crucial questions as: What is the genuine purpose of the specialized dictionary? Who are its intended users? What is their profile? In what situations do they need the dictionary? Specialized dictionaries should be treated as socially constructed products matching dynamically constructed information needs in specific social situations, and not as instruments of recognition of notional facts and relations in the extra-lexicographical world.

We will use the case of a South African dictionary of wine (Van der Merwe 2008) to show that domain-based dictionaries should be transformed into intercultural information tools in order to satisfy the needs of intended user-groups. Wine can be conceptualized as a culture-independent subject field, or series of culture-independent fields, in so far as knowledge about soils, grapes and wine making seems to be shaped and shared universally. But the subject of wine can also be seen as a construction of intercultural knowledge and communication as it is shaped by actors working in it or sharing an interest in it. A user-based approach includes not only experts, but all kinds of experts and non-experts engaged in wine for work or pleasure: producers, oenology students, merchants, professionals of wine tourism, legislation experts, reviewers, consumers and wine lovers. In the *South African Trilingual Wine Industry Dictionary* (SAWIS/Winetec 2012), only terms of the wine industry are lemmatized in English and translated into Afrikaans and Xhosa (**white wine**, Afrikaans *witwyn*, Xhosa *iwayini emhlophe*), although a small number of culture-dependent items are also included (**wine route**, Afrikaans *wynroete*, Xhosa *indlela yewayini*). According to its authors, the dictionary is polyfunctional and aims at heterogeneous user groups (“viticulturists and oenologists, producers and wine makers, viticulture and oenology students and lecturers, public relations officers and information officers, wine marketers, wine writers, wine lovers”), although the fulfilment of this aim remains questionable. The most striking feature is that the lexicographic data (‘physiology and fruit of the vine, winemaking processes’) is not specific to South Africa at all, apart from the fact that the dictionary is made up of English lemmas translated into Afrikaans and Xhosa. In other words there is nothing South African about it. It does not provide any information on local varieties of grapes (Bukettraube, Cape Riesling, Hanepoot, Roobernet, or Pinotage, the local cross between Pinot Noir and Cinsaut, which is unique to South Africa) nor on any other crucial cultural aspects of viticulture (ethical trade, history, soils, climates, wine of origin legislation).

We will also demonstrate the move from subject-based to user-based dictionaries through the study of a French website of wine featuring lexicographic data structures and search options. The website is characterized by a user-based content management system aiming at the satisfaction of heterogeneous information needs. The website can be seen as a source of inspiration for lexicographic innovation as data selection is ruled by a user-based strategy – who’s interested, in what, why and how – being the crucial questions that have been answered in order to cater for the needs of the foreseen user-groups. Furthermore, a few lexicographic adaptations, including the inclusion of database applications in the content management system, would turn it into a powerful information tool that could easily surpass any subject-based dictionary of wine.

## References:

- Leroyer, P., Bergenholtz, H. (2012). Métalexicographie culturelle, fonctions lexicographiques et finalité pragmatique. In: *Etudes de linguistique appliquée* [in press].
- Van der Merwe, M. F. (2008): Wine and Words: A Trilingual Wine Dictionary for South Africa. *Lexikos* 18, 337-348.
- Winetech and SAWIS (2012) *South African Trilingual Wine Industry Dictionary*  
<http://www.sawis.co.za/dictionary/index.php>

## The lemmatization of qualificatives in Setswana and Silozi

John LUBINDA

Department of French, University of Botswana, Gaborone

Lemmatization issues are at the core of practical lexicography in any lexicographic project. It is here that one of the most important decisions for the lexicographic treatment of collected data has to be made, namely how to present, in a dictionary, the selected lemmata derived from the corpus.

One of the grammatical categories in Bantu languages that poses a particular challenge in this regard is that of qualificatives. Prinsloo (1994) discussed the lemmatization of verbs in Northern Sotho. In a subsequent publication, Prinsloo and Gouws (1996) formulated a new dictionary convention for the lemmatization of these verbs. Subsequently, Prinsloo and De Schryver (1999) examined the lemmatization of nouns in African languages, with special reference to Sepedi and Ciluba. Prinsloo (2002) went on to discuss the lemmatization of copulatives in Northern Sotho and later published an article on the lemmatization of adverbs in that language (Prinsloo (2003)). More recently, Prinsloo (2011) analyzed the lemmatization of nouns and verbs in isiZulu. However, our literature review has revealed that, to date, no one has undertaken any study of the lemmatization of qualificatives (relative, possessive, enumerative, quantitative, etc.) in Bantu languages. Yet this is the grammatical category that perhaps presents more challenges in lemmatization than any other. It also challenges our common assumptions about the concept of “word”.

In the present paper the author takes a critical look at the manner in which some qualificatives have been lemmatized in some Setswana and Silozi dictionaries. The problem of lemmatizing qualificatives is further compounded by the choice that one has to make on the orthography: between a conjunctive or a disjunctive system of writing. In a conjunctive system of writing the qualificative stem is never detached from the concord. For instance, the possessive or relative qualificative stem, shorn of its concord, is never considered to be an independent lexical item. Invariably, it occurs with a concord affixed to it. To discuss critically observed current practices in the lemmatization of qualificatives in Silozi and Setswana reference has been made to the following dictionaries, among others :

Jalla, A. (1982) *Silozi – English Dictionary*, United Church of Zambia, NECZAM, Lusaka.

Kgasa, M.L.A & Tsonope, J. (1998) *Thanodi ya Setswana*, Longman , Gaborone.

Mareme, G.B. (2008) *Thanodi ya Setswana*, Nutrend Publishers, Pietermaritzburg.

Matumo, Z. I. (1993) *Setswana English Setswana Dictionary*, Macmillan, Gaborone.

O’Sullivan, O. (1993) *English – Silozi Dictionary*, Zambia Educational Publishing House, Lusaka.

It will be shown that the user’s guide to grammar, in the front matter, can make it easier for the dictionary compiler to lemmatize qualificative stems in a more user-friendly manner in the nomenclature. Thus, with reference to information contained in the table of possessive concords, for instance, one may proceed to lemmatize the Setswana possessive stems (-me/ka,

-gago, -gagwe, -rona, -lona, -bone) which, normally, do not occur as independent word units in this language. It can be explained that the stem remains invariable. What changes is the concordial prefix, depending on the referent and the noun class to which it belongs.

## References

- Prinsloo, D.J. (1994) The lemmatization of verbs in Northern Sotho, South African Journal of African Languages, no. 2, pp. 93 – 102.
- Prinsloo, D.J. & Gouws, R.H. (1996) Formulating a new dictionary convention for the lemmatization of verbs in Northern Sotho, South African Journal of African Languages, Vol. 16, no. 3, pp. 100 - 107.
- Prinsloo, D.J. & De Schryver (1999) The lemmatization of nouns in African Languages, with special reference to Sepedi and Ciluba, South African Journal of African Languages, vol. 19, no. 4. pp. 258 – 273.
- Prinsloo, D.J. (2002) The lemmatization of copulatives in Northern Sotho, LEXIKOS, 12, Bureau of the WAT, Stellenbosch, pp. 21 – 43.
- Prinsloo, D.J. (2003) The lemmatization of adverbs in Northern Sotho, LEXIKOS, 13, Bureau of the WAT, Stellenbosch, pp. 21- 37.
- Prinsloo, D.J. (2011) A critical analysis of the lemmatisation of nouns and verbs in isiZulu, LEXIKOS, 21, Bureau of the WAT, Stellenbosch, pp. 170 – 191.

## Treatment of taboo words in Gabonese lexicography: a case study of five reference works

Ludwine MABIKA MBOKOU

Institut de Recherche en Sciences Humaines (IRSH)

Groupe d'Etudes et de Recherches sur la Communication (GERC)

Libreville, Gabon

Most of the early lexicographic works found in Gabonese languages have a background of religious phraseology. They were largely compiled from the experience of the Christian missionaries with the linguistic communities they encountered. Some of the reference books they produced were closely related to the life of the congregation they pertained to, some others allow their corpora to be more open and, as such, they constitute a great source of information for the compilation of larger corpora.

However, the religious influence on the practice of lexicography has left some marks. Some of the terms of the language have been left out. Some others were judged unacceptable. It is also known that in African languages, some words that are considered offensive are not spoken. For example, it is often observed that the speaker of a particular community will use euphemism or hedging strategies to mention certain body parts. In addition, a word that can appear 'decent' to a European can be offensive to an African. For instance, a few African languages do not mention the term for "*breast*" in discourse, but will use a surrogate sentence instead. As far as this point is concerned, Raponda-Walker, the first Gabonese Catholic Bishop, showed in his various works (lexicography, ethnology, history, etc.) a keen straightforward account of the lexicon despite the social content of a lexical item within a particular culture. It is not surprising that his work set the foundation for modern Gabonese lexicography.

Translating and portraying a language is not an easy task especially when the languages treated do not have any cultural link. Different jokes and anecdotes have been made about translators whose misinterpretation prevents the sense from being accurately conveyed. For example, among the Bapunu people, a well-known story tells how a priest was chased away and associated to witchcraft because his interpreter did not choose the proper word. The biblical text "Jesus walked on water" is translated in French as "*Jésus a marché sur les eaux de la mer*". The word

*eaux* meaning “waters” has a homophone *os* which means “bones”. Likewise, the word *mer* which means “sea, water” has a homophone *mère* which means “mother”. Because of the phenomenon of the liaison, the plural “waters” is “*les eaux [lezo]*”, with the same pronunciation as the plural of the homophone “*os*” “*les os [lezo]*”. The interpreter confused the different homophones and the translation came out as something like “*Jesus walked on the bones of his mother*”. In the Punu culture such a statement is associated with the highest form of witchcraft which involves the murdering of the closest relative and the use of their bones for magical and demonic purposes. As a result the priest was chased away and his interpreter banned from the village.

This story demonstrates the sensitivity of the Punu culture. Thus, any translation/interpreting work should be aware of certain delicate aspects of culture. In this story the misleading factor comes neither from the strong culture nor from the Yipunu language, but most probably from the inaccuracy of the discourse context that might have provoked the miscomprehension of the interpreter. Dictionary compilation with an editorial policy for the treatment of taboos should, therefore, take into account both cultural sensitivity and the discourse context.

Looking at the development of Gabonese lexicography, this paper will, in its first part, use a comparative approach to illustrate how taboo words were treated in different bilingual dictionaries involving Gabonese languages and French. I will show how the lexicographers of these dictionaries included more than a mere linguistic description. Furthermore, a comparison will be made between Raponda-Walker’s works and dictionaries compiled by other missionaries. In a second section of the paper, the focus will be on the role of these bilingual dictionaries in modern Gabonese lexicography and in the establishment of a dictionary culture within Gabonese linguistic communities. The paper concludes on new projects and innovative lexicographic endeavours.

### **A critical analysis of the use of Xitsonga terms in Public Services – electronic and otherwise – in South Africa**

**Pearl N. Magamanana**

Letsitele, Limpopo, South Africa

The paper focuses on establishing the appropriateness of Xitsonga terms used in public electronic applications and services in South Africa, where these services are available in different official languages. These services are currently available at the automated terminal services used by Absa, Nedbank and MTN. Human Language Technology (HLT) in South Africa is an important field because it bridges the technology gap and therefore enables people to access information and services easily and assists in developing languages.

It is therefore important to conduct the research in the field because it was observed that despite the effort by these organisations (Absa, Nedbank and MTN), people (Xitsonga speakers in particular) are not using the Xitsonga option. One would expect them to embrace the opportunity of accessing information and services in the language they know and understand best. However this lack of enthusiasm has been attributed to a number of reasons, ranging from the inappropriateness or incorrectness of terms to the use of foreign terms.

The primary aim of this paper is to assess the appropriateness of the Xitsonga terms used in these automated terminal services, and therefore to discover if the Xitsonga terms used in current

services to the public are indeed correct or appropriate. Another aim is to find out the type of term-formation method used to provide terms for these applications and if the methods used contribute to advance the development of the Xitsonga language.

A corpus of terms was collected directly from these services and analysed, looking at different term-formation methods and the three dimensions of terminology (Cognitive, Linguistic and Communicative) which provide term-formation methods and term-formation rules. The analysis is supported by data obtained through interviews where Xitsonga speaking users of the three automated services were given a chance to express views about the correctness or appropriateness of the terms in the corpus, and to suggest alternatives where necessary.

The key findings are that users consider many terms used in the automated teller machines of Absa and Nedbank as well as those used by MTN customer services helpline to be inappropriate. Some term-formation methods used did not bring any change as far as the development of Xitsonga is concerned, such as direct loan, acronyms and abbreviations. However other term-formation methods used by the three automated terminal services such as paraphrasing, transliteration and compounding, do contribute to the development of the language. The paper will conclude by touching on some suggestions for lexicographic treatment of these terms, in order to consolidate their contribution to Xitsonga language development.

### **A comprehensive and extra-lexicographic process in support of teaching local languages in Gabon**

**Edgard MAILLARD ELLA**

Human Sciences Research Institute/ National Scientific and Technological Research Centre,  
Gabon

Local languages have always been of keen interest to the government of Gabon. Indeed, during the government's last general meeting for education in May 2010, it was firmly decided to introduce them into teaching programmes.

This important decision raises two major issues which directly or indirectly need the full commitment of Gabonese lexicographers. Firstly, they will have to focus on the compilation of the type of dictionary which will be appropriate to this situation. Secondly, Gabonese lexicographers have to be committed to the compilation of language teaching materials, as there is an obvious lack of these.

Regarding the first issue, the decision of the government to introduce local languages in education means that the compilation of dictionaries can no longer be based on emotional and ideological criteria. The new projects have to be well planned, as new dictionaries need be genuine instruments of language and communication according to the real user needs and the real situation.

In this regard, although French is the official language in Gabon, a Gabonese variety of French is arising, which has more meaning and sense of identity for the Gabonese people, who want to be able to use it officially. Consequently, we suggest, as a part of a comprehensive lexicographic process, the compilation of dictionaries which are focused on Gabonese local language bilingual dictionaries, each one of which will be treating Gabonese French and a local language. They would be bidirectional and produced in a pocket format.

Briefly, the compilation of these Gabonese local language bilingual dictionaries involves adapting a French/English bidirectional bilingual dictionary. This lexicographic procedure comprises two main steps: step one is aimed at making the first part of the new dictionary while step two is aimed at building the second part. Each one of these steps consists of two important processes. The first process in step one, is the adaptation of the French macrostructure in order for Gabonese French to become the source language. The second process in this step consists of replacing English texts in the microstructure so that the relevant local language will become the target language.

In step two, the first process is to replace English in the macrostructure with the local language texts. The second process, which may be the easiest one, mainly requires adjusting a few aspects of the remaining French texts in the microstructure.

As far as the second issue referred to above is concerned - that is, the production of local language school books - we think that Gabonese lexicographers have to get involved, as part of an extra-lexicographic process. Indeed, when compiling dictionaries, lexicographers deal thoroughly with important language issues such as conjugation, grammar and orthography. In this regard, they are able to contribute considerably when it comes to making corresponding school books. Within the dictionary making process to compile Gabonese local language bilingual dictionaries, lexicographers have been dealing with all these aspects of these languages. Therefore, they are, in effect, simultaneously making provision for the compilation of language text books.

From the planning and the compilation of a *Gabonese French- Bitam Fang Ntoumou / Bitam Fang Ntoumou- Gabonese French dictionary*, we show in this article, firstly the aspects of a comprehensive lexicographic process which can lead to the compilation of dictionaries of the same type in other Gabonese local languages. Secondly, we demonstrate how from the comprehensive lexicographic process of the compilation of Gabonese local language bilingual dictionaries, it is firmly possible to contribute to the making of school books dealing with conjugation, grammar and orthography in Gabonese local languages.

### **‘Balanced’ and ‘representative’ corpus; A case study of the imbalance and lack of representativeness of Northern Sotho corpora**

**Victor M. MOJELA**

Sesotho sa Leboa NLU, University of Limpopo (Turfloop Campus), South Africa

The Northern Sotho language is made up of more than 27 dialects which are found in almost all the five municipal districts of the Limpopo Province. Of all these dialects, almost half are marginalized ‘languages’ which were never included in the standardized version of Northern Sotho, and as a result, these dialects are not represented in the available Northern Sotho corpora. The best known corpus in this language is the University of Pretoria electronic corpus of about 6,8 million words. This corpus is compiled from published materials and/or from the written language (or written dialects) ; the rest of the dialects, which did not form part of the written (standard) language, are not represented. According to Gouws & Prinsloo (2005:22), a balanced and representative corpus should, ‘*contain large quantities of both spoken and written data. Spoken data is absolutely essential especially for those African languages which do not as yet have many written sources.*’

Standard Northern Sotho is based on the few dialects found in the areas around Sekhukhune, Waterberg districts and the South and South-Western section of the Capricorn district, thereby leaving almost all of the dialects in the North and North-eastern part of the Capricorn district and the whole of Mopani and part of Vembe districts as well as Mapulaneng, in Mpumalanga, out of the written standard language, and thus, out of the recorded corpora. Northern Sotho corpora are based on the following dialects: Sekone, Sepedi, Sekopa and Sekgaga (of Mphahlele), which are dialects spoken in the areas around Sekhukhune, Polokwane, Mokopane, Bela Bela and Mookgophong. With regard to ‘balanced’ and ‘representative’ corpora pertaining to the South African indigenous languages, Gouws & Prinsloo (2005:25) give the following warning to lexicographers:

*‘South African corpus compilers should try to build corpora which are big enough, well ‘balanced’ and ‘representative’ so that valid conclusions for lexicographic purposes can be drawn.’*

Unfortunately, this was not the case with the Northern Sotho corpora, which are neither ‘balanced’ nor ‘representative’, apparently due to the fact that:

- i. Most of the compilers seemed to rely too much on the written standard language, which is, even today, not ‘balanced’ and ‘representative’.
- ii. The marginalized Northern Sotho dialects, which are in the majority, were never converted to written forms and some of these dialects are quite remote from the standardized language.
- iii. The written Northern Sotho language is still based on the ‘outdated’ missionary orthographies which tended to favour the few dialects which happened to have missionary activities in their proximity. The 18<sup>th</sup> and 19<sup>th</sup> century missionary societies concentrated too much around the areas which produced the first written Northern Sotho documents on Christianity.
- iv. Even though some of the corpus compilers, who were mostly non-mother tongue speaking lexicographers, went the extra mile to incorporate researched oral materials in their corpora, most of their researches continued to marginalize the dialects in the Lowveld, including the North, North-East and the Eastern part of Limpopo and Mpumalanga.

As a result, many of the emerging elite groups from these marginalized dialects and a new generation of Magoši (traditional leaders) from the areas such as Bolobedu, Makhutšwe, Botlokwa, Senwabarwana, Mapulaneng have started questioning the validity of standardized Northern Sotho and of incorporating their dialects within it.

The issue of apparently unbalanced and unrepresentative corpora in the Northern Sotho language has also led to two radical consequences, that is, ownership or rejection of the language. Some of the elite groups from the dialects which were favoured by standardization have started claiming ownership of the language, while those groups which were side-lined by the standardization of Northern Sotho started disowning and opting for withdrawal from the standard language because they believe their dialects were misplaced.

This paper will recommend the comparison and analysis of the Northern Sotho corpora, especially the University of Pretoria Corpus, with the contents, structures and vocabularies of the unwritten spoken dialects like Sepulana, Selobedu, Sephalaborwa, Setlokwa, Sedzwabo, and will conclude that there is an urgent necessity to update the standard Northern Sotho language and,

moreover, for corpus compilers and linguists to do intensive research work to re-discover the unrecorded and forgotten dialects of Northern Sotho, and to start finding ways for the inclusion of these dialects in the official Northern Sotho orthography.

### References:

- Gouws, R.H & Prinsloo, D.J.** 2005. *Principles and Practice of South African Lexicography*. Sun Press, Stellenbosch
- Higa, M.** 1980. Sociolinguistic Aspects of Word-borrowing. Mackey, W.F. and J. Ornstein (Eds.). 1980. *Sociolinguistics: Studies in Language Contact*: 275-292. The Hague: Mouton.
- Mojela, V.M.** 1999. Prestige Terminology and its consequences in the development of the Northern Sotho vocabulary. Unpublished DLit et Phil thesis, Unisa, Pretoria
- Mojela, M.V.** 1991. *Semantic Changes Accompanying Loan Words in the Northern Sotho Lexicon*. Unpublished M.A. Thesis. Pretoria: Vista University.
- Monig,** 1988 *The Pedi Northern Sotho Terminology and Orthography No. 4*
- PanSALB.** 2008. *Melao ya Mongwalo le Mopeleto ya Sesotho sa Leboa*

## A Macrostructure for an English-French Dictionary of Phonetic Sciences

**Hugues Steve NDINGA-KOUMBA-BINZA**

Centre for Text Technology, North-West University, South Africa

Ndinga-Koumba-Binza (2011) emphasised the relevance of compiling an English-French bilingual dictionary of phonetics. The core reasons for such a project being that not only are phonetic sciences rarely the topic of bilingual specialized lexicography projects, but also that the development of new technologies has had an important impact on language sound studies with the creation of various new terminologies and lexicons, which often struggle to find proper equivalents from one language to another (English and French, in the case of this study).

This paper falls within the same framework of Ndinga-Koumba-Binza's conceptualization plan for the intended bilingual dictionary of phonetic sciences. It is aimed at gathering further inputs from experts, especially on questions related to the macrostructure in particular, and the nomenclature in general. The question that this paper seeks to respond to is 'What are the main characteristics of the macrostructure for the planned dictionary of phonetic sciences?'

Nielsen (1990) has reviewed the notion of lexicographic macrostructure 'in respect of the degree of complexity of lexicographic macrostructures and the interrelationship between and among the macrostructural components' (Nielsen 1990: 49), and suggested a definition of the term that adds on the traditional definition contained in most lexicographic literature (i.e. lemmata quantity and arrangement, cf. Hartmann 1983: 70 & 225; Svensén 2009: 78 & 368; and Mavoungou 2010: 125). Although this paper adheres to the traditional and widely accepted use of the term 'macrostructure', it also incorporates suggestions made by Nielsen (1990). In fact, Nielsen's perception is that the 'lexicographic macrostructure is not merely the arrangement of lemmata, but a structure which applies to the dictionary as a whole' (Nielsen 1990: 49). For Nielsen (1990: 50), a macrostructure of any given dictionary is 'realised by the arrangement of the separate parts, or components, which collectively make up the entire dictionary'.

Furthermore, Nielsen (1990: 52) indicates from "*available empirical evidence*" that lexicographic macrostructures can be classified into two categories; the simple macrostructure and the complex macrostructure. The former refers to "*a lexicographic macrostructure which is composed of only two macrostructural components*" (Nielsen 1990: 52), and the latter refers to

“a lexicographic macrostructure which is composed of more than two macrostructural components” (Nielsen 1990: 53).

My task in this paper is threefold. First, I will identify the relevant macrostructural components for the planned dictionary. This will then lead to the determination of the appropriate macrostructure type for the dictionary. Finally, the third focus of this paper will be to demonstrate the interrelationship between the selected macrostructural components. The identification of relevant macrostructural elements also adopts Nielsen’s methodology of “empirical evidence” by observing the nomenclature, the structure and the contents of various dictionaries in the field of bilingual specialized lexicography such as Hildebert (2008), Viljoen (2006) and Kaufmann & Bergenholtz (1992).

This paper aims to contribute to the enhancement of bilingual or multilingual specialized lexicography as introduced by works such as Bergenholtz & Tarp (1995), Bergenholtz, Nielsen & Tarp (2009), and Fuertes-Olivera (2009 and 2010) , and in the African context by Carstens (1997 and 1998) and Mpofu & Mangoya (2005). The paper concludes by commenting on the contribution of bilingual specialized lexicography to knowledge and technology dissemination between various cultures.

### References:

- Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of specialised lexicography. The preparation of specialised Dictionaries*. Amsterdam: Benjamins.
- Bergenholtz, H., Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography at a crossroads: Dictionaries and encyclopedias today, lexicographical tools tomorrow*. Bern: Peter Lang.
- Carstens, A. 1997. Issues in the planning of a multilingual explanatory dictionary of chemistry for South African students. *Lexikos* 7: 1-24.
- Carstens, A. 1998. Science through Sepedi: Is Terminologisation a Worthwhile Venture?. *Lexikos* 8: 1-17.
- Fuertes-Olivera, P.A. 2009. Specialised lexicography for learners: Specific Proposals for the construction of pedagogically-oriented printed business dictionaries. *Hermès – Journal of Language and Communication Studies* 49: 167-188.
- Fuertes-Olivera, P.A. (Ed.). 2010. *Specialised dictionaries for learners*. Berlin: Walter de Gruyter.
- Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and practice*. London: Academic Press.
- Hildebert, J. 2008. *Dictionnaire de l’anglais de l’informatique anglais/français français/anglais*. Lonrai, France : Business Management Series.
- Kaufmann, U. & H. Bergenholtz. 1992. *Genteknologisk ordbog dansk-engelsk engelsk-dansk*. Copenhagen: GEC Gads Forlag.
- Mavoungou, P.A. 2010. *A dictionary plan of Yilumbu: Metalexigraphic criteria for the compilation of a trilingual dictionary Yilumbu-English-French*. Saarbrücken, Germany: VDM Verlag.
- Ndinga-Koumba-Binza, H.S. 2011. Compilation of an English-French dictionary of phonetic sciences: An ongoing project. Paper presented at the 16<sup>th</sup> *International Conference of the African Association for Lexicography (AFRILEX)* at University of Namibia in Windhoek, Namibia: 5-7 July 2011.
- Nielsen, S. 1990. Lexicographic macrostructures. *Hermes, Journal of Linguistics* 4: 49-66.
- Svensén, B. 2009. *A handbook of lexicography: The theory and practice of dictionary-making*. Cambridge University Press.
- Viljoen, H.C. (Ed.). 2006. *Kuberwoordeboek Afrikaans-Engels, Engels-Afrikaans. Cyber dictionary Afrikaans-English, English-Afrikaans*. Pretoria: Protea Boekhuis.

## **Standardisation of the principles and methods of terminology development in South Africa: A case of Statistics South Africa.**

**Lufuno NETSHITOMBONI**

Publication Services, Statistics South Africa, Pretoria, South Africa

Most language development processes are dependent on terminology development for concepts and translation of documents into African languages. As a result, the need for a national standard in the terminology development process becomes even more crucial. This is the case since most of the language practitioners who are involved in these processes are not trained terminologists ; they are often appointed as language practitioners responsible for translation, editing and terminology development and in some cases, language planning. The initiatives to develop languages therefore need common sets of principles and a common approach to address the following challenges in the area of terminology development:

- the need to have a systematic method of developing terminology adopted by all language units or institutions involved in terminology development processes in South Africa;
- the need for coordination of all terminology development projects in order to avoid duplication of effort and to ensure uniformity in the terminology development process
- lack of acknowledgement of terminology principles that are not applicable to the South African context, especially for African languages.

The main focus of this paper is to briefly explain the function of SC1 (SABS TC 37, sub-committee 1, Principles and Methods) and to look at examples of the technical-domain terminology development process and a terminology project on general terminology. This will be done to explore the possibility of developing national standards on the methods and principles used for terminology development in South Africa. Examples of methods and principles applied to the development of statistical terminology in the project initiated by Statistics South Africa(StatsSA) will be discussed. The purpose of the discussion will be to present a possible approach or standard that could be adopted in a technical domain terminology development process.

Examples of what could be standardised are taken from the project that is currently being run by StatsSA. The project focuses on the development of statistical terminology into the 10 official languages of South Africa. The terminology used had been developed or excerpted from the surveys that StatsSA conducts. The need for such a project was to ensure that the collection of data that takes place in the diverse linguistic communities in South Africa is more inclusive in nature. That is, it is envisaged that the terminology developed in these languages will alleviate the burden that translators experience when translating questionnaires into African languages. The project was also initiated in order to standardise the concepts that were used in the surveys, and ensure comparability of the statistics by other countries. The secondary phase of the project has been concluded. However, even with the project not finalised, there are some common trends that have been picked up both from this project and the one conducted in 2008. These trends can be used as a basis to develop standards that could assist this process.

This paper contends that terminology development processes are taking place at a very fast pace, but still in an uncoordinated manner. Therefore, the methods that are adopted by different

institutions need to be standardised. A comparative study between the project initiated by StatsSA as an example of technical domain terminology project and projects that were initiated by the Department of Arts and Culture on general terminology will be conducted. This will be done in order to determine the need to standardise domain-specific terminology development methods and general terminology development methods that could also be adopted at a national level, and in some cases, internationally.

## **An exploration of Setswana cattle colour terms and their lexicographic treatment**

**Thapelo J. Otlogetswe**

Department of English, University of Botswana, Gaborone

One of the key features of Setswana's unique indigenous knowledge systems (IKS) is its animal identification system through the lexicalization of animal skin colour terms. The system predates the manual branding as well as the electronic system of cattle identification (bolus ingestion). With individuals owning large herds of cattle, there was pressure on cattle owners to develop a unique system of cattle identification which demanded one to lexicalize many skin colour combinations, develop and lexicalize unique ear marks as well as lexicalize the animal horn shapes. In this paper we present some of the data collected for the centre section of a Setswana monolingual dictionary which deal with the cattle colour terminology. We also consider how such terms have been lemmatized in a number of Setswana dictionaries such as Kgasa & Tsonope (1998), Matumo (1993) and Mareme (2008).

Much of the literature on colour theory distinguishes between primary and secondary colours. Primary colours are basic colours that cannot be created by mixing other colours. These are yellow, red and blue. Secondary colours are those colours that are created by mixing primary colours. These are orange, green and purple. Tertiary colours are on the other hand created by mixing secondary colours together. Berlin & Kay (1969) in their study of Setswana colour terminology observe that colour term inventories are drawn from a severely restricted sub-set of all possible inventories. Davies, Corbett, McGurk & Jerrett (1994) argues that Setswana has a composite term *botala* 'grue' (green & blue) "covers the blue and green regions of colour space". Semantically and lexicographically these are homographs and not a single term with a broad semantic scope since they are semantically distinct. They also observe that Setswana has a basic term for brown. This is true though Setswana has lexicalised shades of brown with terms such as *thokwa* (dark brown), *tshetlha* (light brown), *khunohu* (reddish brown) and others. Otlogetswe & Bagwasi (2008) in their paper: *An analysis of two Setswana colour terms: ntsho & tshweu* consider colour terminology within a European mould in which only pure or mixed colours are lexicalized. Many African languages lexicalize adjacent colours, something not common in European languages. Once this fact is considered, a much more complex colour system in African languages is revealed. To understand Setswana animal colour terms one must therefore consider the following: the sex of the cow, pure colours, the size of the patch on the skin, mixed colours and overlapping as well as adjacent colours. Lexicographically the challenge becomes how to represent this information in a dictionary.

A dictionary's definition is insufficient to capture the complexity of Setswana cattle terminology. Since Kgasa and Tsonope (1998), colours have been treated as part of the dictionary middle section or part of the dictionary back matter. Their inclusion as the central or

back matter of a dictionary is therefore essential to facilitate their identification and recognition by dictionary users. This means that the definition in the body of the dictionary is cross referenced to the specific image for illustration purposes. The challenge however is that not all printing will reproduce the animal colours as seen in the real world. Additionally, there will always be a certain degree of disagreement between the Batswana about Setswana cattle colours. This is in part dialectal; in part it is an argument about whether a patch in a cow is large enough to warrant recognition. Evidently, much needs to be done to document the Setswana cattle/animal terminology in general. Knowledge of animal colour terminology is largely fading in what is becoming an increasing urbanised country and therefore needs to be recorded in dictionaries. The paper therefore discusses the different strategies of representing colour terms in Setswana dictionaries. It argues that photographic images of cattle as part of the centre or back matter of dictionary are an essential element in the comprehension of Setswana colour terminology.

### References:

- Davies, I.R.L., Macdermid, C., Corbett, G.G., McGurk, H. and Jerrett, D. 1992. *Color terms in Setswana: a linguistic and perceptual approach*. Linguistics. Vol 30. No.6. 1065-1103.
- Berlin, B. and P. Kay. 1969. *Basic Color Terms*. Berkeley: University of California Press.
- Kgasa, M.L.A., and Tsonope, J. 1998. *Thanodi ya Setswana*. Gaborone: Longman.
- Mareme, G.B. 2008. *Thanodi ya Setswana*. Pietermaritzburg. Nutrend Publishers.
- Matumo, Z.I. 1993. *Setswana English Setswana Dictionary*. Gaborone: Macmillan.
- Otlogetswe & Bagwasi (2008) *An analysis of two Setswana colour terms: ntsho & tshweu*. Marang: Journal of Language & Literature. Vol. 18. p.1-13.

## A critical evaluation of dictionaries with amalgamated lemma lists

**Danie J. PRINSLOO**

Dept of African Languages, University of Pretoria, South Africa

The aim of the paper is to critically evaluate dictionaries for closely related languages such as Afrikaans and Dutch, the Sotho languages, and the Nguni languages, compiled with a single (amalgamated) lemma list for the related languages. The recently published Groot Woordeboek (Afrikaans and Dutch), henceforth GWAD, will be analysed as a case in point. The focus will be on the advantages and disadvantages of such a single dictionary with Afrikaans and Dutch lemmas presented in a single lemma list compared to two separate dictionaries/sections Afrikaans-Dutch and Dutch-Afrikaans. The five types of lemmas distinguished in GWAD, namely (a) absolute cognates, (b) absolute cognates with different forms, (c) partial cognates, (d) non-cognates and (e) false friends, constitute the scope of the evaluation.

A basic requirement for the compilation of dictionaries with amalgamated lemma lists is substantial overlap in terms of the form and meaning. For Dutch and Afrikaans Gouws et al. (2004) indicate a substantial overlap in terms of frequently used words. For the Sotho languages Prinsloo (2006) compared the top 10,000 words in Sesotho, Setswana and Sepedi and found that these languages have many words in common rendering a single lemma list of 22,537 compared to a 30,000 lemma list in three separate dictionaries, thus a reduction of almost 30%.

GWAD claims to be unique as the first dictionary to be compiled with an amalgamated lemma list. Such a dictionary is in the first place a translation dictionary giving translation equivalents in the target language for words in the source language and vice versa. A fundamental difference, however, is that as a result of the amalgamated approach it is also a contrastive dictionary

reflecting the differences and similarities between the two languages in a more direct way than separate dictionaries, for example, Afrikaans-Dutch and Dutch-Afrikaans, could achieve. This is applicable to all five categories (a) - (e) distinguished above and it will be argued by means of analysis of randomly selected Dutch/Afrikaans examples that the dictionary succeeds in its systematic and consistent presentation of such similarities and differences for these categories. For example, in the case of *knap* sense distinctions such as ‘intelligent’ and ‘excellent’ that are applicable to both Afrikaans and Dutch are treated simultaneously for Afrikaans and Dutch in the A/N (Afrikaans/Nederlands) section, and ‘neat, beautiful, attractive’ applicable only to Dutch versus ‘barely’ applicable to Afrikaans are treated in separate sections marked as N (Nederlands) and A (Afrikaans) respectively.

However, the advantages of an amalgamated dictionary have to be weighed up against a number of possible negative aspects. First, redundancy is inevitable as a result of the need for cross-referencing in categories (b) and (d) where a separate lemma for the other member of the language pair has to be given in its appropriate alphabetical position to guide the user to the lemma where the required information could be found. Secondly, users have to familiarise themselves beforehand with the uncommon article structure in GWAD and especially with a number of non-standard conventions such as ‘=’ (equal), ‘≠’ (not equal), ‘A/N’ (applicable to Afrikaans and Dutch) and the use of roman type (Dutch) versus italics (Afrikaans) which is crucial for understanding the information presented. Thirdly, an amalgamated approach results in relatively long(er) articles simply because treatment is offered for two (or more in the case of the Sesotho) languages in the same article which means that the user has to read through more data in order to find what he/she is looking for. This, in principle runs contrary to current trends in lexicographic guidance to guard against information overload and to guide the user more directly to what they are presumed to be looking for. Fourthly, it could be argued that unlike the section in a given dictionary article treating similarities for Afrikaans and Dutch simultaneously, the separate sections for Afrikaans and Dutch do not contribute to space saving in comparison to separate Afrikaans-Dutch and Dutch-Afrikaans dictionaries.

It will be concluded that, notwithstanding certain deficiencies inherent in the amalgamation approach, GWAD is a dictionary of high lexicographic achievement and a recommended option for Afrikaans and Dutch lexicography.

## References:

- GWAD Pharos Groot Woordeboek. Afrikaans en Nederlands.** (Prisma Groot Woordenboek Afrikaans en Nederlands) 2011. Willy Martin (Ed.). Cape Town: Pharos.
- Gouws, R.H., D.J. Prinsloo and G.-M. de Schryver.** 2004. Friends will be Friends — True or False. Lexicographic Approaches to the Treatment of False Friends. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 797-805.
- Prinsloo, D.J.** 2006. Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: Lexicographic Approaches to the Treatment of False Friends. A Viability Study. *Lexikos*, 16 (193-204)

## **On-screen data presentation of collocations within a framework of lexicographic and usability requirements**

**Hilma A. Reinhardt**

Department of Modern Foreign Languages, University of Stellenbosch, South Africa

Without the right tools, learning a language can be a real challenge. Dictionaries are such tools. However, the success of a tool depends on its appropriateness for the task at hand (cf. Bergenholtz & Bergenholtz; 2011).

Following an interdisciplinary approach, this presentation takes a closer look at an appropriate way to present collocations on screen in an electronic learner's dictionary. On the one hand the paper will highlight lexicographic data relating to collocations to be made available to the intermediate level foreign language learner (FLL) with regards to language production. On the other hand, usability theories and practices will be set to work, in order to provide for a possible on-screen presentation and functionality.

In this presentation 'collocation' is understood to consist of a base and a collocate, in the sense of Hausmann (2004: 311), who stipulates the base to be semantically autonomous and the collocate to be reliant on it and receiving its meaning only when used in conjunction with its base. Collocations are furthermore seen as a single lexicographic treatment unit (Heid & Gouws, 2006: 982), receiving lemma-like status (Heid & Gouws, 2006: 983), and having linguistic properties (like morphosyntactic preferences and syntactic construction) and semantic relations with other collocations or with other treatment units (cf. Heid & Gouws, 2006). The underlying question is, how to make available the collocations and their entries to a FLL who intends to produce and understand language.

I understand the concept of 'electronic dictionary' as being a multi-functional mother-dictionary, as Gouws (2007: 321) calls it, with an underlying electronic source of data (cf. Reinhardt, 2011: 5ff.). This data source could be based on corpora-analysis by use of, for example, the 'Sketch Engine' (see Kilgarriff, Rychly, Smrz, et al., 2004) or any other corpora-analysis, such as described by Evert in his article *Corpora and Collocations* (Evert, 2008).

I believe an electronic dictionary's power lies in the computer's ability to present exactly the data (extracted prior to the presentation) as required in the usage situation (cf. Tarp, 2011) on-screen and in a user-friendly way. For this, I envisage two basic access structures for an electronic dictionary, namely search and browse. 'Search' is mainly to cater for consultation of a dictionary by a FLL to reach a short-term communicative goal, for example during immediate language production when writing an essay in the target language. Its importance is emphasised from the lexicographer's side, for example, by Almind (2005: 39) and from the usability expert's side, for example, by Nielsen (2000: 224). 'Browse' is mainly to cater for consultation of a dictionary to help the FLL achieve a longer-term goal, mainly of cognitive nature, namely exploring the target language in order to learn and internalise it (cf. Tarp, 2011; Reinhardt, 2011: 114, 138). Allowing users to browse the electronic dictionary's content means that they may explore the content, clicking on links and buttons, following information of interest to them. Searching and browsing may also be used as intertwined processes, just as users see fit to reach their consultation goal.

Usability is defined as the extent to which a product can be used by specific users within a specific usage situation in order to achieve specific goals effectively, efficiently and

satisfactorily (DIN EN ISO 9241-11, 1998: 4). The usability of an electronic dictionary is therefore determined by the effectiveness, efficiency and satisfaction experienced by the user during its use. These general objectives of software usability are close in spirit to the more specific objectives of the lexicographic Function Theory (cf. Tarp, 2011; cf. Bergenholtz & Bergenholtz, 2011; cf. Heid, 2011).

During the presentation it is planned to show how a particular selection of lexicographic data and its appropriate on-screen presentation would enable the FLL involved in language production (short-term goal) and acquisition (long-term goal) to consult electronic dictionaries successfully. The learner should be enabled to choose the least complicated access structure, that is, the path that is followed to find the information required.

The method adopted is to show a FLL's possible search entries, being either a base or collocate in either the target or even a source language, and to accompany the user's choice of entry through a series of steps until a useful search result is achieved. The different presentations of data shown during the necessary steps will be accompanied by explanations of the lexicographic and usability requirements which should be satisfied in each case.

The theory underlying the proposed approach towards achieving the on-screen data presentation of a user-friendly electronic dictionary is of a general nature and therefore applicable to other target groups than the FLL; it is applicable for the on-screen presentation of data of African and European languages alike.

## Bibliography:

- Almind, R. 2005. Designing Internet Dictionaries. *Hermes, Journal of Linguistics*, 34:37-54 [Online]. <http://www.idiomordbogen.dk/Lit/Hermes/Hermes-34-3-RAB.pdf> [01.07.2007].
- Bergenholtz, H. & Bergenholtz, I. 2011. A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool, in Fuertes-Olivera, P. A. & Bergenholtz H. (eds.): **e-Lexicography. The Internet, Digital Initiatives and Lexicography**. London/New York: Continuum. 187-207.
- DIN EN ISO 9241-11. 1998. *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten – Teil 11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze (ISO 9241-11: 1998)*. Berlin: Beuth.
- Evert, Stefan (2008). Corpora and collocations. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, article 58. Mouton de Gruyter, Berlin. [extended manuscript Online]. [http://purl.org/stefan.evert/PUB/Evert2007HSK\\_extended\\_manuscript.pdf](http://purl.org/stefan.evert/PUB/Evert2007HSK_extended_manuscript.pdf) [21.05.2012]
- Gouws, R. H. 2007. On the development of bilingual dictionaries in South Africa: aspects of dictionary culture and government policy. *International Journal of Lexicography*, 20(3):313-327.
- Hausmann, F. J. 2004. Was sind eigentlich Kollokationen? In Steyer, K. (Hrsg.). *Wortverbindungen – mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003*. Berlin: de Gruyter. 309-334.
- Heid, U. 2011. Electronic dictionaries as tools: Towards an assessment of usability, in Fuertes-Olivera, P. A. & Bergenholtz H. (eds.): **e-Lexicography. The Internet, Digital Initiatives and Lexicography**. London/New York: Continuum. 287-301.
- Heid, U. & Gouws, R. H. 2006. A model for a multifunctional electronic dictionary of collocations, in Corino, E., Marellò, C. & Onesti, C. (eds.). *Atti del XII Congresso Internazionale di Lessicografia / Proceedings of the twelfth Euralex International Congress, Turin*. Alessandria, Italien: Edizioni dell'Orso. 979-988.
- Kilgarriff, A., Rychly, P., Smrz, P. and Tugwell, D. 2004. The Sketch Engine. In *Proc.* Lorient, France. [Online]. <http://www.kilgarriff.co.uk/publications.htm> [17.05.2012]
- Nielsen, J. 2000. *Designing Web Usability: The Practice of Simplicity*. Thousand Oaks, California: New Riders Publishing.

- Reinhardt, H. A. 2011. *Anforderungen für eine Benutzeroberfläche eines elektronischen Wörterbuches deutscher Kollokationen für afrikaans-sprechende Lerner des Faches Deutsch als Fremdsprache*. Master thesis. Stellenbosch: Stellenbosch University [Online]. <http://hdl.handle.net/10019.1/6470> [28.03.2011]
- Tarp, S. 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In: P. A. Fuertes-Olivera/H. Bergenholtz (eds.): **e-Lexicography. The Internet, Digital Initiatives and Lexicography**. London/New York: Continuum: 54-70.

## **Old wisdom: The surprising and highly relevant lexicographical knowledge we can get from a specialized dictionary published in 1774**

**Sven TARP**

Centre for Lexicography, Aarhus University, Denmark

There is little doubt that lexicography, both as a practice and a theory, has experienced important progress during the years, especially when one compares lexicographical works from the various centuries. However, history – including lexicographical history – should not be viewed as an ever-growing progress but as a process with its ups and downs, and the above statement should therefore not be regarded as something absolute and rigid, especially not because there are important exceptions to this general rule to be found in various lexicographical works, especially those under influence of the Age of Enlightenment which, in some aspects, are extremely advanced even compared with present-day dictionaries. M. Postlethwayt's *The Universal Dictionary of Trade and Commerce* from 1774 is one such example.

The full title of this specialized dictionary (see below) which is in fact more a narrative than a title in the modern sense of the word provides immediate information about its content and its usefulness for the potential users. The paper will give a general presentation of this rare and interesting dictionary which was published in various editions in the 18<sup>th</sup> Century but which, for one reason or another, is not even mentioned in Cowie's (2009) comprehensive *The Oxford History of English Lexicography* although it can still be found and consulted at various libraries, among them the historical *Biblioteca de Santa Cruz* in Valladolid, Spain, where the research behind this paper has been carried out. The presentation will include a detailed description of the nature and structure of the various lexicographical data – inclusive of illustrations – included in this impressive work. The paper will then focus on and analyse the so-called *remarks* which are frequently addressed to the dictionary articles.

These *remarks* constitute something really unique, at least in comparison with modern trade and business dictionaries, as they contain highly useful advice, instructions and recommendations to the intended user group not only in terms of communication in the broad sense of the word but also in terms of behaviour in a specific trade and business environment that may be strange to people coming from beyond the seas. In this respect, the dictionary has, besides the 'normal' communicative and cognitive functions, an important operative function, cf. Tarp (2008), providing assistance to the users when they need to navigate in a different culture with which they are not completely familiar and, in this way, creating the necessary conditions for a meaningful communication and a successful business venture.

On this basis, the paper will analyse the 'interaction' between the various functions displayed by the dictionary and will finally try to systematize the various types of *remarks* according to their specific content and purpose. In this way, some elements of a future theory of 'operative' advice, instructions and recommendations in dictionaries will be formulated.

## References:

- Cowie, A.P. (Ed.).** 2009. *The Oxford History of English Lexicography*. Oxford: Oxford University Press.
- Postlethwayt, M. (Ed.).** 1774. *The Universal Dictionary of Trade and Commerce : With large Additions and Improvements, Adapting the same to the Present State of British Affairs in America, since the last Treaty of Peace made in the Year 1763. With Great Variety of New Remarks and Illustrations Incorporated through the Whole : Together with Anything essential that is contained in Savary's Dictionary : Also, all the Material Laws of Trade and Navigation relating to these Kingdoms, and the Customs and Usages to which all Traders are Subject*. The Fourth Edition. London: Printed for W. Strahan, etc.
- Tarp, Sven.** 2008: Revival of a Dusty Old Profession. *Hermes. Journal of Linguistics* 41: 175-188.

## Developing a mathematics and science school dictionary in isiXhosa: some challenges

**Zola WABABA and Julius DANTILE**

IsiXhosa National Lexicography Unit, University of Fort Hare, South Africa

This paper explores the challenges encountered in the process of defining translation equivalents and examines the translatability of some technical concepts and terms into IsiXhosa. Research has shown that understanding of concepts and terminology in high conceptual learning areas such as mathematics and science by the learners who speak African languages is crucial for them to prosper in their education. Most students who speak isiXhosa as a first language (L1) perform below par in these learning areas, and it is reported that one of the contributing factors is a low level of grasping concepts and technical vocabulary in this language. What do we mean when we refer to the understanding of a concept? Although it has been always a mammoth task to define 'concept', for us, 'concept' refers to a person's mental picture or idea of something. So words that we use are intended to project or to articulate concepts. According to the project entitled 'Concept Literacy in Mathematics and Science' Young et al. (2004: page) argue that "a concept is a mental picture which has a standard and universally accepted meaning".

During the past four years the IsiXhosa National Lexicography Unit (XNLU) has been engaged in the lexical development initiative for Mathematics and Science (defining concepts and providing equivalent terms in isiXhosa) in order to enhance teaching and learning for isiXhosa speaking learners. The terminology list, which is the basis of the dictionary product: Illustrated Mathematics and Science School Dictionary for Grades 1 to 6 was born out of a broad consultative terminology development process that involved subject teachers, subject experts, linguists, translators and terminologists. We used the term 'school dictionary' for this product based on the definition provided by Tarp, S. (2011: 226);

*A school dictionary is a learner's dictionary (or pedagogical dictionary) especially designed to assist school children in learning languages (whether a native or a foreign language) and scientific and practical disciplines.*

Our approach started by defining the intended product, that is, asking what the lexicographical function of this dictionary was. We used the modern theory of lexicographic function (Bergenholtz, H and Tarp, S. 2003: page) which postulates that;

*the subject field of lexicography is dictionaries, a human-made product ... dictionaries are considered utility products that are made in order to satisfy certain human needs... All theoretical and practical considerations must be based upon a determination of these needs.*

The pedagogical and didactic qualities of this dictionary are determined by the comprehensibility and translatability of the translation equivalents to ensure that the mathematical and scientific translations in IsiXhosa are more comprehensive. This functional theory of lexicography provides two models, the knowledge and communicative models, that will be utilised to determine whether this dictionary product takes cognisance of the linguistic matters and rigour of mathematical and scientific concepts in order to ensure that the language used in this product is accessible and meaningful to both the teacher and learner. Furthermore, this dictionary seeks to intervene on conceptual issues between the pedagogical content knowledge and the comprehension by the learner.

In our analysis, which we hope will bring out some new insights into techniques and methods that can be utilised when dealing with concepts and terms from English into IsiXhosa, especially in a language for specific purposes context, we have utilised a number of strategies that include, amongst others:

- a. phonological and grammatical adaptation of concepts from English into IsiXhosa, e.g. **computer** = *ikhompyutha* or **photosynthesis** = *ifotosinthesisi* or **account** = *iakhawunti*;
- b. compounding translation equivalents that are long phrases, e.g. **symmetrical** = *ngokolingano-macala* or **denominator** - *inani eliphezulu kwiqhezu* = *unaniphezulu kwiqhezu* = *unaniphezulu* or **numerator** – *inani eliphantsi kwiqhezu* = *unaniphantsi kwiqhezu* = *unaniphantsi*
- c. providing semantic expansion to IsiXhosa general words to give them specialised meaning; e.g. **matter** = *inkqunto* or **right angle** = *idolo* or **sharp curve** = *idolo*,
- d. blending, e.g. **baseline** = *umgcantsusa* [umgca + intsusa] or **numberline** = *umgcamanani* [umgca + amanani], and
- e. acronym BODMAS = uBODMAS.

Through these strategies, we have managed to deal with the mistranslations and misinterpretations of certain mathematical and scientific concepts to achieve comprehensibility and readability. As a result we argue that this dictionary will be able to mediate on issues of pedagogy and conceptual understanding by the learners.

## Special Session on SABS TC37

**An overview of SABS TC 37 – a South African mirror committee of the international ISO/TC 37**

**Mariëtta ALBERTS**

Research Unit for Languages and Literature in SA Context, Northwest University, Potchefstroom Campus, South Africa

All the aspects related to the standardization of terminology and other language and content resources are encapsulated in the activities of and dealt with by a technical committee of the International Organization for Standardization (ISO) and a South African mirror committee at the South African Bureau of Standards (SABS). The business of these technical committees, ISO/TC 37 and SABS TC 37 is discussed.

The scope of ISO/TC 37 and SABS TC 37 is the standardization of principles, methods and applications relating to terminology and other language and content resources. The South African mirror committee focuses on issues such as globalization, regionalization and cooperation possibilities within the South African Development Community (SADC). It functions according to a framework for national language policy and deals with the localization and centralization of language. SABS TC 37 promotes multilingualism and the values of standardized terminology principles within all working environments, for example business, industry, science and technology. The committee encourages all endeavours in the field of human language technologies (HLT).

The main aim of this paper is to give a brief overview of the activities of ISO/TC 37 and SABS TC 37 in order to create awareness for the standardization efforts on a wide range of language related activities.

The national committee concentrates on four subcommittees (SCs) and the work relating to their working groups (WGs), namely:

- SC1 “Principles and methods” – standardization of principles, methods, and applications relating to terminology and other language and content resources.
- SC2 “Terminology and lexicography” – standardization of the application of principles and methods in terminology work, terminography and lexicography.
- SC3 “Computer applications for terminology” – standardization of models for information processing and of related coding systems applicable to terminology work and terminography.
- SC4 “Language resource management” – standardization of specifications for computer-assisted language resource management.

Since 2003 several delegations from South Africa had already attended the annual international plenary sessions of ISO/TC 37 where they participate in the activities and liaise with international scholars. SABS TC 37 has P-membership status of ISO/TC 37 and participates in all its activities.

The SABS TC 37 members are active participants and are regarded as experts in the different domains and therefore they are able to contribute to the development of international standards. Several international standards have already been adopted as South African National Standards (SANS) and in some cases the international standards were revised and augmented to suit South African conditions. No local standards are undertaken without considering the ISO published standards or the work in progress. It remains the decision of the local committee to adopt an international standard as a national standard or to develop a new standard if the international standard is not suitable or relevant.

Given the particular multilingual situation in South Africa as well as the nature of standardization practices, it is of great importance to continue sending as representative a team of experts as possible to the annual ISO/TC 37 plenary meetings. The annual ISO/TC 37 plenary meetings provide great opportunities for South African experts to meet up with colleagues working in the same domain and sharing ideas. South African delegations consisting of language practitioners, translators, terminologists, lexicographers, text and speech technologists are formally involved in many of the subcommittees and working groups and their influence is clearly recognised at meetings of this nature.

The main reason for limited involvement by South Africans is logistically and geographically related. Most of the European and even American colleagues are in close proximity, and hence interim working sessions are scheduled on a regular basis, most of it being sponsored by their respective institutions. In recent years the South African delegations were some of the larger groups attending the annual ISO/TC 37 plenary sessions, which was good as it was possible to cover a wide range of SC activities. Following various discussions, it became apparent (and was actually pointed out by international colleagues) that the South African committee system of SABS TC 37 and its annual activities (i.e. the committee meetings and annual workshops) were rather unique compared to what is taking place in other countries. England, for instance, does not have any system in place matching the South African system. It is therefore important to keep the local system intact and continue with its work and sending as representative as possible a group of delegates to these international events. The role of South African influence in a direct as well as indirect manner in international standardization processes should not be underestimated.

South Africa, that is, SABS TC 37, will be hosting the next ISO/TC 37 plenary sessions to be held in June 2013 at the SABS in Pretoria and it would be an opportunity for South Africans and other delegates from SADC and the rest of Africa to attend the various plenary sessions to familiarize them with the activities of ISO/TC 37.

### **An overview of SABS TC37 SC3 - Systems to manage terminology, knowledge and content**

**Karen CALTEAUX**

HLT Competence Area, CSIR Meraka Institute, Pretoria, South Africa

The International Organization for Standardization (abbreviated as ISO, from the Greek *isos* meaning 'equal', for standardization purposes) is a non-governmental, worldwide developer and publisher of International Standards. It consists of a network of national standards institutes from 163 countries. Each country is represented by one member (usually from governmental structures or private sector national partnerships of industry associations) [1]. In the case of South Africa, this member is the South African Bureau of Standards (SABS).

The process for developing new standards is typically as follows: a request for a standard to be developed emanates from a stakeholder entity and is communicated to a member institution; the member institution proposes the development of a new standard to the relevant technical committee; committee members vote on the proposal; if the majority are in favour, the standard is developed. The development of a standard follows a six-step process: from a preparatory stage through to a publication stage. [2].

Technical committees (TCs) consist of experts from industry, business and technical sectors who have requested the development of the standard and who will put it to use. Government representatives, researchers and academics often also form part of the technical committees. The technical teams must ensure that they act in the interests of all their stakeholders [2]. ISO TC37 is the technical committee for Terminology and other language and content resources. It has published 32 standards. The TC consists of four subcommittees (SCs) which focus on specific aspects of the TC's work. This paper presents an overview of the work of SC3 of TC37.

## **Introducing ISO TC37 SC3**

ISO TC37 SC3 focuses on *Systems to manage terminology, knowledge and content*. The scope of the SC's work is the standardization of specifications and modelling principles for systems to manage terminology, knowledge and content with respect to semantic interoperability. This SC has published four standards and has three standards under development. The SABS TC37 SC3 is a mirror committee of the ISO committee with the same title and is responsible for ensuring that the standards developed by the ISO committee are made relevant for our local circumstances. South Africa is one of 22 participating countries in this SC. An additional 11 countries have observer status in this committee, that is, they do not vote on the standards under development. SC3 also has four working groups (WGs) which focus on data elements, vocabulary, data interchange and database management respectively.

While the business case for TC37 centres on the optimal use of terminology as a type of language and content resource that represents knowledge in a specific domain, that of SC3 centres on defining standards and best practices for using computers to manage terminology and other language resources [3].

The multilingual nature of society and its rapid development towards a knowledge society necessitates the development of tools to create, access, process and represent language and terminology resources. The standardization of methods for doing this, ensures interoperability in the management of global content, while simultaneously protecting the investment of commercial companies using information technologies to manage such content [3].

## **Application of SC3 standards**

Language service providers are no longer the only ones carrying out language resource and terminology management. This is because the applications in which terminological data is used are growing beyond their traditional use in translation processes. New applications include, for instance, controlled authoring and search engine optimization. Best practice would dictate that software developers would use SC3 standards to design terminology management systems (TMS) and applications for managing other types of language resources that adhere to the principles defined by SC1, the methods defined by SC2, and the structural schemes defined by SC4.

Various entities would benefit from the adoption of these standards as best practices, including governments, commercial entities, individuals and NGOs which maintain terminology or other language data, or plan to do so in the future [3].

## **Business drivers for adopting SC3 standards**

Applying SC3 standards to the design, development and maintenance of terminology databases has three main benefits from a business perspective:

1. Cutting costs, for example by eliminating poor software development practices.
2. Improving quality, for example by interchanging data between different databases and repurposing terminology from databases in other computerised applications.
3. Increasing the competitive advantage of the business, for example by ensuring that the users of the software benefit from a predictable level of quality.

## **Challenges**

Two challenges around the development and application of the TC37 standards are cited most often:

1. A lack of awareness among language service providers and language content creators of the work of the TC and the benefits of adopting standards as best practices in their endeavours.
2. Difficulties in obtaining funding for the development and application of standards, because it is often difficult to quantify the benefits of applying TC37 standards.

This paper forms part of an endeavour by SABS TC37 to address these challenges locally.

## References

- [1] International Organization for Standardization. About ISO. <http://www.iso.org/iso/about.htm>. Accessed on 2012-03-31.
- [2] International Organization for Standardization. How ISO develops standards. [http://www.iso.org/iso/about/how\\_iso\\_develops\\_standards.htm](http://www.iso.org/iso/about/how_iso_develops_standards.htm). Accessed on 2012-03-31.
- [3] International Organization for Standardization. Business Plan of TC37. [http://www.iso.org/iso/standards\\_development/technical\\_committees/other\\_bodies/iso\\_technical\\_committee.htm?commid=48136](http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48136). Accessed on 2012-03-31.

## **The standardisation of the application of principle and methods in terminology, terminography, lexicography, translation and interpreting work – the role of TC37/SC2**

**Hatu MACHABA & Mapula GAFFANE**  
Department of Arts and Culture, South Africa

Subcommittee 2 (SC2) is one of the four subcommittees of ISOTC37. The scope of ISOTC37/SC2 is the standardization of the application of principles and methods in terminology work, terminography, lexicography, translation and interpreting. The main aim of this paper is to outline the activities of ISOTC37/SC2 in order to create awareness of the work done internationally and nationally relating to standardization of the application of principles and methods for terminology, lexicography, translation and interpreting work.

Considering the unique multilingual nature of South Africa and its accelerated development towards a multilingual information and knowledge society, terminology development and lexicography remain the key elements to facilitate knowledge sharing. Terminologies provide the building elements of content for various subject matters, therefore, they play a basic and critical role in education as well as in all situations of professional, commercial and scholarly communication.

TC37/SC2 contributes significantly to the harmonisation of the methods for the compilation, presentation and management of terminological data in order to achieve consistent, high-quality and ready-to-use terminology in all domains of knowledge and professional and commercial environments.

With the current debate in South Africa on how to assess the quality of translations, the translation and interpretation processes developed by TC37/SC2 come at an opportune time for the South African translation community. These processes are oriented towards promoting quality assurance through the various phases of a translation or interpreting project. A new project on assessment of translations with a focus on quality criteria has the potential to address a highly controversial and contested matter of assessment of the quality of translations and will

hopefully address the concern raised by translators of low quality translations which occur due to market pressures.

SC2 standards also contribute to an increased coherence and consistency in definitions of technical terms. This improves the consistency and coherence of all standards, regardless of their technical domain.

Language codes are a core component of the TC37/SC2 programme of work. The ISO639 series of standards specifies machine-readable codes for languages and linguistic regions that are critical to the entire software industry. The local committee's active participation in the development of these two- and three-letter codes ensured the inclusion in these standards of codes representing all South African languages. Given the wide range of topics and aspects of terminology, lexicography, translation and interpreting covered in the TC37/SC2 standards that are applicable to the South African context, continued active participation of the national standardization committee in the development of standards and the increased wider participation by experts from various relevant institutions in the country can never be overemphasized.

## **Ontologies and standardisation in ISO/TC 37**

**Laurette PRETORIUS**

Department of Interdisciplinary Research, College of Graduate Studies, University of South Africa

The challenges posed by the exponential growth in the amounts of information that are available electronically are well known. One such a challenge stems from the fact that this information is mostly encoded in natural language as unstructured text in the many different languages of the world. Making sense of this information not only requires advanced technological support for the processing and understanding of the multitude of natural languages, but also relies on sophisticated knowledge representation. This has led to the intensive research and development of a rapidly increasing range of so-called semantic technologies by means of which sense may be made of the huge and increasing amounts of information available in digital form. The semantic technologies on which we focus in this paper are what have become known as *ontologies*.

The scope of ISO/TC 37 is the “Standardization of principles, methods and applications relating to terminology and other language and content resources in the contexts of multilingual communication and cultural diversity” [1, p. 3]. It is therefore not surprising that ISO/TC 37's focus on language resources should also manifest in a strong interest in standardization initiatives in knowledge representation, and in particular in ontologies. Indeed, the specific scope of ISO/TC 37/SC 3 is “Standardization of specifications and modelling principles for systems to manage terminology, knowledge and content with respect to semantic interoperability” [1, p. 4].

Two important recent ISO/TC 37 ontology initiatives are discussed:

1. WG 8: “Task force on Ontology” is made up of individuals from various ISO/TC 37 working groups with an interest in ontology. Its initial interest was in compiling a taxonomy of ontology projects. It currently exists as ISO/TC37/WG8 “Ontologies - Linguistic, terminological and knowledge organization aspects” [2]. The first meeting of ISO/TC 37/WG 8 took place at the 2011 ISO/TC 37 meeting in Seoul, Korea. At these

meetings its focus was redirected and its purpose is still in flux [3] - a matter for discussion at the 2012 ISO/TC 37 meeting.

2. The ISO/TC 37/SC 3 “Ontology Integration and Interoperability (OntoIOP)” initiative started in April 2011 and currently has the status of newly approved project ISO/NP 17347 [4]. ISO/TC37/SC3/WG3: ISO WD 17347 – “Ontology Integration and Interoperability (OntoIOP) – Part 1: The Distributed Ontology Language (DOL)” [5] is currently under intensive international discussion.

In practical applications, multiple ontologies and logics are usually necessary for achieving semantic integration and interoperability. OntoIOP focuses on interoperability both among ontologies (i.e. concerning matching, alignment, and suitable means to write these down) as well as among a variety of ontology languages (e.g. OWL, UML, Common Logic, etc.). By having all these languages as part of a meta-standard, ontology designers will be able to use their ontologies verbatim while also being able to relate them to other ontologies in suitable and appropriate ways [6]. In particular, the above-mentioned Part 1 of the OntoIOP draft standard provides a meta-language (DOL) for logically heterogeneous ontologies, modular ontologies, formal and informal links between ontologies/modules, annotation and documentation of ontologies [5].

The purpose of this paper is twofold. Firstly, the focus is on ontologies in order to contextualise the ISO/TC 37 initiatives. We briefly address the following questions: (a) What is an ontology? (b) What characteristics differentiate between different kinds of ontologies, as well as between, for example, controlled vocabularies, taxonomies, folksonomies, wordnets, etc.? (c) What roles can ontologies play in knowledge representation and standardisation? Secondly, we provide an overview of the activities, progress and plans of ISO/TC 37/WG 8 and SC 3/OntoIOP as presented in various available reports and explored at the meeting of ISO/TC 37 in Madrid, Spain, 24-29 June 2012.

## References:

- [1] BUSINESS PLAN of TC37: Terminology and other language and content resources. Available at [http://www.iso.org/iso/standards\\_development/technical\\_committees/business\\_plans.htm](http://www.iso.org/iso/standards_development/technical_committees/business_plans.htm)
- [2] ISO TC 37. Available at [http://www.iso.org/iso/standards\\_development/technical\\_committees/other\\_bodies/iso\\_technical\\_committee.htm?commid=48104](http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48104)
- [3] Wright, S.E. (2011). ISO TC 37: Terminology and other language and content resources, OntoIOP Telecon, October, 20. Available at [http://ontolog.cim3.net/file/work/OntologiesAndStandards/2011-10-20\\_ISO-NP-24707\\_OntoIOP/ISOTC37\\_KRRs--SueEllenWright\\_20111020.pdf](http://ontolog.cim3.net/file/work/OntologiesAndStandards/2011-10-20_ISO-NP-24707_OntoIOP/ISOTC37_KRRs--SueEllenWright_20111020.pdf)
- [4] ISO/NP 17347. Available at [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=59599](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59599)
- [5] Mossakowski, T., Kutz, O., Lange, C. (2011). Ontology Integration and Interoperability (OntoIOP) – Part 1: The Distributed Ontology Language (DOL). IAOA/OOR/Ontolog “Ontologies and Standards” mini-series. Available at [http://ontolog.cim3.net/file/work/OntologiesAndStandards/2011-10-20\\_ISO-NP-24707\\_OntoIOP/OntoIOP--TillMossakowski-OliverKutz-ChristophLange\\_20111020.pdf](http://ontolog.cim3.net/file/work/OntologiesAndStandards/2011-10-20_ISO-NP-24707_OntoIOP/OntoIOP--TillMossakowski-OliverKutz-ChristophLange_20111020.pdf)
- [6] Mossakowski, T. (2011). [CL] ISO initiative OntoIOP (Ontology interoperability). Available at <http://philebus.tamu.edu/pipermail/cl/2011-May/002240.html>

## **Standards for Language Resource Management (ISO/TC37/SC 4)**

**Justus ROUX**

North-West University

**Sonja BOSCH**

University of South Africa

The aim of this paper is to create awareness of the rules for language resources as endorsed in the standards of ISO Technical Committee (TC) 37 Sub-Committee (SC) 4 from which all linguistic communities can benefit. These standards promote the development of high-quality language resources such as text and speech corpora, lexica, ontologies and terminologies, as well as language technology applications.

We give a snapshot of current activities related to ISO standards specific to language resource management as established by the TC37/SC 4 by addressing the objectives of the SC4, areas of language resource management, as well as the relevance of language resource management standards in the South African language technology context.

### **Objectives of SC4**

The so-called scope of SC4 is the standardization of specifications for computer-assisted language resource management. Therefore the main objective of SC4 (Business Plan of TC37, 2010) is to “specify principles, methods, and data structures for creating, coding, processing and managing language resources, in particular written and speech corpora, lexical resources, and resources representing morphology, semantics, syntax, and so forth.” Data used in natural language processing applications is also covered by standards emanating from these specifications.

If new developments in language technology, knowledge and information management satisfy the norms of international standardization, the applicability and reusability of language resources are maximized; language resources of different kinds become available in machine-readable format, and innovation leading to development of methods and tools for managing language resources is supported.

According to the Common Language Resources and Technology Infrastructure (2009), standards for the management of language resources and tools can also be used for facilitating the development of conversion tools that translate existing resources and tools with heterogeneous data formats into interchange formats and back. It is emphasized that for the development of new resources and tools, standards should be adhered to as closely as possible in order to avoid the need for data conversion in the first place.

### **Areas of language resource management**

A number of specialized areas of language resource management are addressed by SC4 activities, e.g. word segmentation, annotation (semantic, syntactic, and morphological), named entities, metamodels, and multilingual information frameworks. Typical working groups cover themes such as: basic descriptors and mechanisms for language resources (terminology used in language resources, basic mechanisms and data structures for linguistic representation, and a meta-data representation scheme to document linguistic information structures and processes); workflow of language resource management (standardization of guidelines for language validation and net-based distributed cooperative work) etc. (ISO/TC 37/SC4, *s.a.*). Activities of working groups will be discussed in more detail in the presentation.

## **Relevance to the South African language technology context**

The South African Department of Arts and Culture has set up a National Centre for HLT (NCHLT) to facilitate among others, the development of world-class reusable text and speech resources which are to be managed by an agency that will create and maintain a catalogue of available resources and facilitate “the licensing of these to researchers and developers with a view to overcome the obstacles of resource scarcity in the local HLT domain.” (Department of Arts and Culture, 2012). A key requirement of such an HLT infrastructure is interoperability and reusability of resources as well as language-independent merging and comparison of linguistic information. Therefore language resources and tools need to adhere, as far as possible, to common data formats that are compliant with encoding standards and best practices so that information can be exchanged across different resources and at different levels of analysis (Common Language Resources and Technology Infrastructure. 2009).

For South Africa, with its multilingual environment of 11 official languages, the availability of language resources and information content in standardized structures within the NCHLT, will be one of the main success factors for the processing and managing of information and knowledge in machine-readable as well as human-readable form.

To date, TC37/SC4 has published eight standards and is currently working on eight new standards (ISO/TC 37/SC4, *s.a.*). In the presentation, practical examples will be given of instances of the application of language resource standards in the South African context.

### **References:**

Business Plan of TC37. 2010. Available:

[http://isotc.iso.org/livelink/livelink/fetch/2000/2122/687806/ISO\\_TC\\_037\\_Terminology\\_and\\_other\\_language\\_resources.pdf?nodeid=1160801&vernum=-2](http://isotc.iso.org/livelink/livelink/fetch/2000/2122/687806/ISO_TC_037_Terminology_and_other_language_resources.pdf?nodeid=1160801&vernum=-2)

Common Language Resources and Technology Infrastructure. 2009. Available:

<http://www.clarin.eu/files/standards-text-CLARIN-ShortGuide.pdf>

Department Arts and Culture. 2012. Available:

<http://www.dac.gov.za/functions/ArtsCulturalPromotionalDevelopment.html#NLS>

ISO/TC 37/SC4. *s.a.* Available: <http://www.tc37sc4.org/index.php>

## **ADDENDUM: Abstracts omitted from published conference booklet**

### **Luxdico: A Lëtzebuergesch - German online dictionary: what could it learn from South African online dictionaries?**

**Juliane KLEIN**

University of Leipzig, Trier, Germany

Lëtzebuergesch is the national language of Luxembourg, a small country in Europe (approx. 500,000 inhabitants) and is spoken by around 390,000 people. The language belongs to the High Germanic language family and is closely related to the Moselle Franconian dialects spoken in the border region of Germany. It is mainly used as a spoken language in all domains of life in Luxembourg and is an important factor of integration in everyday life. Nevertheless, code-switching is extremely common and French or German are mainly used when talking to non-native speakers. French is the language of legislation. French, German and Lëtzebuergesch are used in court and are administrative languages.

As Lëtzebuergesch became standardized late in the last century and only gained official status as a national language in 1984, its situation can be compared with the situation of the official African languages of South Africa. There are only a few dictionaries that contain Lëtzebuergesch, the majority of dictionaries being either French or German. Nevertheless, the status of the language is still high because it is seen as a unification factor for the people, which implies that a variety of dictionaries would be very useful for language learners and mother tongue speakers.

The dictionary discussed in this presentation is the *Luxdico*, currently the only available completed online Lëtzebuergesch – German dictionary. It is based on the hard copy version of the dictionary, which was first compiled as a French – Lëtzebuergesch dictionary before the German version was compiled. The dictionary was compiled for German mother tongue speakers who work in Luxembourg and want to be able to communicate in Lëtzebuergesch.

It is compared with two African language online dictionaries that were compiled by dedicated non-professional lexicographers who also wanted to provide a tool that helps people to communicate. The first one is a general isiZulu-English dictionary (*isiZulu.net*), the second is a general isiXhosa-English dictionary (*the Xhosa – English Dictionary*).

First, the macro structure of the dictionaries was analysed. In a second step, the central word list and the entry structure of the dictionary entries were analysed. The focus was on user-friendliness, that is, how useful is the dictionary for a language learner. This not only included the actual usage, for example, how difficult was it to find a word, but also the actual contents of the central word list.

As the outer-texts of the *Luxdico* are minimal, *isiZulu.net* and the *Xhosa – English Dictionary* can be used as examples for useful outer-texts. Although all three dictionaries are quite self-explanatory, only the African language dictionaries provide a user-guide. The *Luxdico* is also missing information on the most important grammatical features of the language. *IsiZulu.net* includes a basic overview of the most important grammatical features of isiZulu, including noun classes and concords. The *Xhosa – English Dictionary* also provides the user with a basic isiXhosa grammar and some facts about the language.

Neither the *Luxdico* nor the *Xhosa – English Dictionary* include information on pronunciation or sound files. Here, *isiZulu.net* could be used as a positive example that is worth copying, as it provides the user with pronunciation information, including sound files for all sounds in both languages.

The entry structure of *Luxdico* is different from the African language dictionaries, as the former only includes a lemma plus the translation, whereas the latter two are more detailed. *isiZulu.net* provides additional information on the part of speech, relevant grammatical information and pronunciation information in IPA. The *Xhosa – English Dictionary* also provides part of speech and grammatical information about the lemma, as well as usage examples for both languages.

Although the *Luxdico* is useful for German learners of Lëtzebuergesch it could be improved by copying some features of African language online dictionaries. Nevertheless, as the *Luxdico* is currently the only freely available online dictionary that covers all letters (there are some that only cover parts of the alphabet) its value for the language and its speakers and learners is not be underestimated and it should indeed be seen as a very useful tool for language learners. The critical remarks here should be understood as ideas to improve the dictionary. The African

dictionaries referred to were chosen as positive examples because the languages are in a similar position and there is no reason why a European language should not learn from other languages on other continents that share the same problems, as those languages will also face similar problems in compiling dictionaries.

#### References:

*isiZulu.net*: <http://isizulu.net/>

*Luxdico*: <http://dict.luxdico.com/deu/>

*Xhosa – English dictionary*: <http://www.xhosadictionary.com/index.php>

### Double Pluralism of class 6 (ama-) and 10 (iiN-) nouns in isiNdebele: A case study in the isiNdebele Bilingual Dictionary

**K. Sponono MAHLANGU**

Isihlathululi-Mezwi SesiNdebele, Department of African Languages,  
University of Pretoria, South Africa

This article focuses on the challenges faced by isiNdebele lexicographers in the lemmatization of double pluralised lemmas of class 6 (ama-) and class 10 (iiN-) nouns. In isiNdebele a stem-based lemmatization strategy was adopted for the lemmatization in the bilingual dictionary of 2006. The IsiNdebele bilingual dictionary is the only dictionary in isiNdebele that has adopted a stem based lemmatization strategy thus far. The article attempts to show that the publication of any dictionary should not only be the result of the preceding compilation activities but has to be regarded as the culmination of a much more comprehensive set of activities involving the so-called lexicographic process (Gouws and Prinsloo, 2005:9). When planning it is important to realise exactly what the lexicographer hopes to achieve with the inclusion of each entry and to know what kind of a contribution such an entry makes in achieving the purpose of a dictionary. The main argument to date has been that stem lemmatization is accepted as the best strategy for conjunctively written languages (Prinsloo, 2011:171). Van Wyk (1995:83) also points out that the stem tradition is characteristic of the Nguni languages.

The demise of isiNdebele class 11 has meant that all nouns that had previously belonged to this class in Southern Ndebele had to be channelled to another class. A striking feature of many of the erstwhile Class 11 nouns that have been transferred to Class 5 in Southern Ndebele is that they have a double plural form, that is, they may either take their corresponding plural form in Class 6 (ama-) or in class 10 (iiN-) or in both classes. Class 5 nouns can form their plural by substituting the singular prefix with either the plural prefix ama- or iiN- that lose their aspiration when transposed to class 10 while they retain their aspiration when they appear in Class 6. The challenge encountered is that the issue of aspiration, de-aspiration as well as the frequency was not considered when lemmatizing these noun classes, thus the inconsistencies.

As an example in the isiNdebele bilingual dictionary of 2006, the lemma is given followed by the prefix/prefixes in brackets. Then the abbreviated part of speech in italics follows. Finally definitions are provided. For example, lemmas appear as follows in the isiNdebele Bilingual Dictionary (2006):

**Ibhigiri** (mug) has been lemmatized as **-bhigiri (i-/ama-, i-/iim-)** *bz. mug*

**Cl.5 (i-) > Cl.6 (ama-); Cl.10 (iim-)**

**Ibhere** (bear) has been lemmatized as **-bhere (i-/ama-)** *bz. bear*

**Cl.5 (i-) > Cl.6 (ama-)**

**Ibherede** (beret) as **-bherede (i-/iim-)** bz. beret, round flat cap with no peak.

**Cl.5 (i-) > Cl.10 (iim-)**

From the examples given it was found that there are inconsistencies in the lemmatization of nouns with double pluralised prefixes of class 6 (ama-), class 10 (iiN-), or both, in the isiNdebele Bilingual Dictionary.

The article will start by conducting an in-depth investigation of the effects of the double pluralism of class 6 and class 10 nouns, and also analyse them. Reference will be made to the studies by Mpungose in *Lexikos* 8, described in his article 'Analysis of the Word-Initial Segment with Reference to Lemmatizing Zulu Nasal Nouns', as well as to de Schryver's article in *Lexikos* 20 entitled 'Revolutionizing Bantu Lexicography' and Prinsloo's article 'A Critical Analysis of the Lemmatisation of Nouns and Verbs in isiZulu', in *Lexikos* 21. Following on from this, the paper will attempt to show that a stem based lemmatization strategy is not an appropriate strategy for nouns with double pluralism. The problems that are encountered can be avoided, if the 'word principle' for lemmatization is adopted since nouns are entered as complete words. The presentation will be supported by examples drawn from the isiNdebele Bilingual Dictionary (2006). Finally, suggestions will be made for the improvement of the dictionary and the article will also assist lexicographers in the maintenance of consistency when lemmatizing double pluralised nouns.