



# **AFRILEX 2011**

## **African Association for Lexicography**

### **Programme & Abstracts**

**16th Annual International Conference  
University of Namibia, Windhoek  
5th – 7th July 2011**



**University of Namibia  
Department of Language and Literature Studies**

**PHAROS**  
WOORDEBOEKE | DICTIONARIES

© 2011 AFRILEX, African Association for Lexicography  
ISBN 978-0-620-50705-9



**AFRILEX**  
**African Association for**  
**Lexicography**

**Programme & Abstracts**

**16th Annual International Conference**  
**University of Namibia, Windhoek**  
**5th – 7th July 2011**

*Hosted by:* *Department of Language and Literature Studies*  
*University of Namibia, Windhoek, Namibia*

*Conference Organiser:* *Dr HL Beyer*

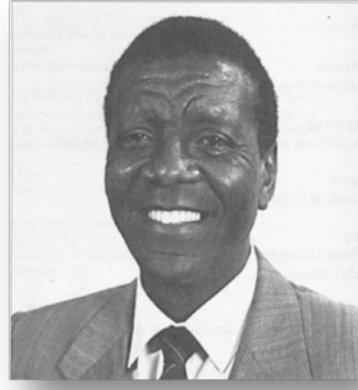
*Abstract reviewers:* *Dr VM Mojela, Dr TJ Otlogetswe, Prof. DJ Prinsloo,*  
*Prof. E. Taljard*

*Abstract booklet editor:* *Ms J Wolvaardt*

## AFRILEX HONORARY MEMBERS



Professor R. H. Gouws



Professor A.C. Nkabinde

## AFRILEX Board 2009 – 2011

<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>Registrar:</b>	Mr JM Motsapi
<b>Organiser:</b>	Ms J Wolvaardt
<b>Members:</b>	Prof. DJ Prinsloo Dr TJ Otlogetswe
<b>Co-opted members:</b>	Dr JM Lubinda (2010) Dr HL Beyer (2011)

## **AFRILEX 2011 – A Few Words from the President**

On behalf of the Afrilex Board, I would like to welcome you to the 16<sup>th</sup> *Annual International Conference of the African Association for Lexicography*, also known as 'AFRILEX 2011'. This year's edition takes place in Windhoek, Namibia, after we were in Gaborone, Botswana, last year, and will go to Pretoria, South Africa, next year. As an association that aims to bring together all lexicographic activities that take place on the African continent, as well as all friends of Afrilex from further afield, the Afrilex Board is particularly pleased to see that Afrilex is indeed going places.

AFRILEX 2011 has been meticulously prepared by a local organising team headed by Dr Herman L. Beyer, assisted by Dr Chrisna M. Beuke-Muir and Ms Petronella F. Genis, from the Department of Language and Literature Studies, University of Namibia. Their dedicated, professional and original conference website (available at <http://afrilex2011unam.webs.com>) will remain a model to all of us for years to come.

This year's abstract adjudication process was not only managed by Jill Wolvaardt, but in the process she also developed, together with Prof Elsabé Taljard, a tracking system that they now wish to make available to us for future conferences. Congrats to them, and thanks to their referees and the members of the programme committee. The *Abstracts Booklet* which you are holding was also edited by Jill Wolvaardt.

AFRILEX 2011 promises to be another stellar gathering, with speakers coming from a dozen different countries in Africa and Europe, namely Belgium, Botswana, Denmark, DR Congo, Gabon, Germany, Namibia, Poland, South Africa, Spain, Uganda, and Zimbabwe. Two workshops precede the conference proper, one presented by Dr Herman Beyer on evaluation criteria for Namibian school dictionaries, the other convened by Prof. Danie Prinsloo on intelligent and dynamic electronic dictionaries. We also continue the tradition of giving the floor to dictionary publishers during a Publishers' Session, slotted in between the workshops and the cocktail party. Our keynote speakers will be Prof. Arleta Adamska-Sałaciak, from Poznań, Poland, who will be talking on lexicographic equivalence, and Prof. Wilfrid Haacke, from Windhoek, Namibia, who will describe the software used to compile his Khoekhoegowab Dictionary.

A special word of thanks is also due to Pharos (NB Publishers) for their continued sponsoring of our association. May this lexicographic gathering be as successful as all previous editions of our annual international conference!

Gilles-Maurice de Schryver  
President: AFRILEX.

# CONTENTS

<b>Programme</b>	<b>6</b>
<b>Keynote Addresses</b>	<b>10</b>
<b>I</b> Equivalence in a Semantic and Lexicographic Perspective: A Case Study <b>Arleta ADAMSKA-SALACIAK</b>	10
<b>II</b> Optimising Data Utilisation in Lexicography: The Case of the Khoekhoegowab Dictionary <b>Wilfrid H.G. HAACKE</b>	11
<b>Parallel Sessions</b>	<b>12</b>
<i>A Flexible Database Model for Multiple Dictionaries</i> <b>Richard ALMIND</b>	12
<i>One data base, four monofunctional dictionaries</i> <b>Inger BERGENHOLTZ</b>	14
<i>The lexicographical treatment of conjunctions in Afrikaans dictionaries,</i> <b>Nerina BOSMAN &amp; Annél OTTO</b>	16
<i>Needs Adapted Data Presentation in e-Information Tools</i> <b>Theo BOTHMA &amp; Henning BERGENHOLTZ</b>	17
<i>The automatic lemmatiser as a tool to enhance access to Setswana dictionaries,</i> <b>J.H. BRITS &amp; Rigardt PRETORIUS</b>	19
<i>Designing a Dictionary for the De la Bat School for the Deaf in Worcester</i> <b>Hanelle FOURIE</b>	21
<i>Equivalent Selection in LSP e-Lexicography: A Case Study with Spanish Accounting Terms</i> <b>Pedro FUERTES-OLIVERA</b>	22
<i>Linking dictionary and corpus data in online language tools</i> <b>Ulrich HEID &amp; Danie PRINSLOO</b>	24
<i>Pronunciation guides for a South African English school dictionary</i> <b>Lorna HILES</b>	26
<i>Comparing two African language online dictionaries from a non-native speaker's perspective</i> <b>Juliane KLEIN</b>	28
<i>Challenges of predictability and consistency in the first comprehensive Sotho dictionary</i> <b>Inge KOSCH</b>	29
<i>Packaging phonetic information in dictionaries of Bantu languages</i> <b>John M. LUBINDA</b>	31
<i>Towards a monolingualised bilingual learner's dictionary</i> <b>Jana LUTHER</b>	32

<i>The Ordering of Lexical Entries in Tshivenḁa Dictionaries with Special Reference to Tshivenḁa/English Ṱhalusamaipfi Dictionary</i>	<b>James MAFELA</b>	34
<i>Legal issues in lexicography: an examination of the handling of brand names in the Isichazamazwi SesiNdebele</i>	<b>Mandlenkosi MAPHOSA</b>	35
<i>Linguistic diversity and genetic relatedness of African and Southern Bantu languages: opportunities and challenges in the area of lexicography and multi-lingual mother-tongue-based education.</i>	<b>Samba Buri MBOUP</b>	36
<i>Purism and inadequacies: A case study of the effects of strict corpus-based dictionary writing in Sesotho sa Leboa.</i>	<b>Victor M. MOJELA</b>	38
<i>What is right or wrong: In Quest of formal Lusoga</i>	<b>Minah NABIRYE</b>	39
<i>Compilation of An English-French Dictionary of Phonetic Sciences: An Ongoing Project</i>	<b>HS NDINGA-KOUMBA-BINZA</b>	41
<i>The treatment of etymology in Setswana dictionaries</i>	<b>Thapelo J. OTLOGETSWE</b>	43
<i>A critical analysis of the lemmatisation of nouns and verbs in isiZulu</i>	<b>Danie PRINSLOO</b>	44
<i>Requirements for an on-screen presentation of an electronic dictionary of German collocations for Afrikaans-speaking learners of German as a foreign language</i>	<b>Hilma A. REINHARDT</b>	46
<i>A global approach for a dictionary of Lingála: from localisation of software to a lemmatisation strategy</i>	<b>Bienvenu SENE MONGABA</b>	49
<i>Meaning explanation in learners' e-dictionaries: Current strategies and their theoretical, functional, user and practical motivation</i>	<b>Piet H. SWANEPOEL</b>	50
<i>Do we need a (new) theory of lexicography?</i>	<b>Sven TARP</b>	51
<i>What do foundation phase learners know about a dictionary?</i>	<b>Michele VAN DER MERWE</b>	53
<i>The impact of borrowing as one of the term creation strategies in South African Indigenous languages: The translator's perspective</i>	<b>Boni P.K. ZUNGU</b>	54

# AFRILEX 2011: PROGRAMME

**TUESDAY 5 JULY 2011**

## PRE-CONFERENCE WORKSHOPS

**09:00 – 11:00 Workshop 1**

*Presenter:*

Herman Beyer (University of Namibia)

*Topic:*

Towards Evaluation Criteria for School Dictionaries for Namibian Schools

**11:00 – 11:30 TEA**

**11:30 – 13:30 Workshop 2**

*Convenor:*

D.J. Prinsloo (University of Pretoria)

*Topic:*

Intelligent and dynamic electronic dictionaries

**13:30 – 14:30 LUNCH**

**14:30 – 16:00 Publishers' Session**

**16:00 – 16:30 TEA**

**19:00 – Cocktail Party**



## WEDNESDAY 6 JULY 2011

**08:30 – 09:15 Registration**

**09:15 – 09:30 Official Opening**

Word from the President of Afrilex

**G.-M. de Schryver**

A word of welcome on behalf of the University of Namibia

### Keynote Address I

<b>09:30 – 10:25</b>	Equivalence in a semantic and lexicographic perspective: A case study <b>Adamska-Salaciak A.</b>
----------------------	--

**10:30 – 11:00 TEA**

### Parallel Sessions

<b>11:00 – 11:25</b>	Meaning explanation in learner's e-dictionaries: Current strategies and their theoretical, functional, user and practical motivation <b>Swanepoel P.</b>	Pronunciation guides for a South African English school dictionary <b>Hiles L.</b>
<b>11:30 – 11:55</b>	Needs Adapted Data Presentation in e-Information Tools <b>Bothma T.J.D. &amp; Bergenholtz H.</b>	Packaging phonetic information in dictionaries of Bantu languages <b>Lubinda J.M.</b>
<b>12:00 – 12:25</b>	Linking dictionary and corpus data in online language tools <b>Heid U. &amp; Prinsloo D.J.</b>	Compilation of An English-French Dictionary of Phonetic Sciences: An Ongoing Project <b>Ndinga-Koumba-Binza H.S.</b>
<b>12:30 – 12:55</b>	Equivalent Selection in LSP e-Lexicography: A Case Study with Spanish Accounting Terms <b>Fuertes-Olivera, P.A.</b>	The impact of borrowing as one of the term creation strategies in South African indigenous languages: The translator's perspective <b>Zungu B.P.K.</b>

**13:00 – 14:00 LUNCH**

### Parallel Sessions

<b>14:00 – 14:25</b>	Requirements for an on-screen presentation of an electronic dictionary of German collocations for Afrikaans-speaking learners of German as a	A global approach for a dictionary of Lingála: from localization of software to a lemmatization strategy <b>Sene Mongaba B.</b>
----------------------	--	--

	foreign language <b>Reinhardt H.A.</b>	
<b>14:30 – 14:55</b>	Do we need a (new) theory of lexicography <b>Tarp S.</b>	<b>To be confirmed</b>

**15:00 – 15:30 TEA**

### Parallel Sessions

<b>15:30 – 15:55</b>	Towards a monolingualised bilingual learner's dictionary <b>Luther J.</b>	Challenges of predictability and consistency in the first comprehensive Sotho dictionary <b>Kosch I.M.</b>
<b>16:00 – 16:25</b>	Designing a Dictionary for the De la Bat School for the Deaf in Worcester <b>Fourie H.</b>	The lexicographical treatment of conjunctions in Afrikaans dictionaries <b>Bosman N. &amp; Otto A.</b>
<b>16:30 – 16:55</b>	The treatment of etymology in Setswana dictionaries <b>Otlogetswe T.J.</b>	What do foundation phase learners know about a dictionary? <b>Van de Merwe M. &amp; Steyn M.</b>

**17:00 – 18:00 Annual General Meeting**

**19:00 Conference Dinner**

## THURSDAY 7 JULY 2011

### Keynote Address II

<b>09:00 – 09:55</b>	Optimising Data Utilisation in Lexicography: the Case of the Khoekhoegowab Dictionary. <b>Haacke W.</b>
----------------------	--

**10:00 – 10:30 TEA**

### Parallel Sessions

<b>10:30 – 10:55</b>	The Ordering of Lexical Entries in Tshivenda Dictionaries with Special Reference to Tshivenda/English T̄halusamaipfi Dictionary <b>Mafela M.J.</b>	Flexible Database Model for Multiple Dictionaries <b>Almind R.</b>
----------------------	---	---

<b>11:00 – 11:25</b>	Purism and inadequacies: A case study of the effects of strict corpus-based dictionary writing in Sesotho sa Leboa <b>Mojela V.M.</b>	One data base, four monofunctional dictionaries <b>Bergenholtz, I.</b>
<b>11:30 – 11:55</b>	Comparing two African language online dictionaries from a non-native speaker's perspective <b>Klein J.</b>	The automatic lemmatiser as a tool to enhance access to Setswana dictionaries <b>Brits J.H. &amp; Pretorius R.</b>
<b>12:00 – 12:25</b>	What is right or wrong: In Quest of formal Lusoga <b>Nabirye M.</b>	A critical analysis of the lemmatisation of nouns and verbs in isiZulu <b>Prinsloo D.J.</b>

**12:30 – 12:40 Closure**

**12:40 LUNCH**

## **FRIDAY 8TH JULY**

### **Post-Conference Excursion**

**NOTE:** The post-conference excursion is only available to delegates (and partners) who registered and paid before 1 June 2011.

### **Programme**

from 09:00	Delegates are transported to a central pick-up point in Windhoek
10:45	Pick-up by the tour operator at the central meeting point in Windhoek
11:00	Depart to Okapuka Ranch (±30km north of Windhoek)
11:30	Arrival at Okapuka Ranch and welcoming drink
11:45	Animal spoor tracking guided activity (one hour; includes walking)
12:45	Return to the lodge, freshen up and prepare for lunch
13:15	Set menu lunch (drinks for delegates' accounts)
14:15	Prepare for game drive
14:30	Guided game drive (2h30min)
17:00	Lion feeding activity (30min)
17:30	Return to Windhoek and lodgings

**NOTE:** The programme is subject to conditions specified by the tour operator and Okapuka Ranch and may be amended or completely replaced without prior notice.

## KEYNOTE PRESENTATION I

### Equivalence in a semantic and lexicographic perspective: A case study

**Prof. Arleta ADAMSKA-SALACIAK**

Department of Lexicology and Lexicography, School of English,  
Adam Mickiewicz University, Poznań, Poland.

As is commonly acknowledged, the biggest problem facing bilingual lexicographers is lack of perfect interlingual equivalence. For centuries, linguists and philosophers alike have been quick to point out that the lexicons of natural languages are not isomorphic. Reasons for this anisomorphism can be sought on three interrelated planes: language structure, extralinguistic reality, and conceptualisation (respectively, Zgusta's (1971) 'difference in the inventory of non-designative items', 'difference in denotata', and 'difference in designata'). Simply put, the differences may reside in the language, the world, or the mind – or any combination of these.

The inevitable result – of the dashed expectations of perfect equivalence on the one hand, and of the practical need for bilingual dictionaries on the other – is that what goes under the name of *lexicographic equivalence* is something highly heterogeneous. Dictionary equivalents come in different shapes and sizes, depending, to some extent, on the language pair involved and, to a much greater extent, on the particular lemma treated in a given entry. Growing awareness of this fact has resulted, over the years, in the creation of several tentative typologies of equivalence. One such attempt, building on existing metalexicographic work, will be presented here. A summary discussion of some auxiliary strategies for overcoming non-equivalence will conclude the theoretical part of the presentation.

The practical part comprises a detailed analysis of a single problem encountered while preparing LSW2, a new edition of a bilingual dictionary for Polish learners of English. The task at hand involved choosing a viable counterpart, from among a series of imperfect equivalence candidates, for a semantic neologism which, having become extremely popular in recent years, qualified for inclusion in the Polish-English part of the dictionary. In the course of discussing the case, reference will be made both to the metalexicographic categories introduced earlier and to such concepts developed by lexical (especially cognitive) semantics which may prove helpful in capturing the subtle differences in meaning between the source-language item and its competing target-language renditions.

The purpose of this micro-scale dissection of a single specimen is to demonstrate that, even with all the analytical tools at our disposal, we are still some way from being able to classify, let alone deal with, all the instances of imperfect interlingual correspondence that come our way in the day-to-day business of bilingual dictionary making. Persisting in the efforts to advance our understanding of the complex issues covered by the blanket term *lexicographic equivalence* thus seems crucial for improving the treatment of meaning in bilingual lexicographic products.

#### References:

- LSW2 = Fisiak, J., Adamska-Salaciak, A., Idzikowski, M., Jagła, E., Jankowski, M., Lew, R. (eds.). 2011. *LONGMAN Słownik Współczesny Angielsko-Polski, Polsko-Angielski*. Harlow: Pearson Education Limited. 2<sup>nd</sup> edition.
- Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

## KEYNOTE PRESENTATION II

### **Optimising Data Utilisation in Lexicography: the Case of the *Khoekhoegowab Dictionary*.**

**Prof. Wilfrid H.G. HAACKE**

Dept. Language & Literature Studies ,University of Namibia, Windhoek, Namibia

Lexicographic practice in Southern Africa has been making rapid strides in the direction of multifunctionality of databases through the advent of generic lexicographic software with powerful retrieval facilities. Yet a considerable number of projects – especially in Khoesan languages – still use or have recently used a word processor with the sole objective of compiling a dictionary. Most of these projects suffer from a long standing legacy of having been started single-handedly as by-products of fieldwork in a language, at a time when affordable dedicated lexicographic software was not yet available.

In view of this situation, a case study of the *Khoekhoegowab Dictionary Project* (formerly *Nama Dictionary Project*) is presented. This venture started in the early 1980s with handwritten index cards but has, since the early 1990s, advanced to custom-made software based on an off-the-shelf database manager using a now obsolete DOS platform and other proprietary software. In 2010, the database was transferred to customised open-source spreadsheet software in order to ensure its continuous utility for future publications and research.

The intention of the project was to compile a bilingual and bidirectional Khoekhoegowab-English/English-Khoekhoegowab dictionary that is marked for (and sorted according to) lexical tone. Hence the software was designed with these and other specific aims in mind. The concept provides for two strictly distinct stages: firstly, a pre-dictionary stage in which the data would be entered into a relational database by means of a commercially available database management programme, *File Express*; secondly, a dictionary compilation stage in which this data is eventually converted into a print-ready presentation lay-out format by means of custom-made compilation and editing software.

The paper demonstrates how the database was configured for particular purposes, mainly to serve both as template for the lemma and article format in the printed Khoekhoegowab-English dictionary, and as multipurpose retrieval tool for extra-lexicographic purposes. It is emphasised by exemplification that the design of the database configuration is of pivotal importance for the extent of potential data utilisation. The present database provides, for instance, for a set-up that specifies input and output data for lexical tone and therewith has supplied the data for the analysis of the tonology of Khoekhoegowab. This analysis in turn was a prerequisite for systematically marking the dictionary for tone.

Further advantages and spin-offs of using a relational database are briefly enumerated for the particular case, as well as facilities that the software design does not offer. Finally it is reported how, after the actual experience of imminent software failure because of obsolescence, the existing database file was recently exported into an OpenOffice CALC spreadsheet. New software was written to eventually produce a print-ready page lay-out of the dictionary. Before the dictionary compilation will be done for a second edition, further dialect data will be added to the database for some years of ongoing fieldwork. The fact that open-source software is used ensures that the database can survive indefinitely.

It is ultimately suggested that the time, cost, and effort spent on converting lexical data that have been compiled in a word processor document into a structured database with fully fledged

retrieval facilities is, in the long run, amply compensated for by the far-reaching elimination of human error in the dictionary, by the automatisisation of processes like language reversal and sorting, and, finally, the significantly enhanced usability of the data for purposes other than dictionary compilation.

Books that have emanated from the database:

- 1999a Haacke, W. & E. Eiseb. *Khoekhoegowab-English / English-Khoekhoegowab Glossary/Mîdi Saogub*. Windhoek: Gamsberg-Macmillan. 278pp.
- 1999b *The Tonology of Khoekhoe (Nama/Damara)*, Quellen zur Khoisan-Forschung No. 16. Cologne: Rüdiger Köppe. 245pp.
- 2002 Haacke, W. & E. Eiseb. *A KHOEKHOEGOWAB DICTIONARY with an English-Khoekhoegowab Index*. Windhoek: Gamsberg-Macmillan. 754pp.
- 2010 Haacke, W., E. Eiseb & C. Gericke *Khoekhoegowab-Afrikaans Afrikaans-Khoekhoegowab Glossarium/Mîdi Saogub*. Windhoek: Macmillan. 407pp.

## PARALLEL SESSIONS

### A Flexible Database Model for Multiple Dictionaries

**Richard ALMIND**

Centre for Lexicography, Aarhus School of Business,  
Aarhus University, Denmark

The most important aspect of a database for lexicographical use is to remember that it is a vessel for data. Underlying the database is a description of how data elements (objects such as fields and tables) are related to each other, often called a relationship diagram. In a lexicographical context, for instance, it describes that data *A* (“definition”) is related to data *B* (“lemma”) through the key *n*. This construction of interrelated data usually consists of alphanumeric data, which is one among many possible data types, but could also be pure numbers or binary data such as images, sound bites, or video clips. Exactly what is stored in a database is for the editor to decide and the programmer to prepare. The database is not a dictionary in itself nor is the editor application that is used to maintain the data.

The underlying database for any modern dictionary is usually sufficiently flexible to allow for most publishing methods and it would be wrong to constrain oneself to thinking in printed versions of dictionaries alone since the electronic options are more natural to databases, especially since the restrictions of fixed media like print are a hindrance to understanding the true possibilities of electronic dictionaries. Whether the output is print or electronic the main reason to use a database is the possibility to act swiftly on new insights be they political, social, linguistic or other. Wikipedia is the preferred standard against which many online information systems are measured and the underlying method of updating data almost as it is generated in real life is key to understanding database-driven information.

Unfortunately, there are two very large obstacles for traditional lexicographers to understand what database-driven lexicography can lead to. The first is their self-imposed limitation to linguistics. Whereas there is nothing wrong in having fixed limitations in a given field and using those limits to explore a set of possibilities, in this case it quickly reaches the limits of what kind of insights can be found by using linguistic phenomena as a case study for lexicography. A lexicographical tool based on dynamic media like the internet can lead to tools that go far beyond the dissemination of linguistic data and reach much further and much more naturally into information sciences than expected. The second obstacle is somewhat harder to dispel. Being used to thinking in fixed media like print it becomes difficult to understand that lexicographical data as defined in a database are interchangeable building blocks and that the original definition of a dictionary article loses its meaning once the user aspect comes into play. In other words: the data in the database has no function until it becomes visible to the user and keeping a fixed article in mind and applying that image to the design of a database is a restriction best overcome quickly since this type of limitation hinders a proper design, sometimes even making it useless.

Using the *Dictionaries of Accounting* as an example it will be shown that flexibly designed databases or data collections extend the lexicographer’s possibilities to let a dictionary evolve from a traditional type, where all data is shown all the time, to a more modern approach where a small set of data is shown when and as it is needed, an approach suitably explained as “less is more”.

The resulting dictionaries have special focus on different user needs, which are not, however, the focus of the presentation. The main focus is on the design of the underlying database itself,

which opens for new uses not originally intended.

The database behind the *Dictionaries of Accounting* is designed to be a one-to-many relationship between languages where English is at the hub and other languages relate to the English definitions of international accounting terms. This allows for any dictionary in language pairs where English is one of the languages, for instance English-Spanish and Danish-English but not Danish-Spanish. This limitation is not a severe hindrance but necessary for various reasons.

However flawed the design might look at first glance, it is useful in regions with many languages and common law and customs. For instance, since the database has been designed to be flexible, its lexicographic objects or “building blocks” such as collocation-tables, synonym/antonym-constructs etc. are interchangeable and it would be possible to use the framework and editing facilities to create a medical dictionary with Afrikaans or English at the hub and any number of other languages attached to it. It might in fact be possible to create two parallel versions, one in English and one in Afrikaans, and keep them synchronised with a relatively low use of resources. The same could be done with other areas of specialised languages such as law, biology, geography, etc.

### **One database, four monofunctional dictionaries**

**Inger BERGENHOLTZ**  
Kolding Music School, Denmark

Recent years have witnessed an ongoing discussion in metalexigraphy about the difference between mono- and polyfunctional dictionaries. If not all, then at least very many printed dictionaries have been polyfunctional ones. The arguments for them have been of an economic nature - you cannot sell enough copies of a dictionary if it has only one function and does not appeal to a broad group of users. This may be true for some small languages with comparatively few dictionary users. It may also apply for many dictionaries for special languages, be it a small or a large language area. Therefore I started my music dictionary, Politiken Music Dictionary, back in the 1990's with a polyfunctional aim addressing a rather large user group and having at least two functions. The first being to help understand musical terms as they are found by laymen or semi-experts in texts on classical music or in sheet music. The second function was to give more detailed knowledge about classical music, about music instruments, genres, and history and translating the often Italian expressions which you find in sheet music. The background of the printed dictionary was established in a dictionary database. Out of this database all elements of the entries were printed. The only problem was to consider in which order the single elements should be shown in order to make the access easy and understandable (cf. Bergeholtz/Bergeholtz 2007).

The idea of transferring the printed work to the internet was a natural one because there was no existing really profound and scientifically supported music dictionary in Danish. The work on the internet dictionary started after 2004 on the same premises as the printed dictionary and used the same database. But after considerations made by the Center for Lexicography at the University of Aarhus two monofunctional and one polyfunctional dictionary resulted (Bergeholtz/Bergeholtz 2011). The setting up was based on the following thoughts: a polyfunctional dictionary is expected to be used by everyone in any kind of user situation with communicative and cognitive problems. But a dictionary is a tool and good tools are not polyfunctional, they are designed and made to fulfil specific tasks. It should be the same for



information tools. A good tool is a tool conceived for a certain function and for a certain user group for certain needs.

In this first version of the internet music dictionary the database worked as a basis out of which the three dictionaries selected their information. But the conception was still based on traditional thinking about how much you should show to the user: much is good, more is better. In particular, the access to the use of foreign expressions in music theory and in printed music was not optimal. The expressions were, so to speak, hidden between other elements in the first two dictionaries which were called *Understanding of Musical Terms* and *Knowledge about Musical Terms*. Ideas took shape to make the dictionaries more specific in order to avoid information overload.

Therefore we have been working on a new setup after thinking more clearly about the searching of the database and the presentation of the single entries, with the result now of four dictionaries corresponding to four different user needs for information about musical terms all of which are still based on the same database (which is of course constantly enlarged and improved). The dictionaries are called *Understanding of Musical Terms*, *Knowledge about Musical Terms*, *Find a Musical Term*, and *Translation Dictionary*.

Such a differentiation needs different sorts of searching. First of all we shall have to differentiate between two sorts of searching in the database: the minimalising and the maximalising method of searching. The first being where the searching stops after having found a hit in one field and then does not go on. The maximalising search goes on searching in all the fields chosen. Furthermore you foresee searching for absolute search strings, that is, only looking up the exact orthographical form of the word. Like the minimalising search it has the advantage of giving fewer results, but the disadvantage of not giving results which might be relevant for the user, for instance results with the searched expression as part of compounds or being inflected. If wanted you could foresee searching with Boolean operators by writing more than one word and combining them with 'and' 'or' 'but not', e.g. chamber music *and* piano *or* violin *but not* flute.

In both the third and the fourth dictionary you will often get more than one hit (and if you write only a part of a term, in the first dictionary, too). If it is more than ten hits the lemmas will be presented as a list from which you can choose the relevant ones. If you click on them you get the possibility to go on into more details.

The first dictionary, *Understanding of Musical Terms* only presents the lemma(s) and a short explanation and a picture, if there is one. The second dictionary, *Knowledge about Musical Terms*, presents the lemma(s), the same short explanation, and often a rather long explanation with history and other background information, synonyms, references to other entries, internet links, sound examples and pictures. In the third dictionary, *Find a Musical Term*, you can look up expressions which are not necessarily lemmas but are part of the text in the single entries or you can look up all the lemmas which have certain elements in common. The fourth dictionary is the *Translation Dictionary*. Here you can get foreign expressions translated into Danish or look for the foreign word(s) for a Danish term or explanation, that is, the translation can go both ways.

The line of thought behind these dictionaries is that the lexicographer - not the single user - decides how the searching works. In the Center in Aarhus the experts are working on methods to introduce individually defined searching methods. I have not adopted them, waiting for the ongoing experiments. The user does not know how the system works. Searching in different fields does not imply automatically how the results are being presented. You can search in fields where the data are not shown and you can show data from fields where no searching took place.

The paper will show tables after which the searching is organised and examples how it is presented in the dictionaries.

### References:

- Bergenholtz Henning/Inger Bergenholtz, Inger: A timeless music dictionary. In: *Lexikos 17*, 2007, 407-415.
- Bergenholtz Henning/Inger Bergenholtz, Inger: A dictionary is a tool, a good dictionary is a monofunctional tool. In: P.A. Fuertes-Olivera/H. Bergenholtz (eds.): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London & New York: Continuum 2011. (in print)
- Politiken Music Dictionary = Bergenholtz, Inger: *Politikens Musikordbog*. København: Politiken 1996.
- The Danish Internet Music Dictionary = Inger Bergenholtz in cooperation with Richard Almind and Henning Bergenholtz: *Musikordbogen*. Ordbogen.com 2011. ([www.ordbogen.com](http://www.ordbogen.com)).

## The lexicographical treatment of conjunctions in Afrikaans dictionaries

Anél OTTO

Dept of Language and Literature, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa

Nerina BOSMAN

Dept of Afrikaans, University of Pretoria, South Africa

Since the article by Gouws (1992) on the lexicographical treatment of co-ordinating conjunctions, very little research has been done regarding the treatment of different types of conjunctions in Afrikaans dictionaries.

Within the newer approaches to grammar (like functional systemic grammar and cognitive grammar) the traditional distinction between lexicon and grammar is less pronounced. Linguistic categories such as conjunctions are to be found somewhere on a continuum between lexical morphemes (Langacker's term) such as *giraffe* and *encyclopedia* and grammatical morphemes such as *-ing*, *of* and *be*. When a closed set like conjunctions behave both semantically and syntactically as a word class of their own, it is reasonable to expect that they too must receive a special type of lexicographical treatment – currently there is no evidence of this in Afrikaans dictionaries.

Conjunctions function to integrate clauses, and on a structural level, it is their syntactic behaviour (the way in which they influence the word order of the clause which is integrated) that forms the basis for the distinction between the following three types of conjunction in Afrikaans: coordinating conjunctions, subordinating conjunctions and conjunctive adverbs. Since subordinating conjunctions and conjunctive adverbs are syntactically marked in Afrikaans (in contrast to coordinating conjunctions such as *en*, *maar*, *want*, *of*) syntactic sub-categorisation is clearly called for when dealing with these conjunctions in a dictionary (Gouws 1992:93). This sub-categorisation and appeal by Gouws is not indicated in the dictionaries consulted in this research.

Of equal importance is the requirement for explicitly mentioning the **function** of the conjunction, not only within sentence boundaries, but also within broader textual units. Semantic relations like reason, cause, etc. can be indicated in a dictionary by making use of the various taxonomies already available in the literature on cohesion (cf. Halliday and Hasan 1976, Carstens 1997) in the definiens slot.

On a semantic level it does not really matter whether these conjunctions are coordinating or subordinating, but the internal polysemic nature of conjunctions (cf. Messerschmidt & Messerschmidt 2007) should receive systematic treatment in the dictionary.

### **Methodology**

The researchers will use insights gleaned from more than one linguistic approach or theory which among others will include structural linguistics, functional systemic grammar and cognitive linguistics - as presented by Halliday and Hasan (1976), Ponelis (1979), Langacker (1987), Matthiessen and Thompson (1988), Feinauer (1990), Gouws (1992), Bosch (1997), Bosch (1998), Taylor (2003), Verhagen (2001). This method of work is admittedly somewhat eclectic but, as such, not unusual within the lexicographic tradition especially where linguistic concepts are involved. In this manner we hope to clarify the linguistic foundations of our proposed treatment of conjunctions - which is more than what appears to be currently the case. Limitations in the treatment of conjunctions in the three explanatory dictionaries *WAT*, *HAT* and *Pharos Verklarende Afrikaanse Woordeboek* and in one learner's dictionary, *Basiswoordeboek van Afrikaans*, will be critically analysed.

Certain general principles and guidelines for treating conjunctions - as a separate and unique word class - will be proposed. These principles will be illustrated by discussing the current treatment of the Afrikaans conjunctions *dat*, *as* and *of* in some detail and by suggesting alternative treatments of these three conjunctions.

## **Needs Adapted Data Presentation in e-Information Tools**

### **Henning BERGENHOLTZ**

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

### **Theo JD BOTHMA**

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

In the current debate about the status of lexicography there are at least three quite different opinions:

1. Lexicography does not have and does not need any kind of own theory but can use all relevant linguistic theories;
2. Lexicography needs a theory specifically for the lexicographical praxis, but this discipline is still a part of linguistics;
3. Lexicography is a genuine part of information science and can use theories and learn from practice in the information society, but it also needs special theories for lexicography.

It is the last standpoint we will follow in this paper, discussing the information needs in the information society partly using the function theory of lexicography. The paper will briefly address issues regarding information overload and information stress showing how commercial systems try to address this by means of relevance ranking based on system relevance. It is argued that this is insufficient as system relevance does not take into account the unique situation of an individual user in a specific information need environment. We argue that a user has different information needs over time in a specific contextualised environment such as a work task. A user profile can be constructed on the basis of a number of variables such as the user's knowledge

base, language expertise in terms of the subject discipline and the language of the document (L1, L2 etc.). Even though certain aspects of a user's profile remain (fairly) constant (e.g. L1 proficiency, domain knowledge) his/her profile can change in terms of the information needs depending on the specific situation. For example, a user may usually require detailed information aimed at the expert if he/she needs information for a standard work task in which he/she is an expert; however, if the user needs to read background information in a discipline in which he/she is only a lay person, the nature of the information needed to satisfy the information need may change to less detailed information aimed at a lay person. The available information can therefore also be characterised in terms of a number of variables, such as complexity, detail, language etc. To satisfy a user's information needs there should be a match between the user's profile in a given situation and the available information. This will be discussed in more detail and illustrated by means of examples.

Dictionaries as information tools tend to provide as much information as the lexicographer deems necessary, without taking into account the actual information need of a user in a specific situation. Within the function theory of lexicography it is suggested that the information provided be filtered in terms of the situation for which the user needs the information, viz. a communicative situation (either text reception or text production) or a cognitive, interpretative or operative situation. The differences between these situations will be discussed briefly. We will show at the hand of examples how this approach to lexicography is transforming dictionaries as information tools in the dictionaries of the Centre for Lexicography (Centlex) at the University of Aarhus.

The principles and theory underlying the Centlex dictionaries therefore make provision for filtering data in terms of the specific user needs in a given situation. The remainder of the paper will show how modern information technologies and techniques can enable the user to customise the information made available at a much more granular level based on the specific information need in a given situation. Technologies and techniques include advanced search and filtering options, adaptive hypermedia, detailed mark-up of data in the database by means of metadata in an RDF schema and open and linked data. The user is empowered to customise the granularity of the information that he/she accesses in any given information need situation and remains in control of exactly how much or how little information is made available to him/her by the set of e-information tools.

We propose that predefined filters be provided for the most common user situations (as is currently the case in the Centlex dictionaries) and that users with different information needs be empowered to use another search order, to do searches in more/different fields, to have the data presented in a different order or to have data presented from different/additional fields, depending on their specific information need in a given situation. In the end there would be the possibility of an unlimited number of different dictionaries based on the customisation principles discussed above and the number of fields in the database. A good tool is a tool which is able to fulfil the needs of a certain user group by giving quick access to the data, by giving relevant and correct data in a way which is understandable for this user group, but which does not send the user into information death by giving a large amount of unnecessary data. Such a good tool must necessarily be a monofunctional tool prepared for a certain user group with a certain type of information need according to a certain type of user situation.

## The automatic lemmatiser as a tool to enhance access to Setswana dictionaries

**J.H. BRITS**

Language Directorate, North-West University (Potchefstroom), South Africa

**Rigardt PRETORIUS**

School of Languages, North-West University (Potchefstroom), South Africa

Lexicographers have the difficult task of balancing user-friendliness, a budget, limited space and a variety of target audiences when doing their work. The task becomes even more difficult when it comes to the lexicographers of Southern Bantu languages. One way of dealing with these problems would be to go the electronic way – as has been successfully done with the Northern-Sotho dictionary. Unfortunately at this stage, Setswana does not have a commercial electronic dictionary available.

At the Potchefstroom Campus of the North-West University, Setswana is taught as foreign language to students who mainly speak Germanic languages, i.e. Afrikaans and English (we also teach Setswana as first language on the Potchefstroom Campus, dictionary use by mother-tongue speakers is not the focus of this paper. For many of them Setswana is the first encounter they have with a Bantu language with a grammar vastly different from theirs. A good learner dictionary is indispensable for the acquisition of a new language, but as we shall argue, students taking Setswana as foreign language find Setswana paper dictionaries difficult to use. However, there are some resources available in the field of natural language processing and in this paper we want to show how morphology and an automatic lemmatiser might be used to make existing dictionaries more accessible for students of Setswana.

Lemmatisation is a natural language processing procedure that determines the lemma of an input word. It can be used for text mining, to make indexes, list concordances and many other applications in corpus-based research. It plays an important role in the bigger picture of natural language processing, but it has the potential to do so much for the ordinary users of Setswana dictionaries.

In this paper we shall discuss how four Setswana dictionaries are structured. For some dictionaries, students need advanced linguistic knowledge to find the right lemma entry. It is therefore important to give an overview of the lemmatisation procedure followed in Setswana dictionaries. We took the following dictionaries: *Setswana-English dictionary* of Brown (1988), *Setswana-Engels-Afrikaanse woordeboek* of Snyman *et al.* (1990), *Setswana English Setswana dictionary* of Matumo [(1993) based on the *Setswana English dictionary* of Brown (1988)] and the *Kompakte Setswana woordeboek* of Dent (1994). We looked up two nouns (“woman” and “axe”) and two verbs (“to answer” and “to open”) in Setswana and we found the following results of lemmatisation procedures in the Setswana dictionaries:

	Setswana Dictionaries			
	Brown (1988)	Snyman <i>et al.</i> (1990)	Matumo (1993)	Dent (1994)
<i>-sadi</i>		✓		
<b>mosadi</b> (“woman”)	✓		✓	✓
<b>basadi</b> (“women”)	✓			✓
<i>-lepe</i>		✓		

<b>selepe</b> (“axe”)	✓		✓	✓
<b>dilepe</b> (“axes”)	✓		✓	✓
<b>araba</b> (“answer”)	✓	✓	✓	✓
<b>arabile</b> (“answered”)			✓	
<b>arajwa</b> (“be answered”)		✓	✓	
<b>bula</b> (“open”)	✓	✓	✓	✓
<b>budisa</b> (“let open”)		✓		
<b>bulaka</b> (“open wide”)			✓	
<b>bulega</b> (“become opened”)	✓	✓	✓	
<b>bulegile</b> (“has been opened”)			✓	
<b>bulegileng</b> (“that has been opened”)				✓
<b>bulela</b> (“open for”)	✓		✓	
<b>butswa</b> (“was opened”)			✓	

Secondly, we shall have a look at lemmatisation as it was applied in the rule-based automatic lemmatiser by Brits (2006). In natural language processing there are two main approaches in developing applications, namely rule based and machine learning. In this particular automatic lemmatiser for Setswana the rule-based approach was followed. It means that the suggested hierarchy in morphological analyses by Krüger (2006) and Kotzé (2005) played a big role in the development of this lemmatiser. In this section the terms “root”, “stem” and “lemma” will be discussed briefly from a morphological point of view.

This brings us to the research question of the paper: can the morphological analysis presented in the automatic lemmatiser make Setswana dictionaries more accessible? We shall then illustrate how we adapted this lemmatiser to help in the students’ search for words and meanings. The user will be able to insert the input (search) word and the lemmatiser will then give as output (almost) all the possible forms of that word to cater for all the lemmatisation techniques followed in paper dictionaries.

We shall then report on the efficiency of the lemmatiser for the students of Setswana. In our experiment there will be two groups made up of students from different study years. The experimental group will answer a general questionnaire about their experiences with Setswana paper dictionaries and look up words (selected from reading texts) with the help of the lemmatiser. The control group will answer the same questionnaire and look up the same words but without the lemmatiser. In the last section we shall then discuss our findings and recommendations.

### References:

BRITS, J.H. 2006. Outomatiese Setswana lemma-identifisering [“Automatic Setswana lemmatisation”]. Potchefstroom : North-West University. (Thesis – MA).

BROWN, J.T. 1988. Setswana-English dictionary. Johannesburg : Pula Press. 593 p.

DENT, G.R. 1994. Kompakte Setswana woordeboek. Pietermaritzburg : Shuter & Shooter. 207 p.

KOTZÉ, A.E. 2005. Towards a morphological analyser for past tense forms in Northern Sotho:

- verb stems with final 'm' and 'n'. *Southern African Linguistics and Applied Language Studies*, 23(3): 245-258.
- KRÜGER, C.J.H. 2006. Introduction to the morphology of Setswana. München : Lincom Europa. 314 p.
- MATUMO, Z.I. 1993. Setswana English Setswana dictionary. 4th ed. Gaborone, Botswana : Macmillan Botswana. 647 p.
- SNYMAN, J.W., SHOLE, J.S. & LE ROUX, J.C. 1990. Setswana-Engels-Afrikaanse woordeboek. Pretoria : Via Afrika. 527 p.

## **Designing a Dictionary for the De la Bat School for the Deaf in Worcester: A case study**

**Hanelle FOURIE**

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

The present study makes use of a case study to determine the needs and characteristics of the learners at the De la Bat School for the Deaf in Worcester, Western Cape, in order to design and propose a theoretical model for a school dictionary. The case study was chosen in this case in lieu of the user study, given the constraints experienced by a single researcher. The De la Bat School for the Deaf was chosen because it provides a representative sample of Deaf learners from all over the country, including all race and socio-economical groups.

The Doofstomme- en Blinde Instituut (Institute for the Deaf, Dumb and Blind) was founded in 1881 by the Dutch Reformed Church. The school grew to such an extent that the Divisions for the Deaf and Blind were separated in 1905, resulting in the Worcester Skool vir Dowes with Mr B.J.G. de la Bat as its principal. The name was changed to De la Bat School at the centenary celebrations of the school in 1981.

The De la Bat School provides specialised education to deaf children. The students range from three years old in pre-school kindergarten classes to 20 years old in high school. They come from the Western Cape, other South African provinces, and other countries like Namibia and Zimbabwe. These learners are deaf or hard of hearing, or have multiple disabilities and are frequently from disadvantaged families. Only 35% of parents are able to pay school fees.

Today the De la Bat School is a state subsidised autonomous structure. The National Institute for the Deaf (NID) acts as the sponsoring body for the school. In 2010 there were 170 learners at the school.

The medium of instruction is Sign Language, with Afrikaans or English as the language for reading and writing. Speech development takes place in the Foundation Phase within the ability of each learner, and (spoken) language instruction is done with the help of the Red Star method which was developed by educators in the Foundation Phase. The Red Star method will be discussed and explained in greater detail, as well as how it can be implemented in the proposed dictionary.

Subject teaching continues from the Intermediary Phase to high school, where learners receive instruction in subjects such as the home (spoken) language, first additional language, mathematics, natural sciences, social sciences, technology, economic and management sciences, life orientation and arts and culture. At the NID College vocational training is also provided through further education, occupational and skills training for deaf students in a variety of study fields, for example, Hospitality, Cosmetology and Beauty, Construction, Upholstery, Welding, Agriculture, Early Childhood Development and Office Administration

Most important to this particular study is the fact that the school is also the first school in the country to implement a Sign Language Curriculum which is systematically being rolled out across the various grades in school. Ultimately, it must be possible for a learner to complete their Grade 12 studies – and beyond – in Sign Language, not only by receiving instruction in Sign Language, but also by taking their examinations in Sign Language. Modern video and computer technology makes it possible to record a signed version of the question paper, and a student's signed answers can then easily be captured with a webcam. In this way literacy in the spoken language will no longer be a prerequisite for academic qualification, and Sign Language will at last (start to be) recognised as a fully fledged natural language, capable of the same things as any spoken language.

In this light it is thought that there is no better time to offer a dictionary as teaching and learning resource, as this will also contribute to learners' sense of pride in their language and award a certain amount of status to it. The proposed model intends to provide a "growing" macrostructure that expands as learners progress through the school grades. Currently the school has no proper dictionary for class use and class vocabulary lists consist of lists written in the spoken language, with no visual aid or any possibility to provide or represent a sign equivalent. The proposed dictionary is aimed primarily as a class aid to the learners, but may also serve the parents and siblings at home. The dictionary will have a triple macrostructure consisting of 1) written entries, 2) sign entries arranged and searchable according to hand shape, and 3) thematic categories accompanied by pictures. All entries will contain video clips of signs, because the dictionary will be electronic.

While dictionaries of spoken languages with a written form are often used as tools of standardisation, it will be explained why this is not the case for the majority of sign language dictionaries, including the one that is proposed. The question of what constitutes a lexeme in a signed language will also be discussed.

## **Equivalent Selection in LSP e-Lxicography: A case study with Spanish accounting terms**

**Pedro A. FUERTES OLIVERA**  
University of Valladolid, Spain

Interest in specialised lexicography has been propelled by the development of LSP communication in academic circles, as well as by the consolidation of function-based approaches to lexicography that have identified the existence of several user types, e.g. experts, semi-experts and interested laypersons, and use situations, typically cognitive-oriented and communicative-oriented (Bergenholtz & Tarp, 2003, 2004). This paper follows suit and elaborates on the selection of Spanish equivalents in a particular dictionary project. It addresses the issue as a lexicographical problem that can be solved by resting on lexicographical principles that take into consideration the nature of lexicography, the technical options the Internet offers, and the defining characteristics of specialised discourse.

The analysis is presented as a case study regarding the selection of Spanish equivalents in the *Diccionario Inglés-Español de Contabilidad* (Nielsen et al., 2009), one of the 'accounting dictionaries' that are extracted from the same 'accounting database' (Nielsen et al., 2010). The accounting dictionaries are a set of two monolingual and three bilingual online dictionaries with the languages Danish, English, and Spanish (a Spanish dictionary and a Spanish-English dictionary are in the pipeline). The theoretical foundation underlying the project gives priority to



lexicographical functions, that is, the help these dictionaries can give to users in specific types of situation where users require knowledge to resolve issues relating to accounting.

The *Diccionario Inglés-Español de Contabilidad* aims to satisfy the needs of translators (primary user group), accountants and financial experts (secondary user group), and students of accountancy, students of translation, journalism and interested laypersons (tertiary user group). It is a polifunctional dictionary that differs from the English and Danish accounting dictionaries in not yet having incorporated functionalities for retrieving dynamic articles with which users can solve their needs in specific use situations (for example, ‘help to understand an accounting term’). These search options, which are already operative in the English and Danish accounting dictionaries, signal the way ahead for e-lexicography, which is mostly concerned with differentiating between an information database and an information tool, upgrading the access process used, and exploring the introduction of Boolean searches, and maximising and minimising searches, to name just a few of the lexicographic characteristics that are being dealt with in state-of-the-art e-lexicography (Fuertes-Olivera & Bergenholtz, 2011).

The *Diccionario Inglés-Español de Contabilidad* also offers innovative lexicographical solutions in issues such as the selection of equivalents. The decisions taken do not follow the theoretical recommendations and practical applications discussed in an abstract way, usually in connection with explaining the concept of equivalence as linguists and translators do (Adamska-Salaciak, 2010). Instead, the discussion adopts a lexicographical approach that focuses on three principles that should permeate the planning and construction of specialised dictionaries: *relevance*; *proscription*; and *recreation*.

The principle of *relevance* refers to the quality of being directly connected with the subject field in question, the function(s) of the dictionary, and the user situation in which the dictionary is intended to be used (Fuertes-Olivera & Nielsen, 2011). For example, if the dictionary aims primarily at meeting the needs a user has when translating a text, it is very important to include a translational equivalent instead of, say, an explanatory equivalent. In this dictionary, users always have *one and only one insertable Spanish equivalent*, as well as an English definition and synonyms (sometimes) per English lemma. This is a lexicographical solution with which translators are relieved from the suffocation effect they suffer when dictionaries offer several options and paraphrases instead of insertable equivalents.

*Proscription* refers to the recommendation lexicographers do as a result of their analysis of the several possible options available (Bergenholtz & Gouws, 2010). The proscriptive approach in specialised lexicography is analysed in connection with the nature of specialised communication, and its interest for having a one-to-one correspondence between a concept and its linguistic representation. In the *Diccionario Inglés-Español de Contabilidad* this principle informs the selection of terms such as the so-called IAS/IFRS, which are exemplars of *equal terms*.

*Recreation* comprises the process of creating a Spanish equivalent whenever we judged necessary, either because there are variations (fluctuation) between competing terms, or because the Spanish existing term(s) were not always insertable as shown in the internal and external subject classification made by the joint efforts of specialists in (Spanish) accounting and lexicographers. This process translated into using English terms as Spanish equivalents (e.g. *rating*), and resorting to Spanish literal translations, especially of multi-word terms such as **business combination involving entities or businesses under common control**, whose Spanish equivalent is the Spanish literal translation *combinación de negocios entre entidades o negocios bajo control común*.

## References:

### A. Dictionaries

Accounting Dictionaries 2010 = Nielsen, S., L. Mourier and H. Bergenholtz. (2010). *Danish Accounting Dictionary; Danish-English Accounting Dictionary; English Accounting Dictionary; English-Danish Accounting Dictionary* URL:

<http://www.ordbogen.com/ordbog/regn/index.php?dict=a007> (Last accessed 20 January 2011)

*Diccionario Inglés-Español de Contabilidad 2009* = Nielsen, S., L. Mourier, H. Bergenholtz, P. Fuertes-Olivera, P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, A. Sastre Ruano, S. Tarp, M. Velasco Sacristán. 2009. *El Diccionario Inglés-Español de Contabilidad* In: [http://www.accountingdictionary.dk/regn/gbsp/regngbsp\\_index.php](http://www.accountingdictionary.dk/regn/gbsp/regngbsp_index.php) (Last accessed 20 January 2011).

### B. Other Literature

Adamska-Salaciak, A. (2010). “Examining Equivalence”. *International Journal of Lexicography* 23(4): 387-409.

Bergenholtz, H. & Gouws Rufus H. (2010). “A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography”. *Lexikos* 20: 26-51.

Bergenholtz, H. and S. Tarp. 2003. “Two Opposing Theories: On H.E. Wiegand’s Recent Discovery of lexicographic Functions”. *Hermes. Journal of Linguistics* 31, 171-196.

Bergenholtz, H. and S. Tarp. 2004. “The Concept of Dictionary Usage”. *Nordic Journal of English Studies* 3, 23-36.

Fuertes-Olivera, Pedro A. & H. Bergenholtz (eds.) (2011). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London & New York: Continuum.

Fuertes-Olivera, Pedro A. & S. Nielsen (2011): “The dynamics of terms in accounting: what the construction of the *accounting dictionaries* reveals about metaphorical terms in culture-bound subject fields”. In Rita Temmerman & M. Van Campenhout (eds.), *The Dynamics of Terms in Specialised Communication. An Interdisciplinary Perspective*. Special Issue of *Terminology. International Journal of Applied Issues in Specialized Communication* 17(1).

## Linking dictionary and corpus data in online language tools

Ulrich HEID

Institut für Informationswissenschaft und Sprachtechnologie, University of Hildesheim, Germany

Danie J PRINSLOO

Dept of African Languages, University of Pretoria, South Africa

**Objectives:** With a view to providing customized lexicographic data for individual users and usage situations, it has often been proposed to combine the structured data from a dictionary database with textual data from the internet or from corpora. Recently, Tarp (to appear) has called such solutions “lexicographic Model T Fords”, which “[...] link to the internet where already existing data is reused in order to satisfy the users’ specific needs”. An early implementation is Køhler-Simonsen’s (2006) specialised dictionary, *ZooLex*. Others will be discussed below.

In this paper, we argue in favour of more dynamically linking corpus data to a dictionary, to illustrate details of the use of multiword expressions (noun+verb-collocations in our examples), with a view to text production needs. We show that, by using computational linguistic tools, corpus examples can not only be displayed, but also linguistically analysed and generalized, to

give the dictionary user a clearer picture of their actual use in texts. This seems to go into the direction of what Tarp (to appear) calls “solutions based upon a recreation and re-representation of the data“, an approach which he sees as needed for the future (“lexicographic Rolls-Royces”).

**Current solutions:** Several online dictionaries provide some kind of access to corpus data. The technically simplest method (and the least useful one for users) is to juxtapose both types of resources in a common graphical interface (cf. DWDS<sup>1</sup>, and the criticism by Asmussen, to appear). An alternative is to provide a portal which links from the dictionary to one or several corpora and to other internet resources; this is done in the 2009 version of Verlinde’s *Base lexicale du français* (BLF<sup>2</sup>), where the DAFLES<sup>3</sup> dictionary is linked, among others, to the OPUS website<sup>4</sup> of parallel corpora (cf. Verlinde, Leroyer, Binon 2009). Preliminary tests with 33 users in a usability laboratory (cf. Bank 2010) showed that most users find the portal function difficult to use: it was not clear to them when and why they ended up on websites not belonging to the dictionary, how to interpret the data given there, and how to navigate back to the dictionary itself.

A more focused approach is followed in *ordnet.dk*, a version of the Danish dictionary *Den Danske ordbog*<sup>5</sup>: in its main user interface, readings and collocations are listed, and a small clickable icon ‘K’ gives access to a KWIC representation of corpus sentences containing the respective collocation.

**Proposed approach:** We intend to follow *ordnet.dk*’s approach, by relating specific lexicographic data on collocations with a search engine for corpus data and an underlying corpus. The collocations contained in the dictionary provide lemmas to search for, as well as the word class and grammatical relation of the collocation base and the collocate. Regular expression search is sufficient for languages with limited word and constituent order freedom. Cf. also recent developments in BLF, Verlinde (2011).

A test on Afrikaans *aandag gee* and *aandag skenk* (“pay attention”) in the 127 m words from Beeld<sup>6</sup> showed that the former is ca. 6 times more frequent than the latter. Similarly, *fela pelo* (“be disheartened”) in the Northern Sotho *Pretoria Sepedi Corpus* (PSC, 5 million words, cf. De Schryver and Prinsloo (2000)) gives a clear distribution over morphological forms, *fela* (302 hits), *fele* (135), *felang* (5), *felwa* (2), and useful information to the user in terms of other frequent collocations of *pelo*, e.g. *beta pelo* (92) (“take courage”), *kwa pelo* (28) (“hear/listen to one’s heart”), *hlomola pelo* (71) (“feel sorry for”) etc. Similarly the most frequent collocations for the user looking up the lemma *ipona* (“to see oneself”) such as *ipona molato* (42) (“see oneself guilty”), *ipona phošo* (23) (“see oneself at fault”) *ipona botlaela* (9) (“see oneself foolish”) are culled from the corpus.

For German, which has much case syncretism and a relatively free constituent order, results are slightly less good, as in addition to true positives also sentences are picked up where the two items in question don’t form a collocation; thus, the use of syntactically analysed texts is preferable. On this basis, Weller and Heid (2010) have proposed to not only extract example sentences for collocations, but also data about the morphosyntactic properties of the base and the collocate, such as number, determination, voice, tense, etc. Such data are extracted along with each example sentence and stored in a database. In a second step, preferences are calculated for

---

<sup>1</sup> <http://www.dwds.de/>

<sup>2</sup> <http://ilt.kuleuven.be/blf/>

<sup>3</sup> DAFLES: *Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde*

<sup>4</sup> <http://opus.lingfil.uu.se/>

<sup>5</sup> <http://ordnet.dk>

<sup>6</sup> We use a section of the Pharos *Media24* Afrikaans corpus, made available to us by Pharos publishers. We gratefully acknowledge Pharos’s contribution to the present work.

each collocation. Such preferences (e.g. *have high hopes* typically in the plural) can be signalled to the user (on demand) in the graphical user interface, along with relevant examples. This also concerns lexical variation in idioms and collocations; for example, the German idiom *keinen Mucks machen* (“not say a word”) has variants like *keinen Mucks geben*, *keinen Mucks tun*.

**Proposed presentation.** In the talk, we intend to present the main argument, the state of the art, as well as results of experiments on Afrikaans, German and Northern Sotho, and mock-up screens of a possible user interface. We also intend to address the possibilities and limitations of the proposed approach.

### References:

- Asmussen, Jörg: “Combined Products: Dictionary and Corpus”, to appear in: Gouws et al. (Eds.): *Dictionaries. An International Handbook*, Vol. 5/4 (Berlin: De Gruyter), to appear.
- Bank, Christina (2010). Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen. Universität Hildesheim: MA-Dissertation, ms. 103 pp.
- De Schryver, Gilles-Maurice, Danie J. Prinsloo (2000). The compilation of electronic corpora with special reference to the African languages. *Southern African Linguistics and Applied Language Studies*, 18(1-4): 89 – 106, 2000.
- Køhler-Simonsen, Henrik (2006). “ZooLex: the wildest corporate reference work in town?”, in: Elisa Corino, Carla Marengo, Cristina Onesti (Eds.): *Atti del XII Congresso Internazionale di Lessicografia*, EURALEX, Torino, (Alessandria: Edizioni dell’Orso) 2006: 787-793.
- Media24: <http://www.media24.co.za>.
- Tarp, Sven: “Lexicographical and other e-tools for consultation purposes: towards the individualization of needs satisfaction”, to appear in: Pedro A. Fuertes Olivera, Henning Bergenholtz (Eds): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. (London/New York: Continuum), to appear 2011
- Verlinde, Serge: “Modeling Interactive Reading, Translation and Writing Assistants”, to appear in: Pedro A. Fuertes Olivera, Henning Bergenholtz (Eds): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. (London/New York: Continuum), 2011
- Verlinde, Serge, Patrick Leroyer, Jean Binon (2009): “Search and you will find. From stand-alone lexicographic tools to user-driven task and problem-oriented multifunctional leximats”, in: *IJL* 23:1 (2009): 1-17.
- Weller, Marion, Ulrich Heid (2010): “Multi-parametric extraction of German multiword expressions from parsed corpora”, in: *Proceedings of LREC-2010, Linguistic Resources and Evaluation Conference*, Malta, 2010 [CD-ROM].

## Pronunciation guides for a South African English school dictionary

Lorna HILES

Cape Town, South Africa

The provision of clear, intuitive, consistent pronunciation guides in a dictionary has always been a challenge to lexicographers. Add a multilingual user base and a general lack of IPA knowledge, and the lexicographer has even more challenges to negotiate.

This paper broadly discusses the issue of pronunciation in the context of a South African English school dictionary. It introduces different aspects of pronunciation guides and suggests topics for further study. In Part 1 I discuss the problems associated with pronunciation guides in a South African English school dictionary, as well as the current methods of representing pronunciation in school dictionaries. I present the pros and cons of each method, as they apply to

South African school dictionaries. Part 2 of this paper looks into the varieties of South African English and how they are represented in school dictionaries, and how this influences the presentation of pronunciation guides in the dictionaries.

### **Part 1**

Problems associated with pronunciation guides in South African English school dictionaries include:

- No standard English pronunciation, which means several variants are acceptable
- Different home languages mean that users bring different prior knowledge to their learning
- Afrikaans sounds, Sotho sounds and Nguni clicks exist in South African English words
- Knowledge of IPA is very limited in South Africa
- Respelling systems are not universal, so each dictionary needs a decoding table

I discuss these problems in greater depth, providing evidence, and solutions where possible.

The two types of pronunciations guides that I look at are:

- IPA (International Phonetic Alphabet)
- Respelling (including as subsets: respelling based on English spelling rules; and respelling based on individual phonemes)

I examine each of these systems and present advantages and disadvantages of each.

The next section of Part 1 of the paper consists of a comparison between different respelling systems in different South African school dictionaries. The guide to the respelling system (decoding table) used in each dictionary is presented, as well as a sample of words, providing a comparison between the dictionaries. As no South African school dictionary currently provides IPA as a pronunciation guide system, I provide an IPA pronunciation guide for each of the sample of words.

To conclude Part 1 of this paper, I present preliminary findings and offer suggestions of further study.

### **Part 2**

In Part 2 of this paper I discuss the varieties of South African English and the problems that they present to the selection of pronunciation guides in a school dictionary. I look at the concept of standard pronunciation, and whether this is applicable to South African English. School dictionaries do not contain the space to represent all variants of pronunciation, and I discuss when one variant is sufficient, and when the words require more than one. I will also look at the difference between written pronunciation guides and spoken ones, as used in online dictionaries and dictionary CD-ROMs.

Questions asked in Part 2 include:

- How many users of an English school dictionary are mother-tongue English speakers?
- Does a spoken pronunciation guide require a South African accent?
- Is there a preferred South African accent?

Part 2 of this paper will conclude with a summary of results and suggestions for further study.

## Comparing two African language online dictionaries from a non-native speaker's perspective

**Juliane KLEIN**

Institute of African Studies, University of Leipzig, Germany

This presentation compares two online dictionaries from a learner's perspective. The focus is mainly on non-native speakers from overseas who have little competence in both languages. Online dictionaries were selected because most of the newer African language hardcopy dictionaries are not available in German university libraries, which implies that students either work with very old dictionaries or have to work with free online resources.

The first dictionary is an isiXhosa-English online dictionary from the US (Xhosa-English dictionary) the other is an isiZulu-English dictionary (isiZulu.net) from Germany. Both dictionaries were selected because they are bilingual, general, user-friendly, online dictionaries from related languages and are available for free. As user-friendliness is an important factor for language learners, the main focus was on how useful those dictionaries are for language learners.

In a first step, the outer-texts of the dictionaries were compared. Both dictionaries offer a short user-guide in English, which provides the user with instructions and illustrates the lemmatization and ordering principle of the dictionary.

The isiXhosa dictionary offers information on isiXhosa grammar, including: noun-classes, verbs, consonant changes, numbers, and demonstratives. Other information includes: the *Universal Declaration of Human Rights* in isiXhosa and English, as well as a short basic introduction to the language family. Another important part of the outer texts is the forum. This forum is available as a Facebook application and allows Facebook users to ask questions about missing words or to discuss about the dictionary entries.

The isiZulu.net offers a very short basic introduction on isiZulu grammar, which shows the most important features at a glance. A pronunciation guide provides pronunciation examples for every isiZulu IPA symbol. The user can either listen to a MP3 file or read the IPA pronunciation. This dictionary also offers a forum where users can discuss missing/incorrect dictionary entries or provide new/additional translations.

In the second step, several dictionary entries are compared. Here, the focus is not only on the translation but also on the entry structure, i.e. the other information that is provided together with the translation.

A typical entry of an isiXhosa noun in the dictionary consists of the English (or isiXhosa) search word, the translation and its lemmatized form in the isiXhosa lemma list, i.e. the noun and plural form according to the noun class and the stem in the lemma list. A typical noun entry in the isiZulu dictionary consists of the English (or isiZulu) search word and the direct matches. The direct matches offer a direct translation and the singular and plural forms according to the noun class, as well as the lemmatized stem form and the pronunciation in IPA.

Both dictionaries are quite straight-forward: the user enters a word in either English or isiXhosa/isiZulu, clicks on search and gets a translation or the link to the respective forum, if the word is not in the lemma list. Both dictionaries are clearly designed for non-native users who are computer literate and have at least a basic knowledge about internet applications and dictionaries and want a free, quick and easy translation.

Having online dictionaries that also provide information about the language may generate a deeper interest in the language and may encourage users to buy larger (printed) dictionaries of

the respective language. Providing a forum where users can discuss and give feedback helps to keep the dictionaries up-to-date and user-friendly. As the users can participate in the compilation through providing new entries, the dictionaries will also grow and provide more entries and will therefore be more helpful in the future. These two dictionaries are not only useful quick references for learners but they also promote the use of the language internationally and are a visible proof that the languages have arrived in the internet age and can be used with “cool” applications like Facebook.

#### **References:**

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

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

## **Challenges of predictability and consistency in the first comprehensive Sotho dictionary**

**Inge KOSCH**

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

This paper focuses on the lexicographic principles of predictability and consistency, with special reference to Endemann’s *Wörterbuch der Sotho-Sprache*, 1911 (Dictionary of the Sotho language). It is a monodirectional dictionary with Sotho as the source language and German as the target language. It can be regarded as the first attempt at a comprehensive dictionary, because the term ‘Sotho’ in the title refers to the three languages which are classified as the Sotho group of languages, namely Northern Sotho (also known as Sesotho sa Leboa or Sepedi), Southern Sotho (Sesotho) and Tswana (Setswana), spoken in South Africa and parts of southern Africa.

For a user to benefit maximally from his or her dictionary consultation experience, the information in a dictionary should ideally be arranged in a predictable and consistent order. This will ensure that the user will find the information where he or she expects to find it, and according to a predetermined pattern (obligatory microstructure). As pointed out by Gouws and Prinsloo (2005:16) a dictionary without a properly planned microstructural programme functions in an unsystematic way and violates the predictability criterion in situations where “the lexicographers decide in a haphazard way to include a certain data category in a specific article and omit it from the next”. Predictability and consistency are requirements which should run like a golden thread through the macro-, medio- and microstructure of dictionary articles. Adherence to these requirements is one of the marks of a user-friendly reference work that will allow for easy access and trouble-free retrieval of required information.

Owing to the agglutinative nature of the Bantu languages, dictionary compilers had to be innovative in their methods of entry, either according to words or word stems or a combination of the two. In their *Comprehensive Northern Sotho Dictionary*, Ziervogel and Mokgokong (1975) followed the stem tradition for nouns and verbs, which means that all entries are entered under the first letter of the stem. Kriel (1983), on the other hand, while also following the stem tradition for verbs, used the word tradition for nouns in his *Pukuntšū Woordeboek*. Endemann also followed a stem approach like Ziervogel and Mokgokong, but with two major differences. The first difference is that in the case of nouns the prefix precedes the lemma (separated from it by a hyphen) instead of following it. The second difference is that, apart from isolated cases, all entries are presented as main lemmata. This affords quicker and easier access to items as one will not have to pick them out from amongst a long list of other sub-lemmata under a specific lemma.

On the other hand, this approach obscures the interrelationship between words which are inflectionally or derivationally related, because they might be positioned far from each other under the same article stretch or they may even appear under another letter of the alphabet altogether. In Endemann's dictionary, a noun such as *mmopi* 'creator', for example, is separated from the verb stem *-bopa* 'create' by 18 other lemmata. For the user to find *mmopi* a higher level of access skill is thus required as well as knowledge of the sound change rules of the language.

The method of entry according to word stems requires more than just basic reference skills of the user and an understanding of the principles underlying this method. Any scholar, especially someone with no prior knowledge of the structure of a Bantu language, will find it difficult to look up the meaning of words without first consulting the explanatory introduction which should be part of any good dictionary. Even though Endemann provides users' guidelines in the introduction, he relegated many observations to the individual entries themselves, making some dictionary articles disproportionately long. He was also not very consistent with the completion of paradigms. In the front matter he mentions that the reader should be able to easily complete the paradigms for him/herself.

Articles in Endemann's dictionary do not always have the same microstructure. Some articles merely consist of a translation. In many cases the relevant sense of a lemma sign is not explained by means of its usage in a context. In other cases lengthy explanations and personal experiences or opinions are recorded as first item under the letter of an alphabet.

The paper aims to establish to what extent the challenges posed by the principles of predictability and consistency were met in Endemann's dictionary and hopefully some interesting insights can be gained from this example of an early attempt at dictionary compilation to assist in continually improving the user-friendliness of dictionaries in the Sotho languages. Endemann's dictionary has hitherto remained at the outskirts of scholarly investigation, the main reason probably being the fact that the target language is German and that, as such, it is not readily accessible to every scholar of the Bantu languages.

### 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. 1983. *Pukuntšu woordeboek (Noord-Sotho – Afrikaans, Afrikaans – Noord-Sotho)*. Pretoria: J.L. van Schaik.
- Ziervogel, D. & Mokgokong, P.C. 1975. *Groot Noord-Sotho woordeboek/ Comprehensive Northern Sotho dictionary*. Pretoria: J.L. van Schaik/Unisa.



## **Packaging phonetic information in dictionaries of languages of the Sotho-Tswana cluster**

**John M. LUBINDA**

Dept of French, University of Botswana, Gaborone, Botswana

A dictionary is said to be “a carrier of text types representing a variety of different texts” (Gouws, 2002 : 55) and information categories of various types. The selection of the information categories to be included in a dictionary is important. But equally important is the manner in which the selected information is packaged and displayed for easy access and retrieval by the target user.

One of the entry components in the central list of a linguistic dictionary is the indication of the pronunciation of each recorded lemma. As Bo Svensén (2009 : 114) notes, “the need for pronunciation information in dictionaries varies between languages [and] ..... between different types of dictionaries”. He points out that in a language such as Finnish (and to a lesser extent, Italian or Spanish) where there is a one-to-one correspondence between spelling and pronunciation there is hardly any need to provide information on pronunciation. But in a language like English, where there is little correspondence between spelling and pronunciation, and where the position of word stress is not predictable, there is need to provide this type of information. He further points out that the need to include information on pronunciation is largely determined by the type of dictionary one wishes to compile, in terms of the intended target user group and dictionary typology. A monolingual dictionary intended primarily for native speakers of the object language hardly needs indications on pronunciation, except for occasional loan words. On the other hand, a learner’s dictionary of production and aural reception needs information on pronunciation at the segmental and suprasegmental levels. Once the question of whether or not to represent pronunciation has been settled, one has to decide on the best way of presenting (“packaging”) this information.

Another important question that needs to be addressed is that of variation in pronunciation (regional, social or stylistic). Is the dictionary compiler going to account for regional, social and stylistic forms of pronunciation as well or simply confine himself/herself to the pronunciation of the variety of the language chosen as the standard variety? Mavoungou (2010 : 98), among others, points out that the absence of indications on pronunciation in Gabonese dictionaries (especially the omission of tone-marking) is one of the weak points of Gabonese lexicography. This is generally true of Bantu lexicography, in languages of the Sotho-Tswana cluster. Until relatively recently (from about the early nineteen-nineties to the present), scant attention was paid to this entry component by dictionary compilers in these languages. However, with the advent of electronic dictionaries with audio recordings, the pronunciation component of the dictionary is increasingly receiving scholarly attention. Moreover, different conventional methods of notation have been tried, including what is often termed “respelling”. Atkins and Rundell (2008 : 206) remind us that “the most common way of showing how a word is pronounced is to use the International Phonetic Alphabet (IPA)”. The pertinent question that arises in this regard, is : Are the target users sufficiently familiar with the IPA or with the Alphabet of the International African Institute, as a useful adjunct to the learning of the pronunciation of the target language ? Can they accurately interpret phonetic transcription ?

In the study that forms the basis of this presentation, a dozen selected dictionary tokens of major languages of the Sotho-Tswana cluster (Setswana, Sesotho, Northern Sotho and Silozi) were examined with a view to finding out how pronunciation is presented and displayed in these dictionaries. Reference details of these dictionaries are given in the actual conference paper,

including a discussion of the manner in which pronunciation information is handled in each of them. Preliminary findings indicate that only one dictionary contains information pertaining to pronunciation, in the form of phonetic transcription, in the nomenclature. The other nine devote only brief sections in the front or mid matter to comments on pronunciation, such as the presentation of vowel and consonant charts of the languages as well as attempting to explain letter values in relation to other languages, like English. Some of them do not even have any comment at all on pronunciation. It is clear, therefore, that so far, scant scholarly attention has been paid to the phonetic form of lemmata by compilers of these dictionaries. The paper is therefore an attempt at dictionary criticism on this particular entry component of the dictionary.

The paper revisits this topic, which is still of current interest, in light of common lexicographic practices observable in some recently published dictionaries of some Bantu languages of the Sotho-Tswana cluster. It examines the various potentialities of the combination of the printed dictionary and audio recordings, in the era of the electronic dictionary and the dictionary on CD-ROM or on-line. The merits and demerits of “respelling”, as a compromise solution are re-examined. At the prosodic level, various conventional methods of marking tone (register and contour tone) are examined in light of current practices, and propositions are made.

### **References:**

- Atkins, S. & Rundell (2008) *The Oxford Guide to Practical Lexicography*, Oxford University Press, Oxford.
- Bo Svensén (2009) *A Handbook of Lexicography : The Theory and Practice of Dictionary-making*, Cambridge University Press, Cambridge.
- Gouws, R. H. (2002) “Using a Frame Structure to Accommodate Cultural Data”, in *Éléments de Lexicographie Gabonaise*, Tome II, Jamacs-Hillman Publishers.
- Mavoungou, P.A. (2010) *Lexicographie et Confection des Dictionnaires au Gabon*, Sun Media, Stellenbosch.

## **Towards a monolingualised bilingual learner’s dictionary**

**Jana LUTHER**

Pearson Southern Africa, Cape Town, South Africa

The significance of bilingual dictionaries for second and foreign language acquisition is today widely recognised. It has been established that the majority of learners prefer bilingual dictionaries to monolingual dictionaries and, internationally, bilingual dictionaries are shown to have improved considerably over the last decades.

Yet such improvements are not universal, and have thus far had little or no impact on the small, inexpensive bilingual dictionaries that many language learners use. In addition, the ideal that a bilingual dictionary ought to be designed to primarily serve the needs of the mother-tongue speakers of one of the two languages is seldom realised.

Consequently, as Martyn Back (2005) points out, “using a bilingual dictionary [... still mostly remains] a tricky business, even for the experienced user. The encoding user is always, to a greater or lesser extent according to his/her level of linguistic competence, stepping into the unknown; translators know that to lift a foreign language term from a dictionary without further cross-checking is fraught with danger, and most teachers have anecdotes about the hilarious misuse of dictionaries. [...] Lack of grammatical knowledge in the students’ own language and scant grasp of interlingual issues (most notably the fact that word for word translation is a recipe

for disaster) lead to widespread misuse of standard dictionaries, with catastrophic results in the classroom.”

Lexicographers worldwide have taken cognisance of this predicament, widely expressing the need for bilingual dictionaries that better fulfil the requirements of their users. This is also the case in the South African context, where bilingual dictionaries dominate the lexicographic environment.

Published in June 2011, the *Longman-HAT Basic Dictionary* (LHBD) is the first of a series of new bilingual English-Afrikaans/Afrikaans-English dictionaries to be launched by the publishers of the *Longman Dictionary of Contemporary English* (LDOCE) and the *Handwoordeboek van die Afrikaanse Taal* (HAT). This new dictionary differs in a number of ways from other English-Afrikaans dictionaries already on the market.

Against the above background, the aim of this paper is to give an outline of some of the key features of the LHBD as a modest step towards a new type of bilingual learner’s dictionary.

The contents of the Longman-HAT series are drawn from existing Longman and HAT dictionaries, the Longman Corpus Network (a group of corpora of authentic English language), the fast-growing HAT corpus of contemporary Afrikaans and the Internet.

The LHBD is based on the words designated as S1 in the Longman Communication 3000 – a list of the 3 000 most frequent words in both written and spoken English as determined by statistical analysis of the 390 million words contained in the Longman Corpus Network. In the Longman Communication 3000 the thousand words considered the most important for verbal communication in English are marked by an S1 symbol. These top thousand most frequent words in spoken English and their primary meanings are at the core of the LHBD.

For each meaning the editorial team endeavoured to choose the most appropriate Afrikaans translation(s). To show how the headword and the translation(s) are used in typical contexts, and to illustrate typical grammatical patterns and common vocabulary, example sentences and phrases were carefully chosen from the English and Afrikaans corpora. Where headwords have more than one meaning, the different meanings are numbered. The entry for each meaning starts on a new line, making it easy for users to find the correct translation (in bold face) quickly. Parts of speech and inflected forms are as easy to find. To make the dictionary even more user-friendly, no abbreviations are used in the metalanguage.

But what really distinguishes the LHBD from other English-Afrikaans dictionaries currently on the market is an important innovation: in the LHBD every translation is introduced by a clear, simple definition in both the source and the target language. Through these bilingual definitions the LHBD additionally serves as an explanatory dictionary with the aim of leading beginner learners of English and Afrikaans directly to the exact translation for each meaning of a headword.

#### **References:**

Back, Martyn. 2005. Bilingual Dictionaries for Learners [online].  
*Kernerman Dictionary News* 13. June 2005. <http://kictionaries.com/kdn/kdn1310.html>  
[accessed 20 March 2010].

## The Ordering of Lexical Entries in Tshivenda Dictionaries with Special Reference to *Tshivenda/English Thalusamaipfi Dictionary*

James MAFELA

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

A lexical entry is a form which represents a word or an expression which is being described in a dictionary. A lexical entry may consist of different forms, for example, a word, stem, phrase, or a sentence. In compiling a dictionary the editors must of necessity make a selection of lemmas from all words comprising the vocabulary of a language (Swanepoel 1989:32). The entries in a dictionary can be ordered in different ways; they can be ordered, for example, according to the letters of the alphabet, or according to the semantic system. Ordering entries in a dictionary is meant to enable the user to find the correct or suitable word in a language without difficulty. Users do not need to struggle to find a word for which they are looking. Many dictionaries in different languages order their lexical entries according to the letters of the alphabet because this system is considered user friendly. Through this system, the first letters of the words are used to order the lexical entries.

Lexical entries in all Tshivenda bilingual and trilingual dictionaries have been ordered alphabetically. However, Tshivenda differs from other languages because it is a language characterised by diacritics that are used to denote some speech sounds. In addition to the normal letters of the alphabet, there are letters that are identified by the use of diacritics. The letters, *ǀ*, *ǁ*, *ǃ*, *ǂ*, and *Ǆ* represent speech sounds that are different from those of other normal letters of the alphabet. The letters *ǀ*, *ǁ*, *ǃ* and *ǂ* are dentals, whereas *Ǆ* is a velar. The number of letters of the alphabet in Tshivenda has been extended in this regard. The letters of the alphabet with diacritics should also be considered when ordering lexical entries in Tshivenda dictionaries.

For example, the lexical entries in *Venda Dictionary: Tshivenda – English* by Van Warmelo and *Improved Trilingual Dictionary: Venda – Afrikaans – English* by Wentzel and Muloiwa, have been ordered according to the letters of the alphabet, including the letters with diacritics. The letters with diacritics are not mixed with the normal letters of the alphabet, as they stand separately. Likewise, lexical entries in *Tshivenda/English Thalusamaipfi Dictionary* by Tshikota are ordered alphabetically, including the letters with diacritics. However, the letters with diacritics have been mixed with the normal letters of the alphabet; they do not stand separately. For example, lexical entries starting with the letter *ǀ* are mixed with lexical entries starting with the letter *l*. These two letters do not represent the same speech sound. The letter *ǀ* is a dental, whereas *l* is an alveolar.

Van Warmelo (1989) and Wentzel and Muloiwa (1982) explain the usage of both normal letters of the alphabet and those with diacritics regarding the ordering of lexical entries in the style guide. Van Warmelo (1989) states that letters with diacritics rank as separate letters in the alphabet in order: *ǀ*, *d*, *ǁ*, *l*, *ǃ*, *n*, *ǂ*, *t*. Wentzel and Muloiwa (1982:149) in their style guide also state that, “The dental symbols *ǀ*, *ǁ*, *ǃ* and *ǂ* precede the normal *d*, *l*, *n* and *t*; while the velar *Ǆ* in

its turn follows the normal *n*. This implies that words which for example, commence with *ǀ* are listed first and separately from words commencing with *d*.” The explanation in the style guide of the dictionaries mentioned above helps users to understand the ordering of lexical entries and to find words without difficulty. Unlike Van Warmelo (1989) and Wentzel and Muloiwa (1982), Tshikota (2006) does not make use of the style guide to explain the ordering of lexical entries in his dictionary. Users are not informed beforehand about the ordering of lexical entries in the dictionary

The purpose of this paper is to discuss the impact of ordering lexical entries on users in *Tshivenǀa/English Bilingual and Explanatory Dictionary*. This will be achieved by comparing the ordering of lexical entries in this dictionary with that of other Tshivenǀa bilingual dictionaries.

### References:

- Swanepoel, P.H. 1989. *Translation Studies (Honours): Diploma in Translation – Only Study Guide for LEKPER-Q*. Pretoria: University of South Africa.
- Tshikota, S. 2006. *Tshivenǀa/English ǀhalusamaipfi Dictionary*. Cape Town: Phumelela Publishers.
- Van Warmelo, N.J. 1989. *Venda Dictionary: Tshivenǀa – English*. Pretoria: J.L. van Schaik.
- Wentzel, P.J. and Muloiwa, T. W. 1982. *Improved Trilingual Dictionary: Venda – Afrikaans – English*. Pretoria: University of South Africa.

## **Legal issues in lexicography: an examination of the handling of brand names in the *Isichazamazwi SesiNdebele***

**Mandlenkosi MAPHOSA**

Department of African Languages and Culture, Midlands State University,  
Senga Gweru, Zimbabwe

This paper is centred on the legally explosive issue of brand names. It investigates the handling of brand names in the *Isichazamazwi SesiNdebele*. In looking at this issue the paper is driven by the realisation that most general purpose dictionaries normally do not include brand names as part of their macrostructural entries. This is a reality that is vividly expressed by the Chief Editor of *Webster's Third New International Dictionary of the English Language* as cited by Higgins (1997:381) where he proclaims that “this dictionary (Webster's) confines itself to generic words and their functions, forms, sounds and meanings... (presumably as distinguished from those that are non-generic, of which brand names are part). In light of this realisation this paper seeks to investigate whether the Ndebele general purpose dictionary, *Isichazamazwi SesiNdebele* (henceforth, ISN) lived within this tradition or deviates from it. Taking note that some words that could be legally (as informed by the laws of the country) considered to be brand names were included in the ISN the paper went further to probe whether there was a rational scientific basis, or lack thereof, in the selection of certain brand names.

In order to contextualise the study the paper draws comparisons with other Nguni language and international English dictionaries on this aspect of handling of brand names. The paper then ponders on the ramifications of the path that was taken by the ISN lexicographers in handling brand names. Thus the paper looks at whether there is some sort of marking of brand names, or

the use of capitalisation to distinguish them from any other word. This part of the discussion mostly dwells on legal ramifications and also looks at implications for lexicographic practice and theory. In looking at the phenomenon of brand names the paper takes into consideration that dictionaries are reference works that carry with them immense authority. When people have arguments over certain concepts the tendency is to use the dictionary as an arbiter. This practice has extended well into the judiciary systems of many literate societies where the dictionary definition in many cases carries the day in courts. Since dictionaries have so much influence on people's perceptions on interpreting the world (at least from the literary realm) it is the interest of this paper to investigate how brand names are handled in the *Isichazamazwi SesiNdebele*.

Methodologically, the paper relies heavily on desk review of related literature which helps in the understanding of the primary source, the ISN. However, the research does not lose sight of the importance of getting first hand information from the authors of the dictionary under the spotlight which explains the use of interviews with senior editors (the chief editor and the deputy chief editor) to try and get their views on why they took the steps they took and to establish whether they were aware of the legal ramifications, that is, whether the inclusion or exclusion of certain brand names was an expression of their opinion on proprietary rights. In examining the dictionary vis-à-vis the treatment of brand names, the article does not lose sight of the actual linguistic behaviour of the Ndebele linguistic community in its treatment of brand names in everyday discourse. In view of this, the study also sort views from the speakers of the Ndebele language, who are also the target users of the ISN. This was done through a structured questionnaire where the respondents were given ten definitions of unidentified lexical items and they were asked to provide possible lexical items for those definitions. This was a strategy to test the status of the so called brand names/trademarks, at least from the viewpoint of the ordinary speakers of the language. Thus the assumption was that if a significant number of the users identified a brand name as the missing lexical item for the definition provided, then that could be used as a pointer to the generification of that particular brand name.

### **Linguistic diversity and genetic relatedness of African and Southern Bantu languages: opportunities and challenges in the area of lexicography and multi-lingual mother-tongue-based education.**

**Samba Buri MBOUP**

Academy of African Languages and Science, University of South Africa, Pretoria, South Africa

The author of this paper is neither a lexicographer nor a curriculum development specialist. He is an academic trained and working in the field of linguistic and cultural anthropology with a focus on Africa and the African Diaspora, and on an option for African renaissance anchored in multi- and inter-disciplinarity as an alternative paradigm for knowledge production and utilisation. Yet he has a keen interest in both lexicography and curriculum development, as an academic and manager of a UNISA strategic project for the creation of an Academy of African Languages and Science (AALS).

AALS can be defined as a knowledge production and application hub that aims to intellectualise and modernise (South) African indigenous languages [(S)AIL]; and thereby introduce these into curricula as mediums for tuition and research. In this regard, a developmental approach will be adopted in which African languages are promoted as languages for science, technology, education and research, to create conditions for greater social integration

and citizenship as well as cultural self-assertiveness for mother-tongue speakers of (S)AIL at the University of South Africa (UNISA) and beyond.

The paper aims at building on lessons stemming from theoretical assumptions on the genetic relatedness between African languages in general and (Southern) Bantu languages in particular, as established by Meinhof (1899; 1906), Guthrie (1948; 1967) and further confirmed by Greenberg (1966), Diop (1974, 1978; 1977; 1988; 1975) and Obenga (1973; 1985; 1993; 2004). From both a methodological and practical perspective, genetic relatedness may have interesting implications in the area of lexicography and terminology development. In particular, it can be highly productive when applied to various lexicographic and terminographic techniques and procedures such as blending, extension of meaning, paraphrasing and word coinage, as well as in the area of lexicographic standardisation.

The principle of genetic relatedness can also be highly cost effective in terms of saving time and resources when implemented for the promotion of multilingual mother-tongue-based education (MLMTBE), particularly in key areas such as curriculum design and transformation, production of quality didactic materials and enhanced pedagogy.

The approach adopted in the paper is more inductive, prospective and thought-provoking rather than it is conclusively demonstrative. This approach also takes into consideration another key assumption: that with regard to African languages and more specifically Southern Bantu, substantive progress is expected in the areas of descriptive as well as historical and comparative linguistics, and in socio-linguistic studies. This would imply, amongst other things, a conscious and decisive move away from the current 'epistemological comfort zone' markedly characterised by a great number of monographic works on individual languages and in some cases their respective sub-units or dialects, into a new epistemological dispensation that would put greater emphasis on works of generalisation and synthesis.

Hopefully, this new trend in research and scholarship would lead not only to the production of theoretical models for, for example, socio-linguistic survey and (re)classification and management of linguistic diversity in an equitable and sustainable way. It will also be of great significance for applied linguistics, for instance in areas such as lexicography and terminology development as well as curriculum transformation, production of quality didactic materials and enhanced pedagogy.

## References:

Diop, Cheikh Anta

(1974, 1978): *The Cultural Unity of Black Africa*, Chicago, Third World Press; (1990):

London, Karnak House

(1975) : *Comment enraciner la science en Afrique (Tijji xelum sàkk)* : in BIFAN, 37, B Series, N° 1

(1977) : *Parenté génétique de l'égyptien pharaonique et des langues négro-africaines*, Dakar, IFAN-NEA

(1988) : *Nouvelles recherches sur l'égyptien ancien et les langues négro-africaines modernes*, Paris, Présence Africaine.

Greenberg, Joseph (1966): *Languages of Africa*, Indiana University, Bloomington and Mouton & Co, The Hague

Guthrie, Malcolm.

(1948): *The Classification of the Bantu Languages*, London, Oxford University Press for the International African Institute,;

*Comparative Bantu*, Gregg Press, London, reprinted in 1967 and 1970

Meinhof, Carl

(1899): Outline of the Phonetics of the Bantu Languages (*Grundriss einer Lautlehre der Bantusprachen*)

(1906): Principles of the Comparative Grammar of the Bantu Languages (*Grundzüge einer vergleichenden Grammatik der Bantusprachen*)

Obenga, Théophile

(1973) : L’Afrique dans l’Antiquité – Egypte ancienne/Afrique Noire, Présence Africaine

(1985) : Les Bantu, Présence Africaine;

(1993): Origine commune de l’égyphtien ancien, du copte et des langues négro-africaines modernes, Paris, L’Harmattan

(2004): African Philosophy – The Pharaonic Period : -2780-330, Per Ankh, Senegal

## **Purism and inadequacies: A case study of the effects of strict corpus-based dictionary writing in Sesotho sa Leboa.**

**Victor M. MOJELA**

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

Sesotho sa Leboa presently has a corpus of about 6,8 million words (the UP corpus). This corpus is built from the already recorded and published materials in this language. The major irregularity with these published materials is that they are derived from a few dialects of Sesotho sa Leboa, while the majority of the terminology does not form part of the written language. This simply explains why it would not be fair for Sesotho sa Leboa lexicographers to use the existing (insufficient) corpus as reliable tool to compile dictionaries in this language, because almost half of the terminology will be left out of the record. The vocabulary might have been more than twice its size if it was not disadvantaged by purism and the inadequacies emanating from excessive exclusion policies of a stringent type of standardisation.

Geographically, the standard Sesotho sa Leboa orthography (the written language) was built from the dialects within the districts of Sekhukhune, Waterberg and a section of the Capricorn district. The Sesotho sa Leboa dialects in the Northern part of the Capricorn district, the whole of Mopani and Vembe districts, as well as Mapulaneng district, in Mpumalanga, are not part of the written Sesotho sa Leboa language. As a result, terminologies from these sidelined dialects do not form part of the present Northern Sotho or Sesotho sa Leboa corpus of 6,8 million words because these terms were not included in the published materials which were instrumental in the compilation of this corpus.

The sidelined dialects include Selobedu, Sephalaborwa, Sekgaga (Maake & Mogoboya), Seroka, Setlokwa, Sehananwa, Sepulana, etc. Words such as the following are not included in the Northern Sotho written orthography, because their source dialects were sidelined, and their inclusion in the vocabulary would have increased the size of the Sesotho sa Leboa lexicon much further:

**kheṭola** (Selobedu & Seroka) ‘frog’, **segwagwa** (standard NS)

**khemake** (Selobedu) ‘cat’, **katse** (standard NS)

**mokhope** (Seroka, Selobedu) ‘marula beer’, **morula** (standard NS)

**lesalabu** (Seroka, Selobedu) ‘watermelon’, **legapu** (standard NS)

**mphekwa** (Selobedu, Seroka) ‘lizard’ **mokgaditswa** (standard NS)

**moṭanare** (Selobedu, Seroka) ‘mopani tree’ **mopani** (standard NS)



**tsheṭa** (Selobedu, Seroka) ‘greedy’ **bojato** (standard NS)

The standardisation system in this language is still dominated by purism, selectiveness and destructive exclusion policies. Purism and selfish standardisation is due to the fact that most influential members in the language boards prefer to standardise their own dialects and sideline all other dialects which are not represented in the standardising committees. One of the major reasons why the development of languages like English is very fast is because purism is minimal, and it is interesting to realise that the English language has lemmatised many lexical items from most South African indigenous languages, including slang and those of the sidelined or ‘stigmatised’ dialects of Northern Sotho, without fear of ‘contamination’. Dictionaries, like *A Dictionary of South African English on Historical Principles* (Oxford:Oxford University Press, 1996) have lemmatised most of the South African indigenous language terminologies as loan words into English, and these terms are now part and parcel of the English corpus. The following are examples in this regard:

<b>moloji</b>	‘witch’ or ‘wizard’ (1996:472)
<b>mampara</b>	‘a fool’ or ‘fools’ (1996:472)
<b>mpimpi</b>	‘an informer’ or ‘an evil collaborator’ (1996:481)
<b>potsotso</b>	‘girls’ tight trousers’ (1996:102)
<b>Zola Budd</b>	‘a type of minibus taxi’ (1996:808)

For the Sesotho sa Leboa lexicographer, corpus lemmatisation means omitting the bulk of the Northern Sotho terminology from dictionaries, thereby facing the problem of producing dictionaries which are not only one-sided, but also dictionaries which are too inadequate for this language. The main objectives in this research can be, therefore, be summarised as follows: to show the inadequacies of corpus-based dictionary compilations in Sesotho sa Leboa, and to emphasise the importance of field research to bring on board all the lexical items from the sidelined dialects into the Sesotho sa Leboa lexicon.

### **Recommendations:**

The Sesotho sa Leboa National Language Body, which replaced the former Northern Sotho Language Board as standardising authority in this language, should avoid and eradicate all sorts of purism in the standardisation processes. The lexical items from all the Sesotho sa Leboa dialects should not only be incorporated in the vocabulary, but should also be standardised. Lexicographers, especially in Sesotho sa Leboa, should not only rely on the existing corpus when compiling dictionaries, but should embark on intensive research and fieldwork to bring all the omitted lexical items on board in order to have a full and accurate record of the vocabulary of the Northern Sotho language.

## **What is right or wrong? – In quest of formal Lusoga**

**Minah NABIRYE**

Department of African Languages and Cultures, Ghent University, Belgium

In September 2005, the Ugandan parliament approved the use of mother-tongue languages both as a medium of instruction and as a subject in primary schools. The Minister of Education and Sports passed on the implementation of the language policy to the National Curriculum Development Centre (NCDC) to start in February 2007 (NCDC 2006: 1). The NCDC conducted a pilot study of a proposed thematic curriculum in the period 2005-2006, but results from this

pilot showed a grim picture for the success of mother-tongue education in Uganda (Nabirye & De Schryver 2010: 333-334). In spite of the limitations cited, mainly hinting at the dire need for mother-tongue educators and relevant literature, the implementation process is ongoing. If Lusoga is to be used in earnest as a language of instruction, however, it needs a formally accepted standard as a point of reference. This has significant implications for dictionary compilation as well as for corpus building, both of which are explored in this paper.

Though preliminary writing guides are available in existing orthographies and dictionaries, there are additional and unclear instances of Lusoga concepts and usages that lack a defined lexicalisation standard. Presently, there is very little concession on what should or should not be Lusoga in the first place, and what the standard should be. Some problems central to formal language specification have not yet arisen, and are therefore not yet a subject of enquiry. Questions as basic as what a standard is and why it is necessary are becoming increasingly vital. Hence the need to look at the criteria required to differentiate cases such as the following:

- 1a. March **Ogwokusatu** (full construction, capitalised, no abbreviation or complement)
- b. a third sth **ogw'okusatu** (not capitalised, needs a noun complement)
- 2a. weeks **saabbiiti** (full construction, not capitalised)
- b. Sunday **Ssaabbiiti** (full construction, capitalised)
- 3a. Ugandan **Omunauganda** written as **(o)muna Uganda / (o)munaUganda**
- b. Musoga **Omusoga** written as **omusoga / omuSoga**

Studying the lexicalisation of the Lusoga names for months as shown in (1a), we find that *March* **Ogwokusatu** is a compounded form which does not require any further information to fulfil its function. It is a full self-standing construction and a proper name which is capitalised. *Third* as in **ogw'okusatu** in (1b), however, requires qualification to specify the object of say *a third month* **omwezi ogw'okusatu** (which is not the same as March) or *a third time* **omulundi ogw'okusatu**. (2a) is a dependent construction requiring modification to arrive at the full meaning. The difference between (1a) and (1b) is based on the semantic interpretation derived from the use of capitalisation and abbreviation.

In (2a) and (2b) the full meanings are realised as a single lexical form, but the meanings are differentiated by capitalisation. Examples (3a) and (3b) each have two parts whose writing is not yet synchronised. Some write words of this nature as two separate concepts and since they are supposed to be proper nouns requiring capitalisation, treatment of the right placement of the capitalisation is haphazard.

Examples (4a), (4b) and (4c) show cases of homography. (4b) only makes a full construction if it is written conjunctively, while in (4c) the form can be repeated as many times as the emphasis allows, e.g. **boona boona boona boona...** but the meaning is not the same as in (4a).

- 4a. also **boona** (single lexical construction)
- b. all **boonaboona** (single lexical construction, written conjunctively)
- c. all (+ emphasis) **boona boona** (compounded construction, written disjunctively)

Finally, abbreviated forms also need to be contextualised. For instance, abbreviations should indicate the status of a form as either formal or informal, as in (5a) and (5b). These should not be mistaken for cases such as (1a) and (1b), which result in semantic change, but should be considered as merely formal lexicalisation.

- 5a. as a person **ng'omuntu** (informal representation)
- b. as a person **nga omuntu** (formal representation)

The compilation of the first monolingual Lusoga dictionary, the *Eiwanika ly'Olusoga* (Nabirye 2009), was subjected to repeated scrutiny from purists who sought only the 'right terms' to qualify as entries. Given few Lusoga reference works existed at the time, and given that the dictionary was compiled without access to a large corpus, there was very little to guide the right or wrong choice. A newly built corpus, however, is now showing different results and revealing new Lusoga concepts and lexicalisation processes that need further clarification. The paper will therefore highlight the type of instances that have been observed in Lusoga, instances which now need to be specified based on the examples found in the corpus. The paper will also show how far the existing Lusoga orthographies and dictionaries have gone in guiding the Lusoga lexicalisation processes and what still needs further clarification, with the aim of guiding the founding of formal Lusoga.

### References:

- Nabirye, M.** 2009. *Eiwanika ly'Olusoga*. Kampala: Menha Publishers.
- Nabirye, M. & G-M de Schryver.** 2010. The Monolingual Lusoga Dictionary Faced with Demands from a New User Category. *Lexikos* 20: 326–350.
- NCDC.** 2006. THEMA. The Newsletter of the Thematic Primary Curriculum. Issue 1. August 2006. Kampala: National Curriculum Development Centre.

## **Compilation of An English-French Dictionary of Phonetic Sciences: An Ongoing Project**

**Hugues Steve NDINGA-KOUMBA-BINZA**

Centre for Text Technology (CTexT), North-West University (Potchefstroom), South Africa

This paper gives an account of a research project which is concerned with the compilation of a bilingual English-French dictionary of phonetic sciences. The research proposal is drawn from various discussions and the experience the author has had at the Centre for Lexicography (CentLex) at the Aarhus School of Business in Aarhus, Denmark.

Bilingual or multilingual specialised lexicography has often been neglected, but it is recognised that CentLex has significantly contributed with cutting-edge specialised lexicography research and the production of specific-domain dictionaries (cf. Bergenholtz & Tarp, 1995; Bergenholtz, Nielsen & Tarp, 2009; Fuertes-Olivera, 2010). In the African context, very few experiences of specialised lexicography exist. One may quote Carstens (1997) about the planning of a multilingual explanatory dictionary of chemistry for South African students, Carstens (1998) on the science terminologisation in Sepedi, and the compilation of the Shona-English biomedical dictionary (Mpofu & Mangoya, 2005). These three projects have the following characteristics in common: they are concerned with natural science and they involve African languages of the Bantu phylum.

However, it can be noticed that language sciences in particular (or phonetics in this case) are barely the topic of specialised lexicography research projects. Nevertheless, the development of new technologies has had an important impact on language sound studies. Phonetic sciences have, in fact, experienced a rapid expansion within an interconnection with other research disciplines and technologies. The outcome of this development within the discipline itself has been not only the creation of many sub-disciplines, but mainly the creation of various new terminologies, lexicons and jargons. (It should be noted that the concept of "phonetic sciences" is herein understood as it is conveyed within the International Congress of Phonetic Sciences, that

is, a broad meaning that includes not only the various sub-fields of phonetics but also related disciplines such as phonology, audiology, sound acquisition, speech synthesis and speech recognition.)

This development, though, is mainly experienced in the English language. It is indeed often acknowledged that countries of other languages, especially French-speaking African countries, are most of the time outdated as far as technologies are concerned. This is particularly the case in the field of phonetics.

As a trained phonetician and a French mother-tongue speaker with research experience in both French-speaking and English-speaking countries, the author has been aware of the serious lack of relevant and/or exact terminologies in the French language as far as new phonetic advances are concerned. Experience also shows a variety of French terms of the same meaning when individual attempts are made to translate the new English terms of phonetics. Thus, a standard translation dictionary from English to French is needed in the field of phonetic sciences.

Furthermore, while many dictionaries and encyclopedias of phonetics (or containing large articles on any phonetic science) exist in English (e.g. Crystal, 2003; Brown, 2006; Smelser & Baltes, 2001) these are hard to find in French. The existing dictionaries in French are most often very general and articles on phonetic sciences are among the less developed (e.g. Dubois, 1994; Ducrot & Schaeffer, 1995). Eventually, an English-French dictionary of phonetic sciences is likely to be the first of its kind in the field study of language sounds.

The main aim of this project, therefore, is to provide standard translation dictionary of phonetic sciences. The planned dictionary should serve as a tool that would help in developing phonetics and enhancing language technologies in various French-speaking countries. The purpose of the planned dictionary is to address two types of differences in terminology, i.e. the differences between English and French in phonetic sciences, and the differences within the French language where the planned dictionary will also serve as a tool for the standardisation of sound studies terms.

This paper comes within the conceptualisation plan for the compilation of the intended dictionary. The discussion will broadly cover the salient aspects of the proposed dictionary, including:

- (i) the theoretical and methodological approach used in the research project,
- (ii) the data collection and the lemma selection
- (iii) the functions and the structures of the dictionary,

The paper concludes on the importance of specialised translation dictionaries for knowledge dissemination in various languages and 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.
- Brown, K. (Ed.). 2006. *The encyclopedia of language and linguistics*. Oxford: Elsevier.
- 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.

- Crystal, D. 2003. *A dictionary of linguistics and phonetics*. London: Blackwell Publishing.
- Dubois, J. 1994. *Dictionnaire de linguistique et des sciences du langage*. Paris: Larousse.
- Ducrot, O. & J.M. Schaeffer. 1995. *Nouveau dictionnaire encyclopédique des sciences du langage*. Paris: Le Seuil.
- 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.
- Smelser, N.J. & P.B. Baltes. (Eds.). 2001. *International encyclopedia of the social and behavioural sciences*. Oxford: Pergamon.

## **THE TREATMENT OF ETYMOLOGY IN SETSWANA DICTIONARIES**

**Thapelo J. OTLOGETSWE**

Dept of English, University of Botswana, Gaborone, Botswana

In November 1857 Richard Chenevix Trench read a paper entitled *On some deficiencies in our dictionaries* to the Philological Society of England. This paper has been considered by Winchester as a formidable critique of the few dictionaries then in existence. Trench argued that dictionaries suffered from a number of shortcomings – grave deficiencies from which the language and, by implication, the Empire and its Church might well eventually come to suffer (Winchester 1998: 70). The paper was significant in that it gave the first impetus to work on the magnificent *Oxford English Dictionary* (OED). Because of the weaknesses outlined, Trench argued that there was therefore a need for *a new dictionary written on historical principles*. Such a dictionary, he argued, would record every word from its birth to its death, carefully documenting its shades of meanings. He considered a dictionary as ‘an historical monument, the history of a nation contemplated from one point of view; and the wrong ways into which a language has wandered or been disposed to wander, may be nearly as instructive as the right ones in which it has traveled: as much as may be learned, or nearly as much from its failures as from its success, from its follies as from its wisdom’ (Richard Chenevix Trench, 1860:7).

Trench’s criticism offers some useful insights to the neglected subject of etymology in Setswana, though we do not argue for Setswana dictionaries to be compiled on historical principles. The treatment of etymology in dictionaries for African languages is an important one, and yet one that is grossly neglected in African language lexicography.

This paper outlines Trench’s key arguments and applies them to the Setswana etymology project, an important part of the Setswana monolingual dictionary currently under compilation. The paper also argues that one of the most neglected elements of Setswana lexicography is the subject matter of etymology which was key to Trench’s argument and which has been definitive to the OED project. The paper therefore assesses the treatment of etymology in five (5) different Setswana dictionaries, namely: *Thanodi ya Setswana ya dikole* (Kgasa, 1976); *Dikišinare ya Setswana English Afrikaans* (Snyman et al, 1990); *Thanodi ya Setswana* (Kgasa & Tsonope, 1995) and *English-Setswana-English* (Matumo, 1993). Kgasa (1976) traces a word to the language of origin but does not give its etymon (a word in the source language). He has 239 headwords marked with etymology. Snyman et al (1990) just like Kgasa (1976) marks the source language of the etymon, but does not give the etymon. Kgasa and Tsonope (1995) do not include any etymological mark-up in the dictionary. Matumo (1993) marks words as being of foreign origin (FOR) but does not mark the source language or the etymon.

Historically Setswana has been, and currently is, in contact with Afrikaans, English, other local and regional languages (e.g. Zulu). The dictionary must therefore attempt to capture the degree of lexical influence from these languages in its pages. As a growing language, Setswana has been creating numerous words through a variety of word formation processes such as coining, blending etc. It is critical that the origins of such terms are preserved within the pages of a monolingual dictionary. We are currently working on a Setswana monolingual dictionary which has about 16,090 entries at present. 1,067 of these are etymologically marked (6.6%). Three strategies are used in the dictionary etymology mark-up. First, the source language as well as the etymon is given. For instance:

**jêmê** /ʤɛmɛ/ [ln. 9. n- ➔ 10. din-] sejo se se borekereke se se botshe se se tshasiwang mo borothong *Ke rata botoro ka gone jeme e na le sukiri e ntsi* ⇐ Sekgoa: jam

The second strategy is an elaborate one, in which a word whose origin is traced to some other Setswana word(s) is treated elaborately. This is especially visible in the etymology of months. For instance:

**Ferikgong** /fɪriq<sup>h</sup>oŋ/ [ln. 1a. Ø ➔ 2a. bo-] kgwedi ya ntlha ya ngwaga ⇐ Kgwedi e e reilwe ka lephoi la mofiri kgotsa kofiri le le a beng le sela dikgonnyana, le aga sentlhaga, go tla le simolola go baya mae. Ka go nna jalo leina le le tswa mo go mofiri le dikgong.

The final strategy is the non-elaborate strategy in which the same word is traced to Setswana as shown in the example below.

**gôpane** /χɔpanɪ/ [ln. 1a. Ø ➔ 2a. bo-] mokgantitswane o motona o o nnang mo sekgweng = kgwathê ⇐ gôpa<sup>2</sup>

The paper hopes to initiate debates on the treatment of etymology in African languages and languages in contact situations.

## A critical analysis of the lemmatisation of nouns and verbs in isiZulu

**Danie J. PRINSLOO**

Dept of African Languages, University of Pretoria, South Africa

The publication of the first dictionary for isiZulu using a word, instead of the traditional stem lemmatisation strategy reopens the debate on stem versus word lemmatisation in African languages. In particular the question is whether the problem of stem identification which proved to be the major stumbling block for learners to find lemmas in isiZulu dictionaries has been solved? To date most publications on lemmatisation in the African languages were contrasting disjunctively written languages (e.g. Sepedi, Setswana and Sesotho) with those with a conjunctive orthography (e.g. isiZulu, Siswati and isiXhosa) in terms of the advantages and disadvantages of stem versus word lemmatisation. It was mainly argued that stem lemmatisation is an accepted, or even the best strategy for conjunctively written languages, but that word lemmatisation is a better option for disjunctively written languages mainly because stem lemmatisation introduces unnecessary problems for the user of a dictionary of a disjunctively written language, for example, to identify nominal stems. The stem tradition, nevertheless, supported by certain assumptions such as being the more scientific option gained such momentum that a number of stem dictionaries were compiled for the Sotho languages. Word

lemmatisation for conjunctively written languages was considered by Van Wyk (1995) and preliminary experiments were conducted at some of the National Lexicography Units in South Africa on the feasibility and possible advantages of word lemmatisation for conjunctively written languages. However, it was only in 2010 with the publication of the *Oxford Bilingual School Dictionary: Zulu and English* (OZSD) that the almost sacred stem tradition of lemmatisation for a Nguni language was broken, using word lemmatisation for an isiZulu dictionary.

The focus of this paper differs from previous publications in the sense that first, the issue of stem identification takes centre stage, and secondly that the advantages and disadvantages of stem versus word lemmatisation will not be described in terms of conjunctively versus disjunctively written languages but in terms of the benefits versus shortcomings of these approaches for the conjunctively written Nguni languages, isiZulu being a case in point. Thirdly, although a selection of examples will be offered, example analysis will be focused on a paradigm of 2,525 occurrences of different words containing the stem *sebenza* 'work' occurring five times or more in the *Pretoria isiZulu Corpus* (PZC).

A consolidation of the most prominent views on stem versus word lemmatisation which lie scattered over a number of publications will also be attempted. Finally, the success or potential of electronic dictionaries to solve stem identification problems which cannot be solved in paper dictionaries, irrespective of the lemmatisation strategy, will be evaluated.

The discussion focuses on strict stem lemmatisation, lemmatising stems and suffixes, left expanded article structures, word lemmatisation and lemmatisation in electronic dictionaries. Consider the following simplified examples where boldface indicates the lemma:

Strict stem lemmatisation	<b>sebenza</b>
Stem plus suffixes	<b>sebenzela</b>
Left expanded	<i>imisebenzi, ukusebenza</i>
word lemmatisation	<b>imisebenzi</b>

It will be concluded that the weakest option for lemmatising nouns and verbs in isiZulu is the strict stem strategy where the lemma is the basic stem: in the case of verbs, verbal stem without suffixes and in the case of nouns, noun stem without nominal prefixes. This lemmatisation strategy is not user friendly, stem identification is a major obstacle, a huge amount of knowledge of morphophonetics is presupposed and the user is often in doubt whether (s)he has successfully retrieved information. Even if the users managed to identify the stem and to look it up, all the additional information conveyed by the affixes have to be 'added back on' and the user will not know for sure whether (s)he came to the right conclusion. Lemmatising verb stems with their suffixes represents a slight improvement. At least the meanings of the suffixes need not be artificially added on as in the case of strict stem lemmatisation.

Lemmatising stems with their prefixes is a better option because the user has the advantage of seeing the full form of infinitive verbs and the full forms of nouns with additional information such as tonal indication. This strategy is more user-friendly but stem identification remains problematic and a substantial amount of knowledge of morphophonetics is still presupposed.

Word lemmatisation applicable to nouns is by far the better strategy because nouns can be looked up under the first letter. For given non-derived nominal forms the problem of stem identification is solved for all nouns. This strategy is especially beneficial for those nouns where stem identification is problematic. The strategy is user-friendly and no knowledge of the grammar is presupposed. However, for nominal and verbal derivations, especially those where nominal and verbal stems occur with huge clusters of circumfixes, the problem of stem/word identification remains unsolved.

The problem of word/stem identification which is present in all of the lemmatisation strategies employed for isiZulu can only be solved in electronic dictionaries. Most electronic dictionaries are mere translated word lists and are not of much use to the target users especially for their productive needs. A clear exception is *isiZulu.net* where the problem of stem/word identification has been solved for most of the frequently used words in isiZulu, but more comprehensive electronic isiZulu dictionaries are required to alleviate the need for stem/word identification for less frequently used words as well.

**Reference:**

(OZSD) De Schryver, G.-M. (Editor). *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. 2010, Cape Town: Oxford University Press Southern Africa.

**Requirements for an on-screen presentation of an electronic dictionary of German collocations for Afrikaans-speaking learners of German as a foreign language**

**Hilma Anka REINHARDT**

University of Stellenbosch, South Africa

After analysing the treatment of collocations in the only available Afrikaans and German dictionary *Woordeboek/Wörterbuch Deutsch-Afrikaans Afrikaans-Deutsch*, 8<sup>th</sup> edition, by Trümpelmann and Erbe (1983), it was concluded in line with Pienaar (2006) and Gouws (1995, 1997) that collocations are not satisfactorily dealt with in this dictionary, especially not with regards to language production. As part of a remedy we propose to develop an electronic dictionary of German collocations.

In view of the envisaged primary users of such a dictionary (Afrikaans-speaking learners of German as a foreign language, GFL learners, cf. Reinhardt, 2011, sections 2.2 and 2.3), its primary dictionary function (cf. Tarp, 2008) should be language production. Furthermore, it should feature elements of the following dictionary types: dictionary of collocations; dictionary for foreign language learning, which is, for pedagogical reasons, mainly monolingual – German (cf. Reinhardt, 2011: 35-37); electronic (online) dictionary.

In this paper we outline a few aspects of a proposed design of such a dictionary. We focus on data presentation (for example, screen layout) and interaction possibilities (search and browsing). The data to be presented may include the collocation (as a treatment unit), its linguistic properties (like morphosyntactic preferences, syntactic construction) and its relations with other collocations or with other treatment units (cf. Heid and Gouws, 2006). For on-screen presentation and interaction possibilities, we take inspiration from international standards for software usability (DIN EN ISO 9241-11: 1998, DIN EN ISO 9241-110: 2008), usability heuristics as suggested by Sarodnick and Brau (2006) and Nielsen (c1995-2010, 2000 and c2005) and from practical tips by Krug (2006).

Considering the above, a set of lexicographic and usability requirements is established aiming at a presentation which should allow the user easy and satisfactory access to the data, without requiring lexicographic or linguistic background knowledge. Based on the above requirements, screen layout and functionality are proposed.

A clear and consistent design is followed through all the on-screen presentation. It allows the user to easily navigate within the dictionary site. Complying with the overall design, the



homepage already informs the user about the kind of information he/she can expect to find in the dictionary. It also provides an insight into the functionalities and interaction possibilities on offer: how access to the data contained in the dictionary can be gained and where different kinds of information can be found.

Following functionality of standard internet browsers, navigation is handled with the forward and back buttons as well as conventional links. For all presented search results or content of the dictionary, a selection of the assumed most useful data will be shown (for example, most common collocations for a particular base). However, the dictionary offers appropriate links to expand such selections to show the maximum range of relevant data available.

Two different kinds of access are suggested as access structure: search (to satisfy immediate / short term user needs in language production) and browsing (catering for the longer term user needs – creating awareness of and facilitating acquisition of collocations).

The search function is facilitated via a search field. For lexicographic and pedagogical reasons we argue to promote two ideal forms of input by the user: the target language base and the target language collocation (cf. Reinhardt, 2011: 120-121), where all elements – bases and collocates – are entered in lemma form. However, input should not be limited to the lemma forms of words, but input of inflected forms should be allowed (Reinhardt, 2011: 122) and spelling errors tolerated (cf. Bank, 2010: 38).

Though not ideal (Reinhardt, 2011: 122), GFL learners should also be allowed to input an Afrikaans base (Reinhardt, 2011: 122) and even an Afrikaans collocate (Reinhardt, 2011: 124-125). Both these inputs cater for the target user group having just completed the beginners language learning level and still using one-to-one translation in their language production (Reinhardt, 124-125; Waring, 2001).

Search results include a definition of the envisaged collocation via a link, which allows users to unfold/refold it for viewing. Where the definition itself contains collocations, these are linked to their respective dictionary entries. Examples of use of the collocation are provided, and linked where appropriate. In order to enhance the learner's paraphrasing knowledge, useful synonymous collocations are offered and linked to their respective entries. If appropriate, combinations of collocations in which the current collocation occurs, are made available via a link.

Browsing the dictionary enables learners to view collocations onomasiologically so that language production skills may be improved on a long term basis (Gellert, 2001: 2). Search and browsing procedures can be combined to find appropriate collocations.

The presentation will provide theoretical background pertaining to design and layout decisions for the on-screen presentation of the dictionary including a walk-through of the search process. Connections to the mentioned usability standards will be highlighted.

The requirements upon which the envisaged collocations dictionary for GFL learners is based are neither limited to the particular user group of GFL learners, nor to the languages Afrikaans and German. Considering especially the multilingual societies in South Africa and Namibia, it must be emphasised that the recommendations made here may just as well be applied to the usage situations of African and any other learner groups and languages.

## References:

- Bank, C. 2010. *Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen*. Master thesis. Hildesheim: Universität Hildesheim.
- 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.
- DIN EN ISO 9241-110. 2008. *Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006); Deutsche Fassung EN ISO 9241-110:2006*. Berlin: Beuth.
- Gellert, A. 2001. *Ein deutsch-deutsches Lernwörterbuch auf der Grundlage des (überarbeiteten) Zertifikats Deutsch - und wie man damit arbeiten kann* [Online]. <http://www.sprachwissenschaft.ch/IDT2001/pdf/Gellert.pdf> [15.08.2005].
- Gouws, R. H. 1995. Woordeboeke met Afrikaans en Duits as taalpaar: 'n metaleksikografies situering binne die Afrikaans vertalende leksikografie. *South African Journal of Linguistics / Suid-Afrikaanse Tydskrif vir Taalkunde*, 25:3-87.
- Gouws, R. H. 1997. Die zweisprachige Lexikographie Afrikaans-Deutsch/Deutsch-Afrikaans: Ein Überblick unter linguistischem Aspekt, in Wiegand, H. E. *Studien zur zweisprachigen Lexikographie mit Deutsch III*. 35-86.
- Heid, U. & Gouws, R. H. 2006. A model for a multifunctional electronic dictionary of collocations, in Corino, E., Marelllo, C. & Onesti, C. (Hrsg.). *Atti del XII Congresso Internazionale di Lessicografia / Proceedings of the twelfth Euralex International Congress, Turin*. Alessandria, Italien: Edizioni dell'Orso. 979-988.
- Krug, S. 2006. *Don't Make Me Think! A Common Sense Approach to Web Usability*. 2. Ausgabe. Berkeley: New Riders.
- Nielsen, J. c1995-2010. *How to Conduct a Heuristic Evaluation* [Online]. [http://www.useit.com/papers/heuristic/heuristic\\_evaluation.html](http://www.useit.com/papers/heuristic/heuristic_evaluation.html) [29.09.2010].
- Nielsen, J. 2000. *Designing Web Usability: The Practice of Simplicity*. Thousand Oaks, Kalifornien: New Riders Publishing.
- Nielsen, J. c2005. *Ten Usability Heuristics* [Online]. [http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html) [29.09.2010].
- Pienaar, S. 2006. *A critical evaluation of Woordeboek/Wörterbuch with regard to source- and target language forms*. Master thesis. Stellenbosch: Stellenbosch University [Online]. <http://etd.sun.ac.za/handle/10019/1160> [01.05.2009].
- 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]
- Sarodnick, F. & Brau, H. 2006. *Methoden der Usability Evaluation -Wissenschaftliche Grundlagen und praktische Anwendung*. Bern: Hans Huber.
- Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Lexicographica Series Maior. Tübingen: Niemeyer.
- Trümpelmann, G.P.J. & Erbe, E. 1983. *Woordeboek / Wörterbuch – Deutsch-Afrikaans – Afrikaans-Duits*. Pretoria: Van Schaik.
- Waring, R. 2001. *How to get your students to use their dictionaries effectively* [Online]. <http://www1.harenet.ne.jp/~waring/vocab/dictionary/dictionary.htm> [23.02.2010].

## **A global approach for a dictionary of Lingála: from localisation of software to a lemmatisation strategy.**

**Bienvenu SENE MONGABA**  
University of Ghent, Belgium

In the context of DR Congo, French is the language of teaching and learning. However, today's school population shows an insufficient command of it. This makes it desirable to use at least one of the four national languages of DR Congo (Lingála, Cilubà, Kikongo and Kiswahili) in diglossia with French as a teaching strategy for better appropriation of knowledge and know-how.

In our sociolinguistics surveys, we observed that the national languages are acquiring stronger and stronger empowerment in the linguistic market of DR Congo. This means that teachers have to use national languages when they are teaching in order to be understood. However, teachers are limited in using those languages because of the lack of teaching aids such as schoolbooks, maps and dictionaries. Dictionaries are particularly important in this context, as sociolinguistic realities show that bilingual dictionaries are needed to explain in the national languages the meaning of the scientific terms which are used in French in the classroom.

Lingála is the national language spoken in Kinshasa, the capital of the country. Even if this language is going through a straightforward empowering process, i.e. it is used now in areas which used to be the preserve of French, such as the media, justice and schools, its progress is still hindered by the lack of terminological and lexicographic tools; hence the need to provide monolingual or plurilingual dictionaries for this language, according to what is a demand of society.

This is why we have undertaken this research on lexicography. It consists in producing lexicons as teaching aids and in making them available to primary users, that is, teachers and students. For that purpose, the approach to lemmatisation is a crucial factor for finding a word easily in a dictionary.<sup>7</sup>

There are three main aspects concerning Lingála: morphology, concordance rules and orthography. First of all, we have to establish how to present derivative words formed from the same stem but having different prefixes that indicate singular, plural, process, state and so on. As for the two last aspects, we must make the choice of whether or not to record all the alternative versions of the same term observed in the corpus.

### **How the study was conducted**

We started off observing the traditional presentation in existing dictionaries and we proposed some adaptations to achieve the best presentation in terms of contents, orthography and register.

We have constituted our corpus from textbooks, existing Lingála dictionaries and web texts. The main material came from internet forums, where people use actual language. We have also used a recent version of the Bible and some existing novels and schoolbooks, as well as an oral corpus put together by ourselves, drawing from everyday conversations and from the media (radio and television).

The tools used are on the one hand Unitex, a corpus software, and on the other hand TshwaneLex, a lexicographic software. Unitex is a corpus processor in which one can put

---

<sup>7</sup> De Schryver G-M 2008, *A new way to lemmatize adjectives in a user-friendly Zulu-English Dictionary* Lexikos 18 (AFRILEX-reeks/series 18: 2008) 63-91.

linguistic resources and use them.<sup>8</sup> It allows one to install one's own language and process text in that language. We began by making an alphabetic file of the Lingála language. Then we created a dictionary of inflected forms and a dictionary of non-inflected forms. This process allowed us to obtain a wordlist, frequency indication, tagged words and collocations. We used these resources to select from the corpus the most frequent lemmas, examples and potential definitions for inclusion in the lexicon.

TshwaneLex allowed us to record data for the purpose of producing dictionaries. The first step consisted in the localisation of software in Lingála, which enabled us to work in a Lingála environment. In this context, localisation means "*that the entire lexicographic process, from initial compilation all the way to final product, may henceforth be conducted in any language of one's choice*".<sup>9</sup> Using TshwaneLex in a Lingála environment, we started recording entries and analyzing what would be the most satisfactory lexicographic approach (lemmatisation, contents...) for recording entries in a Lingála dictionary.

### **Preliminary conclusions**

At this stage, we have localised TshwaneLex and Unitex in Lingála. Working in a target language environment (Lingála) helps us have a global view of what we want to propose as an entry and its presentation.

To illustrate our work, we have created a bilingual (French and Lingála) chemical lexicon for secondary school students of chemistry and a bilingual (French and Lingála) general lexicon for primary and secondary school students.

---

<sup>8</sup> Unitex 2.0 2008, User Manual, p.11

<sup>9</sup> De Schryver G-M 2006, *Internationalisation, Localisation and Customisation: Aspects of TshwaneLex* Lexikos 16 (AFRILEX-reeks/series 16: 2006) p. 223.

## **Meaning explanation in learner's e-dictionaries: Current strategies and their theoretical, functional, user and practical motivation**

**Piet H. SWANEPOEL**

University of South Africa, Pretoria, South Africa

In arguing for the inclusion of double-decker definitions in electronic monolingual learner's dictionaries, Fillmore (2003) makes the general remark that the new generation of e-dictionaries provides lexicographers with enormous space for meaning explanations but that developments in lexical semantics do not spell out what is to be included in these vast spaces and why.

Theoretical considerations and space are, however, not the only variables that have to be taken into account to optimise the efficacy of meaning explanations in learner's e-dictionaries. The main goal of this paper is to indicate how a number of theoretical, functional, user and practical variables determine how lexicographers design meaning explanations in a corpus of learner's e-dictionaries and how they utilise the three major advantages of electronic platforms (Internet, CDROMs) to do so: space (increases on the mega-, macro- and micro-level), hypertextuality (linking), and multimodality.

Understanding the meaning of a lexical item is crucial to a number of the functions that learner's dictionaries are used for, which underscores the need to optimise the comprehensibility of meaning explanations. Learners differ, however, in their knowledge of an L2/L3, and thus in their ability to comprehend meaning explanations, and they also differ in their ability to use

e-dictionaries optimally. Consequently, lexicographers have to take these variables into account in designing meaning explanations, as well as the limitations imposed by practical considerations (e.g., the possibilities of the platform, medium, costs, etc).

Although not always spelt out, lexicographers currently rely on a number of lexical semantic theories in explaining the meaning of lexical items in learner's e-dictionaries. Besides the use of genus and differentia definitions, inclusion of numerous types of sense relations identified in structural semantics (synonymy, antonymy, hyponymy, hyperonymy, meronymy, etc. both as definition types and as information separate from genus et differentia definitions, plus linking to thesauri) have become a standard part of the meaning explanations. Cognitive semantics with its focus on categorisation and category structure has motivated the inclusion of information on prototype meaning, basic level categorisation, and inclusion of more systematic (category internal) information on metaphoric and metonymic sense extensions. Frame semantics forms the theoretical basis for linking meaning explanations to larger dictionary internal encyclopaedic knowledge structures (frames of all kinds), to more encompassing entries in encyclopaedias, and to illustrations, and to an exponential growth in the number of contextualised senses discerned for lexical items. Barselou's kind of multimodal semantics and other frameworks which stress the situatedness and embodiment of meaning open up new venues to maximise the explanatory strategies provided by the Internet and data basis technologies although these possibilities have not been utilised to a great extent.

The paper concludes with a number of recommendations for further research.

#### **References:**

Fillmore, C. 2003. Double-decker Definitions: The Role of Frames in Meaning Explanations. *Sign Language Studies*, 3(3): 263-295.

### **Do we need a (new) theory of lexicography?**

**Sven TARP**

Aarhus School of Business and Social Sciences, Aarhus University, Denmark

Lexicography is in the middle of an important transition from printed to electronic dictionaries. In this process the question has been raised whether we need a new theory that may guide the conception and production of lexicographical e-tools or if we can use the theories already developed in the era of printed works. In order to answer this question it is first of all necessary to give answer to the question whether a lexicographical theory exists, is possible at all or even wanted. In fact, the very concept of theory is widely disputed within lexicographical circles (cf. Tarp 2010). In a recent publication where he discusses the theory of lexicographical functions, Yukio Tono (2010) asks: "Do we really need a 'theory'?" Tono's final answer to his own question is affirmative. At a high level of abstraction, that is, independent of the specific theory in question, this view is shared by the Russian tradition (cf. Sorokoletov 1978), by Wiegand (1989), the function theory (cf. Tarp 2008), and at least part of a Chinese tradition (Yong & Peng 2008), among others. However, other lexicographers, especially those belonging to the Anglo-Saxon tradition, do not support this vision of lexicography. For instance, Sue Atkins and Michael Rundell (2008) "do not believe that such a thing exists" (i.e. a theory of lexicography). And in a recent book, Henri Bejoint writes:

“I simply do not believe that there exists a theory of lexicography, and I very much doubt that there can be one. Those who have proposed a general theory have not been found convincing by the community, and for good reasons. A theory is a system of ideas put forward to explain phenomena that are not otherwise explainable. A science has a theory, a craft does not. All natural phenomena need a theory, but how can there be a theory of the production of artefacts? There are theories of language, there may be theories of lexicology, but there is no theory of lexicography. Lexicography is above all a craft, the craft of preparing dictionaries, as well as an art, as Landau (2001) says. It may be becoming more scientific, but it has not become a science.” (Henri Bejoint 2010)

The paper will argue that this point of view is strongly embedded in an Anglo-Saxon academic tradition, according to which science is only related to natural phenomena and where all other phenomena are referred to the sphere of art and craft. As mentioned, this tradition is widely opposed by the traditions in other parts of the world. For instance, Sorokoletov more than 30 years ago defined lexicography as “the science of the classification processes of word material and its presentation in dictionaries.”

The paper accepts the idea that a craft is neither a science nor a theory, but it argues that, based upon a meticulous study of the corresponding cultural practice, it is both necessary and perfectly possible to make reflections and little by little systemise them into a organised set of ideas or statements, i.e. a theory capable of explaining, guiding and even renovating the existing practice. This was what Scerba (1940) intended to do in his ground-breaking contribution to lexicography, what Wiegand did with his “general theory of lexicography”, and what has been done with the “theory of lexicographical functions”. It may be that these theories “have not been found convincing” by the Anglo-Saxon lexicographical community, but this does not mean that they do not exist, are not possible and even highly needed by those who try to solve the complex problems within present-day lexicography.

Departing from the fact that lexicographical works are multi-faceted cultural artefacts and utility tools which, during the millenniums, have met a wide range of different needs detected in society and covered almost all spheres of human activity and knowledge, the paper will defend the idea that a theory of lexicography should not focus on the differences regarding the specific content of all these works, but on aspects that unite them and are common to all of them. In this regard, some of lexicography’s core characteristics will be discussed together with its complex relation to other disciplines, especially information science with which it seems to have many things in common. As a conclusion, the paper will make a call for the urgent need to renew existing lexicographical theories in order to assist the present transition from printed to electronic dictionaries.

### References:

- Atkins, B.T. Sue and Michael Rundell 2008: *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bejoint, Henri 2010: *The lexicography of English*. Oxford: Oxford University Press.
- Scerba, L.V. 1940: Towards a General Theory of Lexicography. *International Journal of Lexicography*. Volume 8. Number 4, 1995, 315-350.
- Sorokoletov, F.P. 1978: Traditionen der sowjetrussischen Lexikographie. In Werner Wolski, Werner (ed.): *Aspekte der sowjetrussischen Lexikographie. Übersetzungen, Abstracts, bibliographische Angaben*. Tübingen: Niemeyer 1982, 63-88.
- Tarp, Sven 2008: *Lexicography in the Borderland between Knowledge and Non-knowledge*. Tübingen: Niemeyer.

- Tarp, Sven 2010: Reflections on the Academic Status of Lexicography. *Lexikos* 20, 450-465.
- Tono, Yukio 2010: A Critical Review of the Theory of Lexicographical Functions. *Lexicon* 40, 1-16.
- Wiegand, Herbert Ernst 1989: Der gegenwärtige Status des Lexikographie und ihr Verhältnis zu anderen Disziplinen. In Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin/New York: Walter de Gruyter, 246-280.
- Yong, Heming and Jing Peng 2008: *Chinese Lexicography. A History from 1046 BC to AD 1911*. Oxford: Oxford University Press.

## WHAT DO FOUNDATION PHASE LEARNERS KNOW ABOUT A DICTIONARY?

**Michele VAN DER MERWE**

Cape Peninsula University of Technology, Wellington Campus, South Africa

A new phenomenon appeared on the South African lexicography horizon with the publication of the *Longman Foundation Phase Bilingual Dictionary*. It distinguishes itself from other dictionaries published in South Africa, since it was designed specifically with Foundation Phase (FP) learners (grade 1-3) as target users in mind. This uniqueness in design resulted in the dictionary being a picture dictionary as well as an alphabetical dictionary. Official South African languages like Afrikaans, isiXhosa, isiZulu, Sepedi and Setswana are paired with English.

How successful is this dictionary? In order to attempt to answer the set question, it is imperative that we do research on a usage situation. The aim of this paper is to find out more about the user of such a dictionary and if the dictionary fulfils the lexicographical needs of the user it was designed for. Do FP children know how to make use of a dictionary? What are their reference skills and language skills? Can they do successful look-ups? What dictionary skills and language skills do teachers need to teach them? Was there any improvement after dictionary skills has been taught? What are the implications for the lexicographer?

In order to answer the above questions, an experiment was done with about 100 FP children. According to Tarp (2009:287) the purpose of an experiment is to see how the introduction of a certain factor influences the result. In this case the factor was teaching in dictionary usage. Only a few experiments of this kind have been conducted in lexicography. Tono is probably the researcher who has worked most exhaustively with lexicographical experiments (Tarp, 2009:287). Tono (2001:70-72) describes three different types of experiments, namely (1) a pre-experimental design: the one group pre-test/post-test, (2) a quasi-experimental design: the non-equivalent control group design and (3) a true experimental design: the pre-test/post-test group design. According to Zofgen (1994:50) most psychologists consider experiments the 'royal road' of empirical research. The advantage of experiments is that it is possible to measure how the introduction of certain types of data and other lexicographical tools influence the result of the consultation of various types of users.

With regard to experiments Wiegand (1998:1023-1024) contends that one cannot successfully describe metalexical experiments about the relationship between dictionary structure and dictionary use without a detailed theory of dictionary structure. But, if a theory of user actions and a theory of dictionary structure are available, one can empirically prove by means of experiments (in which the independent features can be controlled) how the features of the dictionary structure affect the use of a dictionary. This will enable a scientific concept of user-

friendliness, which can also make it possible to compare dictionaries on empirical grounds, and not just by approximate judgments.

According to Tarp (2009:279) for research into dictionary use to be relevant, it should generate knowledge of how dictionaries are used, who are the users of dictionaries, where, when and why do they use dictionaries and with what result do they use dictionaries. It is thus necessary to do research into: the type of user situations, the type of users, the type of user needs, the users' usage of a dictionary and the degree of satisfaction of the user needs.

How did our experiment work? A group of homogeneous users, in the sense that they were all grade 3 learners, was selected. A pre-test and a post-test were conducted under supervision in the class room. Tests were marked and results were analysed. Between the two assessments teaching on dictionary use took place. Learning activities for use in conjunction with the dictionary were designed and they were done over a period of six weeks in the class room. After the period of activities, where learning took place, a post-test was written. The post-test was slightly altered, but the same kind of questions were posed as in the pre-test.

In this paper the experiment, results of the experiment, and learning activities will be discussed, as well as the lexicographical implications for the design of the dictionary.

#### **References:**

- Smit, M.2001. Herbert Wiegand with a "Metalexical Panga" in the jungle: An unlocking of Wörterbuchforschung. *Lexikos (11)*: 296-310.
- Tarp, S. 2009. Reflections on Lexicographical User Research. *Lexikos (19)*: 275-296.
- Wiegand, E.1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchforschung, zur Theorie, Kritik und Automatisierung der Lexikographie. Band 1*. Berlin: De Gruyter.

### **The impact of borrowing as one of the term creation strategies in South African indigenous languages: The translator's perspective**

**B.P.K. ZUNGU**

Jandrell Secondary School, Standerton, South Africa

Much has been said about the status of African languages in the pre-1994 era, however, today, the focus is on a need to reposition, revalorise and to empower them so that they can be used in wider range of domains i.e. they should be developed for use, particularly in education, mass communication, legislature and technology (Bamgbose 2008: 1&2). Webb in Webb & du Plessis (2006:156) maintains that:

‘As regards the role of language in the intellectualization of African communities, debates and discussion on issues such as African spirituality, philosophy, theology, systems of government and systems of justice are probably only really possible for a large percentage of African people in an African language, that is in a language that enables its users to debate at an abstract level about’.

Due to technological advancement and innovations that frequently take place today, the demand for appropriate terms for each and every invented object exists. South African indigenous languages were mainly used verbally and this furthered the impediment of the development of their technical terminology. According to Alberts (2008:30) ‘terminology is a strategic resource regarding the functional development of languages’. Languages develop or create their terminology by drawing from both internal sources and foreign



acquisition/borrowing (Mtintsilana & Morris 1988:110). English and Afrikaans are the primary foreign languages that most African languages resort to when faced with the problem of closing terminological gap in South Africa (Mahlangu 2007:1).

Translators who translate from a language such as English into the African languages have to deal with a single biggest problem of the lack of terminology in the African languages in the majority of specialist subject fields (Gauton et al 2003:00). The shortage of terminology in the previously marginalised African languages is the main reason for borrowing in these languages. However, words are not borrowed only because there is a hole in the language for a particular object, practice, or idea but also for prestige (Campbell in Schnoebelen 2005:9).

The concern is that knowledge and the ability to use the native isiZulu terms is inadequate today since the situation demands the knowledge of the foreign equivalent of that particular isiZulu term since the youth is familiar with borrowed words. Frequent use and preference given to borrowed words over the actual native terms is influenced by their advantage of being understood by a wider audience, shortness and easily remembered due to identified with the foreign term from which they were borrowed.

In translation the accessibility of the source text (ST) to the target readership is imperative. The type of text that is being translated, as well as the target readership have a great influence on the translator's choice of words or target text (TT) items for the source text (ST) items. The translator can therefore disregard the existing target language term and employ the borrowed word or even go to the extent of paraphrasing that target language equivalent for the source text term for the sake of the accessibility of the text to the target readership.

The aim of this paper is to critically evaluate the role of borrowing from other languages in the development of indigenous African languages such as isiZulu and also to highlight its impact on this language. The paper will therefore argue that a shift from the use of the native words exists in isiZulu. It will be argued further that this is due to the youth and educated people from the urban areas exposure to foreign languages. This empowers them with foreign sounds and vocabulary. The impact of giving preference to borrowed words over the equivalent Zulu words will be highlighted. The uncontrollable nature of borrowing which is reflected through its ability to change the rules or traditions of a language will be brought to the light.

According to Schnoebelen (2005:9) ample borrowing and speakers' familiarity with foreign languages are responsible for some prohibitions that have been lost in Zulu. Instances exist in translation where the source text uses different terms which through isiZulu dictionaries or terminology books share similar equivalents. An example is that of the word 'green pepper' and 'chilli pepper' which share the dictionary equivalent *upelepele*. The majority of the isiZulu speakers normally use the term *upelepele* to refer to hot chillies. This indicates that sometimes isiZulu dictionaries are inefficient and can even confuse users who are non-mother-tongue speakers.

According to Gouton et al (2003:00) the significance of terminology theory and practice for translators is apparent when the translator is faced with a situation where he/she can no longer rely on existing knowledge and /or dictionary and has to conduct a research beyond the dictionary.

The paper will conclude by making recommendation that the usage of the standardised Zulu terminology in schools must be encouraged. Abdulaziz in Onyango (2005:22) is also of the opinion that 'schools are the most important agent for stabilising and standardising language use'.

## References:

- Alberts, M. 2008. Translation-oriented Terminography in the Electronic Age. A paper presented in AFRILEX Conference in 2008.
- Bamgbose, A . 2008. Vision and mission of the African Academy of languages (Unpublished document)
- Gouton, R. & Descreyver, M. 2003. Towards Strategies for Translating terminology into all South African languages; a corpus -based approach. Paper read in TAMA 2003 South Africa Conference proceedings.
- Mahlangu, K.S. 2007. Adoption of loan words in isiNdebele. A document submitted in fulfilment for a Degree of Master of Arts in the Department of African languages. University of Pretoria.
- Mtintsilana, P.N. & Morris R. 1988. Terminography in African Languages in South Africa. South African Journal of African Languages Vol. 8, No. 109-113.
- Onyango, J.O. 2005. Issues in national language terminology development in Kenya. Swahili forum 12 (2005) 219-234.
- Schnoebelen. T. 2005. The Shape of Things to Come: The nominal morphology of Zulu loan words. UNISA
- Van Hysteen, L. 1999. Problems Regarding Term Creation in South African Languages, with special reference to Zulu. South African Journal of African Languages, Vol. 19, No.: 3 149-214.
- Webb, V. & Du Plessis, T. 2006. The Politics of Language in South Africa. Pretoria: Van Schaik Publishers.