

AFRICAN ASSOCIATION FOR LEXICOGRAPHY

20th International Conference

6 - 8 July 2015



University of KwaZulu-Natal,
Durban, South Africa



language**matters**@UKZN



AFRICAN ASSOCIATION FOR LEXICOGRAPHY

Programme and Abstracts

20th International Conference
University of KwaZulu-Natal,
Durban, South Africa
6 - 8 July 2015

Hosted by: University Language Planning and Development Office
University Teaching and Learning Portfolio, Francis
Stock Building, Howard College Campus,
Durban, South Africa.

Conference organiser: Dr. Langa Khumalo

Abstract reviewers: Dr Blanche N. Assam, Prof. Herman L. Beyer, Prof. Sonja Bosch, Dr Hughes Steve Ndinga-Koumba-Binza, Dr Dion Nkomo, Prof. Thapelo J. Otlogetswe, Prof. Annél Otto, Prof. Danie J. Prinsloo, Dr Michele van der Merwe.

Abstract booklet editors: Prof. Sonja E Bosch & Dr Dion Nkomo

TABLE of CONTENTS

AFRILEX HONORARY MEMBERS	5
AFRILEX BOARD (2013-2015)	6
MESSAGE FROM THE AFRILEX PRESIDENT	7
PROGRAMME.....	9
KEYNOTE PRESENTATION 1	14
Moving from print to digital: implications for dictionary policy and lexicographic conventions	
Michael RUNDELL	
KEYNOTE PRESENTATION 2	15
Between a rock and a hard place: A critical reflection on the intellectualisation of African languages	
Mbulungeni MADIBA	
PARALLEL SESSIONS	15
Centre for Political and Related Terminology in Southern African Languages (CEPTSA) – Translating and Explanatory Dictionaries	
Mariëtta ALBERTS.....	15
The consultation potential of lemma candidates for a bilingual dictionary of closely related languages	
Herman L. BEYER	17
The treatment of academic action verbs in English advanced learner's dictionaries	
Herman L. BEYER & Jill N. IZAKS.....	19
Being Inconsistent: Loanwords Spelling in Northern Sotho Dictionaries	
Sekgaila CHOKOE	20
A learner's dictionary for foreign students learning Afrikaans in a university context	
Dané CLAASSEN	22
Linking the fulfilment of translational needs to lexicographical theory in the context of UI translation	
André DU PLESSIS & Marius SWART.....	24
Equivalent relations and their implications for the database of an electronic bilingual sign language dictionary	
Hanelle FOURIE BLAIR	26
Who are the target users of monolingual learners' dictionaries?	
Rufus H. GOUWS	27
Enhancement of the African Wordnets: additional language resources and tools	
Marissa GRIESEL & Sonja E BOSCH	28
To what extent do Learners' dictionaries use defining vocabularies for their examples?	
Lorna HILES	30

Term extraction for an isiZulu linguistic terms dictionary using a corpus linguistic method	
Langa KHUMALO	31
Towards an inclusive dictionary: bridging the gap between standard and non-standard varieties of Setswana	
Baile G. MAREME & Keabetswe N. LEKOPANYE	33
The Inclusion of Collocations in Dictionaries of African Languages: A Case of Northern Sotho	
Kwena G. MASHAMAITE & Mampa L. MPHAHLELE	35
Terminology Development Principles: Towards a Principled Approach to UKZN isiZulu Technical Terms	
Khumbulani MNGADI	36
Evaluation of the performance of a machine learning lemmatiser for isiXhosa	
Lulamile MZAMO & Albert S.J. HELBERG & Sonja E. BOSCH	37
An analysis of defining principles with special reference to <i>Isichazamazwi SesiNdebele</i> (ISN) definitions	
Eventhough NDLOVU	40
Cross-referencing in <i>Isichazamazwi Sezomculo</i> (ISM)	
Eventhough NDLOVU & Menelisi BHEBHE	41
From common conversations around the <i>Oxford Bilingual School Dictionary: IsiXhosa-English</i> to constructive dictionary criticism	
Dion NKOMO	42
Reflections on talking dictionaries in Zimbabwe's indigenous languages: Can we really talk?	
Dion NKOMO & Emmanuel CHABATA	44
The intellectualization of indigenous languages in South Africa through terminology development and dictionary making: A case study of IsiZulu Language at University of KwaZulu-Natal	
Siyabonga NKONTWANA	46
From <i>Idioticon</i> to ANNA: The role of Stellenbosch University in the development of Afrikaans lexicography	
Gerda B. ODENDAAL	47
A critical analysis of multilingual dictionaries	
Danie J. PRINSLOO	48
Gender Sensitivity in <i>Isichazamazwi SesiNdebele</i>	
Sanelisiwe SAYI	49
African Glossonyms: Complexity and Practices in African Dictionaries	
Hughes S. NDINGA-KOUMBA-BINZA	50
Preparing an online English-Spanish dictionary of business communication	
Sven TARP	52
The Use of LSP Dictionaries on Mobile Phones in Higher Education	
Michele VAN DER MERWE	53

Adaptation of A Dictionary of South African English on Historical Principles for electronic platforms

Tim VAN NIEKERK55

To Include or Not: Lexicographic Recording of Neologisms in the Digital Era

Gao YONGWEI55

AFRILEX HONORARY MEMBERS



Prof. R.H. Gouws



Prof. A.C. Nkabinde



Dr. J.C.M.D. du Plessis



Dr. M. Alberts

AFRILEX BOARD 2013 – 2015

President:	Dr M.V. (Victor) Mojela
Vice-President:	Prof. D.J. (Danie) Prinsloo
Secretary:	Prof. H.L. (Herman) Beyer
Treasurer:	Prof. E. (Elsabé) Taljard
Editor Lexikos:	Prof. E. (Elsabé) Taljard
Members:	Prof. S.E. (Sonja) Bosch Prof. R. H. (Rufus) Gouws Dr H.S. (Steve) Ndinga-Koumba-Binza Dr D. (Dion) Nkomo Prof. A.N. (Annél) Otto
Conference organiser:	Dr L. (Langa) Khumalo

MESSAGE FROM THE AFRILEX PRESIDENT

On behalf of the AFRILEX Board, I would like to welcome all of you to the 20th *Annual International Conference of the African Association for Lexicography*, also known as 'AFRILEX 2015'. This year's edition takes place in the warm and beautiful city of Durban in the University of Kwa-Zulu Natal, hosted by the University Language Planning and Development Office, University Teaching and Learning Portfolio, in the Francis Stock Building, Howard College Campus, Republic of South Africa. This conference follows the successful AFRILEX 2014 which was hosted at the Potchefstroom Campus of the North-West University last 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 pleased to see many scholars from outside and within South Africa attending and participating actively in the AFRILEX conferences, in particular, the scholars from Europe, Namibia, Gabon and Botswana. We want to invite many lexicographic scholars from the entire African continent to form part of the membership of AFRILEX. In the past, AFRILEX used to have a considerable number of scholars from Zimbabwe and Gabon participating actively in every conference, but unfortunately, poor financial support seems to contribute much to the reduction of these memberships. We still need the active participation of these members to demonstrate that AFRILEX is an association for lexicography in the African Continent, and not only for South Africans. AFRILEX is always ready to accept invitations to hold conferences outside South Africa if possible.

AFRILEX 2015 has been meticulously prepared by a local organising team under the leadership of Dr Langa Khumalo, as the local organiser. At this moment I also want to thank Prof. Danie Prinsloo, the AFRILEX Deputy President, who came to my rescue by taking charge of all the activities of AFRILEX when I was unable to do anything after losing my wife and daughter in a road accident in January this year. Thank you very much, Prof. Prinsloo. May God bless you.

Once again, the abstract submission and adjudication process for AFRILEX 2015 was expertly managed by Dr Dion Nkomo with Dr Blanche N. Assam, Prof. Herman L. Beyer, Prof. Sonja Bosch, Dr Hughes Steve Ndinga-Koumba-Binza, Prof. Thapelo J. Otlogetswe, Prof. Annél Otto, Prof. Danie J. Prinsloo, and Dr Michel van der Merwe assisting as peer adjudicators. Prof. Sonja Bosch and Dr Dion Nkomo also did commendable work in the compilation of this Abstract Booklet we are holding now. We want to congratulate and thank them for the job well-done. We also want to say thank you once more to Prof. DJ Prinsloo who excellently managed and kept the AFRILEX website up to date and managed the compilation of the programme for this conference, not forgetting Prof. Elsabé Taljard, our reliable treasurer, who continuously keeps the AFRILEX moneys safe.

Just like the previous conferences, AFRILEX 2015 promises to be another stellar gathering, with speakers coming from a dozen different countries in Africa, Asia and Europe, namely Belgium, Botswana, China, Denmark, Gabon, Germany, Namibia, South Africa, Spain, UK and Zimbabwe. Once-more we want to convey an apology for Annette Klosa, from the Institute for German Language, Mannheim, Germany, who was supposed to have been with us here to present a pre-conference workshop on Monday, the 6th July 2015, but unfortunately, she could not make it due to the

recent promotion she has just got to be head of the German Academic Exchange Service (DAAD) information centre in Vietnam, and also as lecturer for German at the Vietnamese State University at Ho-Chi-Minh-City from September this year. Fortunately, she succeeded in getting her colleague, Dr Kathrin Kunkel-Razum from Berlin to give a pre-conference workshop in her place. On behalf of AFRILEX I want to thank Dr Kathrin Kunkel-Razum for agreeing to present the workshop for us in such a short notice, and also to welcome her to South Africa. We hope that she will hence-forth be with us in our Annual AFRILEX International Conferences. We also want to thank Prof. Renuka Vithal, the Deputy Vice Chancellor: Teaching & Learning, here at the University of KwaZulu-Natal, for coming to welcome us in this beautiful University.

We also continue the tradition of giving the floor to dictionary publishers during a Publishers' Session. Our keynote speakers this year are Michael Rundell, from Macmillan, United Kingdom, and Prof. Mbulungeni Madiba of the University of Cape Town, who is also Chairperson of the Pan South African Language Board (PanSALB), here in South Africa. We welcome them to the AFRILEX 2015 Conference.

Maropeng Victor Mojela
President: AFRILEX

PROGRAMME

Monday 6 July 2015

Pre-conference workshop

Venue: Lecture Room, Unite Building

Electronic Lexicography

Presenter: Dr Kathrin Kunkel-Razum

Dudenredaktion, Bibliographisches Institut GmbH, Berlin, Germany

Session 1: 09:00 – 10:25

10:30 – 10:55	Tea Venue: Room 1-3, Unite Building
------------------	--

Session 2: 11:00 – 11:40

Session 3: 11:45 – 12:25

12:30 – 13:25	Lunch Venue: Room 1-3, Unite Building
------------------	--

Session 4: 13:30 – 14:25

Session 5: 14:30 – 15:25

18:00 – Cocktail Party

Venue: Room 1-3, Unite Building

Tuesday 7 July 2015

08:30 – 09:10	Registration Venue: Foyer, Unite Building
------------------	---

09:15 – 09:30	Official Opening Venue: Lecture Room, Unite Building
------------------	--

Victor Mojela – President of AFRILEX

A word of welcome
Professor Renuka Vithal
Deputy Vice Chancellor: Teaching & Learning, University of KwaZulu-Natal

Keynote Address 1		
Venue: Lecture Room, Unite Building		
09:30 – 10:25	Keynote Michael Rundell , Macmillan, United Kingdom Moving from print to digital: implications for dictionary policy and lexicographic conventions	
10:30 – 10:55	Tea Venue: Room 1-3, Unite Building	
Parallel Sessions		
	Venue: Lecture Room	Venue: Room 4/5
11:00 – 11:25	Who are the target users of monolingual learners' dictionaries? Gouws, R.H.	Being Inconsistent: Loanwords Spelling in Northern Sotho Dictionaries Chokoe, S.
11:30 – 11:55	The Inclusion of Collocations in Dictionaries of African Languages: A Case of Northern Sotho Mashamaite, K. J. & Mphahlele, M. L.	Preparing an online English-Spanish dictionary of business communication Tarp, S.
12:00 – 12:25	Evaluation of the performance of a machine learning lemmatiser for isiXhosa Mzamo, L., Helberg, A. S. J. & Bosch, S. E.	The Impact of Culture Reproduction in Dictionaries on French Language Acquisition: A Case Study at the University of the Western Cape Assam, B. Nyangone
12:30 – 12:55	Enhancement of the African Wordnets: additional language resources and tools Griesel, M. & Bosch, S. E.	To what extent do Learners' dictionaries use defining vocabularies for their examples? Hiles, L.
13:00 – 13:55	Lunch Venue: Room 1-3, Unite Building	

Parallel Sessions		
	Venue: Lecture Room	Venue: Room 4/5
14:00 – 14:25	The treatment of academic action verbs in English advanced learner’s dictionaries Beyer, H. L. & Izaks, J.	14:00- 14:55 Special session: Letting the new HAT out of the bag - a discussion of the new and completely revised HAT Luther, J., Pheiffer, F. & Gouws, R. H.
14:30 – 14:55	The development of indigenous languages in South Africa through terminology development and dictionary making: A case study of IsiZulu Language at University of KwaZulu-Natal Nkontwana, S.	
15:00 – 15:25	Terminology Development Principles: Towards a Principled Approach to UKZN isiZulu Technical Terms Mngadi, K.	15:00 -15:55 Special session: Adaptation of A Dictionary of South African English on Historical Principles for electronic platforms Van Niekerk, T.
15:30 – 15:55	A learner’s dictionary for foreign students learning Afrikaans in a university context Claasen, D.	
16:00 – 17:00	Annual General Meeting	
18:00	Conference Reception Venue: Town Hall	
Wednesday 8 July 2015		
Keynote Address 2 Venue: Lecture Room, Unite Building		
09:00 – 09:55	Keynote Mbulungeni Madiba , University of Cape Town, South Africa Between a rock and a hard place: A critical reflection on the intellectualisation of African languages	
10:00 – 10:25	Tea Venue: Room 1-3, Unite Building	

Parallel Sessions		
	Venue: Lecture Room	Venue: Room 4/5
10:30 – 10:55	The consultation potential of lemma candidates for a bilingual dictionary of closely related languages Beyer, H. L.	10:30- 12:25 Special session Taking hands to build a multilingual open education resource term bank for Higher Education Carstens, A., Madiba, M. & Taljard, E. [Facilitators]
11:00 – 11:25	Equivalent relations and their implications for the database of an electronic bilingual sign language dictionary Fourie-Blair, H.	
11:30 – 11:55	Gender Sensitivity in <i>Isichazamazwi SesiNdebele</i> Sayi, S.	
12:00 – 12:25	To Include or Not: Lexicographic Recording of Neologisms in the Digital Era Yongwei, G.	
12:30 – 13:25	Lunch Venue: Room 1-3, Unite Building	
Parallel Sessions		
	Venue: Lecture Room	Venue: Room 4/5
13.30 – 13.55	From common conversations around the <i>Oxford Bilingual School Dictionary: IsiXhosa-English</i> to constructive dictionary criticism Nkomo, D.	The Use of LSP Dictionaries on Mobile Phones in Higher Education Van der Merwe, M.
14:00 – 14:25	Cross-referencing in <i>Isichazamazwi Sezomculo</i> (ISM) (2006) Ndlovu, E. and Bhebhe, M	Term extraction for an isiZulu linguistic terms dictionary using a corpus linguistic method Khumalo, L.
14:30 – 14:55	From <i>Idioticon</i> to ANNA: The role of Stellenbosch University in the development of Afrikaans lexicography Odendaal, B. G.	Linking the fulfilment of translational needs to lexicographical theory in the context of UI translation Du Plessis, A. & Swart, M.
15:00 – 15:25	Reflections on talking dictionaries in Zimbabwe’s indigenous languages: Can we really talk? Nkomo, D. and Chabata, E.	An analysis of defining principles with special reference to <i>Isichazamazwi SesiNdebele</i> (ISN) definitions Ndlovu, E.

15:30 - 16:00	Tea Venue: Room 1-3, Unite Building	
16:00- 16:25	Towards an inclusive dictionary: bridging the gap between standard and non-standard varieties of Setswana Mareme, G. B. & Lekopanye, K. N.	Centre for Political and Related Terminology in Southern African Languages (CEPTSA) – Translating and Explanatory Dictionaries Alberts, M.
16:30 – 17:00	A critical analysis of multilingual dictionaries Prinsloo, D. J.	African Glossonyms: Complexity and Practices in African Dictionaries Ndinga-Koumba-Binza, H. S.
17:00	Closure	
19:00	Conference Dinner Venue: Room 1-3, Unite Building	

Thursday 9 July 2015

Post-conference excursion
The Inanda Heritage Route

KEYNOTE PRESENTTION 1

Moving from print to digital: implications for dictionary policy and lexicographic conventions

Michael RUNDELL michael.rundell@lexmasterclass.com

Macmillan, United Kingdom

Editorial policies in dictionaries have evolved over hundreds of years. A typical dictionary style guide provides guidance on issues such as inclusion (which words get into the dictionary), the content and wording of definitions, the use of citations or example sentences, the placement of idioms, the application of labels, and the ordering of senses. All of these policies developed at a time when dictionaries were printed books of finite dimensions - as they have been for almost the whole of their history. In many cases, styles which we take for granted as “natural” features of dictionaries can be seen merely as expedients designed to compress the maximum amount of information into the limited space available. A simple example is the kind of “recursive” definition found in many English dictionaries (with equivalents in dictionaries of other European languages) where a nominalization (such as *assimilation*) is defined in terms of the related verb (“the act of assimilating or state of being assimilated”), and the user is required to make a second look-up (to the base word). Is this an ideal solution, or was it favoured simply as a less space-intensive alternative to a self-sufficient explanation which does not require the reader to consult another entry in order to get the full picture?

Lexicography is now in a turbulent phase, as dictionaries gradually migrate from print to digital media. As always happens with changes which are driven by technology, the global picture is uneven, and in many parts of the world paper dictionaries still have a healthy future ahead of them. But the direction of travel is clear, and - it will be argued - digital media are better-adapted than printed books as a platform for reference resources of all kinds. Once a dictionary makes the move to digital, space constraints disappear and all sorts of opportunities present themselves for adding multimedia content and hyperlinking to relevant data in other parts of the Web. Some problems simply evaporate. The awkward question of where to locate a phrase such as *shut the stable door after the horse has bolted* (whether you are a lexicographer deciding where to put it, or a user deciding where to look for it) becomes a non-issue. In many online dictionaries, phrases like this are now freestanding “headwords” in their own right, and as long as the dictionary’s search algorithm is well-designed, users will have no problem locating such entries. Equally, the need for abbreviations, tildes and the like no longer exists (though a surprising number of dictionaries maintain these conventions even in their digital versions). These issues are relatively trivial. But the question arises whether we need to revisit, and re-evaluate, the entire range of editorial policies and conventions in the light of changed circumstances. This talk will look at familiar editorial and presentational conventions, and will consider which are no longer appropriate in the digital medium - and what new policies might replace them.

KEYNOTE PRESENTATION 2

Between a rock and a hard place: A critical reflection on the intellectualisation of African languages

Mbulungeni MADIBA (mbulungeni.madiba@uct.ac.za)

University of Cape Town

The democratic change of 1994 and the concomitant language policy of official multilingualism opened up a plethora of initiatives and projects for the intellectualisation of African languages which were marginalised during colonial and apartheid era. However, as several scholars have noted, these initiatives and efforts have achieved little success. Twenty one years into democracy, African languages continue to be on the margins of professional knowledge domains. It is the argument of this presentation that the intellectualisation of African languages failed partly because of the theoretical, methodological and practical approaches adopted by those who have been involved in this endeavour. The purpose of this presentation is to give a critical reflection of the dilemma which African languages are faced with and to propose alternative strategies for their intellectualization. Accordingly, alternative theoretical, methodological and practical approaches will be proposed for the intellectualisation of African languages.

PARALLEL SESSIONS

Centre for Political and Related Terminology in Southern African Languages (CEPTSA) – Translating and Explanatory Dictionaries

Mariëtta ALBERTS (albertsmarietta@gmail.com)

North-West University

The need for unambiguous communication in the theoretical and applied fields of human activity is constantly increasing. Terminologists and subject specialists are focused on the provision of unambiguous source- and target-language terms for well-defined concepts. In order to achieve this aim, it is necessary to determine the precise meanings of terms which enable users to comprehend and use them in a universally accepted manner.

Political and societal changes in the new democratic South Africa have largely transformed the functional role of the indigenous languages. Section 6 of the Constitution further provides for multilingualism and the development of the linguistic heritage. This system of official multilingualism has produced a substantial demand for terminology creation as various subject-related business matters need to be conducted in the different official languages (cf. Government Gazette 2012, 2013). Multilingualism in South Africa is a sociolinguistic fact to be taken seriously (Alberts 1998: 230). A large proportion of South Africa's inhabitants can only be reached by means of indigenous languages. A problem facing multilingual terminology development in South Africa is a shortage of subject specialists with functional linguistic knowledge of the respective official languages. Terminology development is an interdisciplinary activity where the point of departure is subject related, but without linguistic input it is impossible to compile dictionaries for special purposes. The various bilingual and multilingual translating and explanatory political dictionaries compiled

by CEPTSA will serve as examples of interdisciplinary collaboration where subject matter and language(s) successfully meet.

Ambiguity in the specialised information may give rise to confusion and distortion of the communication process. One way to ascertain the exact meaning of a message conveyed through the medium of language is to document and standardise the terminology of languages for special purposes. When everyone in a specific language group working in a similar working environment understands the same message conveyed by a specific concept denoted by a specific linguistic label, that is the term, one can consider the term to be standardised. This is precisely what terminology is all about. It is the task of the terminologist to make sure that basic terminological principles, cultural differences and language attitudes of the professional group are taken into account when denoting concepts and coining terms.

According to Sager (1990: 4), terminology is concerned with "the study and use of the systems of symbols and linguistic signs employed for human communication in specialised areas of knowledge and activities". Terminology is "a representation of an equally coherent, but possibly differently structured system of concepts" (Sager 1990: 114). A term is created when various linguistic labels are used to describe or name a specific object or concept. It therefore refers to a definite concept which is clearly defined within specific parameters. A term is as such the linguistic representation of a mental construct. There is a special interrelationship between the symbol, the concept (that is its mental representation in one's brain) and the various linguistic labels used in different languages to describe the object and concept. If this does not exist, misunderstanding or miscommunication will result (Sager 1990: 57).

Bilingual and multilingual dictionaries act as facilitators in scientific or technical communication processes. This paper deals with the bilingual and multilingual explanatory political dictionaries compiled by CEPTSA. The phases of the project, consisting of different translating and explanatory versions, are discussed.

The aim of CEPTSA is to promote the usage of political and related terminology in Southern Africa. Research is being done on these subject fields, relevant concepts and terms are harvested, defined and translated. The source language is English and Afrikaans was initially the target language. The Centre already disseminated draft term lists of 1000 core terms in Setswana, Sesotho sa Leboa/Sepedi, isiZulu and isiXhosa. The Centre is currently busy with defining a further 1500 core terms and as soon as this process is finalised, the English/Afrikaans terminology list and definitions will be published and the data translated into the official African languages.

The Centre provides a terminological and subject-related service to lecturers and under- and postgraduate students in international politics, political studies and governance, public administration, municipal government and administration, development studies and strategic studies. A service is also rendered to members of parliament, provincial legislature and local authorities, language practitioners and the media.

Translating terms and definitions in the various official languages requires subject knowledge and linguistic expertise. One of the problems facing CEPTSA is the lack of standardised terminology in the political and related fields. CEPTSA has overcome some of these problems by defining the terms in simplified language. CEPTSA was also successful in obtaining translation assistance from the Language Unit, University of Johannesburg (isiZulu and Sesotho sa Leboa/Sepedi) and the Language Centre, University of Stellenbosch (isiXhosa). CEPTSA focuses on finding amicable financial solutions to translate 2500 core terms and definitions into the other official languages.

References

- Alberts, M. 1998. The Harmonisation of Nguni and Sotho: Terminological Perspectives. In: Prah, K. K. 1998. *Between Distinction and Extinction. The Harmonisation and Standardisation of African Languages*: 229-242. Johannesburg: Witwatersrand University Press.
- Alberts, M. 2014. Terminology Development at Tertiary Institutions: A South African Perspective. *Lexikos* 24): 1-26.
- Centre for Political and Related Terminology in Southern African Languages (CEPTSA). 2002. *Nuwerwetse Politieke Woordeboek/Modern Political Dictionary*. Compiled by the Committee for Political and Related Terminology, CEPTSA. Auckland Park: Rand Afrikaans University (RAUpers).
- Centre for Political and Related Terminology in Southern African Languages (CEPTSA). Botha, S., P. le Clus and A. Venter (eds.). 2011. *Modern Political Dictionary/Nuwerwetse Politieke Woordeboek*. CEPTSA. Auckland Park: University of Johannesburg.
- Centre for Political and Related Terminology in Southern African Languages (CEPTSA). 2015. *Database of the Centre for Political and Related Terminology in Southern African Languages*. CEPTSA.
- Centre for Political and Related Terminology in Southern African Languages (CEPTSA). n.d. *Multilingual Modern Political Dictionary: 250 Trial Definitions. Meertalige Nuwerwetse Politieke Woordeboek: 250 Proefdefiniesies. Sesilimi ezahlukeni samagama anamuhla apathelene nezepolitiki: Incasiselo yamagama angu-250 angumzamo. Pukuntšu ya dipolelo tše dintši ya sebjalebjale ye e amanago le mantšu a politiki: Ditlhalošo tše 250 tša mantšu tša boitekelo*. Johannesburg: Centre for Political and Related Terminology in Southern African Languages (CEPTSA), Auckland Park: Rand Afrikaans University.
- Government Gazette, 2012. *Use of Official Languages Act 12 of 2012*. Government Gazette, Vol. 568, No. 35742, 2 October 2012. Cape Town: The Presidency.
- SAGER, J.C. 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.

The consultation potential of lemma candidates for a bilingual dictionary of closely related languages

Herman L. BEYER (hbeyer@unam.na)

Department of Language & Literature Studies, University of Namibia

This paper aims to show (1) an empirical methodology according to which the likelihood that specific types of lemmata will be consulted in a bilingual dictionary of closely related languages can be determined, and (2) that linguistic theory can play a central role in the lemma selection procedure for at least some types of dictionaries.

The dictionary in question is a monofunctional, concise Dutch-Afrikaans dictionary that is currently being compiled, abbreviated as NAWA. Initially the context of this dictionary will be explained, including its (formal) relation to the existing *Groot Woordeboek Afrikaans en Nederlands* (ANNA – Martin 2011). This will be followed by a brief comparison for lexicographic purposes of the vocabularies of the closely related languages Afrikaans and Dutch, based on existing linguistic literature.

Based on the aforementioned, and utilising a typology of lemma candidates derived from the relations of equivalence distinguished in the front matter of ANNA and the relevant linguistic literature, a hypothesis is put forward which states the treatment priority of the identified types of lemma candidates. This statement of a lemmatisation priority is necessitated by the chosen medium of the dictionary: either in printed form (with its traditional space limitations), or as an electronic dictionary that can be released in functional phases as they are completed. The hypothesis is, therefore, a ranking of the potential lexical information needs of the target users of NAWA in macrostructural terms, and it is based on linguistic principles. The hypothesis states the following lemmatisation priority of Dutch lemma candidates in terms of their equivalence relations with Afrikaans equivalents:

1. Non-cognates
2. Irregular verbs in the infinitive and conjugated form
3. Regular verbs in the infinitive and conjugated form
4. Absolute cognates that differ in form
5. Partial cognates
6. Absolute cognates

To test the hypothesis empirically, a simulation was run of the envisaged typical user situation at which NAWA is targeted. An extract from an authentic Dutch text was given to 269 first year students in Afrikaans and Dutch at Stellenbosch University during their second semester. During the preceding semester 142 (52.8%) of the students had followed a module in basic Dutch reading skills while the remaining 121 students (45.0%) followed a module in Afrikaans lexical semantics and had therefore not undergone any introductory instruction in Dutch (the assumption being that they had not been introduced to Dutch before entering the university). The text extract contained words belonging to all the categories of lemma candidates in the prioritised list above. The students were requested to read the text extract for understanding and to circle every word that they would have looked up in a Dutch-Afrikaans dictionary if they had one at hand.

The resultant instances of potential look-up were recorded in terms of the lemma candidate types per respondent. Overall, a total of 4 823 potential look-up procedures were recorded: 3 374 (70%) by the respondents with no prior knowledge of Dutch (Group A), and 1 449 (30%) by the respondents who had completed the Dutch reading skills module during the previous semester (Group B). To make sense of the data, an equation was developed to express each lemma candidate type's *look-up potential index* (LPI). An index value of 1.0 for a particular lemma type would indicate a 100% chance that that lemma type would be looked up by every respondent, and 0.0 would reflect zero look-up potential. The results can be summed up in the following table:

Lemma Candidate Type	LPI for Group A	LPI for Group B
1. Non-cognates	0.23	0.12
2. Irregular verbs	0.20	0.07
3. Regular verbs	0.19	0.06
4. Absolute cognates (\neq form)	0.15	0.05
5. Partial cognates	0.03	0.01
6. Absolute cognates ($=$ form)	0.03	0.01

Since the simulation was run without the actual use of dictionaries, the students marked more words for look-up than they would have looked up in reality, as is suggested by the calculation of a *consultation frequency index*, which will be explained. Nevertheless, it is not the absolute consultation frequency that matters, but rather the relative consultation frequency among the lemma candidate types. Relations are therefore more important than numbers in themselves, leading to the conclusion that the hypothesis is proven true on the basis of descriptive statistics.

Reference

Martin, W. (Ed.). 2011. *Groot Woordeboek Afrikaans en Nederlands*. Cape Town: Pharos.

The treatment of academic action verbs in English advanced learner's dictionaries

Herman L. BEYER (hbeyer@unam.na)

Department of Language & Literature Studies, University of Namibia

Jill N. IZAKS (jnizaks@unam.na)

Language Centre, University of Namibia

This paper reports on a critical evaluation of the paraphrases of meaning of academic action verbs (AAVs) as lemmata in pedagogical dictionaries, particularly the international English advanced learner's dictionaries, and how the criteria and results of the evaluation could possibly be translated into lexicographical theory. AAVs are verbs like *analyse*, *define*, *describe*, *explain*, *identify*, *interpret* and *quote* as they are used in academic discourse (as distinguished from how they are used in language for general purposes (LGP)). These verbs represent lexemes that are classified as academic vocabulary, which refers to a specialised type of vocabulary used in formal education (academic contexts) and can be found in the Academic Word List (AWL) compiled by Coxhead (2000). The AWL is characterised as follows in *Wikipedia* (2015): "The list contains 570 semantic fields [sic: rather 570 *word families*, cf. Coxhead (2000)] which were selected because they appear with great frequency in a broad range of academic texts. The list does not include words that are in the most frequent 2000 words of English (the General Service List), thus many of the words are specific to academic contexts. However, a significant percentage of the vocabulary contained within the AWL is of general use; it is simply not of high enough frequency to be contained within the General Service List. [...] The AWL was primarily made so that it could be used by teachers (especially teachers of English as a Second Language) as part of a programme preparing learners for tertiary level study or used by students working alone to learn the words most needed to study at colleges and universities." The AWL therefore represents a core lexicon for the teaching and learning of academic literacy. It is important to note that, in the words of Coxhead (2000: 214), these words "are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur." For this reason these words are not accorded term status and hence are treated in dictionaries as LGP vocabulary. In these circumstances the research problem presents itself: The lexicographic treatment of AAVs in international advanced learner's dictionaries is not always optimal for the teaching and learning of academic literacy. The research questions that guide the current study are: (1) Do the paraphrases of meaning of AAVs in international English advanced learner's dictionaries optimally support the teaching of academic literacy to their target users? (2) If not, in what ways can the paraphrases of meaning be improved to optimally support the teaching of academic literacy to their target users? (3) What other dictionary structures could be utilised to optimise the treatment of AAVs for the teaching of academic literacy?

To investigate the problem, a pilot study was started with 20 AAVs. The paraphrases of meaning of the representative lemmata in the relevant dictionaries were evaluated against:

- the important distinction between *receptive vocabulary* (i.e. acquired vocabulary utilised in LGP text reception) and *academic literacy*, the latter yielding different lexical information needs from the former;
- the levels of receptive vocabulary and academic literacy among first year students (i.e. typical target users of the dictionaries) at a southern African university as established by recent empirical research;
- the information with regard to each AAV that a lecturer in academic literacy at said university deems as important in the explicit teaching of AAVs.

The empirical research referred to above also concluded that the explicit teaching of academic vocabulary is desirable. As the broad formulation of the evaluation criteria above suggest, this study approaches the research problem from the perspective of the information that the instructor in academic literacy needs to be included in the lexicographic treatment of AAVs in order to facilitate explicit academic vocabulary teaching by means of dictionary-integrated instruction. This approach corresponds to the methods used when technical dictionaries are compiled, i.e. collaboration between lexicographer and subject expert.

The information included in the meaning paraphrases of AAVs in the “Big Five” English advanced learner’s dictionaries were also compared by means of customised information differentiating matrixes before they were evaluated against the above criteria.

In order to translate the applied evaluation criteria and results into lexicographical theory, Bloom’s revised taxonomy of educational objectives (Anderson and Krathwohl 2001) is considered as a potential framework, while taking due cognisance of the potential pitfalls of pursuing generic treatments of AAVs.

In conclusion, consideration is given to the additional dictionary structures that are available to the lexicographer when dealing with AAVs.

References

- Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Anderson, L.W. and D.R. Krathwohl. (Eds.). 2001. *A Taxonomy for Learning, Teaching, and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives*. London/New York: Longman.
- Wikipedia. 2015. Academic Word List. https://en.wikipedia.org/wiki/Academic_Word_List [Accessed 29 April 2015].

Being Inconsistent: Loanwords Spelling in Northern Sotho Dictionaries

Sekgaila CHOKOE (sekgaila.chokoe@ul.ac.za)

School of Languages and Communication Studies, Department of Languages, University of Limpopo

Northern Sotho, like any other language, develops. It is during this development that new concepts and ideas come to the fore. Consequently, new words are formed, either through coinages or through the process of what Mokgokong (1975) calls the Sothoization of foreign words and concepts. During this process, words of foreign origin are given the linguistic attributes of the Northern Sotho language, i.e., they are assimilated to the phonetic, phonological, morphological and lexical systems of Northern Sotho and thereby acquire the distinctive imprint of Northern Sotho (Mokgokong, 1975:137).

Various dictionaries are perused with a view to how they spell Northern Sotho borrowed terms. A list of such adopted words accumulated from various Northern Sotho textbooks and works of art is used as a primary source. *Northern Sotho Spelling and Orthography* no. 4 of 1988, which is still regarded as the only standard authority in this language, is also scrutinised. Northern Sotho adopted words are sometimes spelt so differently in various dictionaries as if they are different entries, e.g. Sesotho sa Leboa (Northern Sotho) - English Dictionary.

ORIGINAL WORD	PUKUNTŠU-TLHALOŠI YA SESOTHO SA LEBOA	SESOTHO SA LEBOA (NORTHERN SOTHO) - ENGLISH DICTIONARY	PUKUNTŠU YA SEKOLO	POPULAR NORTHERN SOTHO DICTIONARY	GROOT NOORD-SOTHO WOORDE-BOEK
klerk	Klereke	Klereke	klereke	klereke tlelereke	klereke tlereke
christmas	-	krisemose kresmose	kresemose	krisemose keresemose	keresemose
gravel		krabole	-	karabole	kerabole
taxi	Thekisi	Thekisi tekisi theksi	taxi thekisi	thekisi tekisi	thekisi
maximum	-	maksimamoma ksemamo	maksimamo	makisimamo makesemamo	-
province	Profense	Porofense	profense	profense porofense	porobense porofense
project	-	Protšekto	projeke	protšekto	-
Straat	-	Seterata Setarata	setrata	setarata	seterata setarata

This paper purports to highlight the problems encountered in the writing of dictionary entries of some Northern Sotho adopted words. It is argued that some Northern Sotho loanwords are not fully adapted into the Northern Sotho spelling orthography. Purists, as Mojela (2010) calls them, resort to neologism, perhaps because there are no strict rules followed in the spelling of loanwords in Northern Sotho. Anybody, as is evident in the above table, spells a loanword the way they like, and this brings about confusion. There should be a standardized uniformity in the spelling and pronunciation of the same loanwords.

Another phenomenon worth of scrutiny in Northern Sotho adopted words is vowel epenthesis. There are some consonant clusters that do not conform to the syllabic principles of the Northern Sotho language when lemmatizing loanwords. Going back to the table above, *Sesotho sa Leboa (Northern Sotho) - English Dictionary* (<http://africanlanguages.com/sdp/> accessed 25 March 2015) for example, has three different lemmata for the word ‘taxi’, viz. theksi, thekisi and tekisi. The first lemma, **theksi**, which has a consonant cluster /ks/, is “illegal”, and is adapted by adding a vowel in the cluster, hence the second lemma, **thekisi**. This strategy is called vowel epenthesis (Lin, 2014).

Nasal strengthening is also touched in this paper. When a syllabic nasal precedes a soft sound, be it a vowel or a fricative, according to the rules of Northern Sotho grammar, the process of strengthening is inevitable. For example, when one forms a deverbative noun for class 9 from the verb stem –*aena* (iron), one is required to put the class prefix /n-/ before the initial vowel /a/, bringing about strengthening. The new word is *kaeno* (ironing). When affixing the object concord 1st person singular, (n-), the initial sound of the root must be strengthened, e.g *aena nna* (iron me) should be *nkaena*. In short, any syllabic nasal should strengthen the following fricative or vowel. On the contrary, the *Northern Sotho Spelling and Orthography no. 4* of 1988 does not follow that rule when it comes to some Northern Sotho loanwords: *kansela*, *profense*, *saense*, etc. The following, however, show adherence to the

rule: *kontase* (from ‘condensed’ milk), *khontesara* (from condenser), where the /d/ has been strengthened to /t/.

The spelling inconsistencies that prevail in some of the Northern Sotho dictionaries will be looked into and if possible, be rectified. Having different spellings for one word, especially for a language like Northern Sotho, is not helpful. Paying special attention to the spelling of adopted words will, therefore, be a milestone in the development of vocabulary.

References

- Department of Bantu Education. 1988. *Northern Sotho Terminology and Orthography* No. 4. Pretoria: The Government Printer.
- Lin, I. 2014. A variation in illusion: Vowel epenthesis after /h/ in Japanese loanwords (http://sapience.dec.ens.fr/cogmaster/www/doc/MEMOIRES/2014_LIN_Isabelle.pdf). Accessed 25 March 2015.
- Mojela, V.M. Borrowing and loan words: The lemmatizing of newly acquired lexical items In Sesotho sa Leboa. *Lexicos* 20: 700 – 707.
- Mokgokong, P.C. 1975. Context as a determinant of meaning with special reference to Northern Sotho. Unpublished Doctoral Thesis, Sovenga: University of the North.
- Sesotho sa Leboa (Northern Sotho) - English Dictionary* (<http://africanlanguages.com/sdp/>). Accessed 25 March 2015.

A learner’s dictionary for foreign students learning Afrikaans in a university context

Dané CLAASSEN (16018583@sun.ac.za)

Department of Afrikaans and Dutch, Stellenbosch University

A great amount of research has been done during the past few decades on the typological category of learner’s dictionaries. Despite all of this research, the role of learner’s dictionaries in tertiary language teaching with foreign learners as the target group has not been researched quite enough, thus the topic of this paper. A learner’s dictionary is a dictionary compiled for an adult learner who is learning a foreign language (Tarp & Gouws, 2004). It cannot be denied that dictionaries play an important role in the language acquisition process. During the process of language learning the facilitator is not always available, thus a well-chosen dictionary is the next best resource for solving lexical problems that may be encountered in the process of language learning (Lew & Adamska-Safaciak 2014: 1). It is important to note that currently there are no monolingual or bilingual learner’s dictionaries specifically targeting foreign learners who are learning Afrikaans at a tertiary level. Educators and facilitators must thus compensate by prescribing learner’s dictionaries such as Pharos’s *Aanleerderwoordeboek vir skole/Learner’s dictionary for schools* and Longman-HAT’s *Afrikaans Dictionary and Grammar for English speakers* for these learners.

The main goals of the undertaken study can be summarized as follows: to measure to what extent these types of “compensating” dictionaries can succeed in satisfying the lexicographical needs of foreign learners at a tertiary level, to get feedback from the target users themselves and to make suggestions for the possible composition of a learner’s dictionary for foreign students learning Afrikaans as third language at university level. The reason for the analyses of these specific two learner’s dictionaries is because they are currently being prescribed for foreign students learning Afrikaans as third language respectively at, the University of Stellenbosch and the Nelson Mandela Metropolitan University. The degree, to which a certain dictionary succeeds in satisfying the lexicographical needs of a specific target group, does not only depend on the lexicographer, but also on the target group itself. This is where a dictionary culture, including institutions

that maintain this culture, plays a role. The current dictionary culture was brought to light through reviewing the Curriculum Assessment Policy Statements (CAPS), which is currently being followed in the school system. It is important to note that these dictionaries do not specify its user(s) as a foreign language learner(s) at a tertiary institution. It is only prescribed due to the gap that presently exists in the dictionary typology.

The study was initially conducted using the Function Theory as point of departure, and later combined with the General Theory of Lexicography (Wiegand) as framework. Language acquisition courses differ from university to university and thus only the general, most important lexicographical needs of these foreign learners were determined. The two dictionaries were critically analyzed and then evaluated in the form of case studies as well as qualitatively through feedback from twenty students enrolled in the course *Afrikaans Taalverwerwing 178* at the University of Stellenbosch. The case studies took on the form of one test being given as pre-test and post-test. This test was conducted from the coursework and study guide of *Afrikaans Taalverwerwing 178*. Both tests were done in the classroom at the beginning of the year, the pre-test was conducted without a dictionary and the post-test was conducted with the help of either Pharos's *Aanleerderwoordeboek vir skole/Learner's dictionary for schools* or Longman-HAT's *Afrikaans Dictionary and Grammar for English speakers*. The students were not able to choose a dictionary, but dictionaries were handed to them through random selection to minimize the effect of individual intellect and dictionary skills. The case studies were not done to form an ungrounded generalization about one of the dictionaries, but rather to investigate the influence of a dictionary as resource in general and to point out the advantages and disadvantages of each dictionary. Students were also asked to write a comment on their general experience of the dictionary that was randomly assigned to them.

A practical evaluation of the prescribed dictionaries determined that although both of the dictionaries show indications of successfully responding to both text reception and text production needs, they do not completely succeed in satisfying the needs of foreign learners who are learning Afrikaans in a university context. The results of the post-test also showed an overall improvement of 20% with the use of a dictionary in comparison to not using one. The importance of dictionaries as resources in language teaching at university level has thus been emphasized again through this evaluation. Reviewing the advantages and disadvantages of each dictionary showed that, with additional adaptations, a combination of these two dictionaries may already satisfy the lexicographical needs of these target users to a greater extent.

Lastly the results of the study brought to light the lack of a dictionary culture currently at school level and thus also at university level. The written evaluations also confirmed this lack of a dictionary culture. One of the students wrote that if he/she had been aware of the fact that the dictionary was a bilingual one, he/she would probably have finished and done better in the post-test. This indicates that without the manifestation/maintenance of a dictionary culture at university level, the compilation of a dictionary for these users would not be successful.

References

- Abecassis, M. 2008. The ideology of the perfect dictionary: How efficient can a dictionary be? *Lexikos* [Online] 18. Available: <http://www.ajol.info/index.php/lex/article/view/47238/33618>.
- Al-Kasimi, A.M. 1977. *Linguistics and bilingual dictionaries*. Leiden: E.J. Brill.
- Bergenholtz, H & Tarp, S. 2003. Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes, Journal of Linguistics* [Online] 31. Available: https://pure.au.dk/portal/files/9895/H31_11.pdf.

- Lew, R & Adamska-Safaciak, A. 2014. A case for bilingual learners' dictionaries. *ELT Journal* [Online]. Available: <http://eltj.oxfordjournals.org/>.
- Longman-Hat Afrikaans Dictionary and Grammar for English speakers. 2012. 1st edition. South Africa: Pearson.
- Pharos Aanleerderswoordeboek vir skole/Learner's dictionary for schools. 2010. 4th edition. Cape Town: NB Publishers.
- Tarp, S & Gouws, R. 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke. *Tydskrif vir Geesteswetenskappe* [Online] 44(4). Available: http://search.sabinet.co.za.ez.sun.ac.za/WebZ/images/ejour/akgees/akgees_v44_n4_a.pdf?sessionid=01-53765-979002001&format=F.
- Tarp, S. 2004. Basic problems of learner's lexicography. *Lexikos* [Online] 14. Available: <http://www.ajol.info/index.php/lex/article/view/51421/40074>.

Linking the fulfilment of translational needs to lexicographical theory in the context of UI translation

André DU PLESSIS (ahdp@sun.ac.za)

Marius SWART (mariusswart@sun.ac.za)

Department of Afrikaans and Dutch, Stellenbosch University

In this preliminary investigation, two specific constraints experienced by users in the field of computer user interface translation are used as a means of comparing three lexicographic reference works. These three works are the *Kuberwoordeboek/Cyber Dictionary* (Viljoen, 2006), the Pharos Afrikaans-English/English-Afrikaans dictionary (2010) and the Microsoft Language Portal (www.microsoft.com/Language). The first is the only explicitly named specialised bilingual dictionary in the computer-related field. The second is the only comprehensive Afrikaans/English English/Afrikaans dictionary readily available in South Africa. The third is a free online terminology list owned by Microsoft, described in literature as a “remarkably multilingual” company that is “represented on the majority of the world’s computers” (Kelly and Zetsche, 2012:250). This list caters specifically for the target user translating user interface text.

According to Gouws (2006:85) a dictionary should never be compiled at random – the user must be the central consideration in all processes of lexicographical production and the lexicographer has to remain aware of the specific situations in which the dictionary will be used. Potgieter (2011:3) furthermore points out that discussions relating to the compilation of specialised dictionaries have traditionally not paid significant attention to translators as target users, despite their being a group with particularly challenging requirements. Therefore, in order to enable users to extract maximum value from a dictionary, it is imperative that data must be presented in the most useful and accessible fashion.

Technological innovation results not only in the creation of new words, but also in the use of existing new contexts or in their obtaining new meanings (Fontenelle, 2013:1097). Furthermore, users of innovations such as computers and other devices need the language these devices speak to be their own (Kelly and Zetsche, 2012:250). For translators of user interface text this typically causes two problems: inadequate recognition of shifts in part-of-speech due to expanded meanings of existing words, and inconsistency in the formation of compound nouns using these and other words.

A list of selected examples (in English) is used to determine the extent to which each of the three works mentioned above fulfil the needs of this target user group when translating into Afrikaans by not being inadequate or inconsistent. This is determined by means of

checking the Afrikaans equivalent(s) of each respective English word in terms of the two constraints mentioned above.

After determining this, an attempt is made to indicate whether the use of meaning discriminators may have contributed to the success or failure of each of the works. Beyer (2009:11) mentions the following meaning discriminators:

- Punctuation, which provides negative discrimination by using different punctuation marks to separate different kinds of equivalents;
- Part-of-speech indicators, which differentiate syntactic functions;
- Paraphrases of meaning, which differentiates between polysemous values in order to determine the primary context; and
- Contextual and co-textual guidance, which highlight the secondary context in which the particular equivalent operates, i.e. its pragmatic potential.

The Afrikaans equivalents of the respective English terms are checked for their inclusion of these discriminators as one possible measure of the success with which they have been treated lexicographically.

Preliminary work has indicated that the three reference works differ significantly in their potential to satisfy the needs of the target user, but that no single work is sufficient in itself. Furthermore, there is significant inconsistency in the treatment of compound nouns derived from both new words and existing words that have had new meanings attributed to them. Finally, it can be indicated that the discriminatory entries can aid in working towards greater fulfilment of the unique needs of this target user group.

In conclusion, suggestions for additions to the existing lexicographic theory on meaning discriminators in specialised dictionaries, such guidance for translators and other expert users and user interface design, are made based on the results of this investigation.

References

- Beyer, H.L. 2009. 'n Teoretiese basis vir kontekstualisering in tweetalige woordeboeke. *Lexikos* 19. 1-22.
- Du Plessis, M., Pfeiffer, F. Smith-Müller W. and Luther, J. (reds.) 2010. *Pharos Afrikaans Engels/English-Afrikaans Woordeboek/Dictionary*. Cape Town: Pharos Dictionaries.
- Fontenelle, T. 2013. Electronic dictionaries for text processing and correction systems in Gouws, R.H., Heid, U., Schweickard, W. and Wiegand, H.E. (eds.). *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: HSK 5.4*. Berlin: De Gruyter.
- Gouws, R.H. 2006. Discipline, Dilemmas, Decisions and Data Distribution in the Planning and Compilation of Monolingual Dictionaries. *Lexikos* 16. 84-94.
- Kelly, N. and Zetsche, J. 2012. *Found in translation. How language shapes our lives and transforms the world*. Perigee: New York.
- Microsoft Language Portal. [Online]. Available: <http://www.microsoft.com/Language/>.
- Potgieter, L. 2011. Die ontwerp van vertalende vakwoordeboeke met vertalers as eikengebruikers: 'n teoretiese model. Ongepubliseerde proefskrif: Universiteit Stellenbosch.
- Viljoen, H.C. (red.) 2006. *Kuberwoordeboek Afrikaans-Engels, Engels-Afrikaans/Cyber Dictionary Afrikaans-English, English-Afrikaans*. Pretoria: Protea Boekhuis.

Equivalent relations and their implications for the database of an electronic bilingual sign language dictionary

Hanelle FOURIE BLAIR (hfourie@sun.ac.za)

Buro van die WAT, Stellenbosch University

Manley, Jacobsen and Pedersen (1988) are of the opinion that both monolingual and bilingual dictionaries are actually a bunch of lies – monolingual dictionaries engage in the activity of creating the impression that words have "meaning" or "definitions" rather than a particular ability to enter into meaningful contexts; bilingual dictionaries, on the other hand, try to indicate the words have "equivalents" in other languages, rather than certain relationships of partial equivalence which are further complicated by the range of linguistic and extralinguistic contexts in which words and their equivalents may appear (1988: 281).

The comment above is probably meant to be humouristic or at least ironic, but it does underline the complex relationships between words and their meaning(s) in context in monolingual dictionaries, and words and their equivalents in context in bilingual dictionaries. This paper will discuss the typical equivalent relations in bilingual dictionaries and the core of the paper then makes suggestions about how the lexicographer could address these equivalent relations in an electronic bilingual dictionary in South African Sign Language (SASL) and Afrikaans for the foundation phase, as proposed by Fourie (2013). There is no source language or target language *per se* in the proposed dictionary, because the treatment units are offered in both languages, as will be explained. However, the equivalent relations between the two languages are still important and will particularly have an impact on how both signs and words are treated in the database.

The comment on semantics in a ("regular") bilingual dictionary must also provide for a treatment procedure that encompasses all the polysemous senses. Polysemy is a word specific feature which implies that there is not necessarily a translation equivalent in the target language to be found that has exactly the same polysemous senses (Gouws and Prinsloo 2005: 151, 152). Therefore it means that the lexicographer often has to provide a separate translation equivalent for each of the polysemous senses of a lemma. The collection of translation equivalents, whether it is one or more, which are offered in the comment on semantics of a single article is known as the **translation equivalent paradigm** of the given article. The lexicographer has to ensure that the target user of a dictionary can successfully extract information from a translation equivalent paradigm. In the case of more than one translation equivalent in the paradigm, the lexicographer cannot simply rely on the intuition of the user to make the correct choice(s) (Gouws and Prinsloo 2005: 152).

Users that consult a bilingual dictionary often expect that this type of dictionary will also offer data about the meaning of the source language item, as represented by the lemma, and therefore see the translation equivalents as "the meaning of the source language word in the other language". They do not often seem to realise that the data that is presented is not a paraphrase of meaning but rather a list of translation equivalents (Gouws and Prinsloo 2005: 153). Translation equivalents should not, however, be regarded as entries which give the meaning of the lemma, but as lexical items in the target language that may be used to replace the source language item in a specific situation or context. The lexicographic treatment in a bilingual dictionary may therefore not isolate translation equivalents from their typical contexts and cotexts, as can be seen in the following example from the *Groot Woordeboek/Major Dictionary*:

vrug, (-te), fruit; result, effect; embryo, foetus ...

From a semantic perspective it is wrong to argue that any of the equivalents like *fruit*, *result*, *effect*, *embryo* and *foetus* can be seen as the meaning of the word *vrug*; *vrug* does not **mean** effect, but can be **translated** with the word *effect* in a particular context.

We therefore start to see that there are several relations of equivalence between lemmas and their translation equivalents, for example complete equivalence, partial equivalence, divergence, zero equivalence, convergence and communicative equivalence. The paper will provide examples of each in the SASL-Afrikaans language pair and then continue to discuss the impact of equivalence relationships on the database of the proposed dictionary, followed by suggestions on how to treat the signs and words in the database in terms of these equivalent relations.

References

- Eksteen, L.C. et al. (eds.). 1997¹⁴. *Groot Woordeboek/Major Dictionary*. Cape Town: Pharos
- Fourie, H. 2013. *'n Leksikografiese model vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans*. Unpublished doctoral dissertation. Stellenbosch: Stellenbosch University.
- Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PreSS.
- Manley, J., J. Jacobsen and V.H. Pedersen. 1988. Telling lies efficiently: terminology and the microstructure in the bilingual dictionary. In Hyldgaard-Jensen, K. en A. Zettersten (eds.). *Symposium on Lexicography III*: 281-301. Tübingen: Niemeyer.

Who are the target users of monolingual learners' dictionaries?

Rufus H. GOUWS (rhg@sun.ac.za)

Department of Afrikaans and Ducth, University of Stellenbosch

Many terms used in the era prior to the emergence of a general theory of lexicography are still in frequent use in the lexicographic practice and in theoretical discussions, albeit that these terms have not all been allocated a definition that accurately reflects their use in lexicography. This paper offers a critical discussion of monolingual learners' dictionaries and emphasises the fact that the term monolingual learners' dictionary has not been defined in a scientifically sufficient way because the nature of the learner in this learners' dictionary has not been appropriately described. These dictionaries are primarily directed at learners of a foreign language and the needs of these users with regard to the foreign language are at the centre of theoretical and practical work in this component of lexicography. However, the users of monolingual learners' dictionaries often have different mother languages. Unfortunately little or no attention is given to the mother language of these users and the implications it has for their successful use of monolingual learners' dictionaries. As a result of this failure the lexicographic practice produces dictionaries that do not take cognizance of some of the real needs of the users of monolingual learners' dictionaries.

A brief discussion of the introduction of learners' dictionaries to the typological spectrum is followed by a discussion of some significant features of this dictionary type. The focus is then shifted to the need for better user-directed monolingual learners' dictionaries. It is shown how the mother language of users can demand more guidance in learners' dictionaries with regard to the presentation and treatment of false friends and culture-bound lexical items. More attention should be given to language influence and language interference between the language treated in the dictionary and the mother language of intended target users.

Proposals are made for a better model for monolingual learners' dictionaries. In printed dictionaries a well-devised data distribution structure should introduce the use of extended obligatory microstructures in order to complement the generic and default data with data directed at the needs of users with specific mother languages. Utilising a drawer structure outer texts could be included with data regarding a selected number of languages that are mother languages of the intended target user group.

Online dictionaries offer better possibilities to solve the problem of users with different mother languages that consult the same generic monolingual learners' dictionary. The planning of these dictionaries could make provision for the use of a single data bank from which different user-directed variants of the dictionary can be retrieved. Following the user profile specified by each user the generic data and treatment can be retrieved. If a user has specified a specific mother language in his/her user profile the generic data will be complemented by data relevant to the specific mother language. For a single guiding element different partial articles can be retrieved as screen shots and the user can switch from a default profile to a language-specific profile. In such a case the default data will be complemented by additional data with unique items presented in an extended obligatory microstructure.

The inclusion of data regarding different mother languages of the users of the monolingual learners' dictionaries leads to a question regarding the monolingual nature of such a dictionary. Following the established notion of monolingual dictionaries with a bilingual dimension this paper argues in favour of a monolingual learners' dictionary with a multilingual dimension.

By making provision not only for the needs of users with regard to the foreign language they have to learn but also for the problems due to different mother languages within a single user group a model can be developed that will enhance the quality and user-directedness of monolingual learners' dictionaries.

Enhancement of the African Wordnets: additional language resources and tools

Marissa GRIESEL (griesel.marissa@gmail.com)

Sonja E. BOSCH (boschse@unisa.ac.za)

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

The African Wordnet Project (AWN) is a prototypical lexical database consisting of words that are grouped into sets of synonyms called synsets. Such synonym sets are linked by conceptual-semantic and lexical relations (Miller et. al., 1995) that go beyond merely an electronic version of a (printed) dictionary. Where electronic dictionaries have a set format and entries can only be accessed as a whole, synsets in a wordnet are structured so that a user may define a more detailed search, based on the different relations that link the synonyms, and get only the relevant information (cf. Miller et. al., 1993). This makes a wordnet ideally suited as reference to the lexicographer, but also necessitates populating the wordnet with valuable and correct information and using the appropriate interface (cf. Kosem & Kosem, 2011, and Vider & Orav, 2013). Currently the AWN contains data for five languages namely Setswana, isiXhosa, isiZulu, Sesotho sa Leboa (Sepedi) and Tshivenda. The AWN depends heavily on the English Princeton WordNet (PWN) (Fellbaum, 1998) for continual and manual expansion and up to now has used Microsoft Excel sheets and the DEBVisDic interface for development and browsing (see <https://deb.fi.muni.cz/clients-debvisdic.php>).

In this presentation we will address our two-pronged approach to enhancing the development of the African Wordnet, firstly with strategies to fast-track the quantity of content and secondly via a development interface to enrich the quality of the data and support

the developers/linguists in their work. By following this approach, we hope to not only speed up the development of the AWN, but also to create a rich resource that can form the backbone for future lexicographic development.

Studies to speed up the time consuming manual work of building wordnets with corpus extractions have been performed for a variety of languages (cf. Oliver, 2014; Montazery & Faili, 2010). The biggest difference between these languages and the African languages, however, is the richness and quantity of data contained in freely accessible dictionaries. In the case of the resource scarce African languages, we have to rely on a few dictionaries or wordlists that are available online or that are available for retro-digitisation. Two of the data types contained in a complete synset can benefit from (semi-) automatic extraction of data:

Basic synsets – a basic synset is made up of a literal, part of speech tag and the relations mentioned above. Our experiments centred around utilising the minimal amount of information available in freely available electronic resources to identify synsets in the PWN from which we could extract this basic structure and include it in the AWN. It involves merging information from the dictionary with the language independent information in the PWN and presenting the potential basic synset to linguists for final approval and inclusion in the wordnets. Even on limited datasets, the semi-automatic extraction showed promising results, for instance in the case of Setswana, roughly 700 new noun synsets could be identified, 1,000 for Tshivenda and around 3,000 potential new noun synsets for Sesotho sa Leboa.

Usage examples – in addition to the basic synset, each sense should be further enriched by a usage example in the target language showing the use of the sense in context. Usage examples to enhance existing basic synsets were extracted from available corpora (e.g. Wortschatz corpora (Universität Leipzig, <http://corpora.informatik.unileipzig.de/>) and the NCHLT corpora (available via the Resource Management Agency on <http://rma.nwu.ac.za/>)) and again presented to linguists for verification and inclusion. First experiments as to the suitability of these examples will be reported on.

As mentioned above, most of the developmental work for the AWN is performed in Microsoft Excel sheets. The Excel sheets make it difficult to see inaccuracies in the data and impossible to envision the links between synsets in this semantic dictionary. The DEBVisDic platform has been used with some success, but technical difficulties it less than ideal for our purposes. Given these challenges, experiments to automatically check the linguistic and technical quality of the data were undertaken (cf. Smrz, 2003 for an example of work in the BalkaNet project) and the development team decided to design and implement a unique integrated system for the AWN. The resulting development interface will not only ease the inclusion of new synsets by simplifying the addition of basic synsets, but will also perform checks and suggest improvements to existing data. During this presentation, a demonstration of the application with its user friendly interface is envisaged.

References

- Fellbaum, C., (ed), 1998. *Wordnet: An electronic lexical database*. The MIT Press, Cambridge, Mass. ISBN 978-0262-06-197-1.
- Kosem, I., & Kosem, K. 2011. *Electronic Lexicography in the 21st Century. New Applications for New Users*. In *Proceedings of eLex 2011*. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.233.1717&rep=rep1&type=pdf>.
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D. & Miller, K. 1993. *Introduction to WordNet: An On-line Lexical Database*. International Journal of Lexicography 3(4):235-244 [Online]. Available: <http://www.isi.edu/isd/kr/5papers.pdf>.

- Miller, G. 1995. *WordNet: A Lexical Database for English*. Communications of the ACM, Vol. 38 No.11. [Online]. Available: <http://dl.acm.org/citation.cfm?id=219748>.
- Montazery, M & Faili, H. 2010. *Automatic Persian WordNet Construction*. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters Volume (COLING '10), pp. 846-850, Beijing, China, 2010. [Online]. Available: <http://www.aclweb.org/anthology/C10-2097>.
- Oliver, A. 2014. *WN-Toolkit: Automatic generation of WordNets following the expand model*. In Orav, H., Fellbaum, C. & Vossen, P. (eds): Proceedings of the 7th Global WordNet Conference 2014 (GWC2014), Demonstration Session, pp. 7-15, Tartu, Estonia, ISBN 978-9949-32-492-7.
- Smrz, P. 2003. *Lessons Learned from Standardization Efforts in the Context of the BalkaNet project and the Global Wordnet Association*. Proceedings of International Roadmap For Language Resources (ENABLER/ELSNET) Workshop, pp. 11-12, Paris, France. [Online]. Available: http://nlp.fi.muni.cz/publications/write2003_smrz/write2003_smrz.pdf.
- Vider, K & Orav, H. 2013. *Nordic and Baltic wordnets aligned and compared through "WordTies"*. In Proceedings of The 19th Nordic Conference of Computational Linguistics, pp. 147-162. [Online]. Available: <http://www.emmtee.net/oe/nodalida13/conference/31.pdf>.

To what extent do Learners' dictionaries use defining vocabularies for their examples?

Lorna HILES (lorna@dictionary.co.za)

Department of Afrikaans and Dutch, University of Stellenbosch

Example sentences are used in dictionaries to supplement and exemplify the information given in the rest of the dictionary article, for example, the paraphrase of meaning or grammatical information such as part of speech and inflections, etc. Much research has shown how valuable examples are to users of learners' dictionaries in helping them understand the meaning of the word being looked up, and how to use it.

Defining vocabularies are used to restrict the vocabulary used in definitions, to ensure that they are accessible by the expected users of the dictionary. A defining vocabulary is a list of words determined to be A) understood by intended users of the dictionary, and B) useful for the construction of paraphrases of meaning. The lexicographers involved in the compilation of a dictionary will use the defining vocabulary to construct paraphrases of meaning that are simple and accessible for users of the dictionary. When words that are not in the defining vocabulary are necessary for a particular definition, the compilers will need to employ various strategies to compensate.

Example sentences are usually exempt from the use of a defining vocabulary, and are compiled or selected from a corpus to illustrate the use of a word. They are also more flexible than paraphrases of meaning, in that one example sentence does not need to convey the whole meaning of a sense.

The role of the example sentence in supporting the definition is to provide information that the paraphrase of meaning is unable to provide for whatever reason. If the reason is that there is a restriction on the words used in the paraphrase of meaning, then a similar restriction on the vocabulary used in the example sentence will be of no use.

In this paper, I will determine the extent to which the examples in learners' dictionaries follow the defining vocabulary. To achieve this, I shall take a selection of example sentences from four learners' or school dictionaries that use a defining vocabulary for definitions. The four dictionaries are: Oxford Advanced Learner's Dictionary 7th edition

(OALD7), Cambridge Advanced Learner's Dictionary 4th edition (CALD4), Longman South African School Dictionary (LSASD) and Oxford South African School Dictionary 3rd edition (OSASD3). The strategy for the use of examples will be discussed – such as, whether examples are used in each entry or not, and whether example sentences or phrases, or both are used.

I shall analyse the example sentences in terms of their apparent use of the restricted vocabulary and provide data on the use across the four dictionaries. In cases where words outside the defining vocabulary are used, I shall examine whether the examples would be better if they did only contain words in the defining vocabulary, and suggest alternatives.

An example of the kind of analysis that will be shown is as follows:

alert (adj) *Suddenly he found himself awake and fully alert.*
Two alert scientists spotted the mistake.
We must be alert to the possibility of danger.

The underlined words are not in the defining vocabulary of this dictionary. (Obviously, the headword needs to be in the examples, whether or not it is in the defining vocabulary.) The only word in these three examples that is not in the defining vocabulary is *spotted*. The noun “spot” is in the defining vocabulary, but not the verb. In the example sentence, it could be replaced by *noticed*, *saw*, *found*, *discovered*, which are all in the defining vocabulary.

In this paper, I shall show that most of the dictionary entries I looked at followed this example in that most of the words are to be found in the defining vocabulary. I shall also provide some recommendations for further research in this area, and practical recommendations for publishers. While this research does not provide the answer to whether example sentences should remain faithful to the defining vocabulary, it does show to what extent example sentences do use defining vocabulary words.

Term extraction for an isiZulu linguistic terms dictionary using a corpus linguistic method

Langa KHUMALO (khumalol@ukzn.ac.za)

Linguistics Program, School of Arts, University of KwaZulu-Natal

The University of KwaZulu-Natal (UKZN) is compiling a series of LSP dictionaries for various specialized subject domains in line with its language policy and plan (Khumalo 2014:1). The focus in this study is the term extraction for words in the linguistics subject domain. This paper advances the use of frequency analysis and the keyword analysis as strategies to extract terms for the compilation of the dictionary of isiZulu linguistic terms. The study uses the isiZulu National Corpus (INC) of 1, 1 million tokens as a reference corpus as well as an LSP corpus of 100,000 tokens as a study corpus. Our study is analyzed through the use of a software tool called WordSmith Tools (version 6). WordSmith Tools (hence forth WS Tools) is an integrated suite of three main programs, which include the WordList, Concord and Keywords, used in analysing words and word patterns in any given text. WS Tools was developed by Mike Scott, who had earlier worked with Tim Jones to develop MicroConcord. WS Tools was first released in 1996 and the current version (version 6.0) was released in 2011. The Wordlist tool can be used to produce wordlists or word-cluster lists from a text and render the results alphabetically or by frequency order. It can also calculate word spread across a variety of texts. The Concord tool can give any word or phrase in context – so that one can study its co-text, i.e. see what other words occur in its vicinity. The Keywords tool calculates words which are key in a text, i.e., used much more frequently or

much less frequently in a given corpus (e.g. the LSP corpus) than expected in terms of a general corpus of the language (e.g. the INC). Using the WS Tools software a lot of qualitative and quantitative research can be done in the language. Central to this study is a computational determination of which words are typical of the linguistic domain in isiZulu and therefore stand out as preferred candidates for headword selection.

The study uses the corpus linguistic method as a basis for theoretical analysis. According to Sinclair (2005) a corpus is “a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research”. The advantage of such a theoretical approach is that a “[...] a corpus [is] stored in a computer, it is easy to find, sort and count items, either as a basis for linguistic description or for addressing language-related issues and problems” (Kennedy, 1998: 11). The LSP corpus used in this study comprises of the two main isiZulu grammar textbooks *Uhlelo lwesiZulu*, and *Izikhali zabafundi nabaqeqeshi*, a collection of lecture notes from academics in the School of Arts and the School of Education at UKZN, and online linguistic documents in isiZulu.

Using the WS Tools software, the study will perform the following. The author will run a frequency list to determine the most frequent words in the LSP corpus. A frequency list provides an array of different types of words, tokens, or forms which make up a corpus. These can appear either listed from the most frequent token to *hapax legomena* (i.e. those forms that occur only once in a given corpus) or vice versa. Frequency lists are a powerful tool in corpus lexicography. They guide lexicographers on which words to include in a dictionary. Frequency lists also provide developers of second language teaching material with the most relevant words, phrases, and expressions to teach. In this study a frequency list sheds more light on the most common words in isiZulu linguistic domain. These words may be the ones which characteristically typify the domain. According to Kilgarriff (1997: 135) “The more common it is, the more important it is to know it.”

Finally, the study will use the keyword analysis in order to identify words particular to the isiZulu linguistics domain through the calculation of keyness, which isolates words which are key to the LSP corpus. According to Mike Scott (2006:92) keyness is “calculated by comparing the frequency of each word in the word list of the text under investigation with the frequency of the same word in the reference word list”. Calculations will be done using the Keyword tool of WS Tools. The paper will thus show that term extraction for the isiZulu dictionary of linguistic terms is done following reliable computational techniques in corpus lexicography.

References

- Kennedy, G.D. 1998. *An introduction to corpus linguistics*. London/New York: Longman.
- Kilgarriff, A. 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10 (2): 135-155.
- Khumalo, L. 2014. Developing and isiZulu Dictionary of Linguistics Terms: Challenges and Prospects. Unpublished Paper Presented at the 19th International AFRILEX Conference. North-West University. 30 June – 3 July 2014.
- Scott, M. 2004-2006. *Oxford WordSmith Tools version 4*. Oxford: Oxford University Press.

Towards an inclusive dictionary: bridging the gap between standard and non-standard varieties of Setswana

Baile G. MAREME (godfrey.mareme@nwu.ac.za)

Keabetswe N. LEKOPANYE (keabetswe.lekopanye@nwu.ac.za)

SEFALA- SETNALEU, North West University

Setswana is one of the indigenous languages which were afforded official status in post-apartheid South Africa, and is also spoken in neighbouring Namibia, Zimbabwe, and Botswana. In South Africa, this language is found in the Free State, the Northern Cape, Gauteng, as well as the North West province where it enjoys a majority of L1 speakers. Naturally, dialects are found in all the above mentioned provinces where Setswana is spoken. The North West province, in particular is home to various regional dialects of Setswana, for example Serolong, Sekgatla and Sekwena, Sehurutshe, Setlhaping, and Setlharo.

The Pan South African Language Board (PanSALB), through its structures, National Language Bodies (NLBs) set out to put measures in place to develop and preserve previously disadvantaged languages. In the case of Setswana, the Setswana National Language Body (SNLB) has been tasked with authenticating the spelling and orthography of Setswana so as to achieve uniformity and standardisation of the language. On one hand, the great strides made by the SNLB to standardize the language cannot be ignored. Through DAC, SNLB (2005-2012) has reviewed and authenticated a lot of terms in the fields of Science and Technology, Trade and Industry, Sports, Health, Law, Commerce, and Government. In addition, the Spelling and Orthography rules booklet was produced by the same NLB. Despite such endeavors, the recorded decline in the growth rate of Setswana in the last few years (Census, 2011) is worrying and cannot be ignored. This has been a worrying factor to almost all African languages in South Africa.

It has also been observed that a big gap exists between vocabularies of the standard variety and that of the non-standard variety of Setswana (Mareme and Mojela, 2014). These authors have argued that such a gap leaves the lexicographers of monolingual dictionaries of South African indigenous languages with a deficiency in vocabulary. This can be attributed to the fact that vocabulary of the non-standard dialects is overlooked and, therefore goes unrecorded, as standard or proper language. Certainly, such neglect leads to an even further underdevelopment and stagnation of the language as a whole. The *status quo* will prevail for as long as the dialects continue to be overlooked as they may be deemed non-standard. However, its fragmented nature can be used to the advantage for the development of Setswana. Considering the important role and functions of a dictionary in developing a language, this paper proposes that vocabulary be drawn from the various dialects of Setswana and be considered for inclusion in the language's monolingual dictionary after they are standardized. The basis for this argument is that in reality the vocabulary exists, and by virtue of its usage should be harvested, documented, and ultimately form part of the standard language. For instance, a word that occurs in one dialect may be needed in another dialect and therefore such word(s) need to be standardised. The authors are of the view that by overlooking the non-standard dialects, a great deal of wealth of words and their meaning in different social contexts is lost. If they are kept and labelled as dialects, users regard them as inferior. Should there be a word that is only found in the Setlharo- Setlhaping dialect, and non-existent in others, it must be authenticated by the SNLB and be standardized and be used in all official documents. Noteworthy, is that the validity of using a non-standard dialect to enrich the vocabulary, particularly of African languages has already been established by Ditsele (2014) and naturally, this paper argues in line with such assertions. It is the right time to ask pertinent questions relating to the inclusion of the non-standard dialects as standard after having gone through all the processes of authentication, in the various Setswana

dictionaries. The rationale behind this is that drawing vocabulary from the non-standard dialects of Setswana would not only preserve, but also develop the language. It is clear the the recently published Setswana Dictionaries e.g. *Thanodi ya Setswana* (Mareme, 2009); *Tlhalosi ya Medi ya Setswana* (Otlogetswe, 2012); *Macmillan-Setswana and English Illustrated Dictionary* (Cole and Moncho-Warren, 2012) do indeed include the non-standard form. However, it is argued in this paper that there should be an era where dialectal terms are elevated to a standard form is upon the horizon. This is to compensate for a lack of words/vocabulary in one dialect by looking towards another dialect; this will contribute to the overall development of Setswana. In such manner, the standard language will see growth.

The authors advance that the consideration of the non-standard dialects is an ideal tool for developing and preserving the language. To these ends, it is posited that the non-standard dialects should soon form part of the standard form or at the very least be included in the dictionaries. The authors recognise the great potential in this method especially in that it can contribute to the preservation of a language which is one of the purposes of the dictionary. A questionnaire will be developed and results thereof used to back up what is already a talk within the circles of the various users of the language. This paper therefore makes a clear case/argument for the rethinking of the approach to developing vocabulary through the inclusion of vocabulary of the non-standard dialects of Setswana (to standardise and not only to include them) so as to augment the already dwindling vocabulary of the language. The research is to be conducted through an engagement of fieldworkers who would be supported through a pilot study or a consultative process with stakeholders, mainly potential dictionary users such as university students, lecturers, experts in the language, and community members. It is anticipated that the outcomes will prove that the inclusion of words considered to be non-standard [to be standardised] will satisfy the tenet of language development, as well as the needs the dictionary users. It also anticipated that such an exercise would also work in favour of the enlargement of corpora especially for monolingual dictionaries as argued by Golele (2006).

References

- Census 2011. Available at <http://www.statssa.gov.za/publications/P03014/P030142011.pdf>.
- Cole D.T. and Moncho-Warren, L.M. 2012. *Setswana and English Illustrated Dictionary*. Northlands: Macmillan Publishers (Pty) Ltd.
- Ditsele, T. 2014. Why not use Sepitori to enrich the vocabularies of Setswana and Sepedi. *Southern African Linguistics Applied Language Studies*. 32 (2): 215-228.
- Golele, N.C.P. 2006. Thoughts and Views on the Compilation of Monolingual Dictionaries in South Africa. *Lexikos* 16: 251-256.
- Mareme G.B. 2009. *Thanodi ya Setswana*. Pietermaritzburg: Nutrend Publishers.
- Mareme, G.B. and Mojela, V.M. 2014. The TED's 1950 Orthography and developments of standard dictionaries in the Sotho languages, a case of Setswana and Sesotho sa Leboa. Paper presented at the 19th International Conference of African Association for Lexicography, Potchefstroom, RSA, 03 June.
- Otlogetswe, T.J. 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.

The Inclusion of Collocations in Dictionaries of African Languages: A Case of Northern Sotho

Kwena G. MASHAMAITE (mashakj@unisa.ac.za)

University of South Africa

Mampa L. MPHAHLELE, (mphahleleml@tut.ac.za)

Department of Applied Languages, Tshwane University of Technology

The aim of this paper is to explore the inclusion of collocations in a variety of dictionaries of African languages of South Africa and in Northern Sotho in particular. The preliminary investigation into the presence of collocations in the various types of dictionaries in African languages of South Africa showed a dearth of or absence of collocations. The various forms of collocations and the types of collocates will be identified and analysed. This analysis will be used to suggest the lemmas or headwords under which the collocation can be accessed in the dictionary. Collocations consist of bases or base words as well as collocators in the form of modifiers or qualifiers. For example, *meno mašweu* (white teeth) consists of the base word “meno” (teeth) and the qualifier “mašweu” (white). There are two types of collocations namely, lexical collocations consisting of a combination of nouns’ verbs and adjectives. For example, “monwana le lenala” (finger and finger nail), “nwa meetse” (drink water) and grammatical collocations in the form of prepositional phrases used either after or before a lexical word. For example “mosegare wa sekgalela” (midday) and “pula ya dikgadima” (thunder storm). It is proposed that lemmatization of collocations would be based on the base word. For example “mosegare wa sekgalela” will be accessed under the word “mosegare”.

Collocations in any language reflect mostly the idiom of that language which in turn reflects the richness and the broad lexicon defining the language in question. Furthermore, collocations like idiomatic expressions provide a specific meaning of a social context that forms the culture of the people speaking that language. Non-mother tongue speakers might not understand the meaning of collocations because they may be taken literally. For example, in Northern Sotho “go faga teye” means “to brew tea” which is often translated as “go dira teye” literally meaning “to make tea”. It is therefore important to preserve collocations as lexical items or lemmas especially in specialized dictionaries such as collocation dictionaries and to provide their proper meanings some of which are cultural in nature.

The younger generation and some adult speakers have a tendency of coining collocations that are dissonant with the cultural use of the language because of the influence of other languages and English in particular. The collocation “kgatha tema” in Northern Sotho is popularly used as “bapala karolo” which is a direct translation of the English collocation “play a part or role”. The frequent use of this translated form is presumed to be the correct collocation in Northern Sotho whereas it is not. The richness and culture-bound meaning of the collocation “kgatha tema” derived from taking part in the tilling of the ground or farm is lost in the translation “bapala karolo”. Another example is “moletlo wa teye” (tea party). From the western culture a tea party is a function where tea and cakes and deserts are served. The practice in the black South African society is that the invitation is titled “moletlo wa teye” (tea party) but a full meal with soft and strong beverages are served. There is therefore a need to preserve this language heritage in a dictionary form.

The study will evaluate the presence of collocations in the available monolingual and bilingual dictionaries in a variety of African languages of South Africa and in Northern Sotho in particular as sources of information. Research articles on collocations will form part of the theoretical framework upon which this study is grounded.

As indicated above, the methodology used for this study is the documentation approach. Content analysis will be the relevant technique used to analyse information gathered.

There is a need for dictionaries of collocations in African languages in South Africa to enable users to not only know the correct combination of words to enrich one's vocabulary but also to know the meaning of such group of words as a meaning unit. The solution to this lexicographic void of collocation dictionaries in African languages of South Africa would be the development of a frame work for the compilation of collocation dictionaries along the likes of the *Oxford Collocations Dictionary of English*.

The outcome of this research would contribute to the body of knowledge required in the development of African languages through the development of specialized dictionaries of collocations and the preservation of our African culture that is imbued in the use of collocations in our social interaction.

Terminology Development Principles: Towards a Principled Approach to UKZN isiZulu Technical Terms

Khumbulani MNGADI (Mngadik2@ukzn.ac.za)

University Language Planning and Development Office, University of KwaZulu-Natal

Terminology development is a worldwide phenomenon especially in bilingual or multilingual countries. In spite of the increasing English hegemony worldwide, countries like South Africa, Croatia and Ireland are amongst those that take pride in their indigenous languages, and still regard them as key tools to their cultural, economic, educational, political and even more importantly, technological growth. These countries have invested personnel and fiscal resources to ensuring that their languages are developed to the level of or even to the level above English. These developments will in turn play a significant role in the currency of the languages that are being developed in the global convergence. The author holds a view that the key to any language development is technical terminology development. It plays a pivotal role in the afore-mentioned spheres; hence the focus of this paper is to look at the principles that govern the technical terminology processes with specific reference to UKZN isiZulu language terminology development initiatives.

The South African constitution of 1996 is renowned for its quality and responsiveness to the plight of the general populace. It is respected worldwide for its balanced and well-crafted democratic values and practices. The field of language is comprehensively covered in the constitution; and there are extremely important excerpts in it that seek to encompass SA linguistic and cultural diversity. It gives recognition to the importance of respecting people's cultural and linguistic preferences, while section 6(1) ensures the need to develop the underdeveloped languages mainly the South African indigenous languages. In these excerpts of the SA constitution on linguistic and cultural rights, the common denominator is that all languages should be afforded parity of esteem regardless of the size and the number of people in a particular speech community. This in turn, entrenches the culture of equality and access to information for everyone thus enabling citizens to contribute meaningfully to societal discourses.

As indicated earlier, at the heart of this important linguistic currency of the indigenous language development is the technical terminology development. There is ample evidence that demonstrate that terminology development is central to every language's development. In fact, countries that take pride in developing their languages are well developed and ordinarily do well in almost all societal spheres i.e. education, economy, technology and

politics to mention but a few. It is for this reason that the author of this paper intends explore ways in which isiZulu language can be properly developed to level of English in the years to come and become the language of administration, teaching and learning, research, invention, innovation and technological advancement.

Arguments for and against indigenous language development have been advanced and are well documented in the literature since the dawn of SA democracy in 1994. Often, these arguments have to a certain extent been marred by emotional and political posturing rather than well researched evidence based positions. This has resulted in a slow pace of processes geared towards developing indigenous languages. The effects of this slow pace have manifested themselves in many ways in that speech communities have gradually lost faith in the ability of their languages to be languages of economy. In turn, most SA communities have sought to embrace English as the language of success. This has invariably deterred language users from realizing their currency. It is often argued that since these languages are underdeveloped they inevitably stand no chance in the global economy. Parents have succumbed to the notion that English is the beginning and the end of it all.

The paper will show that UKZN is advancing towards a principled way of developing technical terminology in isiZulu. The terminology developed is stored in terminology banks that are developed in order to enhance access and dissemination of the standardized technical terminology. The paper will finally argue therefore that it is not enough to develop and avail new technical terminology in a language like isiZulu but a comprehensive program of use and the said terminology in context is equally vitally important. To this end the paper will show that UKZN has created synergies with various media, and statutory bodies in order to achieve this.

Evaluation of the performance of a machine learning lemmatiser for isiXhosa

Lulamile MZAMO (Lula_mzamo@yahoo.co.uk)

North-West University

Albert S.J. HELBERG (Albert.Helberg@nwu.ac.za)

North-West University

Sonja E. BOSCH (boschse@unisa.ac.za)

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

Dictionary compilation has used corpus-based methods for a long time (Biber, Conrad, & Reppen, 2004, p. 21). Lexicographers in Bantu languages have used corpus data to extract dictionary entries in actual language use (de Pauw & de Schryver, 2008) and used word frequencies in deciding the inclusion or not of a lexical item. Simple word frequencies are however misleading for agglutinative languages because agglutinative languages have a high proportion of inflected words hence the need for word lemma frequencies (Perea et al., 2006). To reduce words to lemmas a typical natural language processing tool, the lemmatiser, becomes a useful lexicographical tool. This paper argues that it is possible to build such a lexicographical tool from existing limited computational resources. This study tests this hypothesis on one such agglutinative language with limited resources, isiXhosa. IsiXhosa is one of the under-resourced languages of South Africa (Sharma Grover, van Huyssteen, & Pretorius, 2011). It is spoken by 8.1million mother-tongue users, 16% of the South African population (Statistics South Africa., 2012, p. 25) and is agglutinative (Allwood, Grönqvist, & Hendrikse, 2003).

This study attempts to build a tool from existing data, a field called machine learning (Mitchell, 1997). The data used in this study is the IsiXhosa NCHLT Annotated Text Corpora (South African Department of Arts and Culture & Centre for Text Technology (CTeXT) North-West University, South Africa, 2013), a product of the NCHLT Project on Text Resources conducted by the North-West University's Centre for Text Technology (CTEXT) and the Republic of South Africa's Department of Arts and Culture (Eiselen & Puttkammer, 2014). The data was generated using rules from a study conducted by Bosch et al. (2006). The data is freely available via the South African Language Resource Management Agency website¹. The data consists of two lemma annotated files, a 50000 word corpus of wordform-lemma pairs and a 5000 word corpus of word-lemma pairs for testing purposes.

An analysis of the training data for isiXhosa was done. The exercise identified the affixes to be removed and inserted to generate a lemma, and combined those into transformation classes. The affixes identified to be removed were prefixes and suffixes, together with corresponding insertions required if any. The analysis confirmed that isiXhosa is a primarily prefixal language, identifying 1897 prefixes covering 82.8% of the data. 237 suffixes were identified covering 15.6% of the data. The analysis identified 4109 transformation classes covering 83.7% of the data. The rest of the data (16.3%) is lemmatised to their full wordforms.

A literature study on what an appropriate isiXhosa lemma should be, was conducted to guide the study. Most word categories should be lemmatised to the stems that the words are derived from. For a few word categories that are not based on stems the full word form should be chosen, e.g. original interjections, some conjunctives and demonstrative pronouns. Pronouns, copulative and interjections derived from other word categories should be lemmatised to the stems of those categories. Nouns derived from verbs are lemmatised to the noun stem form, e.g. umhambi (traveller) > hambu.

Armed with this understanding of the language, and the analytical nature of the language as represented by the data, a choice of a freely available lemmatiser was done. The best fitting type of algorithm for isiXhosa is the Ripple-Down-Rules (RDR) algorithm (Plisson, Lavrac, & Mladenic, 2004). The RDR is a hierarchy of transformational rules induced from word-lemma pairs where the one level contains the rules and the following level contains exceptions to each rule, and so on. The CST lemmatiser (Jongejan & Dalianis, 2009), an implementation of the RDR algorithm, was chosen because it handles prefixes, infixes and suffixes, it is free, and is widely regarded as the benchmark in many studies.

The CST lemmatiser was trained with 36535 isiXhosa word-lemma pairs and tested on a separate data set of 3793 word-lemma pairs.

To evaluate the lemmatiser as a lexicographic tool, the processing speed is measured and the accuracy determined. Of interest is lemmatisation accuracy on known and unknown words. Preliminary results show that the CST lemmatiser can lemmatise the test set at speeds of over 25000 words per second on a 2.4 Ghz Quad Core machine at an overall accuracy of 79%. Even though the speed is good the achieved accuracy is still below the 94% achieved by Brits et al. (2005) on Setswana and the 92.4% achieved by de Pauw & de Schryver (2008) on Swahili. The system achieves 99.3% on words that were part of the training set and 46.6% on unknown words. Work is on-going in order to improve the generalising of the system even further.

¹ <http://www.rma.nwu.ac.za>

References

- Allwood, J., Grönqvist, L., & Hendrikse, A. (2003). Developing a tagset and tagger for the African languages of South Africa with special reference to Xhosa. *Southern African Linguistics and Applied Language Studies*, 21(4), 223–237. doi:10.2989/16073610309486345.
- Biber, D., Conrad, S., & Reppen, R. (2004). *Corpus Linguistics: Investigating Language Structure and Use* (4th Editio). Cambridge: Cambridge University Press.
- Bosch, S., Jones, J., Pretorius, L., & Winston Anderson. (2006). Resource Development for South African Bantu Languages: Computational Morphological Analysers and Machine - Readable Lexicons. In *Proceedings on the Workshop on Networking the Development of Language Resources for African Languages. 5th International Conference on Language Resources and Evaluation* (pp. 38–43).
- Brits, K., Pretorius, R., & van Huyssteen, G. B. (2005). Automatic lemmatization in Setswana: towards a prototype. *South African Journal of African Languages*, 1, 37–48.
- De Pauw, G., & de Schryver, G. (2008). Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes, 18(July), 303–318.
- Eiselen, R., & Puttkammer, M. J. (2014). Developing Text Resources for Ten South African Languages. In N. C. (Conference Chair), K. Choukri, T. Declerck, H. Loftso, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 3698–3703). Reykjavik, Iceland: European Language Resources Association (ELRA).
- Jongejan, B., & Dalianis, H. (2009). Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike. In *Proc. 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP. Suntec* (pp. 145–153). Suntec, Singapore.
- Mitchell, T. M. (1997). *Machine Learning*. WCB/McGraw-Hill.
- Perea, M., Urkia, M., Davis, C. J., Agirre, A., Laseka, E., & Carreiras, M. (2006). E-Hitz: a word frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). *Behavior Research Methods*, 38(4), 610–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17393832>.
- Plisson, J., Lavarac, N., & Mladenec, D. (2004). A Rule based Approach to Word Lemmatization. In *Proceedings of the 7th International Multi-Conference Information Society IS 2004* (pp. 83–86).
- Sharma Grover, A., van Huyssteen, G. B., & Pretorius, M. W. (2011). South African human language technologies audit. *Language Resources and Evaluation*, 45(3), 271–288. doi:10.1007/s10579-011-9151-2.
- South African Department of Arts and Culture, & Centre for Text Technology (CTexT) North West University South Africa. (2013). isiXhosa NCHLT Annotated Text Corpora. Retrieved from <http://rma.nwu.ac.za/index.php/isixhosa-nchlt-annotated-text-corpora.html>.
- Statistics South Africa. (2012). *Census 2011: Census in brief*. Retrieved from http://www.statssa.gov.za/Census2011/Products/Census_2011_Census_in_brief.pdf.

An analysis of defining principles with special reference to *Isichazamazwi SesiNdebele* (ISN) definitions

Eventhough NDLOVU (evennthough@yahoo.co.uk)

Department of Language Practice and Linguistics, University of the Free State

Dictionary making in Zimbabwean Ndebele is still at its infancy. *Isichazamazwi SesiNdebele* (henceforth the ISN) is the first monolingual general-purpose dictionary in the language. Given this background, a lot of work lies ahead in terms of dictionary making in Ndebele and reviews of this pioneering work will go a long way in improving the compilation of future Ndebele dictionaries, particularly with regard to user-friendliness, user-needs, user-perspectives, accessibility, acceptability, cultural acceptability. Adherence to the essential principles of defining and good lexicographic practice ensures the formulation of definitions/paraphrase of meaning which meet the afore-mentioned criteria and saves the much needed space since no dictionary is spared the necessity of saving space (Zgusta 1971; Svensén 1993; Landau 2001; Gouws & Prinsloo 2005). According to Landau (2001:157), “occasionally good lexicographic practice must be compromised, either to save space or for some other compelling reasons, but a few basic principles must never be violated, else they defeat the whole purpose of the dictionary”. This article analyses defining principles with special reference to ISN in order to examine the extent to which these principles and practices were adhered to and suggest possible ways of improving *ISN* definitions/paraphrase of meaning to ensure the compilation of future dictionaries that meet the afore-mentioned criteria. In this article, ISN is used as the primary source of data. Data for this article was gathered through an in-depth analysis of selected ISN definitions/paraphrase of meaning, questionnaires and semi-structured interviews with the ISN compilers and users. This was done to gather compilers’ and users’ perspectives on ISN definitions/paraphrase of meaning in view of their adherence to the essential principles of defining and good lexicographic practice as well as the extent to which ISN definitions/paraphrase of meaning satisfy the criteria of dictionary user-friendliness, accessibility, acceptability, cultural acceptability, user-needs and user-perspectives. The key participants of this study were Ordinary and Advanced Level Ndebele majors in two selected secondary schools in Gwanda district, 2013 and 2014 BAA, BA Honours, BA Dual Honours Ndebele and Linguistics majors at the University of Zimbabwe and Ndebele lecturers in the Department of African Languages and Literature at the University of Zimbabwe. Data for this study shows that some ISN definitions/paraphrase of meaning compromise the essential principles of defining, namely the principle of defining the entry, addressing the question, “what is it?,” directly and immediately and avoiding circularity as well as ensuring cultural acceptability. The use of synonym definitions was seen as the major source of the failure to adhere to the principle of defining, addressing the question, “what is it?,” directly and immediately and avoiding circularity. Coupled with this was the observation that, where synonym definitions were used, particularly in defining taboo or offensive words, this led to the formulation of definitions which compromise the afore-mentioned defining principles. In essence, this resulted in a dictionary that is culturally unacceptable, inaccessible and less user-friendly as well as failing to meet the criteria of user-needs and user-perspectives. It was also noted that there are some definitions/paraphrase of meaning which do not correspond to the part of speech of the word being defined and some entries are defined using more difficult words than the entry being defined. It is also argued in this article that Landau’s (2001:157) classification of priority of essence and avoidance of ambiguity as good lexicographic practice needs to be reviewed because it was observed that compromising these two result in the effective violation of some of the principles of defining. It is, therefore, the view of the researcher that it is either these two are classified as defining principles or it is stressed that just like defining principles, good lexicographic practice

should not be compromised, especially given the intricate relationship between some practices and principles.

References

- Hadebe, S. et.al. 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Gouws, R. H. & Prinsloo, D. J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN MeDIA.
- Landau, S. I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Svensén, B. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton; Prague: Academia.

Cross-referencing in *Isichazamazwi SezoMculo* (ISM)

Eventhough NDLOVU (evennthought@yahoo.co.uk)

Menelisi BHEBHE (maziwamo@gmail.com)

Department of African Languages and Literature, University of Zimbabwe

This article examines cross-referencing in *Isichazamazwi SezoMculo* (henceforth the *ISM*). The *ISM* is the first Zimbabwean Ndebele specialised musical terms dictionary. In light of this, a lot of work lies ahead in terms of dictionary making in Zimbabwean Ndebele and reviews of such pioneering works will go a long way in improving the compilation of future Zimbabwean Ndebele dictionaries, particularly in view of the user-oriented approach to dictionary making. The paper focuses on how cross-referencing was employed to save space and to enhance the microstructure, mediostructure and macrostructure of *ISM*. Cross-referencing in this study is examined in view of headword selection, definitions and cross-referencing of synonyms, variants, illustrations and use of examples. According to Gouws and Prinsloo (2005:177), cross-referencing is a lexicographic device that is used to establish relations between different components of a dictionary, save the much needed space in the dictionary, interconnect the knowledge elements represented in different sectors of the dictionary levels of lexicographic description to form a network and to enhance acceptability of the dictionary by including entries from other varieties of the language as variants and/ or synonyms. The following pitfalls of cross-referencing have been identified; dead cross-referencing, failure to utilise cross-referencing where needed, cross-referencing to the wrong address, cross-references that misguide the user in view of information retrieval and use of cross-referencing to avoid a full treatment of the lemma. The paper examines the use of cross-referencing in *ISM* in relation to dictionary user-friendliness, accessibility, user-needs and user-perspectives. Through the findings of this paper, the researchers hope to suggest possible ways of improving the employment of cross-referencing since dictionary making in Zimbabwean Ndebele is still at its infancy. The key participants of the study were Ordinary and Advanced Level Ndebele majors in two selected secondary schools in Gwanda district, 2013 and 2014 BAA, BA Honours, BA Dual Honours Ndebele and Linguistics students at the University of Zimbabwe and Ndebele lecturers in the Department of African Languages and Literature at the University of Zimbabwe. An in-depth analysis of *ISM* and consultation of *ISM* compilers and users was done to gather compilers' and users' perspectives on *ISM*'s employment of cross-referencing in view of how it enhanced dictionary user-friendliness, accessibility and acceptability as well as how it facilitated the satisfaction of user-needs and user-perspectives. A critical analysis of *ISM* and data gathered through interaction with the participants of the study reveals that in some cases cross-referencing was not effectively

employed and this led to wastage of the much needed space in the dictionary and inevitably compromised the dictionary's user-friendliness and accessibility which, in turn, resulted in the failure to satisfy user-needs and user-perspectives. It was observed that in ISM there are cases of dead cross-referencing and unnecessary repetition of definitions of cross-referenced lemmata, which all have a negative bearing on the dictionary's user-friendliness and accessibility and wasted the much needed space since no dictionary is spared the necessity of saving space. An in-depth analysis of ISM and interaction with the participants of the study showed that, to some extent, cases of dead cross-referencing and repetition of definitions can be attributed to lack of thorough editing of the dictionary among other things. This paper stresses the need for compilers to establish from the onset how cross-referencing will be employed in order to ensure consistency in its use and to save the much needed space as well as to enhance the dictionary's user-friendliness and accessibility. It underscores the need for thorough editing to minimise cases of dead cross-referencing and unnecessary repetition of definitions.

References

- Gouws, R.H. & Prinsloo, D.J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN MeDIA.
- Nkomo, D. & Moyo, N. 2006. *Isichazamazwi SezoMculo*. Harare: Mambo Press.

From common conversations around the *Oxford Bilingual School Dictionary: IsiXhosa-English* to constructive dictionary criticism

Dion NKOMO (d.nkomo@ru.ac.za)

School of Languages, Rhodes University

Hartmann (2001: 56) makes a valid point regarding “the need to place lexicographic criticism ... on a more objective footing”. The main purpose of fulfilling this need is not only producing criticism that evens out the achievements and limitations of lexicographic products. As it will be argued in this paper, the basic objective of dictionary criticism is two-fold. Firstly, it should benefit the target users especially of a new dictionary by drawing them to its specific features. Secondly, it should inform the lexicographers who may have to revise the dictionary or produce similar dictionaries in the future. Its endeavour should have a transformative and futuristic dimension, improving lexicographic practice and dictionary use. Unfortunately, this is not always the case, especially in African communities where the interaction between practical lexicography and metalexicography remains limited. With an infant dictionary culture, dictionary criticism tends to be dismissive of new dictionaries.

The present paper has two parts. Firstly, it attempts a metacriticism of the largely negative opinions that were expressed regarding the *Oxford Bilingual School Dictionary: IsiXhosa-English* (henceforth the OXSD) following its publication. The dictionary was published towards the end of 2014 by Oxford University Press (OUP) – Southern Africa as part of a series of bilingual school dictionaries which pair English and other official South African languages. The series already has more or less similar dictionaries for Afrikaans, Sesotho and isiZulu. The publication of the OXSD generated passionate, albeit common, conversations among members of the isiXhosa speaking community, journalists, academics and the publisher. These conversations were recorded in the form of excerpts from speeches by special guests at the launch and expert interviews, dictionary reviews and statements by the publisher, all of which appeared in newspapers, on radio and social media. Since these were remarks by largely individuals without sound lexicographic grounding (those of the

publisher being an exception), they are considered common conversations. However, they are critical and cannot be ignored as they inform public opinion about the dictionary. For the purposes of this paper, the presenter collected the newspaper articles, speeches and the publisher's press statements which contain these conversations. From the conversations, both celebratory and protesting views can be discerned from the community which considers itself representative of the target users regarding the dictionary. However, the general sentiment appeared to be so negative that the publisher was compelled to offer press statements in response to fierce criticism of the dictionary. Although the celebratory voices are not ignored nor accepted as informed, this paper examines the critical concerns and attempts to explain how most of those conversations fall short of fair and useful criticism that benefits the isiXhosa speaking community in general and the target users in particular, i.e. teachers and learners of isiXhosa. The principles of dictionary criticism proffered and modelled by scholars such as Landau (1984), Hartmann (2001), Hadebe (2004) and Swanepoel (2008) are applied for theoretical engagement with such conversations in the analysis. This is hoped to put the criticism of the OXSD into perspective.

Applying the same principles alluded to above, the second part of the paper offers a more rigorous evaluation of the same dictionary as an alternative form of dictionary criticism of the OXSD. The analysis of the dictionary is contextualised within the lexicographic situation of isiXhosa and South Africa in general, which is characterised by absence or a limited range of dictionaries, especially school dictionaries, in African languages, and a very patchy dictionary culture. The functions of the dictionary are particularly brought to the fore in order to interrogate some inclusion/exclusion policies and structural aspects of the dictionary, some of which came under scrutiny in the aftermath of the publication of the dictionary. Regarding inclusion/exclusion, loan words in the OXSD will be particularly discussed as they were topical in the early discussions about the dictionary. The word-based lemmatisation method, which breaks the tradition of arranging lemmata using word-stems (particularly nouns) in isiXhosa dictionaries will also be dealt with. The contents and functions of outer texts will be discussed as a form of integrating dictionary pedagogy for a target user group that is part of a community with a poor dictionary culture. Far from the negative criticism that dominated newspapers, radio and social media, the presentation concludes that the OXSD is a major contribution to isiXhosa lexicography, the learning of isiXhosa and English as additional language learners. Notwithstanding some imperfections, the dictionary should ultimately be judged on its potential and actual success or failure to satisfy user needs in specific situations of dictionary use.

References

- De Schryver, G.-M. et al. (Eds.). 2014. *Oxford Bilingual School Dictionary: IsiXhosa and English*. Cape Town: Oxford University Press Southern Africa.
- Hadebe, S. Improving Dictionary Criticism in Ndebele. Unpublished Paper. University of Zimbabwe
http://ir.uz.ac.zw/jspui/bitstream/10646/28/4/Hadebe_Dictionary_Criticism.pdf.txt).
- Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Landau, S. I. 1984. *Dictionaries. The Art and Craft of Lexicography*. New York/Cambridge: Cambridge University Press.

Swanepoel, P. 2008. Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria. *Lexikos* 18: 207-231.

Reflections on talking dictionaries in Zimbabwe's indigenous languages: Can we really talk?

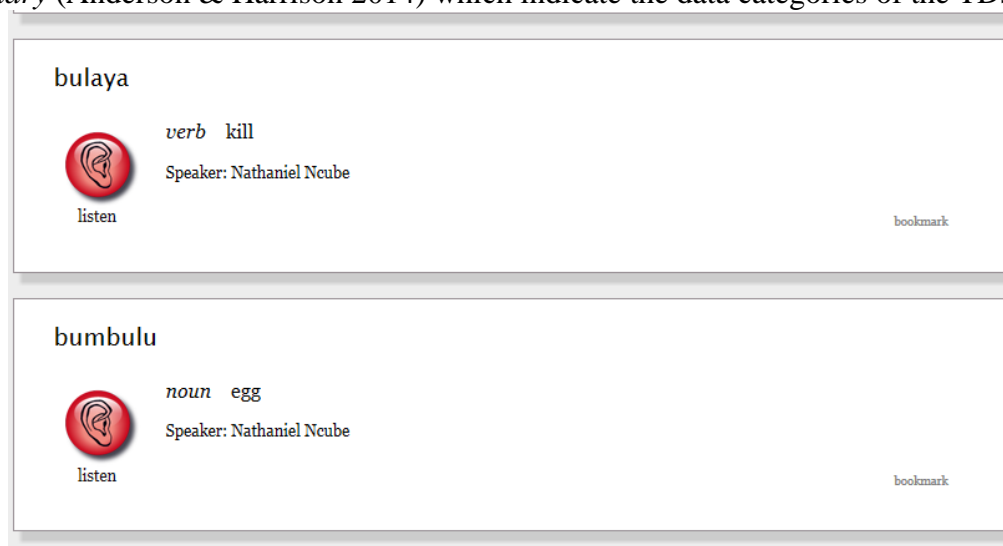
Dion NKOMO (d.nkomo@ru.ac.za)

School of Languages, Rhodes University

Emmanuel CHABATA (emmanuelchabata@yahoo.com)

African Languages Research Institute, University of Zimbabwe

The Enduring Voices Project, spearheaded by the Living Tongues Institute for Endangered Languages, with the support of the National Geographic Society and Swarthmore College, is dedicated to the documentation of the world's endangered languages. The Institute's researchers (linguists) collaborate with mother tongue speakers in the concerned linguistic communities. The language documentation initiative includes the compilations of internet dictionaries called *talking dictionaries* (TDs) for over 80 languages in 15 countries, including Zimbabwe (<http://livingtongues.org/talking-dictionaries/>). In Zimbabwe, TDs are now available for Chikunda, Kalanga, Nambya, Ndaou, Ndebele, Shona and Tshwawo languages. Of these languages, Ndebele and Shona already have general monolingual dictionaries and specialised dictionaries produced during the African Languages Lexical (ALLEX) Project. The proposed presentation seeks to critically reflect on the TDs with a particular focus on those covering Zimbabwean languages. It engages the notion of TDs; why are the dictionaries called as such? In this presentation, the TDs are understood in two ways. The first and more obvious one has to do with the pre-recorded sound clips of lemmata which provide pronunciation guidance to the user. In other words, these dictionaries do talk. The following screenshot is a structural/technical illustration of dictionary articles in the *Kalanga Talking Dictionary* (Anderson & Harrison 2014) which indicate the data categories of the TDs.



However, there is the functional dimension that is linked to the structural/technical dimension of the TDs, i.e. the dictionaries are supposed to support the oral text production function in the respective languages. This is the dimension that is elevated in the presentation as it considers a dictionary as a utility tool with a genuine purpose. On its home page, the Living Tongues Institute for Endangered Languages claims that TD is bilingual “so that speakers of the local dominant language can easily use it, serving as “a resource to help fluent speakers teach words and phrases in their native language to a new generation of speakers” (<http://livingtongues.org/talking-dictionaries/>). Assuming that the TDs are meant to assist

their users to speak, a question is raised on whether or not the ‘talking’ in the TDs refers to the dictionaries’ ability to talk through the sound clips without this being translated into the user’s ability to learn how to speak. A concern is also raised on whether or not the products would be worthy to be called dictionaries if they do not assist the user to speak.

Having elevated the functional dimension in the understanding of TDs, the paper engages with a number of critical questions around the elements of the function theory of lexicography (cf. Tarp 2008), viz. target users, user-situations, lexicographically-relevant problems, lexicographic data and dictionary assistance (lexicographic functions). Around these elements, the following questions are raised and addressed:

- Who does the talking (primary target users)?
- In what situations are they talking (and with who)?
- In the relevant situation(s), what problems would prompt the consultation of a talking dictionary?
- What types of information can the target users retrieve from data types contained in the TDs?
- Is the data sufficient to help the target users talk?

Although the dictionary compilers caution that users should not expect to speak the languages fluently after using the TDs, one would at least expect the user to have obtained sufficient information for basic conversations, if the dictionaries really have any functional value. Assuming positions of the target users of the Zimbabwean TDs, the presenters ask an overall question: Can we really talk?

The Zimbabwean case study is contextualised within the country’s new constitution which promotes multilingualism in which minority language speakers’ rights to use their languages are affirmed while learning languages other than one’s own, is also encouraged (Government of Zimbabwe 2013). Against the background of existing dictionaries in Ndebele and Shona on the one hand and no TDs’ predecessors in the rest of the languages, TDs represent very limited efforts that fail to maximise on previous linguistic research and lexicographic achievements in the respective languages. Their scope (a coverage of less than 200 lemmata) in some languages and limited lexicographic treatment (see screen shot above) mean that the talking dictionaries cannot serve the languages as their repositories, unless more data is added, and neither can they support the oral text production function, which requires much more than pronunciation of a limited number of words. However, the TDs do illustrate the opportunities that are brought by internet in the lexicography of under-resourced languages such as the indigenous languages of Africa. Potential additional data categories and other suggestions are proposed for the TDs.

References

- Anderson, G.D.S. & Harrison, K.D. 2014. *Kalanga Talking Dictionary*. Living Tongues Institute for Endangered Languages. <http://www.talkingdictionary.org/kalanga>.
- Government of Zimbabwe. 2013. *Constitution of Zimbabwe*. Harare: Government Printers.
- Living Tongues Institute for Endangered Languages. <http://livingtongues.org/talking-dictionaries/>.
- Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non knowledge: General Lexicographical Theory with Particular Focus on Learner’s Lexicography*. Tübingen: Max Niemeyer.

The intellectualization of indigenous languages in South Africa through terminology development and dictionary making: A case study of IsiZulu Language at University of KwaZulu-Natal

Siyabonga NKONTWANA (Nkontwana@ukzn.ac.za)

University Language Planning and Development Office, University of KwaZulu-Natal

This paper discusses the development of indigenous languages in South Africa. There are a lot of stages involved in the process of developing languages, which requires careful language planning, and the processes involve the acquisition planning, status planning and corpus planning stages. This paper critically evaluates these stages through looking at the Higher Education Institutions, particularly the development of isiZulu at the University of KwaZulu-Natal. The literature reveals that Africa is rich in indigenous languages hence the number of languages per country ranges from three in Swaziland to over 200 in the Democratic Republic of Congo. Even though this continent is rich in languages most of its countries utilise the languages of the erstwhile colonial master as official languages. To this end English is used in Botswana, Lesotho, Malawi, Namibia, Swaziland, Zambia and Zimbabwe, while French is used in the Democratic Republic of Congo, and Portuguese in Angola and Mozambique, to name but a few.

South Africa has about 25 languages. Among those eleven have been accorded, through the constitution (1996), the official status including English and Afrikaans. However English and Afrikaans still serve as the only *de facto* official languages in South Africa. The University of KwaZulu-Natal has crafted a language policy and plan. The language policy and plan was first approved in 2006 and has, in 2014, been revised and approved by its council. The question that arises is whether or not the UKZN language policy and plan has been embraced by the university community, including students, staff members and the management? Whether or not the language policy and plan has been properly implemented? This paper therefore wishes to *inter alia* answer these questions. At the center of UKZN's language policy is the development of isiZulu language to be at par with the English language. This paper also looks into how UKZN is planning to achieve this policy obligation.

This paper uses the Language Management Theory as a fundamental basis for its discussions, particularly the language policy and planning of isiZulu at UKZN. The paper looks at how the policy was developed, who were the role players and how the policy is monitored and evaluated. This critical evaluation will be done using the apparatus of the Language Management Theory.

The methodology used by this paper is the mixed method whereby both qualitative and quantitative approaches are used to complement each other. The population study for this paper consists of UKZN students, staff members and the management who will be interviewed. The paper will also draw the quantitative data from the official university documentation, reports and archived information.

The paper will explain the process of isiZulu intellectualization at UKZN and its expected products. This process includes isiZulu terminology development wherein standardized terminology will be utilized to create bilingual dictionaries which are the final stage of the process. At the end of this paper it should emerge that UKZN has a solid policy based on a sound theoretical basis, which serves as a vehicle to accomplish the full intellectualization of isiZulu be used at par with English as a language of administration, teaching, science and innovation. The paper will finally argue that UKZN's language policy model be utilized as a reference for other institutions of higher learning wanting to develop indigenous languages of South Africa.

From *Idioticon* to ANNA: The role of Stellenbosch University in the development of Afrikaans lexicography

Gerda B. ODENDAAL (godendaal@sun.ac.za)

Department of Afrikaans and Dutch, Stellenbosch University

In many respects Stellenbosch University can be regarded as the cradle of Afrikaans lexicography. It is home to one of most comprehensive and lengthy lexicographic projects in Afrikaans, the *Woordeboek van die Afrikaanse Taal* (Dictionary of the Afrikaans Language). Being one of the first university departments in South Africa to appoint a Chair for Afrikaans, the Department of Afrikaans and Dutch especially played an important role in developing Afrikaans lexicography to an international standard. But although it was a lecturer of Dutch who first alluded to the fact that Afrikaans is distinct from Dutch through his lexicographic work on Afrikaans, everyone was not convinced from the outset that Afrikaans could or even should make its mark in the field of lexicography. This gave rise to great contention between Afrikaans and Dutch during the early years of the Afrikaans lexicography. It resulted, amongst other things, in lexicographers often relying heavily on Dutch in the lexicographical embodiment of Afrikaans. Through the contribution of the Department of Afrikaans and Dutch in the field of practical as well as theoretical lexicography, however, Afrikaans lexicography gradually freed itself from the influence of Dutch and started to stand on its own feet. Today lexicography is one of the areas of focus in the Department of Afrikaans and Dutch and the Department has a good international profile in the field of lexicography.

This paper on the historical lexicography of Afrikaans examines the role of Stellenbosch University, and specifically the Department of Afrikaans and Dutch, in the development of Afrikaans lexicography. Firstly, this paper shows why the dictionary *Proeve van een Kaapsch-Hollandsch Idioticon met Toelichtingen en Opmerkingen betreffende Land, Volk en Taal* by the Dutch lecturer at Stellenbosch, Nicolaas Mansvelt, is regarded as the beginning of Afrikaans lexicography. This is followed by an overview of the initial struggle in the Afrikaans speech community between proponents of Afrikaans and Dutch. On the one hand there was a pro-Dutch movement, led by Stellenbosch lecturer W.J. Viljoen, which sought to introduce a simplified form of Dutch through the publication of various dictionaries. On the other hand the pro-Afrikaans movement, under the leadership of J.J. Smith, campaigned for Afrikaans to be adopted as medium of instruction at Stellenbosch University. After Smith was appointed as the first lecturer of Afrikaans at Stellenbosch University, it was decided to compile a dictionary for Afrikaans. With the compilation of the *Woordeboek van die Afrikaans Taal* the struggle between Afrikaans and Dutch continued, however, as Smith often stayed too close to the Dutch spelling. This paper goes on to show how Dutch had a far-reaching influence on the development of Afrikaans lexicography during the twentieth century as lexicographers continued to rely strongly on Dutch dictionaries in the compilation of various monolingual, bilingual, special-purpose and other dictionaries for Afrikaans.

Under the influence of various lexicographers with ties to the Department of Afrikaans and Dutch at Stellenbosch University, Afrikaans and Afrikaans lexicography gradually freed itself from Dutch and developed an independent Afrikaans lexicography of international standard, on practical as well as theoretical level. Finally, this article shows how Afrikaans lexicography has come into its own, as symbolised by the publication of the *Groot Woordenboek Afrikaans en Nederlands* (Large Dictionary Afrikaans and Dutch), also known as ANNA (Afrikaans-Dutch, Dutch-Afrikaans). In this bilingual dictionary the two languages are not presented separately, but information is rather brought together or amalgamated, thus emphasising the differences between Afrikaans and Dutch by confronting speakers of Afrikaans with the existence of ‘Dutchisms’ in the language. This shows that Afrikaans

lexicography exists alongside that of Dutch and is recognised internationally as such, as symbolised in the name of the Department of Afrikaans and Dutch.

A critical analysis of multilingual dictionaries

Danie J. PRINSLOO (danie.prinsloo@up.ac.za)

Department of African Languages, University of Pretoria

This paper questions the lexicographic value of multilingual dictionaries. Dictionaries covering three or more languages spoken in South Africa are taken as a case in point. An attempt will be made to reflect on their virtues and shortcomings as reference works and learning tools. Dictionaries such as the *Concise Multilingual Dictionary* (CMD), *Multi-Language Dictionary* (MLD) and *Multilingual Illustrated Dictionary* (MID) will be analysed as typical examples.

Among the merits of such dictionaries count the impressive presentation of specific everyday themes such as ‘my body’, ‘family and friends’, ‘clothing’, ‘my home’, etc. Such a thematic approach naturally reflects appropriate contextualisation of lexical items that belong together which would otherwise have been scattered all over the dictionary if an alphabetical ordering was followed. Furthermore, most of these dictionaries are appealing to the eye, attractive to use and encourage learning and dictionary use. Such dictionaries normally score high marks in terms of comparison between different languages. The user can see at a glance how the translation equivalents in the other languages treated compare to his/her source and target language knowledge. The merits of these dictionaries will be briefly outlined.

In contrast, however, it will be argued that the multilingual model underpinning the compilation of multilingual dictionaries is problematic on many levels resulting in products of inferior lexicographic quality. This is aggravated by various degrees of inconsistency, the lack of cohesion in the treatment of the different languages and even basic or sloppy errors in the dictionaries. Problematic aspects to be explored are (a) poor and inconsistent covering of lexical items, (b) impoverished/skeleton dictionary articles and lack of sense distinction, (c) lack of parity between translation equivalents, (d) absence of examples of usage, (e) incomplete translation equivalent paradigms and (f) risk of misguiding the user through inappropriate text/speech production or communicative guidance.

Poor and inconsistent covering of lexical items is problematic since these dictionaries can hardly be used to look up common or frequently used words most likely to be looked for by their target users.

Impoverished dictionary articles are rooted in the model because in dictionaries treating *seven* languages, there is hardly any space available beyond the listing of a single translation equivalent for each lemma. No comment on form other than the lemma itself or comment on semantics giving at least basic data types such as examples of use, idioms, etc. e.g. are presented. So, for example, treatment of the lemma *interest* in MID is limited to the single-word translations *inzalo*, *phaello*, *inzala*, *morokotso*, *rente* and *tswalo* in isiZulu, Sesotho, isiXhosa, Setswana, Afrikaans and Sepedi respectively.

Lack of sense distinction is for example evident from the treatment of the lemma *with* (prep.) in MLD for Sepedi. Here only the sense of “together with / accompanying” is given and not the equally important sense of “with something, i.e. as an instrument”- thus an instance of an incomplete translation equivalent paradigm.

Lack of parity between translation equivalents of the different languages vary from basic mismatches in terms of register, mismatches in terms of semantic or grammatical paradigms where at least the translations are correct, to mismatches in sense distinction and ultimately to mismatches in homonym distinction. So, for example, there is a mismatch in

terms of senses for the abovementioned lemma *with* between Sepedi and Sesotho where only the sense of “together with” is treated but in the case of Setswana, isiXhosa and isiZulu only the sense “with as an instrument” is treated. More serious mismatches are instances where homonyms are confused in the treatment for the different languages. So, for example, is the lemma might (n) = power/strength in CMD translated for its nominal meaning in isiZulu and isiXhosa as *amandla* and in Sepedi and Sesotho as *maatla* and *matla* respectively but in Setswana as the modal verb *ka* ‘can/may/might’ instead of the nominal meaning *maatla*. The inexperienced user can incorrectly conclude that *ka* is the word for power in Setswana and produce incorrect phrases such as **o tshwanetse go nna le ka* ‘you should have power/strength’.

In CMD circumflexes are only indicated for Setswana. This creates the impression that, e.g. Sepedi and Setswana have the same word for *begin*, but that the pronunciation differs, i.e. *thoma* versus *thôma*. The risk of the user being misguided by the above-mentioned shortcomings and inconsistencies is very high and refutes one of the basic principles of good lexicography, i.e. that the user should be guided by the dictionary not to make mistakes in especially text and speech production.

References

- (CMD) *The Concise Multilingual Dictionary. English, Xhosa, Zulu Northern Sotho, Southern Sotho Tswana, Afrikaans.* Johannesburg: Ad Donker.
- (MID) *Multilingual Illustrated Dictionary. English, IsiZulu Sesotho, IsiXhosa, Setswana, Afrikaans, Sepedi.* Cape Town: Pharos. 2010.
- (MLD) *South African Multi-Language Dictionary and Phrase Book. English, Afrikaans, Northern Sotho, Sesotho, Tswana, Xhosa, Zulu.* Cape Town: reader’s Digest.

Gender Sensitivity in *Isichazamazwi SesiNdebele*

Sanelisiwe SAYI (sayisanelisiwe@gmail.com)

Department of African Languages and Literature, University of Zimbabwe

This paper examines the objectivity or lack thereof in the treatment of gender lexical items in *Isichazamazwi SesiNdebele* (henceforth ISN). The focus will be on the paraphrases of meanings of selected gender-related terms in the dictionary. The ISN is the first and so far only monolingual general purpose dictionary in Zimbabwean Ndebele. Correctness, objectivity and/or neutrality are important qualities of good paraphrases of meaning (cf. Landau 1984; Gouws & Prinsloo 2005). A review of the ISN is relevant in this respect for purposes of informing future dictionary-making in the language.

The data for this paper was gathered through content analysis of selected gender-related lemmata in the ISN, focusing on their paraphrases of meaning. A questionnaire survey was also conducted among potential ISN users who include the following: Ndebele high school students, Ndebele university students, Ndebele teachers and Ndebele university lecturers. Semi-structured interviews were also conducted with some ISN compilers to establish the compiler perspective regarding the treatment of gender related lexical items in the dictionary.

It was found that in the treatment of gender lexical items in ISN, adultery, prostitution and socially-unacceptable behaviours are associated with the female gender even in cases where they would apply to males as well. This was revealed by the content analysis of the paraphrases of meaning of selected terms. For example paraphrases of meanings for lemmata like, *iwule* (prostitute), *izanka* (a frivolous woman), *isifebe* (adulteress), *impumangingene* (a woman of loose morals) associate women while exonerating their male counterparts who

display similar behavioural traits. On the other hand, the male gender lexical items are portrayed with objectivity, if not glorification, in respect of behavioural traits for which women are condemned. It appears that the ISN compilers were lenient and biased when treating gender terms. This is unfortunate given that a dictionary is supposed to be neutral and objective. Users consult a dictionary largely to get meaning information and if the paraphrase of meaning is biased this tends to mislead the dictionary user, hence the dictionary itself becomes less objective. The treatment of gender-related terms in the ISN perpetuates gender stereotyping among the dictionary users who consult it for information and in society in general. The questionnaire survey revealed mixed feelings which are dominated by a disapproval of the manner in which gender-related terms were defined.

This paper argues that derogatory remarks about women are derived from the society and the same biased notions are further perpetuated in dictionaries, whose production is dominated by males (Mawema 2006). Although a corpus was used in the compilation of the ISN, the fact that the corpus was based on mainly male-dominated Ndebele literature implies that the corpus was not balanced in terms of gender. It is proposed that a corroboration of corpus evidence with a linguistic survey, e.g. in the form of questionnaires plus intuition may produce a gender sensitive lexicographical treatment of gender terms in terms defining. The Ndebele corpus that was used as a major source for selection of lemmata was not just created on a neutral basis and might capture some male chauvinistic ideologies that are dominant in African societies such as the Ndebele. In the same vein as de Schryver and Lepota (2001), Hadebe (2002) and Mawema (2006), the present researcher recommends that lexicographers should not only rely on the corpus in selection and explanation of lemmata since it might be gathered from people who are biased or simply who maintain long standing prejudices about certain issues in society. This should also be done in order to produce functional and acceptable dictionaries which have one agenda, of empowering their users by giving unbiased paraphrases of meaning and defying all manner of prejudices associated with certain uses of terms in society.

References

- De Schryver G.-M. and Lepota, B. 2001. The Lexicographic Treatment of Days in Sepedi, or When Mother Tongue Intuition Fails. *Lexikos* 11:1-37
- Gouws, R. H. and Prinsloo, D. J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN Press.
- Hadebe, S. 2002. The Ndebele language corpus: A review of some factors influencing the content of the Corpus. *Lexikos* 12: 159-170
- Hadebe, S. et al. 2001. *Isichazamazwi SesiNdebele*. College Press: Harare.
- Landau, I. 1984 *Dictionaries: The Art and craft of Lexicography*. Cambridge: Cambridge University Press.
- Mawema, M. B. 2006, *Gender Sensitivity in Shona Language Use. A lexicographic and corpus based study of words in context*. ALLEX Project, University of Zimbabwe/University of Oslo.

African Glossonyms: Complexity and Practices in African Dictionaries

Hughes S. NDINGA-KOUMBA-BINZA (22602569@nwu.ac.za)

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

African linguists often point out that it is an ethnoscientific requirement that African languages should be named or addressed to in any other language (i.e. English, French, etc.)

according to the native speakers' intuition of the correctness of their glossonyms. In other words, an African language should be designated according to the way it designates itself.

It is believed that this view seems to have influenced the writers of the current Constitution of the Republic of South Africa (henceforth "the Constitution") in the designation of African languages which were elevated to official status in 1996. Indeed, the Constitution clearly states: "*The official languages of the republic are Sepedi, Sesotho, Setswana, siSwati, Tshivenda, Xitsonga, Afrikaans, English, isiNdebele, isiXhosa and isiZulu*" (Constitution of the RSA, Chapter 1, Article 6).

Thus, the constitutional designation and spelling of African glossonyms are imposed at both administrative and academic levels (except for Sepedi which is now often being designated as *Sesotho sa Leboa* in most recent research papers and academic books). It has also been experienced several times at academic conferences in South Africa that a speaker is corrected for not using the "appropriate" glossonym, despite speaking in English or Afrikaans, while designating an African language of South Africa.

This practice in South Africa towards African glossonyms is progressively being adopted in various Sub-Saharan African countries such as Angola, Botswana, Congo, DR Congo, Gabon, Kenya, Malawi, Mozambique, Namibia, Tanzania, Uganda, Zambia and Zimbabwe. The following examples are noted in the few countries mentioned above:

- (i) Oshiwambo, and no more **Ovambo** or **Wambo** (Angola and Namibia)
- (ii) Kilambya, and no more **Lambya** (Malawi, Zambia and Tanzania)
- (iii) Chichewa, and no more **Chewa** or **Cewa** (Zambia, Malawi and Mozambique)
- (iv) Chinambzya, and no more **Nambya** (Zimbabwe and Botswana)
- (v) Civili, and no more **Vili** (Gabon, Congo and Angola)
- (vi) Kiswahili, and no more **Swahili** (Kenya, Tanzania, Uganda and DR Congo)
- (vii) Lembaama, and no more **Obamba** (Gabon and Congo)
- (viii) Otjiherero, and no more **Herero** (Namibia and Botswana)
- (ix) Sekgalagadi, and no more **Khalahari** (Botswana), and
- (x) Yipunu, and no more **Punu** (Gabon, Congo and DR Congo)

Most of these languages belong to the Bantu phylum of the Niger-Congo family of African languages. It can be observed that the way their glossonyms are required to be addressed to or written imply the grammatical and lexical structures in their respective original forms. For instance, most languages spoken in South Africa have their glossonyms starting with the noun prefix which indicates the noun class system as observed in all Bantu languages.

Meanwhile, the use of glossonyms as promoted by African linguists and as designated in the Constitution fails to account for a number of issues. The following matters can be noted among others:

- (i) the distinction between grammatical categories,
- (ii) the medium used in a particular conversation or discourse,
- (iii) glossonym translation from one language to another, and
- (iv) foreign word adoption (or foreign word nativization) as experienced in most languages of the word

It has been shown that these matters are of great importance when it comes to dictionary compilation in African languages (cf. Drame 2000, Mahlangu 2007 & 2014, Mojela 2010, Ncube 2005, and Nong, de Schryver & Prinsloo. 2002). The present paper points out that the

ethnoscience requirement, which forms part of an anthropological research methodology, is rather more complex than it seems as far as the glossonyms (and not only African glossonyms) are concerned.

This paper seeks the following two aims. First, it pursues to demonstrate the complexity of the ethnoscience requirement within glossonyms in consideration of the issues mentioned earlier, i.e. the identification of grammatical categories, the language used as medium of communication, the designation or translation procedure of glossonyms in a particular language, and foreign word nativization processes.

The second aim is to observe current practices in available African dictionaries (especially bilingual and/or multilingual dictionaries). The main focus will be to identify how the abovementioned issues are dealt with in these dictionaries. Selected dictionaries from various languages will be examined with specific focus on the macro- and microstructures. In the macrostructure, attention will mainly be paid on the lemmatization of glossonyms in terms of the spelling and the morphological structure of the each lemmatized glossonym. In the microstructure, the main interest will consider among others the translation, the definition and the grammatical information of each (lemmatized) glossonym.

This paper should contribute to the glossonym debate in South Africa with practical recommendations.

Preparing an online English-Spanish dictionary of business communication

Sven TARP (st@bcom.au.dk)

Centre for Lexicography, School of Business and Social Sciences, Aarhus University

This paper will discuss the plan for an online English-Spanish dictionary of business communication, a dictionary which will be produced in collaboration between experts and scholars from various countries. It will focus on the philosophy behind the overall dictionary concept, the methods applied to develop this concept, some of the requirements to the corresponding database, as well as the foreseen user interface which may lead the user to 10 completely different articles for each word (string) entered in the search field.

The dictionary project is based on the functional principle combined with the option to individualise, i.e. a lexicographical Model T Ford taking the first modest steps towards a future Rolls Royce according to the classification proposed by Fuertes-Olivera & Tarp (2014). The project is inspired by the experience obtained from a similar project including four printed bilingual dictionaries between Danish, on the one hand, and English and Spanish, on the other, as well as the experience from more than 25 years of teaching foreign-language business communication.

The planned functions of the new dictionary project are L2 text-production, L2 text-reception, L1-L2 translation, L2-L1 translation and L1 text-production for first-language speakers of both English and Spanish, i.e. a totality of 10 lexicographical functions expressed in a combination of monolingual and bilingual lexicographical solutions.

After explaining the general idea of the project, the paper will then continue with a description of the overall method by means of which the lexicographical data to be included in the dictionary are determined. This includes delimitation of the topic to be covered and determination of user needs taking into account relevant user characteristics and extra-lexicographical situation where lexicographical relevant needs may occur.

The paper will then discuss some of the necessary requirements to the underlying database. The overall principle is that the database (and the dictionary as such, understood as the totality of data that may be visualized in the totality of hypothetical consultations) should include as much data as possible relevant to the topic as well as to the detected user needs, whereas the individual articles displayed on the screen should contain as few data as possible

in order to avoid information overload, i.e. the data required to satisfy the user's needs in each case, no more no less. By means of the classical method of analysis-synthesis, the data categories to be covered will first be separated into their smallest relevant parts (giving each of them a number in the database) and then combined in order to establish their relevant mutual relations (which are also reflected in the database). In this respect, the concept of relevance used will also be explained.

Finally, the paper will show the foreseen user interface and provide examples of which data categories may be displayed for each of the 10 planned functions. The idea is that the user, as a rule, does not need to carry out more than two actions (entering a search string and clicking on a specific option) in order to get quick and easy access to the data required in each case. In this connection, the paper will also discuss some of the techniques that may be used to further reduce the amount of data displayed in the first instance in order to avoid information overload and adapt the dictionary to smaller devices like tablets, iPads, mobile phones, etc.

References

- Andersen, Birger, Karin Balsgart, Morten Pilegaard and Sven Tarp 1999: *Dansk-engelsk erhvervsordbog*. Aarhus: Systime.
- Andersen, Birger, Karin Balsgart and Sven Tarp 2006: *Engelsk-dansk. Erhvervsordbog*. Copenhagen: Gyldendal.
- Fuertes-Olivera, Pedro A. and Sven Tarp (2014): *Theory and practice of specialised online dictionaries: Lexicography versus terminology*. Berlin, Boston: De Gruyter.
- Hvid, Virginia, Anne Lise Laursen and Sven Tarp 1999: *Dansk-spansk erhvervsordbog*. Aarhus: Systime.
- Hvid, Virginia, Anne Lise Laursen and Sven Tarp 2003: *Spansk-dansk erhvervsordbog*. Copenhagen: Gads Forlag.

The Use of LSP Dictionaries on Mobile Phones in Higher Education

Michele VAN DER MERWE (michelevdm@sun.ac.za)

Department Curriculum studies: Afrikaans, Faculty of Education, Stellenbosch University

MALL (Mobile Assisted Language Learning) is employed all over the world as technology is advancing. In South Africa the use of mobile phones has increased dramatically and most university students are equipped with mobile phones.

According to Naismith (2004) mobile technologies can be used in the design of six different types of learning:

- Behaviourist learning; where quick feedback or reinforcement can be facilitated by mobile devices because they are at hand;
- Constructivist learning; where learners build up new concepts perhaps engaging with their physical and social environment;
- Situated learning, where learners take a mobile device into an educationally relevant real-world location and learn from that;
- Collaborative learning; where mobile devices are an essential means of communication and electronic information sharing for learners in groups outside their institution;
- Informal and lifelong learning; possibly unconstructed or opportunistic, driven by personal curiosity, chance encounters and the stimulus of the environment.

- Supported learning, where mobile devices monitor progress.

Design of content could be the following: Learner-created content: if students are expected to construct some of the content as part of their learning, this can be done in various locations and mobile devices can facilitate it. Personalized content: learners can receive, assemble, share and carry around personally useful and appropriate resources. Updated content: updates may be more easily delivered to mobile devices. Timed or scheduled content: learners can engage with content frequently, repetitively using a mobile device. Prioritized content: some content can be made available on mobile devices in such a way to prioritize it over other content. Aural content: if listening is important, delivering audio via a personal mobile device can be engaging and convenient. Flexible content: learners may appreciate having the option of mobile access to learning material and resources, as an alternative to desktop content.

A project to act as support for language learning within the Department of Curriculum Studies of the Faculty of Education at the University of Stellenbosch was undertaken. A LSP dictionary, consisting of subject terms to be used in the first year of B.Ed. studies, was designed to be used on a mobile phone. MobiLex was designed and it fitted into the framework of multilingualism at the university.

Motivation for the project lies in the following: Investigations into possible reasons for low throughput rates at universities have revealed a low level of competency in the language of teaching and learning as one of the main reasons for the lack of academic success among SA undergraduate students. (Leibowitz, 2001). Many students at entrance level at university have not reached an adequate level of academic literacy. Academic literacy could be defined as “the reading, writing, speaking and thinking skills, dispositions, and habits of mind that students need for academic success.” (Warshauer et al, 2004) Academic support for first year students is essential for success at university level. As Read and Ambrose (1999) indicate, the key to accessibility in academic subjects is vocabulary and it is on that basis that the commonly-used academic and word lists by Nation & Coxhead were developed. As Pam Peters indicated in 2008: “Although there are technical glossaries on the internet, their ad hoc definitions do not target the needs of novices in the discipline, or those with limited English.”

A study was conducted on the perceptions of B.Ed. 1 students on the use of dictionaries on mobile phones. Research was done via the method of a questionnaire and empirical data was collected. Ethical clearance was obtained of the university to conduct the study. A total of 80 students responded to the questionnaire of those 51 were Afrikaans speaking students, 6 Xhosa speakers and 23 English speakers. An overview of the questionnaire will be presented during the presentation and noteworthy results regarding the perceptions on LSP dictionaries will be discussed. An overwhelming majority of students indicated that they would make use of a dictionary on a mobile phone and they also confirmed the need for such a dictionary on a mobile device. Students were asked to indicate subjects for which they needed support for vocabulary and their answers supported the notion that subjects they had no prior knowledge of and with which they were confronted with for the first time in their academic career, scored highly on the needs’ list.

Interesting answers were supplied on the question of information categories, as students expressed the need for certain information categories that lexicographers would not normally associate with LSP dictionaries.

References

Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.

- Leibowitz, J. 2001. Students' prior learning and their acquisition of academic literacy at a multilingual South African university. Unpublished doctoral dissertation. Sheffield: University of Sheffield.
- Peters, P. 2008. Termfinder: creating online termbanks of technical terms for early university study. *Journal of Applied Linguistics* 3.2: 219-248.
- Warshauer, M., Grant, D., Del Real, G. & Rousseau, M. 2004. Promoting academic literacy with technology: successful laptop programmes in k-12 schools. *System* 32: 525-537.

Adaptation of *A Dictionary of South African English on Historical Principles* for electronic platforms

Tim VAN NIEKERK

Dictionary Unit for South African English, Rhodes University

In July 2014, the Dictionary Unit for South African English (DSAE) published a pilot version of its flagship dictionary *A Dictionary of South African English on Historical Principles* (Oxford University Press, 1996) online at dsae.co.za. With limited resources, the aim of this project was to make the out-of-print dictionary accessible with basic functionality. In 2015, a collaboration between the DSAE and the University of Hildesheim will begin to address the need for a thorough adaptation of the dictionary text to support publication on multiple electronic platforms, in the process reviewing the design of the dictionary and incorporating current e-lexicography research. This session is intended as a precursor to the research project, with a view to obtaining informed input from practicing lexicographers and all interested participants attending the AFRILEX conference. After a brief initial presentation contextualising the dictionary and providing an overview of the kinds of adaptations envisaged (see summary document provided in conference pack) the remainder of the session will be open for suggestions, comments and discussion.

To Include or Not: Lexicographic Recording of Neologisms in the Digital Era

Gao YONGWEI (ywgao@fudan.edu.cn)

Fudan University, China

It was conservatively estimated that the number of new words that appear in the English language each year reached 800 (Landau 1989). But as the estimation was made in the pre-Internet days, this number was no doubt underestimated. Thanks to technological advances in recent decades and the spread of the Internet in particular, new words are coined more often than before, and they are circulated much easier and more widely. The fact that the editorial team of *Oxford English Dictionary* includes approximately 2000 new words each year through quarterly updates is a potent indication that the English language now boasts more new words than before. What's more, the data released by Texas-based Global Language Monitor even suggest that the language is increasing at a rate of thousands of lexical additions a year, and as this figure may include ephemeral lexical items and even proper names of various kinds, no serious lexicographer will give much credence to it. No matter what the number is actually is, there is broad consensus among dictionary-makers that they now have more new-word entries to write. With people's increasing awareness of and interest in new lexical kids on the block, dictionary editors in English-speaking countries seem to be sparing no efforts in recording as many neologisms as possible in their dictionaries, and the advertising of some select new terms in the blurbs or through other means (e.g. newspaper articles, blogs) has become the norm.

What kinds of neologisms should lexicographers record in their dictionaries? Allan Metcalf (2002) suggests the FUDGE factors in identifying new words, namely frequency of use, unobtrusiveness, diversity of users and situation, generation of other forms and meanings, endurance of the concept. In reality, although an unpredictable mixture of these factors has been used to decide which terms that might be able to stand the test of time, different dictionary editors may rely on their own set of criteria, one of which is frequency of use which is usually indicated with the help of their respective corpus or corpora. For example, editors with Oxford Dictionaries have recourse to the Oxford English Corpus and the Oxford Reading Programme. But if the criteria for inclusion adopted by dictionary editors are put under the microscope, we may be able to find discrepancies and even problems, and this is what the author of this paper is going to do. The 12th edition of *Collins English Dictionary* (2014) will be used as the source of new words to be discussed in this paper. Firstly, the author will discuss the different categories of new words that have been covered in the Collins dictionary, and such categories will be arranged first by subject field and then by word-formation process. Secondly, the paper will make a special effort in discussing the new senses that have been added, and the different processes through which these new senses were acquired will be touched on as well. Thirdly, the major problems with Collins' inclusion of neologisms will be pored over, and such problems may be summarized as the following: the inclusion of short-lived words (e.g. *cyberathletics*, *spimming*); the inclusion of compounds whose meanings are easily understood (e.g. *capped-rate*, *regulatory risk*, and *safe surfing*); the different treatment of words of the same category (e.g. *cyber mosque*); imbalanced inclusion of neologistic antonyms or derivatives (e.g. the absence of *cyberphobe*), the failure in including some of the latest words (e.g. *McTwist*, *misery lit*, and *mummy porn*), etc. Finally, the author, as a bilingual dictionary-maker, will suggest a practical method of identifying candidates for new-word entries based on cross-checking the currency of such words in news archives and/or news aggregation websites.

References

- Algeo, John. *Fifty Years "Among the New Words": A Dictionary of Neologisms, 1941-1991*. Cambridge: Cambridge University Press. 1991.
- Ayto, J. 2006 *Movers and Shakers: A Chronology of Words That Shaped Our Age*. Oxford: Oxford University Press.
- Barrett, G. *The Official Dictionary of Unofficial English*. New York: McGraw-Hill. 2006.
- Collins English Dictionary*. 2014. 12th Edition. London: Harper Collins Publishers.
- Crystal, D. 2004. *A Glossary of Netspeak and Textspeak*. Edinburgh: Edinburgh University.
- Dent, S. 2004. *Larpers and Shroomers: The Language Report*. Oxford: Oxford University Press.
- Dent, S. 2005. *Fanboys and Overdogs: The Language Report*. Oxford University Press.
- Dent, S. 2008. *Words of the Year*. Oxford: Oxford University Press.
- Hargraves, O. 2004. *New Words*. Oxford University Press.
- Landau, S. I. 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- McFedries, Paul. *Word Spy: The Word Lover's Guide to Modern Culture*. New York: Broadway Books.
- McKean, E. 2003. *The Oxford Essential Dictionary of New Words*. Berkley: Berkley Publishing Group.
- Metcalf, A. 2002. *Predicting New Words: The Secrets of Their Success*. Boston: Houghton Mifflin Company.
- Ostler, R. 2003. *Dewdroppers, Waldos, and Slackers: A Decade-by-Decade Guide to the Vanishing Vocabulary of the Twentieth Century*. Oxford University Press.

Peckham, Aaron. 2007. *Mo' Urban Dictionary*. Kansas City: Andres McMeel Publishing.

