

AFRICAN ASSOCIATION FOR LEXICOGRAPHY

23rd International Conference

27-29 June 2018



**University of the Western Cape,
Cape Town, South Africa**



AFRICAN ASSOCIATION FOR LEXICOGRAPHY

23rd International Conference

**University of the Western Cape,
Cape Town, South Africa
27-29 June 2018**

Abstracts and programme

Hosted by: Department of Language Education, University of the Western Cape, South Africa

Conference coordinator: Dr Hughes Steve Ndinga-Koumba-Binza

Abstract reviewers: Prof. Herman Beyer, Prof. Sonja Bosch, Mrs Charlene de Kock, Mr André du Plessis, Prof. Rufus Gouws, Dr Langa Khumalo, Dr Phillip Louw, Prof. Paul Mavoungou, Mrs Lorna Morris, Dr Gerda Odendaal, Dr Eventhough Ndlovu, Prof. Dion Nkomo, Prof. Danie J. Prinsloo, Mr Marius Swart, Prof. Elsabé Taljard, Dr Michele van der Merwe, Mr Tim van Niekerk.

Abstract booklet editors: Mr André H. du Plessis & Prof. Sonja E. Bosch

© 2018 AFRILEX, African Association for Lexicography
ISBN 978-0-6399494-0-6

Table of Contents

AFRILEX HONORARY MEMBERS	5
AFRILEX BOARD	6
MESSAGE FROM THE AFRILEX PRESIDENT.....	6
CONFERENCE PROGRAMME	8
KEYNOTE PRESENTATION 1	
What does the dictionary have to offer in terms of cultural intelligence, lexical priming, and world Englishes?	
Vincent B Y OOI.....	11
KEYNOTE PRESENTATION 2	
The long and the short of a comprehensive dictionary as a long term project	
Willem BOTHA.....	13
PARALLEL SESSIONS.....	14
Macrostructural Features for a French-Fang School Dictionary	
Blanche Nyangone ASSAM.....	14
Aranea Africana: Large Corpora of African Varieties of English and French	
Vladimír BENKO & Anna BUTAŠOVÁ.....	15
Innovations in the access structures of online and mobile dictionaries	
André H. DU PLESSIS.....	17
Cross-Language Dictionary Alignment for Bantu Languages	
Thomas ECKART, Dirk GOLDHAHN, Uwe QUASTHOFF & Sonja BOSCH.....	18
Usage examples in a multilingual online term bank: why and how?	
Michelle GOOSEN & Elsabé TALJARD	20
The data distribution structure and data access in online dictionaries	
Rufus H. GOUWS.....	21
Dictionary posters in !Xun and Khwedam: A learning experience	
Kerry JONES	22
Hedges in definitions: A historical look at English dictionaries	
Mariusz KAMIŃSKI	24
“Boys will be boys:” an example of biased usage	
Steven M. KAPLAN.....	25
Towards an isiZulu National Corpus	
Langa KHUMALO	26
UX (User Experience) design applied to electronic lexicography: some guiding principles deployed in the digitisation of <i>A Dictionary of South African English</i>	
Bridgitte LE DU & Tim VAN NIEKERK	29
Does the isiNdebele Terminology developed today have any significant impact?	
Sponono K. MAHLANGU	30
French morphogrammic system and lexicographical treatment of homophony in Gabon: the distinction between the masculine and feminine gender in personal pronouns	
Edgard Maillard ELLA.....	32
Theoretical Perspectives for a Comprehensive Dictionary of Gabonese French	
Paul A. MAVOUNGOU, Virginie OMPOUSSA & Hugues Steve NDINGA-KOUMBA-BINZA.....	34
Do we need dictionary skills in the digital era?	
Lorna Hiles MORRIS.....	35

Lexicography in Zimbabwe: Prospects and challenges post the 2013 Constitution of Zimbabwe Amendment No.20 Act and the Zimbabwean Education Blueprint 2015 – 2022 Curriculum Framework for Primary and Secondary Education	
Eventhough NDLOVU	36
Analysis of the French-English Bilingual dictionaries in use at the Junior Secondary Schools in Nigeria	
Rosemary N. OSSAI & Carol C. OPARA	38
The Effects of the Multimedia Package in Teaching Vocabulary within the Nigerian French Language Curriculum	
Rosemary N. OSSAI & Carol C. OPARA	39
Stem lemmatization and phonetic ordering of lemmas in Sotho dictionaries from a user perspective	
Danie J. PRINSLOO & Tello THELETSANE	40
Lemmatization of kinship terminology in Setswana	
Fannie SEBOLELA & Danie J. PRINSLOO	42

AFRILEX HONORARY MEMBERS



Prof. R.H. Gouws



Prof. A.C. Nkabinde



Dr J.C.M.D. du Plessis



Dr M. Alberts

AFRILEX BOARD 2017 – 2019

President:	Prof. H.L. (Herman) Beyer
Vice-President:	Prof. D.J. (Danie) Prinsloo
Secretary:	Prof. D. (Dion) Nkomo
Treasurer:	Prof. E. (Elsabé) Taljard
Editor Lexikos:	Prof. E. (Elsabé) Taljard
Members:	Prof. S.E. (Sonja) Bosch Dr H.S. (Steve) Ndinga-Koumba-Binza Dr L. (Langa) Khumalo Mr A.H. (André) du Plessis
Conference organiser:	Dr H.S. (Steve) Ndinga-Koumba-Binza

MESSAGE FROM THE AFRILEX PRESIDENT

On behalf of the Board of the African Association for Lexicography (AFRILEX), it is my honour and privilege to welcome you to the 23rd Annual International AFRILEX Conference, hosted this year by the University of the Western Cape (UWC) in the iconic city of Cape Town. A special word of welcome is due to all honorary members of AFRILEX present, and to our keynote speakers, Prof. Vincent Ooi from the National University of Singapore, and Dr Willem Botha, Editor-in-Chief of the *Woordeboek van die Afrikaanse Taal*.

This is the second time that AFRILEX holds its annual international conference on the beautiful campus of UWC, the previous occasion having been the 14th annual conference in 2009, when we were warmly welcomed by the Department of Xhosa. On behalf of the entire AFRILEX I would like to sincerely thank the UWC Rector and Vice Chancellor, Prof. Tyrone Pretorius, for allowing us to convene here once again.

It would not have been possible for AFRILEX to have this year's conference in these excellent surroundings if it were not for the kind invitation on behalf of the UWC by AFRILEX Board member Dr H. Steve Ndinga-Koumba-Binza. We would also like to heartily thank the members of Dr Ndinga-Koumba-Binza's interdepartmental team for their dedication and hard work in preparing for this academic gathering. They are Dr Blanche Nyangone Assam (Dept. of Foreign Languages), Mr Hardi Zacharias (Dept. of Institutional Advancement), Ms Zahra February, Ms Molly Naketsana, Ms Someka Ngece, Ms Nonny Shandu-Omukunyi (all from the Dept. of Language Education) and Dr Sebolelo Mokapela (Dept. of Xhosa).

A brief look at the programme reveals the broad spectrum of topics that will once again be dealt with at this conference. They include advances in digital technology for lexicography, issues of text and access, user issues, lexical issues, the framing of language in lexicographical communication, sociolinguistic issues, dictionaries in different media, and dictionaries in language education. Thank you to every local and international paper and session presenter for putting in the time, effort and cost to share your valuable research with us, and for choosing AFRILEX as forum. I would like to take this opportunity to encourage

presenters who have not yet done so to develop their papers into article manuscripts and submit these for peer review to our journal *Lexikos* for possible Gold Open Access publication (Go to: <http://lexikos.journals.ac.za>).

We are grateful to the abstract reviewers credited on the title page of this booklet for their valued service in academic quality assurance.

I would like to thank the Vice President, Prof. Danie Prinsloo, for driving the preparations for this year's conference, for managing our website and for compiling the programme. For ensuring that AFRILEX remains in a financial position to hold its annual conference, we thank our Treasurer, Prof. Elsabé Taljard. Prof. Sonja Bosch and Mr André du Plessis smoothly managed the abstract review process and the editing of this fine booklet – thank you for a sterling job! I also wish to express my appreciation to the rest of the Board, in the persons of Dr Langa Khumalo and Prof. Dion Nkomo (Secretary), for their work and support in the various activities leading up to this conference.

Finally, I must tender my sincerest apologies for my absence from this year's conference. The annual AFRILEX conference is probably the academic highlight of my year, and I can assure you that I regret my inability to attend this year much, much more than AFRILEX does! These circumstances place me even deeper in the debt of our Vice President, Prof. Danie Prinsloo, who will assume the duties and responsibilities of the President during this conference.

I wish one and all a successful and stimulating 23rd Annual International AFRILEX Conference!

Herman L. Beyer
President: AFRILEX



CONFERENCE PROGRAMME

Wednesday 27 June 2018	
Venue: Auditorium, New Life Science Building (NLSB)	
13:00 – 13:55	Registration
14:00 – 14:25	Official Opening
14:25 – 14:30	Tribute to H.E. Wiegand Rufus H. Gouws
Special Sessions	
Venue: Auditorium, New Life Science Building (NLSB)	
14:35 – 15:30	Oxford Global Languages: making lexical content available online for digitally under-represented languages Louw, P.
15:35 – 15:55	Tea
16:00 – 16:55	South African Centre for Digital Language Resources (SADiLaR) De Lange, A., Eiselen, R., Taljard, E. and Bosch, S.E.
17:00 – 17:25	Global association for Lexicography (Globalex) Prinsloo, D.J.
Venue: 1st Floor Foyer, New Life Science Building	
17:30 – Cocktail Party	

Thursday 28 June 2018	
Keynote Address 1: Venue: Auditorium, New Life Science Building (NLSB)	
09:00 – 09:55	Vincent Ooi , Department of English Language and Literature, National University of Singapore What does the dictionary have to offer in terms of cultural intelligence, lexical priming, and world Englishes?
10:00 – 10:25	Tea

Parallel Sessions		
	Venue: NLSB, Seminar Room 1A	Venue: NLSB, Seminar Room 1F
10:30 – 10:55	Do we need dictionary skills in the digital era? Morris, L.	Does the isiNdebele Terminology developed today have any significant impact? Mahlangu, K.S.
11:00 – 11:25	Dictionary posters in !Xun and Khwedam: A learning experience Jones, K.	“Boys will be boys”: an example of biased usage Kaplan, S.
11:30 – 11:55	Cross-Language Dictionary Alignment for Bantu Languages Eckart, T., Goldhahn, D., Quasthoff, U. and Bosch, S.	Innovations in the access structures of online and mobile dictionaries Du Plessis, A.H.
12:00 – 12:25	French morphogrammic system and lexicographical treatment of homophony in Gabon: the distinction between the masculine and feminine gender in personal pronouns Ella, E.M.	Lemmatization of kinship terminology in Setswana Sebolela, F. and Prinsloo D.J.
Venue: Feedem Pitseng Restaurant, New Life Science Building		
12:30 – 13:25	Lunch	
Parallel Sessions (continued)		
	Venue: NLSB, Seminar Room 1A	Venue: NLSB, Seminar Room 1F
13:30 – 13:55	Usage examples in a multilingual online term bank: why and how? Goosen, M and Taljard, E.	Macrostructural Features for a French-Fang School Dictionary Assam, B.
14:00 – 14:25	Towards an isiZulu National Corpus Khumalo, L.	Aranea Africana: Large Corpora of African Varieties of English and French Benko, V. and Butašová, A.
14:30 – 14:55	UX (User Experience) design applied to electronic lexicography: some guiding principles deployed in the digitisation of A Dictionary of South African English Van Niekerk, T. and Le Du, B.	Analysis of the French-English bilingual dictionaries in use at the junior secondary schools in Nigeria Ossai, R.N. and Opara, C.C.
15:00 – 15:25	Tea	
Venue: Auditorium, New Life Science Building		
15:30 – 17:00	Annual General Meeting	
Venue: Feedem Pitseng Restaurant, New Life Science Building		
18:00	Conference Dinner	

Friday 29 June 2018		
Keynote Address 2: Venue: Auditorium, New Life Science Building (NLSB)		
09:00 – 09:55	Willem Botha, Bureau of the WAT The long and the short of a comprehensive dictionary as a long term project	
10:00 – 10:25	Tea	
Parallel Sessions		
	Venue: NLSB, Seminar Room 1A	Venue: NLSB, Seminar Room 1F
10:30 – 10:55	The data distribution structure and data access in online dictionaries Gouws, R.H.	The effects of the multimedia package in teaching vocabulary within the Nigerian French language curriculum. Ossai, R.N. and Opara, C.C.
11:00 – 11:25	Stem lemmatization and phonetic ordering of lemmas in Sotho dictionaries from a user perspective Prinsloo, D.J. and Theletsane, T.	Lexicography in Zimbabwe: Prospects and challenges post the 2013 Constitution of Zimbabwe Amendment No.20 Act and the Zimbabwean Education Blueprint 2015 – 2022 Curriculum Framework for Primary and Secondary Education Ndlovu, E.
11:30 – 11:55	Theoretical Perspectives for a Comprehensive Dictionary of Gabonese French Mavoungou, P., Ompoussa, V. and Ndinga-Koumba-Binza, H.S.	Hedges in definitions: A historical look at English dictionaries Kamiński, M.
Venue: Auditorium, New Life Science Building		
12:00 – 12:25	Closure	
Venue: Feedem Pitseng Restaurant, New Life Science Building		
12:30 – 13:30	Lunch	
Saturday 30 June 2018		
09:00 – 17:00	Post-conference excursion: Sightseeing Tour Bus to Cape Aghulas “Where the 2 oceans meet”	

KEYNOTE PRESENTATION 1

What does the dictionary have to offer in terms of cultural intelligence, lexical priming, and world Englishes?

Vincent B Y OOI (vinceooi@nus.edu.sg)

Department of English Language and Literature, National University of Singapore, Republic of Singapore

The dictionary is a valuable document that mediates between the ‘collective lexicon’ of the wider community of users (i.e. in the case of English, a variety or world varieties of English) and an individual user’s lexicon (i.e. ‘lexical priming’). Going beyond linguistic awareness, it will be a worthwhile dream to have the dictionary be a ‘one-stop’ document that tells people not only the structure and rules of a language but also offer a more complete ‘awareness’ of how to function effectively in situations and avoid cross-cultural miscommunication (i.e. ‘cultural intelligence’). In this talk, I relate these concepts especially to Asian contexts that I have come across and examine their encapsulation in various online dictionaries of English.

If we regard the dictionary as a ‘lexicographic tool’ (a term that seems increasingly favoured—see Ooi 2017), online dictionaries, Wikipedia, terminological knowledge bases and increasingly smartphone Apps can serve as resources for the acquisition of cultural intelligence. ‘Cultural intelligence’ is a business term that subsumes other types of intelligence (metacognitive, cognitive, motivational and behavioural): ‘the more aware individuals are of differences in their cultural environments, the more they will assess these differences and respond accordingly’ (Gooden et al 2017: 228).

As far as linguistic knowledge is concerned, individuals rely on their respective lexicons and are ‘primed’ subconsciously to notice ‘such facts that native speakers make a mental note of when they encounter a word’ (Hoey 2009, 2014). For Hoey, lexical priming means subconsciously noticing such factors as the following: (i) the words that the term occurs with; (ii) the grammatical patterns it occurs in; (iii) its associated meanings; (iv) its usage in terms of politeness vs rudeness, humour vs seriousness; (v) the style it tends to occur in; (vi) the registers that it occurs more often in, (vii) its usage by someone older or younger’ etc.

Hoey’s stipulation of the ‘right phrase at the right context at the right time’ for the user seems a tall order, given that English is arguably the leading global language with the most functional load. The existence of various ‘Englishes’ with unequal statuses (‘Inner’, ‘Outer’, and ‘Expanding Circles’ of English – Kachru 1985) means that there are many potential conflicting lexical primings among countries in which English is regarded as native language (i.e. the UK, U.S, Australia, New Zealand, Canada etc). In addition, ‘Outer Circle’ countries (cf. Kachru 1985) such as Singapore, Malaysia and Hong Kong have formed their own endonormative standards that may conflict with native speaker ones. In these Asian contexts that I am familiar with, there is additionally the presence of diglossia: the H-igh (local educated, formal) variety and the L-ow (informal, colloquial) variety that is known variously as ‘Singlish’, ‘Manglish’ and ‘Hongklish’/‘Konglish’ respectively.

In order to manage this pluralized/pluricentric nature of English and the various lexical primings that exist in a community, I suggest a model of English that I have found useful to delineate the national variety of English (Ooi 2001) vis-à-vis the more international variety, and triangulating the information in a dictionary vis-à-vis a resource such as the GloWbe corpus (Davies and Fuchs 2015). It is important to balance ‘etic’ (outsider) and ‘emic’ (insider) perspectives in the lexicographic endeavor. I also compare the codification of some world English lexicons in terms of varietal and other labels for a number of lexical entries in various print/electronic dictionaries of English nowadays (Ooi 2018).

References

Davies, M, Fuchs, R. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus. *English World-Wide* (36:1), pp.1-28.

- Gooden, DJ, Creque, CA, Chin-Loy, C. The impact of metacognitive, cognitive, and motivational cultural intelligence on behavioral cultural intelligence. *International Business and Economics Research Journal* (16:3), pp.233-230.
- Hoey, M. 2009. Lexical priming. MED (Macmillan English Dictionaries) Magazine, Issue 52. Available: <http://www.macmillandictionaries.com/MEDMagazine/January2009/52-LA-LexicalPriming.htm>. Accessed: 01/05/2018.
- Hoey, Michael. 2014, Words and their neighbours. In J R Taylor (ed) *The Oxford Handbook of the Word*, Oxford: Oxford University Press. DOI:10.1093/oxfordhb/9780199641604.013.39
- Kachru, B. 1985. Standards, codification and sociolinguistic realism: The English language in the Outer Circle, In R. Quirk & H. Widdowson (eds.) *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press, pp.11-30.
- Ooi, VBY. 2001. Globalising Singaporean-Malaysian English in an inclusive learner's dictionary. In B Moore (ed) *Who's Centric Now?: The Present State of Post-Colonial Englishes*, Oxford: Oxford University Press. pp. 95-121.
- Ooi, VBY. 2017. Using the Web for lexicographic purposes. In P. Fuertes-Olivera (ed) *Routledge Handbook of Lexicography*. London: Routledge. pp.684-700
- Ooi, VBY. 2018. Lexicography and World Englishes. In Low, EL and Pakir, A. (eds) *World Englishes: Rethinking Paradigms*. London: Routledge. pp.164-182.
- Pakir, A. (eds) *World Englishes: Rethinking Paradigms*. London: Routledge. pp.164-182.

KEYNOTE PRESENTATION 2

The long and the short of a comprehensive dictionary as a long term project

Willem BOTHA (wfb@sun.ac.za)

Executive Director and Editor-in-Chief, Woordenboek van die Afrikaanse Taal.

There are no short cuts in the compilation of a multi volume overall-descriptive dictionary. The term *multi volume overall-descriptive* is indicative of the fact that most comprehensive dictionaries were started as printed dictionaries and eventually migrated to an electronic version. That epic journey has proved to be painful in many instances and has led to heavy financial losses to the companies undertaking the electronification. Usually the electronification is postponed until the letter Z has been completed, but there is one exception to the rule: the *Woordeboek van die Afrikaanse Taal*.

The ultimate goal of all National Lexicography Units in South Africa is the compilation of a comprehensive defining dictionary. Presently the *Woordeboek van die Afrikaanse Taal* fits into this category of the dictionary typology, but the initial intention was not to compile a comprehensive dictionary. What was envisaged as a standard dictionary modelled on the Dutch *Van Dale Groot Woordenboek der Nederlandse Taal* in two volumes, organically developed into a comprehensive Afrikaans defining dictionary without any clear, recorded decision about a typological deviation.

Time and money are the determining factors in the average compilation time of a comprehensive dictionary of almost a hundred years. The WAT had to reinvent itself thrice in its long history spanning ninety years in order to remain relevant in its time. The first phase of the reinventing started in 1990 when it became clear that it would take more than 200 years to complete the dictionary if drastic measures were not taken to accelerate the project. Over a period of two years the dictionary was completely redesigned involving lexicographers from the theoretical and practical spheres.

The second reinvention of the WAT was sparked by a financial crisis in 2004 when the state's allocation was halved and the unit had to reposition itself once again. The editorial staff was drastically cut and various other money saving measures were implemented. Marketing and fundraising became focus areas of crucial importance. The staff toiled on for another fourteen years after which it became clear that the time for a third radical decision had come. A revolutionary ten year plan was formulated and is being implemented from July 2018. The plan entails the completion of the WAT within ten years, the appointment of five additional lexicographers and finding the projected required funding over the next ten years.

PARALLEL SESSIONS

Macrostructural Features for a French-Fang School Dictionary

Blanche Nyangone ASSAM (bassam@uwc.ac.za)

Department of Foreign Languages, University of the Western Cape, South Africa

Research has shown that the language situation of Gabon (cf. Ndinga-Koumba-Binza 2005 & 2007) as a whole compels the compilation of school dictionaries (cf. Ella 2002 & 2007; Mabika Mbokou 2001 & 2006; Nyangone Assam 2002 & 2006). These dictionaries, which have been indicated as urgent in the “Gabonese language learning and teaching context” (Nyangone Assam 2006: 139), would enhance and reinforce the integration of native languages in the education system of Gabon (cf. Nyangone Assam 2017).

The present paper aims to introduce a research project, which pursues the compilation of a French-Fang bilingual school dictionary. The approach adopted in this project is to emphasise the use and/or initiation of native languages in the early years of the process of the children’s language acquisition during their first years at school. The children referred here are children who are immersed from birth and grow in a multilingual environment (in communities, homes and mostly at school). This process is an advantage when it comes to the acquisition of languages in that multilingual setting. The process of learning other languages would then coincide with the acquisition of their mother tongue, whether in an oral or in a written form. It is assumed that the children would grow with both languages and would master them at an almost equal level. This process of children’s bilingualism in multilingual contexts is somehow described in Mabika Mbokou (2012).

The current paper focuses on a number of macrostructural characteristics of the planned dictionary. The main project for the school dictionary compilation uses the Fang language as a case study in which methods and principles for a learner’s dictionary should be applied. It is a fact that dictionaries differ in size and purpose. Equally, the structure and the content of the article differ accordingly. The contents of the article of a learner’s dictionary may not be the same as in a general reference dictionary. The types of definitions, the number of illustrative examples and collocations differ between larger and smaller dictionaries.

Lexicographers too often neglect the importance of a well-designed macrostructure as a functional component of the total linguistic content of a dictionary. This is proven by the restriction they commit in their attempts to enhance user-friendliness to the macrostructural level. Restriction in the number of pages, restriction in the content, restrict in space, restriction of the lexicon. Dictionaries are instruments of linguistic and communicative empowerment. Lexicographers therefore have to ensure that their intended target users receive optimal linguistic information. At the same time, the way in which the macrostructural elements are presented should correspond to the linguistic competencies of the pupils. Thus, the decision about the form to include as macrostructural component would have to take into consideration the theoretical status attributed to that form.

In this paper, attention will be given to every macrosuctural categories and features that may appear relevant for a learner’s dictionary. These macrostructural categories and features consist of the collection of lexical items included as heads word or lemmas in the dictionary and the elements of the overall structure and organisation of the dictionary. The determination of these categories and features will require a prior understanding of macrostructure types and their respective components. Before presenting the intended school dictionary macrostructure, the paper will first review a few theoretical issues pertaining to the definition of a school dictionary and how it differs from the learner’s dictionary, to the user profile, the dictionary functions and to the data distribution structure. In conclusion, the present study will have to answer the question about the planning of a school dictionary macrostructure.

References

- Ella, E.M. 2002. Un dictionnaire scolaire de langue de spécialité. *Eléments de Lexicographie Gabonaise Tome 2*, edited by J.D. Emejulu. New York: Jimacs-Hillman Publishers. 305-325.

- Ella, E.M. 2007. *A theoretical model for Fang-French-English specialize multivolume school dictionary*. PhD thesis. Stellenbosch: University of Stellenbosch.
- Mabika Mbokou, L. 2001. Le role du Dictionnaire dans le système éducatif. *Eléments de Lexicographie Gabonaise Tome I*, edited by J.D. Emejulu. New York: Jimacs-Hillman Publishers. 206-222.
- Mabika Mbokou, L. 2006. *A model for a macro- and microstructure of a Yipunu-French school dictionary*. PhD thesis. Stellenbosch: University of Stellenbosch.
- Mabika Mbokou, L. 2012. A survey of bilingualism in multilingual Gabon. *Language science and language technology in Africa: a festschrift for Justus C. Roux*, edited by H.S. Ndinga-Koumba-Binza & S.E. Bosch. Stellenbosch: SUN Press. 163-175.
- Ndinga-Koumba-Binza, H.S. 2005. Considering a lexicographic plan for Gabon within the Gabonese language landscape. *Lexikos* 15: 132-150.
- Ndinga-Koumba-Binza, H.S. 2007. Gabonese language landscape: survey and perspectives. *South African Journal of African Languages* 27(3): 97-116.
- Nyangone Assam, B. 2002. Le dictionnaire pédagogique scolaire: quelques aspects du traitement lexicographique. *Eléments de Lexicographie Gabonaise Tome 2*, edited by J.D. Emejulu. New York: Jimacs-Hillman Publishers. 292-304.
- Nyangone Assam, B. 2006. *Dictionaries as teaching instruments for mother tongue education: the case of Fang*. PhD thesis. Stellenbosch: University of Stellenbosch.
- Nyangone Assam, B. 2017. Planning a dictionary for mother tongue education: a conceptual framework for Gabonese languages. *Lexikos* 27: 378-402.

Aranea Africana: Large Corpora of African Varieties of English and French

Vladimír BENKO (vladimir.benko@uniba.sk)

UNESCO Chair in Plurilingual and Multicultural Communication, Comenius University in Bratislava, Slovakia

E. Štúr Institute of Linguistics, Slovak Academy of Sciences, Bratislava, Slovakia

Anna BUTAŠOVÁ (anna.butasova@uniba.sk)

UNESCO Chair in Plurilingual and Multicultural Communication, Comenius University in Bratislava, Slovakia

During the last decade, creation of web corpora has been recognized as an effective way of obtaining language data in situations where building traditional corpora would be too costly or too slow (Baroni et al., 2007; Jakubíček et al., 2013; Schäfer & Bildhauer, 2013). The recently released open-source and free tools for this purpose made it possible that even the low-funded educational and research institutions can undertake projects of creating large web corpora.

Within the framework of our *Aranea* Project (Benko, 2014), a family of Gigaword web corpora for two dozens of languages is being created. These corpora have the same size, are being compiled from data downloaded in approximately the same time, processed and annotated by (possibly) the same tools and are made available via the same web interface. This is why we believe that they (to a certain extent) deserve the designation of being “comparable”.

The present stage of the project is targeted to region-specific variants of some languages, such as African English and French. English is spoken in 25+ African counties with 450+ million speakers, and French in 30+ African countries and territories with a total of 200+ million francophone speakers. There are, however, not many language resources available that would allow for studying African varieties of English and French. Our corpora try to fill this gap.

Data have been downloaded from the web in 2016 and 2017 by means of *SpiderLing* crawler (Suchomel & Pomikálek, 2012), the initial seed URLs were harvested by *BootCAT*¹ (Baroni & Bernardini, 2004), and the top-level domain (TLD) restriction has been set to consider the relevant African TLDs only. We admit that identification of counties based on the respective TLDs is far from being perfect as lots of African web sites are hosted on servers with generic TLDs, such as *.com*, *.org*, *.net*, etc. We want to solve this problem by using a more sophisticated geolocation based on IP addresses in the future.

Both our African corpora have been linguistically annotated (morphosyntactically tagged, i.e., enriched by information on lemma a part-of-speech category for every word) by *TreeTagger*² (Schmid, 1994; 1995) and made freely available via a web interface under the *NoSketch Engine*³ corpus manager (Rychlý, 2007) at our Web Corpora Portal⁴. We believe that these corpora can provide a valuable resource for studying African varieties of both languages.

Our presentation will introduce the results of the first round of corpus processing yielding to 1+ Gigaword of English and 350 Megaword of French data, and will discuss the possible use of this language resource in lexicography (e.g., studying the collocations typical for African English), as well as in other areas of linguistic studies and education.

Fig. 1 shows the result of an example query: “crown birthday”, a South-Africanism, is a rather rare expression usually not observed in very large corpora of “global” English.

Query crown, birthday 10 (0.01 per million) ⓘ	
essenceofcapetown.co.za	It was my crown birthday December last year (I know that's practically 3 months ago,) and now I'm 28 *gasp*.
rabiaadamson.co.za	Traditionally, in South Africa, particularly amongst the Coloured race (or so it seems), you celebrate your Crown birthday on the day that you turn the age that corresponds with the date on which your birthday falls. ¶
rabiaadamson.co.za	Tomorrow it is my friends Christine Nicola Taphel #chrissycrown25 Crown Birthday as she will be turning 25 years old.
blog.webafrica.co.za	I've been looking at The Mac Book Air for ages, dropping hints for my parents to get it for me for my 21st / crown birthday next month.
akonandungane.co.za	For my 26 th birthday just over a week ago, I decided to spend the weekend in Cape Town for sun, sea, food, friends and cocktails, oh and of course the Cape Town International Jazz festival which was especially co-ordinated to coincide with my crown birthday celebrations (Thanks ESPAfrika).
acidice.co.za	Gabby's Crown Birthday Nails & Party ¶
acidice.co.za	Since it was Gabby's crown birthday , I took my cue from that when I did the mani.
essenceofcapetown.co.za	It was my crown birthday December last year (I know that's practically 3 months ago,) and now I'm 28 *gasp*.
blendeventlife.co.za	Sweet 16, 18th Birthday, 21st, Crown Birthday , 30th, 40th, 50th, 60th, anniversary – you name it, we have got you covered. ¶

Fig. 1. South-Africanisms in *Araneum Anglicum Africanum*: “crown birthday”

Acknowledgements

This work has been, in part, financially supported by the Slovak VEGA and KEGA Grant Agencies, Project Nos. 2/0017/17, and K-16-022-00, respectively.

References

- Baroni, M., & Bernardini, S. 2004. BootCaT: Bootstrapping corpora and terms from the web. In: *Proceedings of LREC'2004*. 2004.Lisbon, Portugal.
- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In: *Language Resources and Evaluation* 43 (3), pp. 209–226.
- Benko, V. 2014. Aranea: Yet another Family of (Comparable) Web Corpora. In: *Text, Speech, and Dialogue. 17th International Conference, (TSD 2014)*. Brno, Czech Republic.

¹<http://bootcat.dipintra.it/>

²<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

³<https://nlp.fi.muni.cz/trac/noske>

⁴<http://unesco.uniba.sk/guest/>

- Jakubíček, M., Kilgariff, A., Kovář, V., Rychlý, P., & Suchomel, V. 2013. The Ten Ten Corpus Family. In: *International Conference on Corpus Linguistics*. 2013. Lancaster, UK.
- Rychlý, P. 2007. Manatee/Bonito – A Modular Corpus Manager. In: *1st Work-shop on Recent Advances in Slavonic Natural Language Processing*. 2007, Brno, Czech Republic.
- Schäfer, R., & Bildhauer, F. 2013. *Web Corpus Construction. Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers.
- Schmid, H. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: *Proceedings of International Conference on New Methods in Language Processing*. 1994. Manchester, UK.
- Schmid, H. 1995. Improvements in Part-of-Speech Tagging with an Application to German. In: *Proceedings of the ACL SIGDAT-Workshop*. 1995. Dublin, Ireland.
- Suchomel, V., & Pomikálek, J. 2012. Efficient Web Crawling for Large Text Corpora. In: *Proceedings of the seventh Web as Corpus Workshop (WAC7)*. 2012. Lyon, France.

Innovations in the access structures of online and mobile dictionaries

André H. DU PLESSIS (ahdp@sun.ac.za)

Department of Afrikaans and Dutch, Stellenbosch University, South Africa

The access structure of any dictionary is an important structure as it provides users with a method to obtain the required data/information they seek. This structure overlaps with almost all the other structures in a dictionary. Whether it be the macro, micro or medio structure, they form part of the route a user takes to eventually satisfy his/her needs. In a printed dictionary the access structure is already visible on the cover of the dictionary as it tells a user what type of dictionary it is and if it will be able to aid in satisfying their lexicographic needs. Gouws and Prinsloo (2005:64) refer to two different types of access structures, namely the outer access structure (the search route leading the user to the lemma sign) and the inner access structure (the article internal search route leading the user to the relevant data entry). With online and mobile lexicographic resources, the traditional description and functions of dictionary structures have been adapted to suit the digital user's expectations and needs, and it becomes increasingly difficult to tell the types of structures apart from one another (Müller-Spitzer, 2013:368-369). This does however mean that electronic dictionaries have a certain amount of freedom in leading the user to the required data in different and innovative ways.

As Gouws (2014) notes, the departure from a traditional macrostructure leads electronic dictionaries to provide a central outer access structure through the search engine of the dictionary. Other trends in the outer access structure are “recent searches”, “favourite searches” and “voice searches” (Du Plessis 2015:54). Du Plessis (2015:54) also notes that the inner access structure of online or mobile dictionaries is more akin to those of traditional printed dictionaries. Yet recent trends rely heavily on technological advancement where features such as the swipe or touch functions on mobile phones or the now standard hyperlink are used to provide new search routes to lead the user to the relevant data in an entry.

Although the features of a traditional access structure are prevalent in the many electronic dictionaries, there are some that have taken advantage of the new online environment by integrating standard or new web/mobile practices to the access structure. This paper therefore investigates and poses to highlight innovations in the various forms of access provided to users in popular online and mobile dictionaries. Before this can be accomplished it is necessary to provide an overview of the evolution of the access structures of electronic dictionaries and how, as mentioned, online dictionary structures are sometimes indistinguishable from each other. Some of the dictionaries that form part of this investigation and overview are Merriam-Webster's mobile dictionary application, Muiswerk Woordenboek Pro, PharosOnline, ANW, WAT, the pilot version and original version online DSAE and Larousse's mobile dictionary application. The usability and sustainability of these innovations¹ as well as how they affect other structural elements in these dictionaries will be discussed and briefly evaluated. These evaluations will be based on metalexicographic principles, user interaction/usability standards and practical online or mobile norms set out in, among others, Tarp (2013), Du Plessis (2015), Sharp, et al. (2007) and the ISO Standard 9241 (2011).

Lastly, it is also important to take note of the ever-changing landscape of web and mobile development. Many of these access routes have been around for many years as part of existing online data structures. Practical lexicographers are slowly becoming more versed in web development and IT principles because they cannot provide access to lexicographic data without understanding the components of an online or mobile platform. It is therefore imperative that the metalexicography does not fall behind and try to merely describe these trends, but rather take on a nomadic nature and borrow from or integrate existing computational fields or standards to create a stronger link between theory and practice. Therefore, this paper will try to show through the aforementioned investigation and overview how the metalexicography can learn from the solutions found in practical electronic lexicography.

References

- Algemene Nederlandse Woordenboek*. [Online] <http://www.anw.inl.nl>.
- Du Plessis, A.H. 2015. 'n Analise van die selfoon-WAT: 'n grondslag vir die verbetering van selfoonwoordeboeke. Unpublished master's thesis. University Stellenbosch: Stellenbosch.
- Gouws, R.H. and Prinsloo, D.J. 2005. Principles and Practice of South African Lexicography. Stellenbosch: AFRICAN SUN MeDIA.
- Gouws, R.H. 2014. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3): 481-504.
- International Organisation of Standardisation. 2010. ISO 9241. [Online] <https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-4:v2:en>.
- Larousse. 2018. *Larousse French dictionary*. Mobile application.
- Merriam-Webster. 2016. *Merriam-Webster Dictionary*. Mobile application.
- Müller-Spitzer, C. 2013. Textual structures in electronic dictionaries compared with printed dictionaries: A short survey, in Gouws, R.H., Heid, U., Schweickard, W and Wiegand, H.E. (eds.). *An International Encyclopaedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin: De Gruyter. 367-380.
- PharosOnline. 2017. [Online] <https://www.pharosonline.co.za/home>
- Schijf, T.T.A. 2015. *Muiswerk Woordenboek Pro*. Mobile application.
- Sharp, H., Rogers, Y. and Preece, J. 2007. *Interaction design: beyond human-computer interaction*. Chichester: John Wiley & Sons.
- Tarp, S. 2013. Lexicographic functions, in Gouws, R.H., Heid, U., Schweickard, W and Wiegand, H.E. (eds.). *An International Encyclopaedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin: De Gruyter. 460-466.

Cross-Language Dictionary Alignment for Bantu Languages

Thomas ECKART (teckart@informatik.uni-leipzig.de)

Natural Language Processing Group, University of Leipzig, Leipzig, Germany

Dirk GOLDHAHN (dgoldhahn@informatik.uni-leipzig.de)

Natural Language Processing Group, University of Leipzig, Leipzig, Germany

Uwe QUASTHOFF (quasthoff@informatik.uni-leipzig.de) Natural Language Processing Group, University of Leipzig, Leipzig, Germany; Department of African Languages, University of South Africa, South Africa

Sonja BOSCH (boschse@unisa.ac.za)

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

Despite recent efforts, the current situation relating to lexicographical data for a number of the Bantu languages remains unsatisfactory. This holds not only for the amount and quality of data for monolingual dictionaries, but also for their interconnection to form a *network of dictionaries*. This contribution describes a prototypical implementation for aligning isiXhosa [xho] and Zimbabwean

isiNdebele [nde] language dictionaries based on their English translations using simple string matching techniques and an advanced word embedding approach.

The landscape of Bantu dictionary data is diverse and heterogeneous. The use of open and well-documented standards is a cornerstone for the long-term availability and reuse of existing resources, and their efficient retrieval. For example, lexicographical data for isiXhosa was recently prepared and converted using a dedicated OWL ontology and is now available for all kinds of applications via standard retrieval mechanisms (Bosch et al. 2018). However, many other resources are already available in a heterogeneous digital format. One such valuable source is the *Comparative Bantu OnLine Dictionary* (CBOLD), which offers Bantu language dictionaries under an open licence, including data for Zimbabwean isiNdebele (Pelling 1971).

Zimbabwean isiNdebele is spoken by approximately 1.6 million people in Zimbabwe, Botswana and Zambia (Ethnologue 2005). IsiXhosa, with approximately 8.1 million speakers, is spoken predominantly in the Eastern Cape and Western Cape regions of South Africa. IsiXhosa and Zimbabwean isiNdebele, as members of the Nguni language group, share many linguistic features both with each other and with other Nguni languages – for instance, they are structurally agglutinating and are therefore characterised by words usually consisting of more than one morpheme, and they are characterised by a conjunctive writing system. Like most Bantu languages, Zimbabwean isiNdebele is considered a resource-scarce language, implying that linguistic resources such as large annotated corpora and machine-readable lexicons are not available. Moreover, academic and commercial interest in developing these is limited.

The pairwise alignment of dictionary entries for dozens of different languages is tedious work and requires extensive human and financial resources. However, much available dictionary data already contains translations into another language – in many cases, including that of both sample dictionaries, this is a well-resourced language such as English. This allows the use of common translations to facilitate the task of dictionary alignment. The CBOLD Zimbabwean isiNdebele data, which is currently limited to a subset consisting of nouns and verbs, was converted using the same procedures as for the isiXhosa dictionary described in Bosch et al. (2018). All data was imported “as seen”, in other words, no additional quality assurance procedures were conducted. Unfortunately, the quality of translations in the isiNdebele dictionary is suboptimal: the vocabulary and format are inconsistent, and in some cases the material is incomplete and contains misspellings (cf. the translation for “udadewethu”: “1, my/our sister; 2, my/our stster-in-law (” and “u-phini”: “tiring stick”). Nevertheless, many alignment candidates were identified using a simple string matching approach: around 30% of verbs and noun stems could be aligned to at least one entry of the other language (e.g. 537 of 1773 isiXhosa noun stems). Standard string similarity algorithms will increase the recall to some extent, especially in cases of misspellings.

As an alternative approach, modern word embedding algorithms can help in identifying alignment candidates: based on the distributional hypothesis, words characterised by similar contexts tend to have similar or related meanings. Unlike in under-resourced languages, many of those similarity candidates can be extracted based on large bodies of text material for languages such as English. In our case, cooccurrences profiles of words were used to extract potential word alternatives and to identify further alignment candidates. For example, the Ndebele word “itshuni” has the English translation “tune”, which does not occur among the Xhosa translations. Based on a corpus of South African English, the most similar terms for “tune” are “song”, “melodies” and “ballads”. Of these, “song” is contained in the Xhosa dataset as a translation for “i/amaculo”, and is therefore a natural candidate for aligning both dictionary entries.

Although an isiXhosa–Zimbabwean isiNdebele e-dictionary might not seem to be an everyday necessity, this project is intended to be a proof of concept which could prove very useful as further dictionary/lexicography data for other African languages becomes available.

References

- Bosch, S., Thomas, E., Klimek, B., Goldhahn, D., Quasthoff, Uwe. 2018. Preparation and Usage of Xhosa Lexicographical Data for a Multilingual, Federated Environment. In: *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki (Japan), 2018.

Ethnologue. 2005. Available: http://www.ethnologue.com/ethno_docs/distribution.asp?by=area#1. Accessed on 23 March 2018.

Pelling, J.N. 1971. *A Practical Ndebele Dictionary*. Available: <http://www.linguistics.berkeley.edu/CBOLD/Data/Ndebele.Pelling.1971.txt>. Accessed on 23 March 2018.

Usage examples in a multilingual online term bank: why and how?

Michelle GOOSEN (michelle.goosen@up.ac.za)

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

Elsabé TALJARD (elsabe.taljard@up.ac.za)

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

According to Bowker and Pearson (2002:16) it is imperative that users of terminology resources such as LSP dictionaries and online term banks know how to use terms: “In addition to information about what a term means, they also need information about how to use that term in a sentence. This information can be provided by presenting the terms in context instead of isolation.” By providing the users of terminology resources with usage examples, the information in a definition is complemented, it shows how a term is used in context, it demonstrates typical collocations and indicates the register. Fuertes-Olivera and Arribas-Baño (2008:129) identify three approaches to usage examples in modern lexicography, namely: usage examples that are invented by a lexicographer, usage examples that are extracted (semi-) automatically from a corpus, and usage examples that are extracted from a corpus and modified to a certain extent by the lexicographer, based on the knowledge of his/her mother tongue. With regard to corpus-sourced examples, Rundell (1998:334-335) sums it up as follows: “The corpus provides natural and typical examples that clearly illustrate the points that need to be made, there is no conceivable reason for not using them.”

This paper firstly considers extracting such usage examples semi-automatically from LSP corpora using *Good Dictionary Example eXtractor* (GDEX), part of the suite of functions of *Sketch Engine*, and secondly, presenting these usage examples to South African LSP users on a platform, such as a multilingual online term bank.

Sketch Engine uses GDEX to extract usage examples, firstly, by sorting concordance lines, and secondly, by using the algorithm based method (www.sketchengine.co.uk). By following the first technique, the terminologist determines the criteria according to which the concordance lines are sorted, namely: attributes, structures and number of concordance lines. When sorting the concordance lines of a specific term, the “best” usage examples – according to GDEX – will appear in the first few concordance lines. The end product is thus a (sorted) list of potential example sentences in which the terminologist will most probably find a good, suitable usage example. It is however still the responsibility of the terminologist to scrutinise these concordance lines to determine which sentences are the most appropriate usage examples.

To determine the efficiency of the given technique, three random terms were chosen from a list of academic terms extracted from an English LSP corpus: *ambiguous* (adjective), *convince* (verb) and *rhythm* (noun). The term *ambiguous* produced 27 concordance lines; within the first ten concordance lines, six usable usage examples were found. The rest of the concordance lines produced three usable usage examples. The term *convinced* produced 72 concordance lines where the first ten concordance lines contained eight usable usage examples. Thirteen sentences were found in the rest of the concordance lines which can be used as usage examples. It was confirmed that the GDEX function does sort the best usage examples within the first few concordance lines. It was also confirmed that these usage examples must be modified before presenting them to the target user. The term *rhythm* however did not present promising results when used as a search term within the GDEX function – the first 10 concordance lines only presented one useable example sentence. This is probably not because of any shortcomings of the software, but rather because of the content and small size of the corpus used.

In presenting these extracted usage examples to South African users of a multilingual online term bank, the next obvious step is to translate the given usage examples from the source language

(English) to the target language (one of the 10 official African languages). Quality checks revealed that translated usage examples often did not contain the term that the usage example is supposed to illustrate. Usage example should therefore rather be extracted from a corpus of the target language. Such examples can however not be extracted for languages that do not have sufficient resources such as LSP corpora. For the South African context it is consequently suggested that usage examples should ideally be created in the target language by the terminologist, in co-operation with a mother-tongue speaker and subject specialist.

References

- Bowker, L. and Pearson, J. 2002. *Writing with Specialized Corpora: A practical guide to using corpora*. London: Routledge.
- Feurtes-Olivera, P.A. and Arribas-Baño, A. 2008. *Pedagogical Specialised Lexicography: The representation of meaning in English and Spanish business dictionaries*. Amsterdam: John Benjamins Publishing Co.
- Rundell, M. 1998. Recent Trends in English Pedagogical Lexicography. *International Journal of Lexicography*, vol. 11, no. 4, pp.315-342.
- Sketch Engine. 2017. Available: www.sketchengine.co.uk. Accessed on 19 December 2017.

The data distribution structure and data access in online dictionaries

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

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

Although various dictionary structures, including the data distribution structure, are relevant to both printed and online dictionaries there may be significant differences in the application of these structures in the respective environments. This paper presents a discussion of some aspects of the occurrence and application of one dictionary structure, the data distribution structure, in online dictionaries. Some implications for the mediostructure and the access structure will also come to the fore. Certain similarities in the occurrence of the data distribution structure in printed and online dictionaries are indicated, but the main emphasis is on the telling differences compelled by the nature of the online environment.

Printed dictionaries function as carriers of text types. Where the dictionary displays a frame structure the central list is complemented by outer texts in the front and back matter sections and the data distribution structure is employed to find a landing slot for each item, data type and text type in this “big text”. In this regard it is important to realise that a dictionary as carrier of text types contains more data than can be found in the central list and that an outer text should be regarded as more than a mere cosmetic add on. It is a functional and often integrated component of the dictionary.

The distribution of data in online dictionaries also goes beyond the borders of the alphabetical section. Data can be accommodated in texts and other venues that complement the treatment in the different article stretches, albeit that the outer texts, or, in terms of Klosa & Gouws (2015), the outer features, are not presented in a traditional frame structure.

The present paper will look at various types of outer features that lexicographers can use to enhance the distribution and presentation of data and to increase the successful retrieval of information from the data on offer. It will be argued that lexicographers of online dictionaries should not restrict their users’ access to data to the data included in the dictionary and its different outer features. The article stretches, constituting the traditional central list, along with the outer features introduced for the distribution of data to suffice the needs of the intended target users still form the bulk of the contents and form of an online dictionary. However, access to data should not be restricted to the article stretches and outer features. As a container of knowledge, the dictionary should also function as a gateway to data not included as a formal component of the dictionary. As an information tool the online dictionary should present its users with links to dictionary-external venues and to other reference sources, either within a portal to which the specific dictionary belongs, or to complementing venues where the user can retrieve additional information to ensure optimal consultation success.

In the planning of the data distribution programme lexicographers should negotiate the complementing roles of the data distribution structure and the mediostructure of their dictionaries. A new mediostructural approach to dictionary-external cross-reference addresses will integrate the dictionary into a much broader knowledge domain where users can identify complementing data venues that could assist them in their endeavours to retrieve the information determined by both communicative and cognitive functions of the dictionary.

This paper intends to show that the planning of the data distribution structure of online dictionaries needs to take cognizance of innovative possibilities of accessing additional dictionary-external data. By establishing a user profile each user can consult the dictionary for a variety of data types to satisfy a variety of lexicographic functions. From the perspective of dictionary structures, it will show a new and exciting interactive relation between the data distribution structure, the mediostructure and the access structure.

Online dictionaries can yet again fulfil the assignment that Amos Komensky (1631) identified for his dictionary the *Ianua Linguarum Reserata*, that is to unlock the gate of tongues.

References

- Klosa, A. & Gouws, R.H. 2015. Outer features in e-dictionaries. *Lexicographica* 31: 142-172.
- Komensky, A. 1631. *Ianua Linguarum Reserata*.

Dictionary posters in !Xun and Khwedam: A learning experience

Kerry JONES (jonesleekerry@gmail.com)

Linguistics Department, University of KwaZulu-Natal, Howard College Campus, South Africa;
Director, African Tongue, Cape Town, South Africa

Selection of an “appropriate” orthography for indigenous minority languages in South Africa has unfortunately become a political struggle, especially for Khoe and San languages. Speakers’ practical needs are not given enough attention and the slow or lacking action by policy makers and the Department of Education have not resulted in a successful government-led application of mother tongue literacy for !Xun and Khwedam speakers. As part of my ongoing engagement with the !Xun and the Khwe of Platfontein, basic ‘dictionary posters’ were developed in collaboration with mother tongue speakers. The ‘dictionary posters’ were created in a bilingual format !Xun/Afrikaans or Khwedam/Afrikaans with accompanying illustrations to facilitate understanding.

The link between language attitudes of !Xun and Khwedam speakers from Platfontein and language vitality (Jones, 2017) was explored in my PhD study using a tripartite perspective: knowledge, emotion and behaviour (Mckenzie, 2010:22). The results showed a correlation between positive language attitudes and healthy language vitality in an African context, as previously shown in different contexts (Austin & Sallabank, 2013:313). However, changes in lifestyle, from a hunter-gatherer way of life to a ‘westernised’ existence has had adverse effects in the continued use of cultural and traditional practices specific to the !Xun and the Khwe. Notwithstanding such changes, the !Xun and the Khwe have retained their mother tongue across generations. Yet without institutional support in the form of mother tongue literacy efforts, !Xun and Khwedam will not be assimilated into new language settings such as formal education, literacy and the media. Without intervention, the use of !Xun and Khwedam will likely be reduced in exchange for an increased use of Afrikaans, which is the socially and politically dominant language of the area.

Results further indicated a consistent desire for access to mother tongue literacy in !Xun and Khwedam. Speakers were aware that orthographies existed in their mother tongue yet the majority had never seen such orthographies before. It was evident that mother tongue literacy was hindered by a combination of factors, namely a) lack of access to mother tongue learning materials, b) lack of trained mother tongue teachers and c) lack of supporting teaching and literacy infrastructure.

Drawing on the insights gained from the co-creation of the publication *Ju|'hoan Da'abi!oa N#omtciasi Kokxuisi ǀXanua, Ju|'hoan Prentewoordeboek vir Kinders, Ju|'hoan Children's Picture Dictionary* (Cwi & Jones, 2014) and at the community members' request, the dictionary posters were produced thematically. The selected themes were as follows:

!Xun	Khwedam	Afrikaans	English
'Ho	ǀ	Gesig	Face
Mi n o n!ang	Ti x'e a	My liggaam	My body
Platfontein	Platfontein	Platfontein	Platfontein
Nǀaihmhe	Tc'ao-kx'oxo	Wilde diere	Wild animals
Nǀaihmhe	Xe-xei ya ǀa xo dji, kx'o-xo dji nu encekte dji ta	Goggas	Goggas
Skola N!ang	Cure o ki	In die klaskamer	In the classroom

The themes and lexical entries selected for each poster were based on everyday encounters for !Xun and Khwe children in Platfontein. Along with basic concepts of interest, such as 'Wild animals' and 'Goggas'. As is evident from the table above, !Xun and Khwedam are totally different languages. There were instances of lexical borrowings from either Afrikaans or English, e.g. "Skola" deriving from the Afrikaans "Skool" or English "School". Some concepts overlapped in that the same term was used, e.g. "Insects" and "Wild animals" in !Xun are both "Nǀaihmhe" and in other cases one had to explain a concept rather than have a single lexical entry or phrase for a concept, e.g.: "Goggas" in Khwedam is "Xe-xei ya ǀa xo dji, kx'o-xo dji nu encekte dji ta" literally translated means "Small things that come out at night and can bite you, animals that eat people and insects". All illustrations for the community posters were developed by local community members in Platfontein.

I presented to community members the existing orthographies in both !Xun and Khwedam. Through extensive discussions, the community members made it clear which orthography they preferred to use. Their decisions were based on practical and user-friendly criteria as well as accessibility. The dictionary posters were primarily aimed at children, however they could be used for adult literacy education too. Through the development of these dictionary posters a trial session of mother tongue literacy classes was provided to community members of Platfontein.

Although the materials were well received, without ongoing teaching infrastructure and training for mother tongue !Xun- and Khwedam-speaking teachers, it is unlikely that mother tongue literacy will be adopted by the !Xun- and Khwe-speaking community in Platfontein. Such lack in support hinders the use of the mother tongue in language settings such as formal education, literacy and the media and ultimately allows for language shift to Afrikaans. Without active intervention, the use of !Xun and Khwedam in the families investigated will be limited to private settings such as the home which is likely to lead to progressive language loss.

References

- Austin, P., K., and Sallabank, J. 2013. Endangered languages: An introduction. *Journal of multilingual and multicultural development*, 34 (4): 313-316.
- Cwi, T., & Jones, K. 2014. *Ju|'hoan Da'abi!oa N#omtciasi Kokxuisi ǀXanua, Ju|'hoan prentewoordeboek vir kinders, Ju|'hoan children's picture dictionary*, Pietermaritzburg: University of KwaZulu-Natal Press.
- Jones, K. 2017. *Language attitudes as a change agent for language vitality: A case study of two Khoesan languages in Platfontein (RSA)*. Ph.D Thesis, University of KwaZulu- Natal, Howard College, South Africa.
- McKenzie, R. 2010. The study of language attitudes. In *The social psychology of English as a global language: attitudes, awareness and identity in a Japanese context*. *Educational Linguistics*, 10: 19-40. Amsterdam: Springer Netherlands.

Hedges in definitions: A historical look at English dictionaries

Mariusz KAMIŃSKI (mariusz.kaminski@pwsz.nysa.pl)

Basic Study Centre, University of Applied Sciences, Nysa, Poland

A language has an inventory of metalinguistic devices that are used to comment on what is being said. They are used to communicate that what is being said is not exactly precise or complete. Lakoff refers to such words as hedges, explaining that their “meaning implicitly involves fuzziness” and their “job is to make things fuzzier or less fuzzy” (1972:471).

Hedges express different degrees of category membership. As Lakoff demonstrates, a hedge *par excellence* indicates the most central member of a category (as in *A robin is a bird par excellence*), while *sort of* points to a peripheral example of the category (as in *A penguin is sort of a bird*). What is more, *loosely speaking* does not merely exclude the central member but points to “things that would not ordinarily be considered members” (as in *Loosely speaking, a telephone is a piece of furniture*) (Lakoff 1972 Taylor 1995:77). The above findings, known as prototypical effects in cognitive semantics, provide support for the conception of prototypically organized categories, namely that categories display degrees of typicality, and that their boundaries are blurred (Lewandowska-Tomaszczyk 2007:145).

The theory of prototype is relatively new in the history of linguistics, but it has recently attracted a great deal of attention from dictionary makers and researchers (Swanepoel 1992; Van der Meer 2000; Hanks 1994). The *New Oxford Dictionary of English* (1998) is perhaps one of the first dictionaries to refer to the prototype model as the main principle on which senses are organized (Pearsall 1998; see also Geeraerts 2006).

Given the usefulness of prototype theory for lexicographers, one should expect to find hedges in definitions in large numbers. Assuming that hedges are deliberately used in contemporary lexicography to indicate prototypical features by lexicographers who are informed by cognitive semantics, the question arises whether they have always been used equally frequently in practical lexicography. This is what we address in this paper. This paper aims to examine the frequency and distribution of hedges in definitions in major English dictionaries published over the past centuries.

This study was conducted on seven dictionaries published from 1785 to 2011: Johnson's *Dictionary of the English Language* (1785), Webster's *American Dictionary of the English Language* (1865), the *OED* (1888–1928), *Chambers's Twentieth Century Dictionary* (1952), *Merriam-Webster's Collegiate Dictionary* (2003), the *Longman Dictionary of Contemporary English* (2005), and the *Concise Oxford English Dictionary* (2011). Definitions were extracted by taking a random sample of 30 to 60 pages from each dictionary and converting the images into electronic files. 76 hedges were pre-selected on the basis of Lakoff (1972), and were then identified and counted up in each sample.

The study showed that hedges are distributed unevenly in English dictionaries. They are most frequently used in the *Collegiate*, the *OED*, and *Webster*, and least frequently in *Johnson* and *Chambers*. The *OED* ranks second in terms of the hedge frequency.

Among the reasons for the *OED* being rich in hedges is arguably the need to provide a comprehensive account of meaning of lexical units in the historical perspective. Historical lexicographers uncover the full path of semantic change by highlighting what is typical in language rather than by searching for necessary conditions for categories of meaning. Hedges are indispensable in this task.

The *Collegiate* has the largest number of hedges. They are used as markers of category membership not only within definitions (e.g. *a sprinkling with water esp. in religious ceremonies*) but also between definitions to indicate sense relations (e.g. *sprinkle; esp : to sprinkle with holy water*). Although the *Collegiate* is not strictly speaking a historical dictionary, it employs a developed system of sense division. The use of hedges as sense dividers can be traced back to *Webster*, but the *Collegiate* has exploited their full potential more extensively.

References

- Geddie, W. 1952. *Chambers's Twentieth Century Dictionary*. (Second edition). Edinburgh: W. & R. Chambers.
- Geeraerts, D. 2006. *Words and other wonders*. Berlin: Mouton De Gruyter.

- Hanks, P. 1994. Linguistic norms and pragmatic exploitations, or why lexicographers need prototype theory, and vice versa. In Ferenc Kiefer, Gabor Kiss, Julia Pajzs (eds.), *Papers in Computational Lexicography*. Budapest: Hungarian Academy of Sciences, 89-113.
- Johnson, S. 1785. *A Dictionary of the English Language*. Vol. 1 and 2. London: W. Strahan.
- Lakoff, G. 1972. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 8: 183-228.
- Lewandowska-Tomaszczyk, B. 2007. Polysemy, prototypes, and radial categories. In Dirk Geeraerts and Hubert Cuyckens (eds.). *The Oxford handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 139-169.
- Mish, F. 2003. *Merriam-Webster's Collegiate Dictionary*. (Eleventh edition). Springfield: Merriam-Webster Inc.
- Murray, J. A. H. 1888–1928. *A New English Dictionary on Historical Principles*. Vol. 1–10. Oxford: The Clarendon Press. (OED)
- Pearsall, J. (ed.) 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Stevenson, A. and M. Waite. 2011. *Concise Oxford English Dictionary*. (Twelfth edition). Oxford: Oxford University Press. (COD)
- Summers, D. 2005. *Longman Dictionary of Contemporary English*. (Fourth edition with writing Assistant). Harlow: Longman. (LDOCE)
- Swanepoel, P. 1992. Linguistic motivation and its lexicographical application. *South African Journal of Linguistics* 10: 49-60.
- Van der Meer, G. 2000. Core, subsense, and the *New Oxford Dictionary of English*. In Ulrich Heid, Stefan Evert, Egbert Lehmann, Christian Rohrer (eds.). *Euralex 2000 proceedings*, 419-431. Stuttgart: Universität Stuttgart.
- Webster, N. 1865. *An American Dictionary of the English Language*. Royal quarto edition. Revised by Chauncey A. Goodrich and Noah Porter. Springfield: G.&C. Merriam.

“Boys will be boys:” an example of biased usage

Steven M. KAPLAN (lorero@gmail.com)

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

Inclusive and bias-free communication are the key to truly effective interpersonal relations, as language, culture, and society are inextricably intertwined. Exclusive or biased expression may lead to individuals and groups of people being dehumanised, oppressed, stereotyped, and so on.

Despite there being widespread biased and exclusive expression in the English language, general language and learning dictionaries do not do an adequate job of alerting users to this. Anyone accessing these dictionaries is usually only getting a part of the full picture, information that ignores these aspects, or even definitions which promote further biased and exclusive usage.

The common phrase boys will be boys is defined by most of the best known general and learning English dictionaries. I will use this idiom to help illustrate how a seemingly innocuous phrase is nonetheless laden with bias. To accomplish this task I will dissect the definitions provided by several dictionaries, including those from Oxford, Cambridge, American Heritage, Collins, and Farlex (from thefreedictionary.com.)

The Oxford Living Dictionaries online defines boys will be boys so: “Used to express the view that mischievous or childish behaviour is typical of boys or young men and should not cause surprise when it occurs.”

Briefly examining this analysis, one can note that Mischievous or childish behaviour sounds rather harmless, and if the term were to always be used for that, then that would be one thing. The reality, however, is that this phrase is also often utilised to “justify” all sorts of atrocious behaviours of males of any age. This latter aspect is not even hinted at. Typical and should not cause surprise when it occurs are meant to convey that a) it happens a lot, b) there is no need to make a fuss when it does, and c) boys and young men are “hard-wired” to “inadvertently” cause “harmless” trouble, so there is no point in addressing this conduct in any manner.

A further thorough analysis on how this and other dictionaries perform from the “biased usage” perspective, some of the consequences of such usage, and a suggested definition from an inclusive and bias-free expression dictionary will also be presented.

The definitions from these dictionaries will then be contrasted with how an inclusive and bias-free expression dictionary might deal with this phrase:

boys will be boys – A term indicating that rowdy behaviour is not just characteristic of boys (and frequently men,) but tolerable and even welcome. Serves to reinforce stereotypes and traditions that promote harmful behaviour in boys and males of any age, while girls and females in general are encouraged (likelier forced) to be subservient and willing to be victimised. As such, it unequivocally nurtures and promotes the patriarchal society, along with sharply defined gender double standards. This phrase has been employed to mitigate and trivialise actions ranging from horseplay through gang rape.

The assertions made in the definition provided by the inclusive and bias-free expression dictionary are backed up by references such as:

Sanday (2007), who identifies any number of ways in which boys will be boys is utilised to encourage and defend male sexual violence and rape, including how this phrase serves as a mantra employed to excuse gang rape in middle-class environments, while Scales (2008) emphasizes the role played by the “hard-wiring” defence to rape provided by boys will be boys.

Mansfield, Beck, Fung, Montiel, & Goldman (2017) furthermore stress that boys will be boys is so often utilised to excuse subtler forms of sexual harassment that it becomes ingrained in people’s minds to the extent that acts of this ilk are considered natural or acceptable. As such, this dissuades victims from saying anything about these incidents, and in cases where an aggrieved person does report about it they are usually told they are being too sensitive.

There is plenty of room for improvement among the general language and learning English dictionaries from the “biased usage” perspective, and this paper will help illustrate their current deficiencies through a concrete example, along with providing an alternative inclusive definition.

References:

- Mansfield, K.C., Beck, A.G., Fung, K., Montiel, M. and Goldman, M., 2017. What Constitutes Sexual Harassment and How Should Administrators Handle It?. *Journal of Cases in Educational Leadership*, 20(3): 37-55.
- Oxford Living Dictionaries online. 2018. Available: https://en.oxforddictionaries.com/definition/boys_will_be_boys. Accessed on 7/05/2018.
- Sanday, P.R., 2007. *Fraternity Gang Rape: Sex, Brotherhood, and Privilege on Campus*. New York, USA: NYU Press.
- Scales, A., 2008. Student Gladiators and Sexual Assault: A New Analysis of Liability for injuries Inflicted by College Athletes. *Mich. J. Gender & L.*, 15: 205.

Towards an isiZulu National Corpus

Langa KHUMALO (khumalol@ukzn.ac.za)

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

This paper discusses the context and process of developing the IsiZulu National Corpus (henceforth the INC) at the University of KwaZulu-Natal (henceforth UKZN). A corpus is a judiciously designed and systematically collected natural language data from a variety of text types and sources following a particular set of principles, which constitutes a sample that statistically reflects the use of that particular language, and is stored and accessed by means of computers (McEnery & Hardy, 2011). The size of the corpus and the source from which it is created depends on the intended purpose. It will be evinced that there are many corpora of varying sizes and typologies that have been developed for a variety of languages spoken globally. Some of these corpora are accessed through a repository called the Sketch Engine platform (Kilgarriff, Baisa, Bušta, Jakubíček, Kovář, Michelfeit, & Suchomel, 2014; Kilgarriff, Rychly, Smrz, & Tugwell, 2004). Sketch Engine is a repository for corpora

management, corpora analysis and creation of corpora, that is accessed through subscription. Sketch Engine supports and has corpora in ninety (90) languages, with some languages like English having multiple corpora. Of these ninety languages, only six are African (Afrikaans, Arabic, Igbo, Setswana, Swahili and Yoruba).

It will be (briefly) argued that there is scarcity of corpora in African languages. There are some corpora that have been developed for African languages that do not appear on Sketch Engine. Furthermore, some corpora that have been developed for African languages are not openly accessible and require permission for access, and this is compounded by the fact that most of these corpora do not reside in the continent, but are hosted and kept in Europeans institutions (Khumalo, 2015:24) as shown in **Table 1** below. In the South African context, it will be demonstrated that the University of Pretoria (henceforth UP) led a momentous initiative to develop corpora. In the work that started in the early 1990s, UP developed corpora in all the official languages in South Africa. **Table 2** below shows these corpora and their sizes.

Thus, this paper discusses the challenges in the development of the INC. These challenges include *inter alia* logistical challenges such as whether UKZN is the suitable site for the INC, IsiZulu orthography, access to potential corpus materials and attendant copyrights issues, data processing and storage, and how these challenges were mitigated in the process. The paper further discusses the advantages of using the INC in language teaching and learning. However, it will be evinced that there are challenges and limitations. The main challenge is the technical problems, while others are pedagogical, i.e., is the INC pedagogically appropriate and relevant for language teaching and learning?

It will also be argued in this paper that the development of the INC is motivated as an important precursor to the development of IsiZulu as a scientific language. It is thus, an important resource in the development of IsiZulu Human Language Technologies. The paper outlines the impetus to develop the INC and the current processes of its development. Section 1 discusses briefly the University's language policy and plan, which articulates the impetus to develop IsiZulu. Section 2 outlines the decision leading to the identification of UKZN as the site for the INC. Section 3 discusses the planning processes from the initial corpus linguistics workshop, which sought to explicate the importance of the corpus in the growth of a language and for posterity. Section 4 discusses the criteria for selecting the first materials that formed the basis for INC. Section 5 outlines the boardroom pitches in various organizations that produced IsiZulu textual materials and the synergies that developed therefrom. Section 6 outlines the statistical growth of the INC as shown in **Table 3** below. We discuss the role of the INC in the development of current technologies and its future role in language teaching, and conclude in section 7. It is our conclusion that the OCR presented the biggest challenge in processing some of the corpus files. It is also our conclusion that the development of the INC was crucial in the development of the IsiZulu Spellchecker (Keet & Khumalo, 2017). Finally it can be concluded that while the INC is crucial in language research and documentation, it is currently a challenge to use it as a resource in language teaching and learning because of technical and pedagogical limitations.

Table 1: Some corpora in Africa Languages (Khumalo, 2015).

Language	Name	Acronym	Place(s)	Size
Kiswahili	Helsinki Corpus of Swahili	HCT	Helsinki	25 million
Kiswahili	Kiswahili Internet Corpus	KIC	Pretoria/Ghent	+20 million
Shona	ALLEX – Shona Corpus	ALLEX	UZ/Oslo/Goteborg	2.2 million
Ndebele	ALLEX – Ndebele Corpus	ALLEX	UZ/Oslo/ Goteborg	0.7 million
Nambya	Nambya Corpus	ALLEX	UZ/Oslo	0.3 million

Ciluba	Recall's Ciluba Corpus	RCC	Ghent	0.3 million
--------	------------------------	-----	-------	-------------

Table 2: Corpora in South African languages (De Schryver and Prinsloo, 2000)

LGP Corpus Name	Acronym	Size
Pretoria isiNdebele Corpus	PNC	1, 959, 482
Pretoria siSwati Corpus	PSwC	4, 442, 666
Pretoria isiXhosa Corpus	PXhC	8, 065, 349
Pretoria isiZulu Corpus	PZC	5, 783, 634
Pretoria English Corpus	PEC	12, 799, 623
Pretoria Afrikaans Corpus	PAfC	11, 602, 276
Pretoria Xitsonga Corpus	PXic	4, 556, 959
Pretoria Tshivenda Corpus	PTC	4, 117, 176
Pretoria Setswana Corpus	PSTC	6, 130, 557
Pretoria Sesotho sa Leboa Corpus	PSC	8, 749, 597
Pretoria Sesotho Corpus	PSSC	4, 513, 287

Table 3. IsiZulu National Corpus (INC) Statistics (Khumalo, 2018)

	File name (IsiZulu)	File name (English)	Number of files	Word tokens
1.	Inoveli	IsiZulu novels	487	9, 679, 532
2.	Isolezwe	Isolezwe newspaper	489	7, 289, 832
3.	UmAfrika	UmAfrika newspaper	76	5, 735, 962
4.	Ilanga	Ilanga newspaper	970	4, 361, 605
5.	Izindaba zabantu	Izindaba zabantu newspaper	43	2, 376, 468
6.	Ezasegagasini	Metro ezasegagasini newspaper	61	782 377
7.	Ibhayibheli	Bible	1	435 481
8.	Umthetho	The Hansard	194	367 998
9.	Zulumanuscripts	Dissertations	19	287 684
10.	Umngenelo	Literature competition short stories	16	130 698
11.	Ingede	Ingede newspaper	14	125 972
12.	Uhlelo	Zulu grammar textbooks	2	84 416
13.	Zulusimama	kznonline newsletter	4	34 525
14.	Umthethosisekelo	Constitution of the Republic	1	32 410
15.	Amanothi	Notes	4	27 447
16.	Zuluplay	The play amaseko	1	16 694
17.	Amantshontsho	Bible lesson	1	4 974
18.	Inganekwane	Folktales	1	1 709
Total number of files in the INC			2 384	
Total number of tokens in the INC				31, 775, 784

References

De Schryver, G-M. & D.J. Prinsloo. 2000. The compilation of electronic corpora, with special reference to the African Languages. *Southern African Linguistics and Applied Language Studies* (18): 89-106.

- Khumalo, L. Forthcoming. Corpora as agency in the intellectualization of African languages. In Kaschula & Wolff (forthcoming). *African Languages in Knowledge Societies: The transformative power of languages*. Cambridge University Press, Cambridge, UK.
- Keet, C. M. & Khumalo, L. 2017. Evaluation of the Effects of a Spellchecker on the Intellectualisation of IsiZulu. *Alternations* (Special Issue), 75 -97.
- Khumalo, L. 2015. Advances in Developing corpora in African languages. *Kuwala*, 1(2), 21-30.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Suchomel, V. 2014. The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. 2004. Itri-04-08 the sketch engine. *Information Technology*, 105, 116.
- McEnery, T., & Hardie, A. 2011. *Corpus linguistics: Method, theory and practice*: Cambridge University Press.
- Reppen, R. (2001). Review of MonoConc Pro and WordSmith Tools. *Language Learning & Technology*, 5(3), 32-36.

UX (User Experience) design applied to electronic lexicography: some guiding principles deployed in the digitisation of *A Dictionary of South African English*

Bridgitte LE DU (b.ledu@ru.ac.za)

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

Tim VAN NIEKERK (t.vanniekerk@ru.ac.za)

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

Dictionary quality rests ultimately on content, but lexicographers face a sobering reality when their products go online: no matter how unique or authoritative their data, their product will see limited use unless it meets high standards of design. In dictionaries with more complex structure – e.g. a wider variety of data-types, diachronic presentation of polysemy, evidential citations and dense mediostructural relations – the challenge is increased.

At the same time, in a superficial sense the Internet has made dictionary publishing easier. Dictionaries can be put online with fewer resources than print projects require. This often results, however, in lack of project support in an area where it is later found it is most needed, namely design. This role, traditionally fulfilled by print publishers, is sometimes, for pragmatic reasons, assumed by programmers or lexicographers doubling as designers on electronic dictionaries, with mixed results. Lexicographical theory is not always helpful in this situation. The shift from the design principles appropriate to print lexicography or static “p-works” (Tarp, 2012), to principles applicable to dynamic, functional lexicographic “e-tools” (Tarp, 2012) is often discussed but has not yet resulted in guidelines sufficiently directed at electronic dictionary designers’ actual use cases.

This presentation shows examples of how these challenges can be met with reference to industry-based UX (User Experience) guidelines, as part of an evolving design process in the thorough print-to-electronic adaptation of *A Dictionary of South African English on Historical Principles* (Silva et al., 1996). This project converts a compressed but one-dimensional print microstructure and layout to what, in the electronic medium, becomes a multi-layered interactive web application exposing database-like dictionary functionality to a wider target audience (now including non-specialist users). Initial development phases in collaboration with the Universities of Hildesheim and Stellenbosch produced (1) dataset enhancements to support selective querying and data presentation changes (Van Niekerk et al., 2016); (2) prototypes of micro- and macro-visualisation devices using corresponding micro- and macro-aggregation of data, new to historical dictionaries (Van Niekerk et al., 2016; Le Du & Van Niekerk, 2017); and (3) initial prototypes and wireframes following an overall design review focused on simplified but more functional and accessible navigation features, as well as aesthetic and structural layout changes (Du Plessis & Van Niekerk, 2016).

Whereas designers of concise print dictionaries produced “a miracle of compression” (Atkins & Rundell, 2008:21), this compression is essentially linear. Assimilating the current project’s presentational components (1-3 above) into a single interface calls for design strategies which

accommodate a different kind of structural density. Visual compression of both the *navigation functions* of the dictionary interface (searching, sorting, accessing Help) and its *data* (individual entries) is achievable through layering strategies. Navigational design patterns foreground persistent user interface elements according to priority as defined by the current context. Data or entry design strategies extend beyond layout adaptation and content expanders, to the creation of a visual hierarchy by manipulating colour, contrast and typography that guides users to the most relevant information first.

Situating dictionaries within the context of generic reference-oriented web applications, and then adapting UX principles to lexicographic requirements, is an important step towards effective dictionary design. Users no longer accord dictionaries “the status of a kind of Bible” (Zaenen, 2002); dictionaries have become one of many types of online resources consulted daily and unwieldy design is less likely to be accommodated. This presentation shares practical applications of design principles and invites feedback and discussion.

References

- Atkins, B.T.S. & Rundell, M. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Du Plessis, A. & Van Niekerk, T. 2016. Adapting a Historical Dictionary for the Modern Online User: The Case of the *Dictionary of South African English on Historical Principles*'s Presentation and Navigation Features. *Lexikos* (26): 82-102.
- Le Du, B. & Van Niekerk, T. 2017. Data visualisation in the online Dictionary of South African English. Presentation: AFRILEX 22nd International Conference. 26-29 June, 2017. Grahamstown, South Africa.
- Silva, P., Dore, W., Mantzel, D., Muller, C., Wright, M. 1996. *A Dictionary of South African English on Historical Principles*. Cape Town: Oxford University Press.
- Tarp, S. 2012. Theoretical challenges in the transition from lexicographical p-works to e- tools. In Granger, S. & Paquot, M. (eds) *Electronic Lexicography*, Oxford: Oxford University Press.
- Van Niekerk, T., Stadler, & H., Heid, U. 2016. Enabling Selective Queries and Adapting Data Display in the Electronic Version of a Historical Dictionary. In: *XVII EURALEX International Congress*. 6-10 September, 2016. Tbilisi, Georgia: 635-646. Available: <http://euralex2016.tsu.ge/publication.html>. Accessed on 21/02/2017.
- Zaenen, A. 2002. Musings about the Impossible Electronic Dictionary. Available: https://www.researchgate.net/publication/228569358_Musings_about_the_Impossible_Electronic_Dictionary. Accessed on 21/02/2018

Does the isiNdebele Terminology developed today have any significant impact?

Sponono K. MAHLANGU (sponono.mahlangu@up.ac.za)

iZiko lesiHlathululi-mezwi sesiNdebele, University of Pretoria, Pretoria, Republic of South Africa

Although not supported here by a case study and while acknowledging the prevalence of code switching in any multilingual context, this article explores the extent to which terms created by terminologists are perhaps not always understood and used by isiNdebele speakers. IsiNdebele is one of the Nguni languages that is more under-resourced when compared to the other official languages in the Republic of South Africa. IsiNdebele as a young language is going through a metabolic process of constant change. These changes also affect, inter alia, the formation of the coined words, e.g. the usage of a hyphen in compound terms, as will be discussed in this article. IsiNdebele is faced with the influx of newly coined words every day. These terms come from natural sciences, mathematics, science and technology, HIV and AIDS terms derived from foreign languages, especially Greek, Latin, English and Afrikaans. In 1994 the Department of Arts, Culture, Science and Technology (DACST) initiated the employed African language terminologists to develop and document terminology in African languages (IsiNdebele included) in a variety of subject fields.

In order to develop and protect language rights, the Pan South African Language Board (PanSALB) has created thirteen National Language Bodies (NLBs) as advisory structures to take care

of among other things, standardisation (e.g., spelling or orthography rules), terminology development, verification and authentication of terminologies. In the past, the National Terminology Services (NTS) used to work in collaboration with the old Language Boards. Currently terminologists of the National Language Service (NLS) work in consultation with the National Language Bodies. This was done to ensure the standardisation and stabilisation of terms as well as with popularising terms.

Due technological advancements across the globe, neologisms i.e., new terms, have to be coined for new situations, articles, inventions and environments. Some of the multilingual terminology lists that were compiled are: weather terms, basic health terms, HIV/AIDS terms, building terms, election terms, banking terms, commercial and financial terms, computer terms, mathematical terms, natural science terms, soccer terms, water and sewerage terms and the like.

In coining these terms, the IsiNdebele Language Board (iKhwezi, i.e. the name of the isiNdebele Language Board) and the terminologists had challenges in terms of adhering and applying international guidelines, strategies and principles for terms. The following are the guidelines that the terminologists should take into account when coining and developing terms:

There should be a one-to-one relation between the term and the concept that the term represents, which means that the term must not be ambiguous; it must refer to one concept only.

Terms that refer to related concepts must be similar in one way or other in order to show the similarities between the related concepts.

A term should conform to the spelling, morphology and pronunciation rules of the language for which the term is intended.

Terms should be brief and they should not contain unnecessary information.

Terms should be self-explanatory and transparent.

The meaning of terms should not be confined to a particular context; it should be independent of the context.

Terms should be capable of providing deverbatives.

Once a term used has gained popularity and is generally accepted by the speakers of that language, it should not be changed and be substituted without persuasive reasoning. (Sager 1990:89; Taljard 2008:90)

Some of the challenges experienced during terminology development in isiNdebele occurred during the term creation processes where inconsistencies in terminology were encountered. The usage of a hyphen in the newly coined terms is not in accordance with the prescripts of the spelling or orthography rules of isiNdebele. Some of the newly coined terms are very long and are not hyphenated whereas others are long and hyphenated. For example, on one hand the lexical item Ithungelelwanohlanganiso is a compound referring to an 'intranet' is long and is not hyphenated. On the other hand, the lexical item 'Umbiko-mthethokambiso' is also a compound referring to a 'white paper' is long but is hyphenated. The newly coined terms 'ithungelelwanohlanganiso' and 'umbiko-mthethokambiso' are not generally always understood nor are they generally known by non-professional speakers. Thus, they pose a challenge in isiNdebele

This study examines isiNdebele terminology and specifically some of the new terms found in the Information and Communication Terms (2003), the Multilingual Mathematics Dictionary (2005) and the Multilingual Soccer Terminology (2009). After examining IsiNdebele terminology and specifically some of the new terms found in the above-mentioned terminologies, solutions for the challenges caused by these neologisms are afforded. Specific examples of the challenges are articulated in the full-length article. Finally, the discussion aims to provide some input on amendments regarding aspects of these terms as mentioned earlier, that were coined in isiNdebele terminologies through compounding for example. It is hoped that this study will make a contribution in terms of suggesting possible ways for developing user-friendly and acceptable isiNdebele terminology.

References

Department of Arts and Culture. 2003. Multilingual Information and Communication (ICT) terms. Pretoria: Department of Arts and Culture.

- Department of Arts and Culture. 2005. Multilingual Mathematical Dictionary. Pretoria: Department of Arts and Culture.
- Department of Arts and Culture. 2009. Multilingual Soccer Terminology. Pretoria: Department of Arts and Culture.
- Sager, J.C. 1990. *A Practical Course in Terminology Processing*. Amsterdam/ Philadelphia: John Benjamins.
- Taljord, E. 2008. *Issues in Scientific terminology in African Languages*. In M. Lafon & V. Webb (eds.) *IFAS Working Paper Series*, 88-92.

French morphogrammic system and lexicographical treatment of homophony in Gabon: the distinction between the masculine and feminine gender in personal pronouns

Edgard Maillard ELLA (ellamaillard@yahoo.fr)

Applied Languages, Literature and Communication Center/Human Sciences Research Institute,
Scientific and Technologic Research National Center, Libreville, Gabon

According to existing works, there is no distinction between the masculine and feminine gender in personal pronouns of Gabonese local languages when referring to the third persons singular and plural. These words are treated as referring equally to both genders. In Fang for example (dialect « Ntumu » of Bitam), a French/Fang monodirectional or monoscopal bilingual treatment of « il » (he) and « elle » (she) when they are verb subjects referring to a person goes as follows:

- Français/Fang

il /il/ PRON (*sujet du verbe ► se référant à un homme, un garçon*) A ; **il mange** A dzi
il et elle identiques ► **il ou elle mange** A dzi

elle /ɛl/ PRON (*sujet du verbe ► se référant à une femme, une fille*) A ; **elle (il) mange** A dzi ► **il**

- (French/Fang)

he /he:/ PRON (*subject of a verb ► referring to a man, a boy*) A;
he eats A dzi
he and she identical ► **he or she eats** A dzi

she /ʃi:/ PRON (*subject of a verb ► referring to a woman, a girl*) a; **she (he) eats** A dzi

We would agree with this treatment if the nonappearance of modification in the pronunciation of these pronouns when referring to a man or a woman was not one of the reasons why Gabonese local cultures are regarded in general as having a lack of knowledge regarding the distinction between the masculine and feminine gender. This distinction being a common sense based on sex criteria and therefore clearly known in these cultures, a noun can be masculine or feminine when referring to a man or a woman. In Fang, man and woman for example, are then indicated by *fàme* /fàm/ and *minenga* /minəŋgà/ and not by a same word.

We think this problem can be solved by considering that we are dealing with homophones as there are many others in these languages such as *otàne* /otàn/ meaning 1) umbrella and 2) bat in Fang, and not with a lack of knowledge of this gender. For a suitable lexicographical treatment, we suggest two forms not altering the pronunciation but aimed at distinguishing the dealt pronouns in writing when referring to a man or a woman.

In this regard, we use the French morphogrammic system which consists in adding at the end one or more graphemes called morphograms aimed at not being audible but to give grammatical information. The S of the plural is therefore a grammatical morphogram which distinguishes for example, the personal pronouns of the third person singular « Il » /il/ (He) from the plural « Ils » /il/ (They, referring to men, boys) or « Elle » /ɛl/ (She) de « Elles » /ɛl/ (They, referring to women, girls).

In Fang, for example, from the forms regarded as masculine *A* and *Ba* of the personal pronouns of the third person singular and plural, we will add a grammatical morphogram *a* to the final to have the feminine ones in *Aa* and *Baa*. This will result in the following presentation: Français/Fang

il /il/ PRON (*sujet du verbe ► se référant à un homme, un garçon*) a ; **il mange** A dzi

elle /ɛl/ PRON (*sujet du verbe ► se référant à un femme, une fille*) a ; **elle mange** Aa dzi

- (French/Fang)

he /he:/ PRON (*subject of a verb ► referring to a man, a boy*) A; **he eats** A dzi

she /ʃi:/ PRON (*subject of a verb ► referring to a woman, a girl*) Aa; **she eats** Aa dzi

In conclusion, referring to the morphogrammic system of French is an exchange of methods between international communication and indigenous languages. The last ones starting to develop from a scientific perspective, this will have an impact in dictionaries. Considering that dealing diversely and equally with language and culture is a lexicographer-specific competence request in any language dictionary project (Kalonji, 1993: 39), the morphogrammic system of French can help with a better depiction of local culture knowledge such as the distinction between the masculine and feminine gender in Gabon. This system reminds that even in African languages, writing must be an analytical effort of conceptual clarification (Houis, 1971: 190). Indeed, writing is not a mere record of sounds (Olson, 1999:7), by fixing thoughts, it is the formulation and the specification of what can only be orally suggested (Lurçat, 1988:1).

References

- Houis, M. *Anthropologie linguistique de l'Afrique noire*. Paris, Presses Universitaires de France, 1971.
- Kalonji, M. T. *La lexicographie bilingue en Afrique francophone : L'exemple français-cilubà*: Paris, L'Harmattan, 1993.
- Lurçat, L. 1988. De la nécessité de l'écriture manuscrite comme écriture première. *Psychologie scolaire*, n°66, avril, <http://www.sauv.net/lurcat3.htm#1>
- Olson, D.R 1999. L'Univers de l'écrit. Comment la culture écrite donne forme à la pensée. *Echange et lien social*, N° 93-Avril. Paris: Editions Retz.

Theoretical Perspectives for a Comprehensive Dictionary of Gabonese French

Paul A. MAVOUNGOU (moudika2@yahoo.fr)

Département des Sciences du Langage, Université Omar Bongo, Libreville, Gabon

Department of Language Education, University of the Western Cape, South Africa

Virginie OMPOUSSA (vompoussa@yahoo.com)

Département des Sciences du Langage, Université Omar Bongo, Libreville, Gabon

Hugues Steve NDINGA-KOUMBA-BINZA (nkbinza@uwc.ac.za)

Department of Language Education, University of the Western Cape, South Africa

The variety of the French language as it is spoken in Gabon, i.e. Gabonese French, has variously been described in a number of linguistic fields. It has subsequently been recognized not only as one of the varieties of this worldwide language, but also as one of the local languages of culturally multilingual but officially monolingual Gabon. In the field of lexicography, attention has particularly been given to the lower mesolectal forms (popular Gabonese French), basilectal forms (Gabonese Matitis French) as well as to slang forms (Toli-bangando) with the production of a series of dictionaries. A recent debate on the content of these dictionaries has reached the conclusion that these dictionaries actually do not represent the full scope of Gabonese French (henceforth abbreviated GF). Within the inception of GF lexicography as a core component of the emerging Gabonese lexicography, issues have been raised whether Gabon should stop importing French dictionaries from France and rather start compiling its own dictionaries.

In response to these issues, the (by now quite progressed) project for the planning of a comprehensive monolingual Dictionary of Gabonese French on Historical Principles (henceforth abbreviated DGFHP) was launched by the *Centre de Recherche en Etudes Germaniques et Interculturelles* at Omar Bongo University in Libreville, Gabon. The dictionary will be a comprehensive one in the sense that it will list as many as possible words and expressions attested in GF, namely acrolectal forms (standard French), upper mesolectal forms (common Gabonese French), popular Gabonese French, Gabonese Matitis French and Toli-bangando. In fact, the aim of the projected dictionary is to present the full spectrum of the lexicon of GF.

Thus, the most significant feature of the DGFHP is the lack of a purist bias. Within the Gabonese language situation, GF has a central position. GF has a standard form but also numerous varieties. So-called acrolectal forms of Gabonese French are used in official circles (administration, schools, media, etc.) and they will be included within the central list of the prospective dictionary only if they occur in Gabon. This principle of selection will avoid the repetition of everything that is said in any monolingual dictionary of standard French. The standard variety of GF, upper and lower mesolectal forms (Gabonisms especially) as well as GF slang recorded in the planned dictionary will be extracted from both written and oral sources.

Language dynamics and the daily contact between Gabonese languages have a constant influence on ever-changing Gabonese French. As a matter of fact, French is spoken differently in the nine provinces of the country. Although being a single volume dictionary, the DGFHP will present a quite comprehensive selection of words and expressions from the lexicon of GF. Focus will primarily be given to words and expressions which are particular to Gabonese. This includes words and expressions borrowed from the many languages of Gabon. Some of the words and expressions in DGFHP are not Gabonese in origin but have a particular significance for Gabonese. Some of these items are from Senegal (*essencerie* “petrol station”), Côte d’Ivoire (*avenir radieux* “young women breasts”), Cameroon (*kongossa* “gossiping”), etc.

The macrostructure of DGFHP not only gives evidence of the massive set of data from earlier dictionaries of Gabonese French (Boucher and Lafage 2000), but it describes the maximum of words and expressions attested in Gabon. The different types of microstructural categories presented in the treatment of lemmata include pronunciation, part of speech indicators, labels, variant spellings, morphological data, etymology, idioms and collocations, encyclopaedic notes and citations. The lemmata received a comprehensive treatment and each information category is dealt with in a systematic way. With regard to etymology, the word or the expression in the source language as well as the name of the source language will be given.

Based on the methodology proposed by Wiegand (1998: 151) and Gouws (2001: 65-72), the planning of lexicographic activities in any country should ideally be governed by a *lexicographic*

process, i.e. all the activities leading to the publication of a dictionary as a text. A distinction is usually made between the *primary comprehensive lexicographic process* and the *secondary lexicographic process*. The establishment of a lexicographic process leads to the formulation of a *dictionary plan* encompassing two main components, namely *the organisation plan* (directed at the logistics of the project as well as managerial aspects) and *the dictionary conceptualisation plan*. The dictionary conceptualisation plan can be divided into five subdivisions, i.e. *the general preparation phase*, *the material acquisition phase*, *the material preparation phase*, *the material processing phase* and *the publishing preparation phase*. As part of the secondary lexicographic process, the present dictionary project will give some attention to all these five subdivisions.

The aim of this paper is to present the abovementioned metalexicographical (theoretical) perspectives of the planned dictionary. The paper will cover three main focus areas. It will first highlight the project background and rationale as well as the need for the planned dictionary. The area of focus concerns the research methodology and the data collection. Finally, the paper deals with the two core components of the dictionary nomenclature, the macrostructure and the microstructure. The planned dictionary as well as the current paper will contribute to the theoretical groundwork of Gabonese French lexicography (cf. Nyangone Assam et al. 2016, Mavoungou 2013). Equally, the availability of such a dictionary may be an important step towards the codification of the French variety of Gabon.

References

- Boucher, K. & S. Lafage. 2000. *Le Lexique Français du Gabon (entre Tradition et Modernité)*. *Le Français en Afrique* 14. Special Issue. Nice: Institut de Linguistique Française.
- Gouws, R. H. 2001. Lexicographic Training: Approaches and Topics. Emejulu J.D. (Ed.). 2001: *Éléments de lexicographie gabonaise. Tome I*. New York: Jimacs-Hillman Publishers. 58-94.
- Mavoungou, P.A. 2013. Gabonese French Dictionaries: Survey and Perspectives. *Lexikos* 23: 255-272.
- Nyangone Assam, B., H.S. Ndinga-Koumba-Binza & V. Ompoussa. 2016. What French for Gabonese French Dictionaries? *Lexikos* 26: 162-192.
- Wiegand, H.E. 1998. *Wörterbuchforschung*. Berlin: Walter de Gruyter.

Do we need dictionary skills in the digital era?

Lorna Hiles MORRIS (lorna@dictionary.co.za)

Department of Afrikaans and Dutch, Stellenbosch University, South Africa

Are people using the internet to look up words that they would previously have used a dictionary to find? Are people even using dictionaries anymore?

This paper will explore the answer to the question, “do we need dictionary skills in the digital era?” by identifying traditional dictionary skills and identifying which ones are applicable to alternative looking up options. The alternative looking up options that are discussed are electronic dictionaries, as well as online resources, including search engines and online dictionaries.

A short survey has been done amongst local teenagers to determine whether they use dictionaries for their word lookups, or whether they turn to the internet. The results of this survey are discussed in this paper. The implications of these results will also be discussed. Teenagers were selected as the focus group due to their familiarity with digital resources. This group will also have been through primary school, where they would have been taught dictionary skills.

Preliminary results, which will be discussed further, are that most of the teenagers surveyed go straight to Google to search for a word. They then typically use the first search result that comes up. Some (very few) teenagers do still prefer to use a dictionary, as they are more likely to trust the results.

Electronic dictionaries, which come in the form of an app on a cellphone or software for a computer, require users to employ some of the same strategies as printed dictionaries, and some different strategies: an electronic search requires different skills to a hard copy consultation. Another alternative is to simply Google the word. The pitfalls of these alternatives are discussed in this paper,

as are the advantages. Some examples of dictionary look ups and internet searches will be provided, compared, and discussed. This paper will also make a case for the use of dictionaries as opposed to alternative reference sources such as the internet.

Some of the skills that will be discussed are: knowing alphabetical order, knowing how to use the printed dictionary's access structure, knowing basic spelling rules and conventions. One also needs to know what grammatical and other indicators and abbreviations mean. Skills such as knowing alphabetical order are no longer necessary if one is searching for a word – one types in the word to search for it instead of looking it up. But skills such as sense determination in a polysemous dictionary article would still be necessary.

Different authors provide a variety of taxonomies of dictionary skills, for example Nesi (1999: 55), Lew and Galas (2008: 1274), and Chi (1998: App I). This paper will discuss these skills and which ones, if any, are transferable to alternative sources such as the internet.

Dictionary skills are traditionally taught in the classroom, and help users to get the most out of their dictionaries to make each consultation as successful and quick as possible. If users are no longer using dictionaries, and turn instead to search engines on the internet, will dictionary skills still help them?

This paper will also include a discussion of the relevant literature, which looks at dictionary skills as well as the future of lexicography.

Grefenstett (1998, quoted in Rundell (2012:17)) asks whether there will be lexicographers in the year 3000, with reference to the corpus building and analysing tools and other technological advances that exist in computational linguistics. Rundell (2012:29) rephrases the question and asks instead, “will there be dictionaries?”

This paper follows on from that question by asking, essentially, “will there be dictionary users?”

References

- Chi, M.L.A. 1998. Teaching dictionary skills in the classroom. In: *Euralex '98 Proceedings*. August 1998. Belgium: pp. 565-577. Available: <https://euralex.org/publications/teaching-dictionary-skills-in-the-classroom/>
- Lew, R & Galas, K. 2008. Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-school-level Polish Learners of English. In: *Proceeding of the 13th EURALEX International Congress*. July 2008. Spain:1273-1285.
- Nesi, H. 1999. The Specification of Dictionary Reference Skills in Higher Education. In R.R.K. Hartmann (ed.). *Dictionaries in Language Learning: Recommendations, National Reports and Thematic Reports from the TNP Sub-Project 9: Dictionaries*. Freie Universität Berlin: pp53-67
- Rundell, M. 2012. The road to automated lexicography: An editor's viewpoint. In S. Granger & M. Paquot (eds.). *Electronic Lexicography*. Oxford: Oxford University Press.

Lexicography in Zimbabwe: Prospects and challenges post the 2013 Constitution of Zimbabwe Amendment No.20 Act and the Zimbabwean Education Blueprint 2015 – 2022 Curriculum Framework for Primary and Secondary Education

Eventhough NDLOVU (evennthough@yahoo.co.uk)

Unit for Language Facilitation and Empowerment, University of the Free State, Bloemfontein, South Africa

This paper examines the prospects and challenges of the growth of lexicography in Zimbabwe post the 2013 Constitution of Zimbabwe Amendment No.20 Act and the Zimbabwean Education Blueprint 2015 – 2022 Curriculum Framework for Primary and Secondary Education. It seeks to examine whether or not there are prospects to increase the number of languages in which the African Languages Research Institute (henceforth: ALRI) compiles its dictionaries and the challenges which are likely to be encountered. Section 6 of the Constitution of Zimbabwe states that there are 16

officially recognised languages. The section further obliges the state, its institutions and agencies of government at all levels to treat equitably and promote and advance the use and development of the said languages. It also obliges the state, its institutions and agencies of government at all levels to take into consideration the language preferences of people affected by government communication or measures. Given these provisions, a critical discourse analysis of the policy documents in question is useful in examining the prospects and challenges of the growth of lexicography. Semi-structured interviews with publishers, policy makers and selected language practitioners at ALRI and in the Departments of African and Foreign Languages and Literature in four selected state universities as well as critical discourse analysis of the policy documents in question, show that the provisions of the new Constitution and the new curriculum framework do not provide an enabling environment for the growth of lexicography in Zimbabwe, especially for the previously marginalised officially recognised languages. It was noted that the provisions of these policy documents are weak and neutralised and constitute a form of declaration without implementation. However, interactions with practitioners in the translation and interpretation industry showed that Sections 6 and 70 of the Constitution exhibit a strong awareness of the need to intensify efforts to develop the previously marginalised indigenous languages and to promote multilingualism and multilingual service provision in Zimbabwe.

It was further observed that there is a clear wish to promote and create conditions for the development and use of all the officially recognised languages as well respect citizens' language preferences and linguistic human rights expressed through the obligatory must; a leeway which might pave way for the growth and development of the language mediation industry which will in turn necessitate the compilation of various dictionary types in all the officially recognised languages, given that dictionaries are indispensable tools for language mediators. Acknowledging the close-knit relationship between lexicography and interpreting, language mediators expressed very strong doubts about the growth of court interpreting, due to Sections 6 and 70 of the Constitution. They noted that the reluctance and delay to align other court policy documents with the provisions of the Constitution is clear evidence that the Constitution falls largely in the realm of declaration without implementation, and the prospects for compiling legal dictionaries of various types and in a variety of languages are very limited. From the viewpoint of the education sector, dictionaries are valuable teaching and learning materials across all subjects in the curriculum and there is a great need for lexicographic aids in order to improve native language proficiency and facilitate and support the learning and mastery of foreign languages. Noting the roles of dictionaries in the education sector, educationists indicated that the new curriculum framework falls largely in the realm of declaration without implementation in as far as mother tongue education, use of all officially recognised languages and teaching of foreign languages are concerned. They argued that in as much as the framework foregrounds mother tongue education and teaching of foreign languages, it does not specify the necessary implementation guidelines for the successful implementation of its provisions. Some indicated that the policy was declared which in the circumstances cannot be implemented, and the policy makers are aware of this, especially given that the policy is not accompanied by the requisite seven areas of policy development for successfully implementing a language-in-education policy. As such, they concluded that new curriculum framework is not likely to stimulate and promote the compilation of new dictionary types and dictionaries in the recently acknowledged previously marginalised languages and newly introduced foreign languages.

References

- Government of Zimbabwe. 2013. Constitution of Zimbabwe Amendment (No. 20) Act. Harare: Government Printers.
- Ministry of Primary and Secondary Education. 2015. Zimbabwe Education Blueprint 2015 – 2022 Curriculum Framework for Primary and Secondary Education. Harare: Government Printers.

Analysis of the French-English Bilingual dictionaries in use at the Junior Secondary Schools in Nigeria

Rosemary N. OSSAI (rnossai.unilag.edu.ng@mail.com)

Department of Arts and Social Sciences Education, Faculty of Education: University of Lagos, Lagos, Nigeria

Carol C. OPARA (copara@unilag.edu.ng)

Department of Arts and Social Sciences Education, Faculty of Education: University of Lagos, Lagos, Nigeria

A dictionary is a reference source that gives users information on the meanings, definitions and usages of words. Dictionaries do vary in size, scope, content and type such as monolingual, bilingual and specialized dictionaries. Dictionary strategy helps learners of French as a Foreign Language (FFL) to understand word meanings and usages better. There is no doubt that the stock of individual vocabulary plays a key role in attaining communicative competence. According to Schmitt (2000) lexical knowledge is key to communicative competence. The use of a dictionary gives learners quick and handy access to information either for independent or guided study Walz (2002). Wright (1998) also notes that a dictionary is one of the most readily accessible educational resources that provides a wealth of information to users. Dictionaries can give users additional information which is not readily found elsewhere such as giving the grammatical class of the words: noun, adjective, verb forms, gender etc. Some dictionaries can go as far as using the words in context. According to Nation (2001) the use of dictionary goes beyond providing information to understanding and giving a detailed analysis of a text. Laufer (1990) encourages learners to use dictionaries as a last resort in situations where they are confused about the meaning of a word. Despite the useful role dictionaries play in learning words, learners at junior secondary school level have difficulties using their dictionary either because the dictionaries in use are not suitable or the users do not know how to use the dictionary.

To investigate these, a survey was carried out to analyze the dictionaries students use at junior secondary schools in Nigeria in order to determine their suitability in helping students to learn French vocabulary. In the course of study, six research questions were asked and four hypotheses tested: RQ1. What type of dictionary do you use? RQ2. Which is more practical to check a new word, dictionaries or glossaries? RQ3. Do you find the words in your dictionary ambiguous? RQ4. Does your dictionary contain phonetic symbols? RQ5. Are you motivated by the page layout? RQ6. Does your dictionary contain relevant socio-cultural traits to enhance vocabulary acquisition? Hypotheses: H1. There will be no significant relationship between ambiguity and the effective vocabulary acquisition. H2. The existence of phonetic symbols in the dictionary will not have a significant relationship with the learning of vocabulary. H3. There will be no significant relationship between the page layout and the motivation of the learners. H4. There will be no significant relationship between socio-linguistic traits in the dictionary and the learning of vocabulary.

This study is significant in its bid to improve the students' French vocabulary acquisition and communicative competence as well as provide more resources for the teaching and learning of French. For the data collection and statistical analysis, about three hundred teachers and students were used from three private and two public schools in the Lagos metropolis. Information was gathered through random sample questionnaires, classroom observation, and interviews with students, and teachers. The questionnaire was designed according to the Likert scale. Simple percentage score was used considering 70% as significant since the Nigerian educational system bases excellence at 70%. Results: About 80% of students are faced with problems of ambiguity when they check a new word, 83% of students and 80% of teachers prefer to learn French with coloured picture dictionary. About 80% of teachers said it is necessary to teach phonetics while 33.3% of students do not see the necessity, 31.3% of students say they easily retain words they consult in the dictionary, 65% of students and 40% of teachers say the dictionary plays a more vital role in checking a new word than the glossary of the textbook. From the result analysis the dictionaries in use, also lack socio-cultural factors and aesthetic values to motivate students. This study found lacunas in the dictionaries currently in use and recommends the compilation of a new French-English Picture Dictionary for Schools.

References:

- Laufer, B. 1990. "Ease and difficulty in vocabulary learning: some teaching implications". *Foreign Language*. Vol. (23) issue 2: 147-155.
- Nations, I. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Schmitt, N. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Walz, J. 1990. "The dictionary as a secondary source in language learning". *The French Review*. Vol. 64 No 1: 79-94.
- Wright, J. 1998. *Dictionaries: Resource books for teachers*. Oxford: Oxford University Press.

The Effects of the Multimedia Package in Teaching Vocabulary within the Nigerian French Language Curriculum

Rosemary N. OSSAI (rnossai.unilag.edu.ng@mail.com)

Department of Arts and Social Sciences Education, Faculty of Education: University of Lagos, Lagos -Nigeria

Carol C. OPARA (copara@unilag.edu.ng)

Department of Arts and Social Sciences Education, Faculty of Education, University of Lagos, Lagos-Nigeria

Learning French in a Nigerian linguistic environment is characterized by socio-linguistic problems such as learning French in an environment where English language is predominantly spoken. Emordi (1986) notes that secondary school learners of French as a foreign language are unable to communicate in simple everyday French. Inability to communicate effectively can be associated with the dearth of vocabulary. Folse (2004) affirms that it is necessary to have a certain level of vocabulary in order to communicate. Some research works investigate the use of e-dictionary strategies and references on guessing meanings and usages as was identified by Koyama (2009:131-150). Koyama again in (2015) enquires about the effectiveness of pocket electronic dictionaries in the English as a Foreign Language class through reading comprehension. Nation (2001) also acclaims that e-dictionaries are an effective tool in second language acquisition.

Cajkler (1993) remarks that the full benefits of IT for education will be realised only when teachers at different levels and in every subject exploit it in order to meet their curriculum objectives. Owhotu (2006:146-147) added that ICT is a vital and active learning tool in the delivery and assessment of high quality curricula in order to enhance students' inquiry, interpretation and sustainable engagement. Rod & Chang (2016) and Qiufang (2018) found that text repetition had a positive effect on both comprehension and vocabulary acquisition. Brief vocabulary instruction after the listening activity normally will lead to more effective recall than listening only. Tar, (2013) recommended the development of instructional materials for teaching vocabulary throughout the French language curriculum. Sokmen (1997) indicated that a learner comes across the word between 5 to 16 occurrences in order to assure its retention.

This study developed a multimedia package which integrates some recommendations from past research findings such as developing an instructional material to increase the frequency of learner's intentional vocabulary acquisition which will motivate learners to learn more vocabulary in a context using the e-dictionary strategy. It also investigates the effectiveness of the package. The innovative multimedia package consists of twenty audio listening comprehension passages based on the West African Examination Council, glosses, vocabulary exercises, and the Harrap's Multimedia e-dictionary.

Four research questions and hypothesis are developed. The study adopts a pre-test and post-test control group quasi-experimental design. The population for the study covers all public school students offering French as a subject in senior secondary schools II in Lagos. Intact-non randomized classes will be used for a duration of eight weeks. Two schools will be taken as the experimental group taught using the multimedia package, while the other two, the control group will be taught using the lecture method. Reliability of the instruments is content validated by experts in statistics,

measurements and evaluation, by linguists and by lexicographers. The main instrument used for data collection is the French Vocabulary Achievement Test Multiple Choice (FVAT) developed by the researcher. Kuder Richardson Formula 20 (KR20) is used to establish the reliability of the instrument. The internal consistency of the instrument is 0.76. The result of a pre-test post-test carried out after two weeks in a preliminary survey shows a significant difference between post vocabulary test scores of students taught with the innovative multimedia package.

References:

- Cajkler, W. 1993. In Owhotu, V.B. *An introduction to Information Technologies in Education*. Lagos: Sibon Books.
- Emordi, F. 1986. Report on The French Programme Inspection of The Bendel State Universities Primary and Secondary Schools. Cited in *African Higher Education Review* (6) 2012.
- Folse, K. 2010. Is Explicit Vocabulary Focus the Reading Teacher's Job? *Reading in a Foreign Language* (22)1:139-160.
- Koyama, T. 2009. How effectively do good language learners use handheld electronic dictionaries. *Language Education and Technology* (46) 131-150.
- Koyama, T. 2015. Investigating the effectiveness of the uses of electronic and paper-based dictionaries in promoting incidental word learning. *ICHL 2015 Hybrid learning: Innovation in Education practices*, pp 59-69.
- Rod & Chang 2016. The Effects of Inference-Training and Text Repetition on Chinese Learners' Incidental Chinese. *Journal of Applied Linguistics vocabulary acquisition while listening* In Ed. by Wen, Qiufang. 2018. *Chinese Journal of Applied Linguistics* (41) 1.
- Nation, I. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Owhotu, V.B. 2006 *An introduction to Information Technologies in Education*. Lagos: Sibon Books.
- Wen, Qiufang. 2018. *Chinese Journal of Applied Linguistics* (41) 1. Assessment Literacy of Secondary Teachers: Evidence from a Regional EFL Test
- Sokmen, A. 1997. Current Trends in Teaching Second Language Vocabulary. Tar, M. 2013 *L'acquisition du Vocabulaire Du Français Langue Etrangère dans Les Universités Nigériennes*. Eprints.covenantuniversity.edu.ng.

Stem lemmatization and phonetic ordering of lemmas in Sotho dictionaries from a user perspective

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

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

Tello THELETSANE (theletsanett@ufs.ac.za)

Sesiu sa Sesotho National Lexicography Unit: African Languages, University of the Free State, Bloemfontein, South Africa

This presentation reopens the debate on stem versus word lemmatization and alphabetical ordering of lemmas in paper dictionaries with the focus on the so-called Sotho languages Sepedi, Setswana and Sesotho. Much debate has been ongoing in respect of the Nguni languages, cf. Prinsloo (2011). As far as the Sotho languages are concerned, haphazard references to stem versus word lemmatization have been made by Prinsloo (2009). As a point of departure a brief summary of existing viewpoints in respect of lemmatization and alphabetic ordering will be given followed by a comprehensive account of the full impact that especially the *combination* of stem lemmatization and phonetic ordering have on the users of dictionaries for the Sotho languages.

It will be argued in support of Van Wyk (1995) that stem lemmatization brings no gain but it imposes an unnecessary burden on especially the inexperienced user to find words. Exactly the same goes for a phonetic instead of an ordinary alphabetical ordering. When stem lemmatization and phonetic ordering are combined it is even worse and even experienced users struggle to look up words

in such dictionaries and often have to revert to a guidance page, if provided, or can even incorrectly conclude that the word is not in the dictionary.

The *Dikišinare ya Setswana English Afrikaans* (Snyman et al.:1990), *Southern Sotho-English dictionary* (Mabille & Dieterlen:1988) and the *Comprehensive Northern Sotho Dictionary* (Ziervogel and Mokgokong:1975) will be analyzed for Setswana, Sesotho and Sepedi, respectively. Consider a single example from each of the three dictionaries:

In *Dikišinare ya Setswana English Afrikaans* the user is informed in the preface to the dictionary that two different look-up strategies have to be followed. In some cases even the singular versus plural forms of the same word, e.g. *mmutla* ‘hare’ have to be looked up on its first letter within the stretch M but its plural form *mebutla* ‘hares’ under the third letter of the word in the alphabetical stretch B.

If the user wants to look up *mohla* ‘day’ in the Southern Sotho-English dictionary (s)he won’t find it under M because the user is supposed to know (without any guidance in the front matter of the dictionary) that since a stem lemmatization strategy is followed, the prefix *mo-* has to be removed and that the lemma that should be looked up is *-hlwa*. However, the lemma is not found under H. This category runs from stems beginning with *ha-*, *hi-*, *ho-*, ... *hwi-* but words starting with *hl-* are not listed – they are given under a following main stretch HL as *mo.hla*.

The same situation prevails in the *Comprehensive Northern Sotho Dictionary* but at least the compilers realized that the phonetic ordering is problematic and provided a guiding page in the introduction:

A	1	P	934
B	45	PH	997
BJ	143	PS	1075
D	146	PSH	1076
E	174	PŠ	1078
F	195	PŠH	1079
FS	267	R	1082
FŠ	268	S	1153
G	269	Š	1199
H	361	T	1245
HL	386	TH	1308
I	454	TL	1375
J	470	TLH	1409
K	475	TS	1418
KG	579	TSH	1456
KH	710	TŠ	1475
L	734	TŠH	1493
M	784	U	1517
N	834	W	1524
NG	875	Y	1529
NX	895	Z	1533
NY	896	Bylae	
O	915	Appendix	1534

From this table it is clear that a single main stretch T in the dictionary is presented as no less than eight separate main stretches and for P no less than six main stretches.

More examples of problematic stem lemmatization coupled by phonetic ordering will be discussed.

To put it bluntly – the user is at a loss, in depth morphological and phonetic knowledge of the language is a prerequisite to using these dictionaries and they are regarded as user unfriendly. It will be concluded in reference to De Schryver & Prinsloo (1999) that stem lemmatization is unnecessary and unwanted for disjunctively written languages and that future compilers of paper dictionaries should stick to word lemmatization and to completely abandon the phonetic ordering of lemmas.

References

De Schryver, G.M. & Prinsloo, D.J. 1999. The lemmatization of nouns in African languages with special reference to Sepedi and Cilubà. *SA Journal of African Languages*, 19(4) 258-275.

- Mabille, A. & Dieterlen, H. 1988. *Southern Sotho-English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.
- Prinsloo, D.J. 2009. Current Lexicography Practice in Bantu with Specific Reference to the Oxford Northern Sotho School dictionary. *International Journal of Lexicography*. 2009; Oxford University Press (UK) 22(2) 151-178.
- Prinsloo, D.J. 2011. A critical analysis of the lemmatisation of nouns and verbs in isiZulu. *Lexikos* 21. 169-193.
- Snyman, J.W. Shole, J.S. & Le Roux, J.C. 1990. *Dikšinare ya Setswana English Afrikaans Dictionary Woordeboek*. Pretoria: Via Afrika.
- Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.
- Ziervogel, Dirk & Pothinus C. Mokgokong. 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa – Seburu/Seisimane / Groot Noord-Sotho-woordeboek, Noord-Sotho – Afrikaans/Engels / Comprehensive Northern Sotho Dictionary, Northern Sotho – Afrikaans/English*. Pretoria: J.L. van Schaik.

Lemmatization of kinship terminology in Setswana

Fannie SEBOLELA (fanniesebolela@gmail.com)

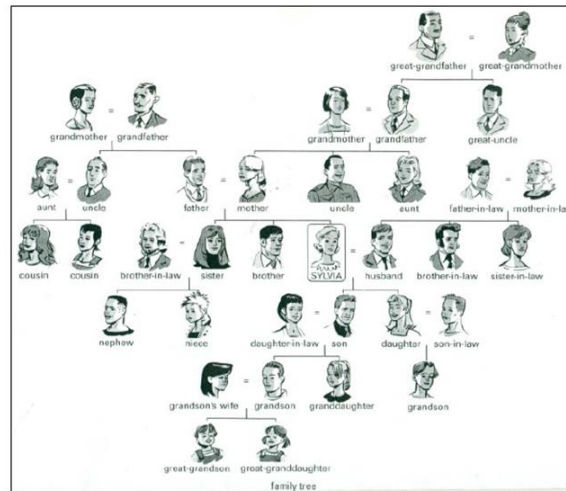
Curro Academy, Soshanguve, South Africa

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

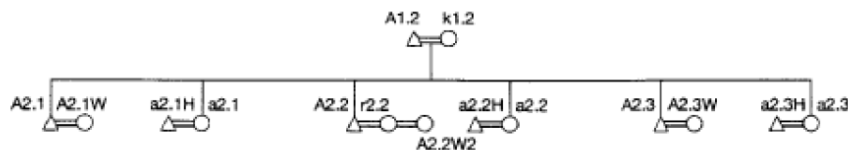
Department of African Languages, University of Pretoria, Pretoria, South Africa

The inclusion and treatment of kinship terminology in African languages, and in particular for Setswana, pose a number of challenges to the lexicographer. Firstly, (s)he has to deal with a complicated system where in contrast to English, for example, several terms exist for a specific relative, e.g. brother or sister depending on his/her position in the family tree as e.g. a relative on father's or mother's side, of the person speaking, age and also if the relative is addressed or simply referred to. Secondly, kinship terms in Setswana occur as single words, compounds and phrases counting up to thousands of instances which ideally should be included and treated in a dictionary but which is practically impossible in a paper dictionary. Lemma selection of single words is in itself problematic in the compilation of a small paper dictionary but could for instance be done on the basis of frequency of occurrence, i.e. to lemmatize the most frequently used ones. For compounds and phrases, however, the situation is more problematic and it is suggested in terms of Prinsloo (2012) that a specific dictionary convention be designed to cater for all elements preceding and following the basic terms as well as for phrases.

For lexicographic purposes, in order to progress in a systematic way, the lexicographer typically departs from a matrilineal or patrilineal approach taking a specific man or woman as the center and describes the relatives of this person in a systematic way by meticulous consideration of each generation. Consider the following diagram from the *Macmillan English Dictionary* (Rundell:2002) as a typical example of the family tree of a woman called SYLVIA.



This paper departs from a brief description of a much more complicated generic family tree for African language terminology adapted by Van Wyk (1988). This complex schematic family tree is interpreted in terms of single-word Setswana kinship terms versus phrases of Setswana kinship. The tree revolves around a male person indicated as EGO “I” and consists of 18 clusters linked to EGO, for example the brothers “Δ” and sisters “O” of a man A2.2 as illustrated in the following section of the family tree:



F2-generation *	Ego A2.2 (male)	Ego r2.2 (female)
-----------------	-----------------	-------------------

Older generation: Eldest brother and his wife: A2.1, A2.1W, Eldest sister and her husband: a2.1, a2.1H, Younger generation: Youngest brother and his wife: A2.3, A2.3W, Youngest sister and her husband: a2.3, a2.3H, etc. (Molalapata 2004:29)

A comprehensive list of Setswana kinship terms was compiled from field work (Molalapata:2004) supplemented by the authors of this paper. This list was then compared to a Setswana corpus giving an indication of the frequency of use of the different terms.

As a next step the kinship terms are studied for inclusion in or omission from four bi-/multilingual Setswana dictionaries: the *Oxford English-Setswana*, *Setswana-English School Dictionary* (Otlogetswe:2013), *English-Setswana Dictionary* (Otlogetswe:2009), *Thanodi ya Temepedi Setswana-English* (Mareme :2015) and *Thanodi ya Kgatho ya Motheo* (Ramagoshi: 2010).

An evaluation of the quality of treatment of the kinship terms will be offered. It will be argued that future Setswana dictionaries should lemmatize more kinship terms and offer enhanced treatment of such terms especially for terms such as *kgaitsadi* where a complex set of meanings is applicable.

It will be attempted to design a few model entries which could serve as a guideline.

Finally a lexicographic convention for the lemmatization of Setswana kinship terms which occur as phrases, *mogatša nkgonne yo mogolo* ‘eldest brother’s wife’, *mogatša kgaitsadiake yo mogolo* ‘my oldest sister’s husband’ or compounds, e.g. *mogatsake* ‘my wife’ will be suggested. This lexicographic convention is based upon the design approach of Prinsloo (2012:283) for Sepedi where an attempt was made to cater for all elements preceding and following the basic term rendering conventions such as: **ga/bo/mma/mogatša/~ake/ago/agwe/abo(na)/ake/atšo(na)/alena** or **bo/mma/mogatša/~ago/agwe**

It will be concluded that although kinship terms are lemmatized and treated in the four Setswana dictionaries studied, more terms need to be included and the quality of treatment should be enhanced.

References

- Mareme, G.B. (Ed.). 2015. *Thanodi ya Temepedi Setswana-English*. (Setswana National Lexicography Unit.) S.p.: South African Heritage Publishers.
- Molalapata, B.T. 2004. The treatment of Kinship terminology in Sotho dictionaries, with special reference to Setswana. Unpublished MA dissertation. University of Pretoria.
- Otlogetswe, T.J. 2009. *English-Setswana Dictionary*. Second Edition. Gaborone: Pentagon.
- Otlogetswe, T.J. 2013. *Oxford English-Setswana, Setswana-English School Dictionary*. Oxford: Oxford University Press.
- Prinsloo, D.J. 2012. Lemmatisering van verwantskapsterminologie in Sepedi. *Lexikos* 22. 272-289.
- Ramagoshi, R. 2010. *Thanodi ya Kgatho ya Motheo*. Cape Town: Maskew Miller Longman.
- Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners*. CD-ROM. London: Macmillan Education. Second Edition. 2007 op *CD-ROM*. London: A&C Black Publishers.
- Van Wyk, J.J. 1988. Verwantskapsterminologie: 'n Etnografiese benadering. *South African Journal of Ethnology*. (11)3.

