

AFRICAN ASSOCIATION FOR LEXICOGRAPHY

26th International Conference

27-29 June 2022





AFRICAN ASSOCIATION FOR LEXICOGRAPHY

26th International Conference

27-29 June 2022

Abstracts

Hosted by: Stellenbosch University

Conference coordinator: Dr Michele van der Merwe

Abstract reviewers: Prof Sonja Bosch, Mr André du Plessis, Dr Hanelle Fourie-Blair, Prof Rufus Gouws, Ms Megan Hall, Dr Phillip Louw, Dr Lorna Morris, Dr H Steve Ndinga-Koumba-Binza, Dr Eventhough Ndlovu, Prof Dion Nkomo, Dr Gerda Odendaal, Prof Annél Otto, Prof Danie Prinsloo, Mr Richard Slater, Prof Elsabé Taljard, Mrs Esmaré van der Merwe, Dr Michele van der Merwe, Mr Andries van Niekerk, Mr Tim van Niekerk

Abstract booklet editors: Mr André du Plessis, Dr Lorna Morris and Prof Sonja E. Bosch

Table of Contents

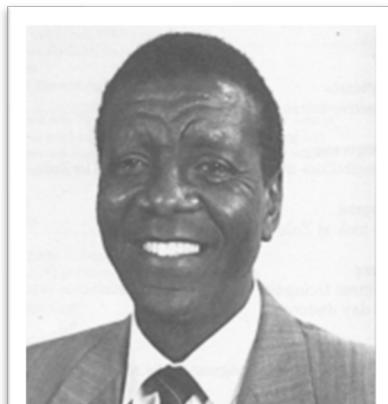
AFRILEX HONORARY MEMBERS.....	1
AFRILEX BOARD.....	2
MESSAGE FROM THE AFRILEX PRESIDENT	2
CONFERENCE PROGRAMME.....	4
KEYNOTE PRESENTATION 1	7
From agric to zol: African Englishes in the Oxford English Dictionary Danica SALAZAR	
KEYNOTE PRESENTATION 2	7
Trying to see the forest for its trees: Lexicography as communication Herman L. BEYER	
Documentary film: Dreaming of Words	9
Special Sessions	9
South African Centre for Digital Language Resources.....	9
SANLU and OUPSA Session	10
PRESENTATIONS.....	12
Promoting linguistic legitimacy via web-based dictionaries Alex BARATTA	12
A prototype Afrikaans online dictionary for academic editing purposes Maret BLOM DE WET.....	13
Experimental methodology – compiling a multilingual online dictionary Sonja BOSCH, Marissa GRIESEL & Elsabé TALJARD.....	15
Implementing dictionary skills in the language classroom Mari Carmen CAMPOY-CUBILLO	17
Improving learners' skills by compiling dictionaries: A case study Valeria CARUSO.....	19
Ancient Egyptian lexicon in the Semitic languages of Ethiopia: New paths in etymological lexicography Letizia CERQUEGLINI.....	21
Search and you will find: innovative access structures in mobile dictionaries André H. DU PLESSIS.....	22
Frequency or keyness? Zorica ĐUROVIĆ	24
The concept of <i>lexicographic condensation</i> : Review and perspectives in digital lexicography María Teresa FUENTES MORÁN.....	25
The Concept of Lexicographic Data: Some Lexicographic, Economic and Technical Reflections Pedro A. FUERTES-OLIVERA	27
Aspects of Vernacular Dictionaries: A Case Study of a Hakka Dictionary Audrey HEIJNS	28
In progress: the Survey of South African Sign Language Place Names Chrismi-Rinda LOTH, Jani DE LANGE, Susan LOMBAARD, Emily MATABANE, Gideon KOTZÉ & Annalene VAN STADEN.....	29

Towards an Integrated Approach to IsiZulu Terminography and Lexicography Njabulo MANYONI	31
The Inclusion of Lexical Particularities in the Comprehensive Dictionary of Gabonese French Paul A. MAVOUNGOU, Hugues Steve NDINGA-KOUMBA-BINZA, Virginie OMPOUSSA & Blanche NYANGONE ASSAM	33
The historical account and contribution of the <i>Greater Dictionary of IsiXhosa</i> to African Language Lexicography Buyiswa MINI	35
Help! A user guide to user guides in electronic dictionaries Lorna MORRIS.....	36
Dictionary Base and Lemmata for a Trilingual Dictionary of Fauna Hugues Steve NDINGA-KOUMBA-BINZA	38
An examination of the reception of the dictionaries compiled by the University of Zimbabwe's African Languages Research Institute (ALRI) Eventhough NDLOVU	40
How great is thy dictionary? Cross-referencing conditions and practices in <i>The Greater Dictionary of (isi)Xhosa</i> Dion NKOMO, Bulelwa NOSILELA & Wanga GAMBUSHE	42
The lemmatisation and treatment of verbal suffixes in Sepedi Danie PRINSLOO & Elsabé TALJARD	44
Turning Bilingual Lexicography Upside Down: Realization of an old dream Sven TARP	45
Using machine translation to compile bilingual lexicographical databases: An unexpected but highly welcome research outcome Sven TARP	47
To see and understand: visual elements in Foundation Phase Dictionaries Michele VAN DER MERWE	48
Towards a corpus of South African Sign Language: Sociolinguistic and Phonological Challenges Andries VAN NIEKERK.....	51
An analysis of a corpus-based approach for purposes of developing a bilingual multidisciplinary e-dictionary for students at the University of KwaZulu-Natal Tholakele ZUNGU.....	52

AFRILEX HONORARY MEMBERS



Prof R.H. Gouws



Prof A.C. Nkabinde



Dr J.C.M.D. du Plessis



Dr M. Alberts



Prof D.J. Prinsloo

AFRILEX BOARD

2021 – 2023

President:	Prof L. (Langa) Khumalo
Vice-President:	Prof S.E. (Sonja) Bosch
Secretary:	Prof D. (Dion) Nkomo
Treasurer:	Prof E. (Elsabé) Taljard
Editor Lexikos:	Prof E. (Elsabé) Taljard
Members:	Mr A.H. (André) du Plessis Dr P. (Phillip) Louw Dr L. (Lorna) Morris Dr H.S. (Steve) Ndinga-Kouumba-Binza
Conference organiser:	Dr Michele van der Merwe

MESSAGE FROM THE AFRILEX PRESIDENT

On behalf of the Board of the African Association for Lexicography (AFRILEX), I wish to officially welcome you to the 26th Annual International AFRILEX Conference. It is particularly such an honour to welcome you to this conference that we are having in person after a two-year hiatus because of a devastating Covid-19 global pandemic. I must hasten to add that we still need to adhere to necessary protocols that have been put in place during the course of our conference.

I want to thank Prof Mbulungeni Madiba for his kind words of welcome as the representative of the host university. Stellenbosch University and the town has a special place in the -Lex Family, with its very rich history in lexicography in this country. We are privileged to be back in the home of the DICTIONARY.

I wish to extend a special welcome to our keynote speakers, Dr Danica Salazar and Prof Herman Beyer. Dr Salazar is the Executive Editor (World Englishes) at the Oxford University Press, in the United Kingdom. Prof Herman Beyer is the head of the Department of Humanities and Art at the University of Namibia, Namibia. He is also the immediate past President of AFRILEX, who holds the distinction of leading the AFRILEX Board in holding its first virtual AFRILEX International Conference in 2021.

May I also extend my words of welcome to Dr Tim van Niekerk, the Director of the Dictionary of South African English at Rhodes University, South Africa, who will be facilitating the pre-conference workshop. We look forward to DSAE evolving and innovative

approaches to online dictionary presentations. A special welcome also to all our esteemed honorary members who are in attendance.

Our conference programme for this year has 27 presentations covering a broad spectrum of very interesting research topics. It also has two special sessions, one from a research infrastructure that focuses on developing digital language resources (SADiLaR Session) and another session that has a special focus on dictionary publishing (Publishers' Session). I would like to invite every delegate to attend and listen to all these engaging sessions because our programme allows us to do that. I want to thank Prof. Danie Prinsloo in this regard for compiling the programme. I would like to take this opportunity to invite all the presenters to develop their presentations into article manuscripts and submit them for possible publication in our peer-reviewed journal Lexikos (Go to: <http://lexikos.journals.ac.za>). I want to further express my gratitude to the abstract reviewers for providing the academic peer-review exercise as noted in this booklet.

I wish to thank the AFRILEX Board members for their tenacity and hard work in organizing and making this conference a success. A special thank you to Dr Michelle van der Merwe, our local organizer at Stellenbosch University for a sterling job. Thank you to our Vice President, Prof Sonja Bosch and Dr Lorna Morris, for the sterling work on compiling and editing the abstracts booklet. Thank you also to Mr André du Plessis for tirelessly looking after the Afrilex website. We thank our Treasurer, Prof Elsabe Taljard for ensuring that Afrilex remains in a sound and sustainable financial position. I also wish to express my appreciation to the rest of the Board, Prof Dion Nkomo (Secretary), Dr Phillip Louw and Dr Steve Ndinga-Kouumba-Binza, for their hard work and support in the various activities leading up to this conference.

Finally, I wish you all an engaging, stimulating and successful 26th Annual International AFRILEX conference. Please do enjoy yourselves, network and grow after two years of lockdown. Do keep safe and adhere to all Covid-19 protocols!

**Prof. Langa Khumalo
President: AFRILEX**

CONFERENCE PROGRAMME

PLEASE NOTE: EXCEPT WHERE INDICATED OTHERWISE, ALL ACTIVITIES WILL TAKE PLACE IN THE GG CILLIE BUILDING, ROOM 3001

Monday 27 June 2022		
09:00 – 12:30	Pre-conference workshop <i>A short, sharp introduction to the creation and presentation of online dictionaries.</i>	Tim VAN NIEKERK, Dictionary of South African English, Rhodes University.
12:30 – 13:20	Conference registration: Foyer of GG Cillie Building	
13:20 – 13:40	Official Opening	Prof. Mbulungeni MADIBA, Dean: Faculty of Education
13:40 – 13:55	Welcome Address by AFRILEX President	Langa KHUMALO
SESSION 1	Coffee station available from 15:00 – 15:30	
14:00 – 14:25	<i>An analysis of a corpus-based approach for purposes of developing a bilingual multidisciplinary e-dictionary for students at the University of KwaZuluNatal</i>	Tholakele ZUNGU
14:30 – 14:55	<i>Search and you will find</i>	André DU PLESSIS
15:00 – 15:25	<i>The concept of lexicographic condensation: Review and perspectives in digital lexicography</i>	María Teresa FUENTES MORÁN
15:30 – 15:55	<i>A prototype Afrikaans online dictionary for academic editing purposes</i>	Maret BLOM DE WET
16:00 – 16:25	<i>Ancient Egyptian lexicon in the Semitic languages of Ethiopia: New paths in etymological lexicography</i>	Letizia CERQUEGLINI
16:30 – 16:55	<i>The Inclusion of Lexical Particularities in the Comprehensive Dictionary of Gabonese French</i>	Paul A. MAVOUNGOU, Hugues Steve NDINGA-KOUMBA-BINZA, Virginie OMPOUSSA, Blanche NYANGONE ASSAM
17:00 – 18:00	SADILAR SESSION	
18:00	Welcome Cocktail, offices of the WAT, Banghoekweg 115	

Tuesday 28 June 2022		
09:00 – 10:00	Keynote: <i>From agric to zol: African Englishes in the Oxford English Dictionary</i> Session Chair: Langa KHUMALO	Danica SALAZAR, Oxford University Press, Oxford
10:00 – 10:25	Coffee	
SESSION 2		

10:30 – 10:55	<i>Using machine translation to compile bilingual lexicographical databases: An unexpected but highly welcome research outcome</i>	Sven TARP
11:00 – 11:25	<i>Turning bilingual lexicography upside down: Realization of an old dream</i>	Sven TARP
11:30 – 11:55	<i>Towards an Integrated Approach to IsiZulu Terminography and Lexicography</i>	Njabulo MANYONI
12:00 – 12:25	<i>Lemmatisation of verbal suffixes in Sepedi</i>	Danie PRINSLOO and Elsabé TALJARD
12:30 – 12:55	<i>Experimental methodology – compiling a multilingual online dictionary</i>	Sonja BOSCH, Marissa GRIESEL, Elsabé TALJARD
13:00 – 13:55	Lunch: La Sala, 44 Ryneveld Street, Stellenbosch	
SESSION 3	Coffee station available from 15:00 – 15:30	
14:00 – 14:25	<i>Dictionary Base and Lemmata for a Trilingual Dictionary of Fauna</i>	Steve NDINGA-KOUMBA-BINZA
14:30 – 14:55	<i>To see and understand: visual elements in Foundation Phase Dictionaries</i>	Michele VAN DER MERWE
15:00 – 15:25	<i>Promoting linguistic legitimacy via web-based dictionaries</i>	Alex BARATTA
15:30 – 15:55	<i>In progress: the Survey of South African Sign Language Place Names</i>	Chrismi-Rinda LOTH, Jani DE LANGE, Susan LOMBAARD, Emily MATABANE, Gideon KOTZÉ, Annalene VAN STADEN
16:00 – 16:25	<i>Historical account and contribution of the Greater Dictionary of IsiXhosa to Lexicography</i>	Buyiswa MINI
16:30 – 16:55	<i>Towards a corpus of South African Sign Language: Sociolinguistic and Phonological Challenges</i>	Andries VAN NIEKERK
17:00 – 18:00	PUBLISHERS' SESSION – SANLU and OUPSA	
18:30	Conference dinner Middelvlei Wine Estate, Stellenbosch	

Wednesday 29 June 2022

09:00 – 09:15	Announcements	
09:15 – 10:10	Keynote: <i>Trying to see the forest for its trees: Lexicography as communication</i> Session Chair: Sonja Bosch	Herman BEYER, University of Namibia, Namibia

10:10 – 10:25	Coffee	
SESSION 4		
10:30 – 10:55	<i>How great is thy dictionary? Cross-referencing conditions and practices in The Greater Dictionary of (isi)Xhosa</i>	Dion NKOMO, Bulelwa NOSILELA, Wanga GAMBUSHE
11:00 – 11:25	<i>Help! A user guide to user guides in electronic dictionaries</i>	Lorna MORRIS
11:30 – 11:55	<i>An examination of the reception of the dictionaries compiled by the University of Zimbabwe's African Languages Research Institute (ALRI)</i>	Eventhough NDLOVU
SESSION 5	Virtual Presentations: 4 MINUTE executive summary of presenter's video + 8 MINUTE discussion Zoom Room opens 30 minutes before the start of the session	
12:00 – 12:12	<i>The Concept of Lexicographic Data: Some Lexicographic, Economic and Technical Reflections</i>	Pedro A. FUERTES OLIVERA
12:12 – 12:24	<i>Implementing dictionary skills in the language classroom</i>	Mari Carmen CAMPOY-CUBILLO
12:24 – 12:36	<i>Frequency or keyness?</i>	Zorica ĐUROVIĆ
12:36 – 12:48	<i>Aspects of Vernacular Dictionaries: A Case Study of a Hakka Dictionary</i>	Audrey HEIJNS
12:48 – 13:00	<i>Improving learners' skills by compiling dictionaries. A case study</i>	Valeria CARUSO
13:05 – 13:15	CLOSING OF CONFERENCE	
13:15 – 14:00	Lunch on the go	
14:00 – 15:00	Annual General Meeting	

Thursday 30 June 2022 Conference excursion

09:00 Pick-up at the back of the GG Cillie Building

KEYNOTE PRESENTATION 1

From agric to zol: African Englishes in the Oxford English Dictionary

Dr Danica SALAZAR

Oxford Languages, Oxford University Press, Oxford, United Kingdom

Short biography: Dr Danica Salazar is Executive Editor for World Englishes for Oxford Languages, where she leads editorial projects for world varieties of English. Prior to joining Oxford University Press, she was the Mellon Postdoctoral Fellow in English Language Lexicography at the English Faculty and Hertford College of the University of Oxford. She researches and writes World English entries for the *Oxford English Dictionary*, and represents Oxford Languages in lectures and conferences, as well as in international print and broadcast media. She holds a doctorate degree in Applied Linguistics (*sobresaliente*) from the University of Barcelona, a Master's degree in Teaching Spanish as a Foreign Language (*sobresaliente cum laude*) from the University of Salamanca and a Bachelor of Arts degree in European Languages (Spanish and French, *magna cum laude*) from the University of the Philippines-Diliman. She publishes and lectures regularly on lexicography, phraseology, World Englishes and Spanish- and English-language teaching. Dr Salazar is the author of *Lexical Bundles in Native and Non-native Scientific Writing* (2014), co-editor of Biomedical English: A Corpus-based Approach (2013), and co-author of several language textbooks for learners of Spanish.

Abstract: In some of its recent quarterly updates, the *Oxford English Dictionary* has published particularly large batches of new and revised entries from South African English, Nigerian English, and East African English. This talk will be a detailed discussion of the editorial work behind these updates for African varieties of English, whose distinctive vocabularies the *OED* is currently taking steps to cover more widely. The presentation will explain how words are selected for inclusion, how new entries are researched and written, how existing entries from previous editions are revised and modernized, how new sources of textual evidence such as social media give *OED* editors greater insight into African Englishes, and how the dictionary's African consultants contribute to ensuring the accuracy and authenticity of the *OED*'s definitions, etymologies, and quotation evidence.

The talk will also discuss how the *OED* has diversified its coverage of African and other World Englishes by removing the bias towards colonial-era flora and fauna terms and exotic loanwords, and giving more attention to contemporary lexis and other types of lexical innovation beyond simple borrowing. New region-specific labels were also created for these African varieties, as well as new pronunciation models, and audio pronunciations consisting of natural-sounding recordings of words as pronounced in the accent of the relevant variety.

The presentation will conclude with some remarks on the future of *OED*'s on-going World English project and its impact on how African Englishes are reflected in the world's foremost authority on the English language, with a view to opening communication channels and avenues of collaboration between lexicographers and language scholars in Oxford and Africa.

KEYNOTE PRESENTATION 2

Trying to see the forest for its trees: Lexicography as communication

Prof. Herman L. BEYER

Department of Humanities and Arts, University of Namibia, Windhoek, Namibia, and
Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa

Short biography: Herman Beyer is associate professor in Afrikaans and head of the Department of Humanities and Arts at the University of Namibia, where he teaches Afrikaans linguistics and professional communication. His primary research field is theoretical lexicography. Currently, his research interests include school dictionary use in the classroom, the lexicography of adjectives in Afrikaans monolingual dictionaries, comments on cotext in dictionaries, and a lexicography of the "Woordelys" of the *Afrikaanse Woordelys en Spelreëls*. His experience in practical lexicography includes roles as co-editor of the *HAT Afrikaanse Skoolwoordeboek* (2009), editor of *Van Dale Miniwoordenboek Afrikaans-Nederlands / Nederlands-Afrikaans* (2010) and contributor to the 11th edition of the *Afrikaanse Woordelys en Spelreëls* (2017), the latter thanks to two terms of service on the "Taalkommissie" of the South African Academy of Science and the Arts (2011-2017). Herman is a research associate in the Department of Afrikaans and Dutch at Stellenbosch University, and former Board member and President of Afrilex. He was born, raised and schooled in Namibia, studied at Stellenbosch University, returning afterwards to Swakopmund, Namibia, as a high school Geography and Afrikaans teacher, where he served for four years before his appointment as lecturer at the University of Namibia in 1997.

Abstract: If someone cannot see the forest for the trees, it means that they fail to see the big picture because they focus too much on the details (cf. McLeod, 2022). Hopefully, this paper will not be a case in point. Rather, the desired outcome would be to have shared an understanding of the forest by looking a bit closer at some of its trees (if some metaphor stretching could be pardoned). The forest, of course, is lexicography. Perhaps it is rather something more of a jungle. Nevertheless, if a pine forest is made up (largely) of pine trees, then lexicography as a forest is made up of all things lexicographic. The question, then, is what falls under the notion of "all things lexicographic". This would seem to depend on the angle from which one observes the forest and its trees, i.e., one's theoretical perspective. In this paper, a particular perspective on the nature of lexicography will be shared, namely that of the theory of lexicographical communication (TLC).

Firstly, the genesis of the theoretical paradigm will be explained by referring to developments in theoretical lexicography which stimulated the need for an alternative theory, in particular Wiegand's theory of lexicographic texts within his broader general theory of lexicography (e.g. Wiegand, 1990), and the modern theory of lexicographic functions developed at the Centre for Lexicography at Aarhus University (e.g. Tarp, 2008). This will be followed by briefly outlining the somewhat eclectic and therefore interdisciplinary nature of the TLC as a theory informed by general communication theory, professional communication theory, theoretical linguistics and document design, in addition to the existing body of lexicographic knowledge.

Secondly, the main tenets of the TLC will be outlined. These have implications for the definition of core terms like *lexicography*, *dictionary* and *lexicographic text*. These terms will be redefined within the paradigm of the TLC.

Thirdly, an overview of some of the fruits that the TLC has yielded, will be presented. These include the following: (a) a reappraisal of dictionary purposes, informed by general and professional communication theory; (b) the development of a new dictionary typology from a textual point of view, informed by text linguistics; (c) the development of text grammars for dictionary articles from a theoretical linguistics perspective; (d) the formalisation of *messages* in lexicographic texts (from the semantic-pragmatic domain of a text grammar), which shows

promise for instruction and assessment in dictionary pedagogies, for user guide development, and for dictionary evaluation; and (e) a new empirical methodology for dictionary evaluation, based on the regulative principles of textuality taken from text linguistics.

In conclusion, an attempt will be made to demonstrate how the insights gained from the TLC might facilitate the identification and description of the trees that define the forest of lexicography.

References

- McLeod, D. 2022. How to Use "Cannot See the Forest for the Trees" Correctly. *Grammarist*. Available: <https://grammarist.com/usage/cannot-see-the-forest-for-the-trees/> Accessed on 24/04/2022.
- Tarp, S. 2008. *Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner's lexicography*. Tübingen: Max Niemeyer.
- Wiegand, H.E. 1990. Printed dictionaries and their parts as texts. An overview of more recent research as an introduction. *Lexicographica, International Annual of Lexicography*, 6:1–126.

Documentary film: Dreaming of Words

Synopsis: Njattyela Sreedharan, a fourth standard drop-out, compiles a dictionary connecting four major Dravidian languages. Travelling across four states and doing extensive research, he spent twenty five years making the multilingual dictionary. This unique dictionary offers a comparative study of Malayalam, Kannada, Tamil and Telugu. 'Dreaming of Words' traces Sreedharan's life, work, love for languages and the struggles to get the dictionary published. The film also explores the linguistic and cultural diversity in India.

To view the film: <https://drive.google.com/file/d/1tXnLMx-j8SCqD-KKXLe4YVuTdkAN4wR-/view?usp=sharing>

Special Sessions

South African Centre for Digital Language Resources

The South African Centre for Digital Language Resources (SADiLaR) is a national Centre supported by the Department of Science and Innovation (DSI) as part of the new South African Research Infrastructure Roadmap (SARIR). It is a multi-partner entity with the North-West University (NWU), functioning as host as well as hub of a network of linked nodes which are based at different academic institutions. The nodes are

- University of Pretoria (Department of African Languages);
- University of South Africa (Department of African Languages).
- CSIR (HLT Research Group).
- North-West University (Centre for Text Technology).

- Inter-Institutional Centre for Language Development and Assessment (ICELDA); and
- Child Language Development Node (Department of General Linguistics at Stellenbosch University)

The role of SADiLaR is to focus on all official languages of South Africa, supporting research and development in the domains of language technologies and language-related studies in the humanities and social sciences. The Centre supports the creation, management and distribution of digital language resources, as well as applicable software, which are freely available for research purposes through the Language Resource Catalogue.

To ensure that SADiLaR fulfills its mandate, it is important to collaborate with scholars in different fields of including e.g., lexicography. SADiLaR has various resources on its repository and the aim of this abstract is to showcase the kind of lexicographic resources and tools available and to also get input from lexicographers on how we can be of better use to them and to the discipline in order to help intellectualize the South African languages so that they can have a functioning role that goes beyond the home. During the session we will also engage the participants on what specific support needs they have that SADiLaR as a national research infrastructure could potentially provide. The idea that language planners are increasingly agreeing that developing languages, such as African languages, is a need in order to improve the process of intellectualization (Finlayson and Madiba, 2002) is a win because it means that more resources can be propelled towards languages. Most importantly, working on intellectualization of languages will have a direct impact on the speakers of the language to derive the pride, self-assurance and resourcefulness in the (new) ability to discuss the most complex of issues ranging from the mundane to academic and beyond (Khumalo 2017:254), the kind of pride that dictionaries bring to their speakers.

References

- Finlayson, R. and M. Madiba. 2002. The Intellectualization of the Indigenous Languages of South Africa: Challenges and Prospects. *Current Issues in Language Planning* 3(1): 40-61.
 Khumalo, L. 2017. Intellectualization through Terminology Development. *Lexikos* 27 (AFRILEX-reeks/series (27): 252-264
 South African Centre for Digital Language Resources. n.d. <https://sadilar.org/index.php/en/>
 Accessed on 12 April 2022

SANLU and OUPSA Session

The African Association for Lexicography (AFRILEX) was established as a platform for facilitating interaction between lexicographic theory and lexicographic practice especially for African lexicographers. The complementary relationship between the two components of lexicography as a discipline may never be overemphasized. Lexicographic practice needs theoretical insights and principles in order to produce dictionaries with enhanced utility value and user-friendliness. Yet lexicographic theory can only be as good as the practice that inspires its insights as well as improved lexicographic processes and products that derive guidance from it. The publishers' session of the 26th International AFRILEX Conference provides an opportunity for both practicing lexicographers, including publishers, and theoretical lexicographers to engage on dictionaries as objects on mutual interests, with the satisfaction of

user needs being a primary priority.

This publishers' session features the South African National Lexicography Units (SANLU) and Oxford University Press – Southern Africa (OUPSA) as major players in the field of lexicographic practice in South Africa. Over the years, SANLU has emerged as the unrivalled dictionary publisher for the national lexicography units (NLUs), which were established to promote the country's official languages through compiling dictionaries, particularly comprehensive monolingual dictionaries. In line with the motivation behind the establishment of NLUs, SANLU derives its inspiration from the Constitution of the Republic of South Africa that stipulates that all official languages must be treated equally. As such, SANLU works with the NLUs in advancing the latter's legislative mandate and vision. This includes exploring avenues of committing government departments, especially the Department of Education, to officially recognize the dictionaries produced by the NLUs not only as resources but also as core learning material. Another example is the SANLU-inspired strategic plan that is aligned with the UNESCO decade of indigenous languages (2022-2023) which will be conceived to yield more school dictionaries in the official African languages. SANLU's contribution in this publisher's session offers more insights into the publisher's role and vision in promoting NLUs' dictionaries as integral towards the attainment of multilingualism in South Africa.

For lexicographic practice to thrive, a societal dictionary culture is required as it guarantees lexicographers and dictionary publishers a reliable market. OUPSA, just like other publishers, is not only grappling with an endeavor to produce user-friendly dictionaries but also lexicography as a complex and difficult investment area that has been characterized by a market that took a huge downturn over the past decade. In this publishers' session, OUPSA delves into detail about the South African dictionary market and its complexities, including the continually debated issue of skepticism around monolingual dictionaries for African languages. Notwithstanding the challenges, OUPSA will share how the publisher has enjoyed notable successes over the past few years due to the following reasons:

- having a clear strategy
- rigorous adherence to catalogue submission and procurement deadlines
- building a diversified portfolio for key account and government market success
- investing heavily in bilingual dictionaries for African languages and English
- working with language experts, expert theoretical and practical lexicographers
- having strong sales and marketing support and infrastructure.

Furthermore, OUPSA will reflect on the implications of the digital revolution for the South African dictionary market and the publisher's vision for the future of lexicography at OUPSA – emphasizing the value of a versatile lexicographical dataset that can be used for multiple outputs. The publisher believes the biggest future revenue generator will be the licensing (for lexical and non-lexical use) of data from those datasets via APIs or bulk data dumps, as illustrated by successes already enjoyed from bilingual datasets in the Oxford Global Languages project. Also considered are dictionary lookups that are fully integrated into all educational products and platforms, making them indispensable assets in the future of commercial publishing.

PRESENTATIONS

Promoting linguistic legitimacy via web-based dictionaries

Alex BARATTA (alex.baratta@manchester.ac.uk)

Manchester Institute of Education, University of Manchester, Manchester, UK

Kermas (2012:75) argues that “lexicographers need to address the culture-specific dimension of knowledge sharing in today’s global village and broaden their cultural viewpoint”, addressing the relative lack of vocabulary within dictionaries (such as the OED) that reflects non-inner circle Englishes, with Korean English under discussion here. Compilations devoted solely to the vocabulary of these Englishes are often housed within online dictionaries, reflecting current lexicographical development within the framework of the Fourth Industrial Revolution. However, there are three issues. First, the status of online dictionaries has been questioned based on not necessarily being compiled by professional lexicographers. Second, they are argued to be of dubious quality, given “inadequate and even incorrect information” (Tarp, 2019:227). Finally, non-inner circle Englishes are often marginalised, based on lexis which diverges from prestigious English varieties such as the inner-circle standard (Galloway & Rose, 2015). However, in addressing these points, we should consider that online dictionaries are a modern approach to lexicography, emphasising “democracy and equal access to meaning-making rights” (Damaso, 2005:4). Further, individuals responsible for creating Korean English online dictionaries have accurate knowledge, as they are *users* of Korean English, and are often EFL teachers in Korea who seek to educate people regarding this variety (Baratta, 2021). Finally, all varieties of English conform to a systematic use of lexis which serves its speakers well, regardless of circle (Kachru, 1991). Thus, the aim is to discuss how both expanding circle Englishes, here Korean English, and online dictionaries used to house such Englishes, are indeed legitimate. Linking to the modern theory of lexicography, one postulate is the need for dictionaries to reflect the needs of a certain group of people in specific social situations (Bergenholtz & Tarp, 2003). One particular group comprises individuals with a need to access the vocabulary of expanding circle Englishes within the extra-lexicographical situation of Internet usage. Online dictionaries are arguably the sole means to do so in the first instance, with the focus here on what might be termed ‘individually-constructed online dictionaries’. While compiled by laypersons, they nonetheless reflect accuracy in their compilation of lexis.

The methodological approach is based on two studies (Baratta, 2019; Baratta, 2021). Study one obtained the views of 36 participants representing fifteen nationalities from all three circles of English, with 35 individuals regarding expanding circle Englishes as legitimate varieties, and 32 participants advocating their use in EFL classrooms. For example, participants discussed online dictionaries as an EFL pedagogic tool, involving the *translation* of Korean English vocabulary to inner-circle varieties (and vice versa), as opposed to ‘correcting’ Korean English vocabulary. This approach reflects current EFL pedagogy (Matsuda, 2003), a means to consider the “socio-cultural functions of global English” (Lvtsevich & Sokolov, 2020:703). Study two conducted an internet search for Korean English online dictionaries, with the search terms *Korean English*; *Konglish*; *Korean English dictionary*; and *Konglish dictionary* (with *Konglish* a necessary search term given its common societal use). From here, the entries deriving from the first page of each search result were

selected, resulting in a total of nineteen websites whose content will be presented as evidence of online codification (Cotter & Damaso, 2007).

The presentation, then, argues that a more inclusive attitude, and approach, to World Englishes can be forged through online dictionaries. It will explain how the proliferation of online dictionaries of expanding circle Englishes, such as Korean English, can help individuals to regard both positively. Further, these dictionaries can be used as a pedagogic tool, thereby reflecting the needs of EFL teachers and students.

References

- Baratta, A. 2019. *World Englishes in English language teaching*. London: Palgrave Macmillan.
- Baratta, A. 2021. *The societal codification of Korean English*. London: Bloomsbury.
- Bergenholtz, H. & Tarp, S. 2003. Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Cotter, C. & Damaso, J. 2007. Online dictionaries as emergent archives of contemporary usage and collaborative codification. *Queen Mary's OPAL #9 Occasional Papers Advancing Linguistics* 1-10.
- Damaso, J. 2005. *The new populist dictionary: A computer-mediated, ethnographic case study of an online, collaboratively authored English slang dictionary*. Master's dissertation, Queen Mary, University of London.
- Galloway, N. & Rose, H. 2015. *Introducing global Englishes*. London: Routledge.
- Kachru, B.B. 1991. Liberation linguistics and the Quirk concern. *English Today* 7 (1): 3-13.
- Kermas, S. 2012. Culture-specific lexis and knowledge sharing in the global village. In R. Fachinetti (Ed.), *English dictionaries as cultural mines* (pp. 73-94). Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Lovtsevich, G.N. & Sokolov A.A. 2020. World Englishes and learner lexicography: View from the Expanding Circle. *Russian Journal of Linguistics* 24 (3): 703-721.
- Matsuda, A. 2003. International understanding through teaching World Englishes. *World Englishes* 21 (3): 436-440.
- Tarp, S. 2019. Connecting the dots: Tradition and disruption in lexicography. *Lexikos* 29: 224-249.

A prototype Afrikaans online dictionary for academic editing purposes

Maret BLOM DE WET (maretblom@sun.ac.za)

Department of Afrikaans and Dutch, Stellenbosch University, South Africa

The academic writing skills and research skills of postgraduate students at higher education institutions in the South African context are inadequate (Van Aswegen, 2007:1141), and therefore there is a growing demand for the services of editors of specifically academic texts. The editors of Afrikaans academic texts, however, experience problems in terms of reference sources (style guides or standardisation sources) that they can use to ensure consistency in the academic texts. The current Afrikaans dictionaries and style guides are outdated and only contain parts that are useful for the academic editor (Blom, 2020:8). Consequently, these editors have a need for an Afrikaans dictionary that is specially aimed at the needs of academic editors (Blom, 2020:18).

In a previous study Fuertes-Olivera and Tarp's (2014) function theory for specialised

online dictionaries was used to set up a model for the design of an Afrikaans online dictionary for academic editing purposes (Blom's 2018 dictionary model). This model was set up as the first part of a larger project that aims to facilitate the work of academic editors and ensure consistency in the editing of Afrikaans academic texts with the help of a comprehensive online Afrikaans academic editing dictionary (Blom, 2020:8). In order to implement Blom's 2018 dictionary model to compile a complete dictionary, the model needs to be refined and updated by adding more data that the academic editors will find useful. The model also needs to be made available to the target usersⁱ to test the functionality, set up improvement guidelines and continuously update the dictionary (Blom, 2020:25).

In this study Blom's 2018 dictionary model is used as the basis to compile a prototype Afrikaans academic editing dictionary, so that the academic editors can test the functionality of the dictionary's data presentation and data description in a follow up study. Firstly, Fuertes-Olivera and Tarp's (2014) function theory for specialised online dictionaries was used to determine the functions and datatypes of the prototype academic editing dictionary. Thereafter, Gouws's (2014a; 2014b; 2018a; 2018b; 2018c; 2018d) research on possible structural adjustments for electronic lexicography was used to make decisions about the structure of the prototype dictionary, and the principles of the ISO standards (ISO-standard 9241-110, ISO-standard 9241-11 and ISO-standard 9241-12) and the interaction design (Sharp, Rogers & Preece, 2007) were used to select the technological features and usability of the online dictionary. The use of this integrated theory of lexicographic and usability principles enabled the compilation of the prototype dictionary's data presentation/layout and data description/content. The layout of the prototype's home page, sections, dictionary articles and user guide was determined by the structural adjustments for electronic lexicography and the principles of the ISO standards and interaction design, which emphasises the presentation of the data and the user's experience with the software. The prototype's functions (communicative, cognitive and operative) and the academic editors expectations of an academic editing dictionary was used to select the content of the prototype dictionary and divide it into 16 different sections, including abbreviations and acronyms; academic reference systems; most common errors in academic texts; general dictionaries; numbers and symbols; uppercase and lowercase letters; punctuation; italics and Roman numerals; spelling; the specialist field of academic editing; tables and graphs; subject terminology; subject dictionaries; science, mathematics and computers; laws and references to laws and mathematical notation.

References

- Blom, M. 2020. 'n Teoretiese model vir die ontwerp van 'n aanlyn Afrikaanse Akademiese Redigeerwoordeboek. *Tydskrif vir Nederlands en Afrikaans*, 27(1):5-29.
- Fuertes-Olivera, P.A. en Tarp, S. 2014. *Theory and practice of specialised online dictionaries: lexicography versus terminography*. Berlyn/Boston: Walter de Gruyter GmbH.
- Gouws, R.H. 2014a. Article structures: Moving from printed to e-dictionaries. *Lexikos*, 24:155-177.

ⁱ The target users for the online Afrikaans academic editing dictionary are professional academic editors, as well as second-year, third-year, short-course and honours students who have to do academic editing as part of their study obligations.

- Gouws, R.H. 2014b. Makrostrukturaanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe*, 54(3):481-504.
- Gouws, R.H. 2018a. Accessibility, access structures and access procedures, in V. Jesenšek & E. Milka (eds.). 2018. *Wörterbuchstrukturen zwischen Theorie und Praxis*. Berlin: De Gruyter. 3556.
- Gouws, R.H. 2018b. Internet lexicography in the 21st Century, in S. Engelberg, H. Kämper & P. Storjohann (eds.). *Wortschatz: Theorie, Empirie, Dokumentation*. Berlin: De Gruyter. 215236.
- Gouws, R.H. 2018c. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos*, 28:177-195.
- Gouws, R.H. 2018d. Dictionaries and access, in P.A. Fuertes-Olivera & A. Pedro (eds.). *The Routledge Handbook of Lexicography*. London: Routledge. 43-58.
- Sharp, H., Rogers, Y. & Preece, J. 2007. *Interaction design: Beyond human computer interaction*. Chichester: John Wiley.
- Van Aswegen, E.S. 2007. Postgraduate supervision: The role of the (language) editor: Sed quis custodiet ipsos custodes? *South African Journal of Higher Education*, 21(8):1139-1151.

Experimental methodology – compiling a multilingual online dictionary

Sonja BOSCH (seb@hbosch.com)

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

Marissa GRIESEL (griesm@unisa.ac.za)

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

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

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

The availability of multilingual electronic dictionaries for the South African languages is essential, not only for direct target users, but also for indirect target users. The need for high quality electronic resources is especially beneficial for languages with scarce resources such as Tshivenda. Direct target users include users who consult the dictionary to find translation equivalents of, for example, Tshivenda lemmas in English or Afrikaans, or any variation of this configuration. Such a basic translation dictionary would be useful for Afrikaans or English speakers who want to learn Tshivenda, and also for Tshivenda speakers wanting to learn Afrikaans or English. The online nature of the dictionary makes it a useful multidirectional tool, allowing the user to search for lemmas in any of the three languages. Indirect users would be interested in the machine-readable database underlying the dictionary, for the development of human language technology applications such as machine translation systems.

This presentation explores the optimal use of free electronic/online resources for compiling a trilingual e-dictionary for Tshivenda, English and Afrikaans.

The approach followed involves an experiment in which the compilation process was automated as far as possible to bring about savings in terms of time and human effort. English is used as a bridge for the translation between the source language, Tshivenda, and the target language, Afrikaans. The free available resources used, include:

- a Tshivenda-English dictionary (Murphy, 1997), sourced from the Comparative Bantu

OnLine Dictionary project (CBOLD) (<http://www.cbold.ish-lyon.cnrs.fr/>) under an open license;

- four English-Afrikaans machine translation systems, namely Google Translate (<https://translate.google.com/>), Bing Microsoft® Translator (<https://www.bing.com/translator/>), Yandex.Translate (<https://translate.yandex.com/>) and english-afrikaans.co.za (<https://english-afrikaans.co.za/>);
- an Afrikaans spell checker, WSpel, which is currently the most comprehensive Afrikaans spell checker available for free;
- corpus query programs, which were found to be suitable for both Tshivenda and Afrikaans texts, namely
 - AntConc: <https://www.laurenceanthony.net/software/antconc/>;
 - LanksBox: <http://corpora.lancs.ac.uk/lanksbox/>; and
 - Voyant Tools: <https://voyant-tools.org/>.
- a Tshivenda corpus of 1.4 million orthographic words ('tokens') used for refinement. This is a raw corpus, without any annotation of, for example, word categories, and consists of digitized texts across a wide variety of genres.
- a dictionary compilation program, *Lexonomy*, that carries no installation, programming knowledge, or expensive costs to create a basic online dictionary, thus enabling more users to set up their own dictionaries (Méchura, 2017). The platform can be used to compile, format, customize and eventually publish a monolingual or multilingual dictionary online.

This proof-of-concept experiment is based on 10% randomly selected dictionary entries in the Tshivenda-English dictionary (Murphy, 1997) and includes four parts of speech, namely nouns, verbs, adverbs and adjectives. Closed word classes are excluded. In short, the English entries were translated to Afrikaans using the machine translation systems mentioned above. Manual evaluation of the automatically generated translations by first language speakers was also performed on this data and we report on the most salient challenges with regards to the machine translation from English to Afrikaans. Correct Afrikaans translations were then added to the existing English and Tshivenda equivalents and this newly created trilingual dictionary was imported to *Lexonomy* for easy browsing and future expansion.

The general findings of the experiment are that - as expected - there are certain limitations to such a semi-automated process with free resources, which require a certain amount of human intervention. Although the composite e-dictionary cannot be considered a final product, the cloud-based, open-source dictionary writer and publisher *Lexonomy* offers the opportunity for human experts such as lexicographers, linguists, etc., to make the necessary updates in a user-friendly manner. This is due to the adaptability and easy layout of *Lexonomy*.

A significant advantage of the semi-automated process described in this presentation is the saving of person-hours utilized compared to time normally spent by lexicographers on the development of dictionaries. The draft proposal formulated (also see Bosch, Griesel and Taljard (forthcoming)) is useful for the creation of online multilingual dictionaries, compiled by using available online or electronic resources. The resulting trilingual dictionary is available online as proof of concept on which further work can be built. To access the online dictionary, a free profile can be registered on the *Lexonomy* platform at <https://www.lexonomy.eu/> after which a draft version of the dictionary can be seen at <https://www.lexonomy.eu/POCVenEngAfr/>. The fact that the database underlying the dictionary is available in a machine-readable format,

namely XML, is important for indirect target users for reuse to develop electronic resources, especially for resource-scarce languages.

References

- Bosch, S.E., Griesel, M and Taljard, E. (forthcoming). Optimalisering van gratis elektroniese/aanlyn hulpbronne vir woordeboeksamestelling – ’n drietalige woerdeboekeksperiment. (Optimization of free online / electronic resources for dictionary compilation - a trilingual dictionary experiment). *Lexikos* 32(2):1-26.
- Měchura, M. 2017. Introducing Lexonomy: an open-source dictionary writing and publishing system. In *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 Conference*, pp. 19-21.
- Murphy, M.L. 1997. Venda: CBOLD. <http://www.cbold.ish-lyon.cnrs.fr/>

Implementing dictionary skills in the language classroom

Mari Carmen CAMPOY-CUBILLO (campoy@uji.es)

English Studies Department, Universitat Jaume I, Castelló, Spain

In this presentation I am going to look at how online dictionary skills can be related to different language proficiency levels. I will provide task examples and possible ways to assess dictionary skills. The questions behind this presentation are:

- (1) What is the role of online dictionaries as (self-)learning tools?
- (2) Are online dictionaries introduced in the classroom in terms of their (multimodal) affordances?
- (3) Can online dictionary skills be part of language proficiency instruction?
- (4) Can online dictionary skills be assessed as part of a language proficiency level framework? What is their place in such a framework?

Dictionaries have been present in language learning environments for a long time. Their role within these environments, however, has not always been clearly defined. The main four language skills (reading, writing, listening, speaking) have been thoroughly described in terms of content, aims, proficiency levels and assessment means. Dictionary skills, on the other hand, need to be formally included within these skills. Furthermore, this inclusion needs to be put into practice with a certain level of international agreement so that different research proposals for dictionary skill training methodologies and implementation in curricula may be discussed from a shared theoretical and practical perspective. The different dictionary skills need to be aligned with specific language skills, particularly those of reading and writing which are most commonly related to dictionary use.

Implementing dictionary skills in an educational context means that formal teaching of these skills is given, and that a practice and assessment of those skills are planned and carried out as part of the syllabus implementation process. In the context of European university Foreign Language Learning and English for Special and Academic Purposes sessions, students may access online dictionaries in their computers (language labs), tablets or mobiles while solving a foreign language task. Assessment of dictionary skills as instrumental competence is important because assessment means testing that pedagogy worked and that learning took place.

The cycle of dictionary skill implementation is shown in Figure 1 below:

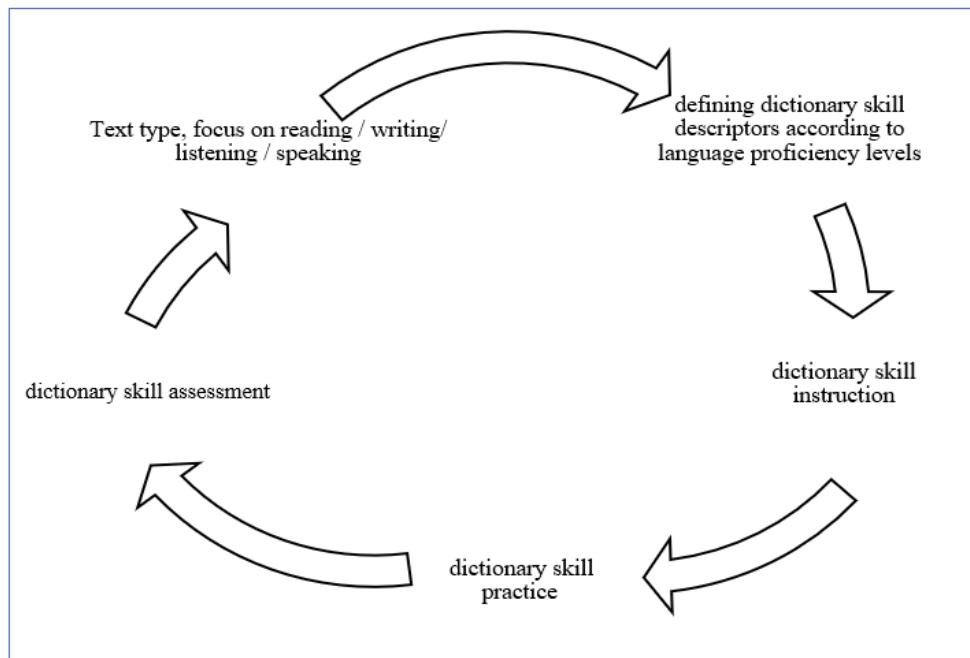


Figure 1. Dictionary skill development and implementation process

As suggested in Gouws and Prinsloo (2005: 39) “Good dictionaries are products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are.” Part of this best usage can be enhanced by dictionary skill teaching. It is the role of teachers to be familiarized with online dictionary affordances and to select appropriate dictionaries for their students.

Language learners tend to google their information needs and tend to assume that any result is a good result. They obtain and combine answers from a variety of online sources, as navigation makes this relatively easy. One possible source type is online dictionaries. Bothma and Gouws (2020:49-50) state that: “Members of digital societies have become used to different possibilities to obtain information on demand. Their demand often exceeds the extent of information contained in a single source. Lexicographers need to take cognizance of this and also of the fact that a growing number of their potential target users are digital natives” and that they “move from one source to the next if their need is not satisfied completely”. This lack of satisfaction when dictionaries are the source is sometimes due to the fact that learners are not informed about all the information types that a specific dictionary may provide.

The dictionary skills implementation proposal aims to (i) make those skills and the possibility to question “first result” validity part of the process, and (ii) provide learners with the opportunity to interact with a variety of dictionaries on a regular basis making them aware of online dictionary affordances. This implies explicit dictionary skill teaching so that learners may make more informed searches and may be taught on possible information types that would otherwise be ignored or not found.

Teachers’ dictionary choices need to be informed about dictionary affordances and the different means of access to different types of information they provide. Online dictionary micro- and macrostructure may differ considerably from one online dictionary to another and

this needs to be considered when designing dictionary skill training sessions to provide descriptors that can be included in a syllabus (Campoy-Cubillo 2015). This presentation will show dictionary skill descriptors applied to the use of specific dictionaries and will provide dictionary skill tasks that can be assessed using such descriptors to highlight what learners with a particular language proficiency level can do with specific online dictionary types.

References

- Bothma, T. J., Gouws, R. H. 2020. e-Dictionaries in a Network of Information Tools in the e-Environment. *Lexikos* 30: 29-56.
- Campoy-Cubillo, M. C. 2015. Assessing dictionary skills. *Lexicography*, 2(1), 119-141.
- Gouws, R. H., & Prinsloo, D. J. 2010. Principles and practice of South African lexicography. Stellenbosch; African Sun Media.

Improving learners' skills by compiling dictionaries: A case study

Valeria CARUSO (vcaruso@unior.it)

Department Department of Literary, Linguistic and Comparative Studies, University of Naples 'L'Orientale', Naples, Italy.

This paper reports on an experiment in dictionary compilation carried out with *Lexonomy*, a web-based tool for publishing dictionaries originally developed by Michal Měchura (2017) and later updated (Rambousek, Jakubíček & Kosem, 2021) to be included in the ELEXIS (Krek et al., 2018) project infrastructure. Despite the vast literature focused on the importance of teaching lexicographic skills (for a summary see for example Gavriilidou (2013)), cursory attention has been paid to the types of activities to be used for either improving users' understanding of dictionaries and for acquiring better language skills through the use of dictionaries.

With this aim in mind, one hundred university students in Linguistic Mediation were guided through compiling a multilingual Covid-19 dictionary. The activities, held in virtual rooms of video conferencing web platforms during the pandemic lockdown, profited from online resources like *Lexonomy*, which ensure data crowdsourcing from authorised contributors. Besides, using *Lexonomy* students should easily understand the microstructural organisation of dictionary articles and thus learn to distinguish the different types of linguistic items to be found in general language dictionaries and how they are described, and arranged, in the entry. This learning goal is expected to be achieved because *Lexonomy* offers an intuitive XML visual editor of the entry structure, given in the form of a clickable hierarchical tree, through which users can add data directly in its nested elements. The task of compiling entries should therefore be easy to learn and engaging to accomplish. In the experimental protocol here described, the learning activities were specifically meant to achieve different goals, such as to practise metalinguistic and metalexicographical knowledge as well as enhancing linguistic and translation skills.

The first attempt to compile the *Dizionario Multilingue del Covid19 - Covid19 Multilingual Dictionary - Covid19 Mehrsprachiges Wörterbuch* was carried out during class hours of a course in Lexicology and Lexicography. Students collected articles from major newspapers for the languages involved in the project: Chinese, Dutch, English, German, Italian,

Polish, Portuguese, Spanish and Russian. Texts were shared in online folders and were used to build comparable corpora using the *Sketch Engine* tools.

Later on, students were given the article microstructure schema, defined by the teacher, for compiling the multilingual dictionary. The article is made up by a meaning explanation and one example of usage of an Italian lemma, followed by its collocates and licensed prepositions. The corresponding lemmas in the other languages covered by the dictionary are given on the same page as separate entries, arranged in the same structural organization as their Italian counterpart. Direct access to foreign words is possible by performing free searches in the search box since the multilingual macrostructure is not reversible.

As regards dictionary data, translation equivalents were identified by comparing the single and multi-word terms extracted from the comparable corpora collected. Yet additional searches on Google news or in the *Sketch Engine* “Covid-19 Corpus” were necessary when the language data proved to be insufficient.

With the aim of evaluating students’ performance, specific assessment parameters were used:

1. accuracy in compiling the dictionary entries,
2. metalinguistic and metalexicographical skills acquired,
3. effectiveness of peer learning.

In two of the three parameters analysed, the students achieved good results, whereas metalinguistic and metalexicographic skills, assessed by means of an open-ended questionnaire, were poor. Only a fraction of the students (around 30%) were able to explain fully some key concepts for the task, such as what a dictionary entry, a microstructure or a collocation is.

The presentation will provide further details on the experimental compilation and the results achieved by the students.

References

- Dizionario Multilingue del Covid19 - Covid19 Multilingual Dictionary - Covid19 Mehrsprachiges Wörterbuch. <https://www.lexonomy.eu/hpg58cpv>. Accessed on 20/03/2022.
- European Researchers’ Night. <https://ec.europa.eu/research/mariecurieactions/event/2021-european-researchers-night>. Accessed on 20/03/2022.
- Gavriilidou, Z. 2013. Development and validation of the Strategy Inventory for Dictionary Use (S.I.D.U.). In: International Journal of Lexicography, pp. 135–153.
- Krek, S., Kosem, I., McCrae, J.P., Navigli, R., Pedersen, B.S., Tiberius, C. & Wissik, T. 2018. European lexicographic infrastructure (elexis). In Proceedings of the XVIII EURALEX International Congress on Lexicography in Global Contexts. pp. 88-892.
- Lexonomy. <https://www.lexonomy.eu>. Accessed on 20/03/2022.
- Měchura, M. B. 2017. Introducing Lexonomy: an open-source dictionary writing and publishing system. In Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference, 19-21 September 2017. Leiden, The Netherlands, pp. 662-689.
- Rambousek, A., Jakubíček, M., & Kosem, I. 2021. New developments in Lexonomy. In: *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*. Brno: Lexical Computing CZ.

Ancient Egyptian lexicon in the Semitic languages of Ethiopia: New paths in etymological lexicography

Letizia CERQUEGLINI (cerqueglini@tauex.tau.ac.il; [lolaluna2017@gmail.com](mailto:lolahuna2017@gmail.com))

Affiliation: Department of Hebrew Language and Semitic Linguistics, Tel Aviv University, Tel Aviv, Israel

This is an analysis of a collection of one hundred etymologically related and semantically similar non-cognate words (NCWs) found in Ancient Egyptian and the Ethiopian Semitic languages (henceforth, ESLs). I set out to retrace their origin and diffusion via historical phonological comparisons. My data is from Old and Middle Egyptian textual sources (Allen, 2015; Faulkner, 1981) and Ge'ez and Amharic dictionaries and lexicographical works (Appleyard, 1977; Bulakh & Kogan, 2013; Leslau, 1976, 2006; Weninger, 2005). This study is part of a project for the creation of an etymological dictionary of foreign words in ESLs. Despite the geographical proximity and evidence of historical contact between ancient Egypt and the Horn of Africa since the third millennium BCE (Phillips, 1997), the linguistic contact between Ancient Egyptian and ESL has received insufficient attention. The present contribution thus opens a new path to investigating the prehistory of the Semitic languages of Africa through Egyptian lexical evidence.

The etymological stock shared by Ancient Egyptian and ESLs is very broad, as both Egyptian and Semitic linguistic families belong to the Afroasiatic phylum (Takács, 2012). Yet the NCWs under examinations here cannot be considered cognates, i.e., kindred words derived independently from identical roots. NCWs are etymologically related, yet not attributable to common ancestors via systematic phonological changes. Furthermore, while cognate words can semantically diverge during their independent evolutions (Kogan, 2015), NCWs are so semantically specific and similar in Egyptian and ESLs as to be considered ‘cultural words’ (Goddard, 2018).

Both the phonological and semantic clues suggest that these words spread among Ancient Egyptian and ESLs via contact. Three hypotheses can be formulated regarding the origin of NCWs and the direction of borrowing: (1) NCWs were originally Egyptian and borrowed by the ESLs; (2) NCWs passed from ESLs to Ancient Egyptian; (3) Both Ancient Egyptian and ESLs borrowed NCWs from a third language. Hypothesis (1) is valid only for some of the words in question (thirty-nine words). The roots of the other sixty-one words are found also in other Semitic languages (Akkadian, Hebrew, Arabic, etc.) with similar meanings, as lexicographical comparisons demonstrate, so their etymological stock is doubtlessly Semitic. For these sixty-one NCWs, Hypothesis (2) could be valid, yet they do not reflect the phonological rules leading from proto-Semitic to ESLs. Hypothesis (3) is thus of interest. Indeed, Ancient Egyptian absorbed a vast number of Semitic words in both its formative (Takács, 2012) and historical phases (Borg, 2021; Muchiki, 1999; Noonan, 2016).

To conclude, among the one hundred words selected for this analysis, thirty-nine NCWs are originally Ancient Egyptian, and sixty-one NCWs originated within the Semitic milieu in the Middle East after ESLs had separated and reached the Horn of Africa across the Red Sea (Kogan, 2015). These sixty-one NCWs seem to have entered ESLs through Ancient Egyptian: Indeed, their meanings are almost identical in Ancient Egyptian and ESLs and they do not reflect the phonological rules leading from proto-Semitic to ESLs. They rather show pronunciations similar to those mirrored by the Ancient Egyptian spellings.

References

- Allen, J. 2015. *The Ancient Egyptian Pyramid Texts*. Atlanta: SBL.
- Appleyard, D. 1977. *A Comparative Approach to the Amharic Lexicon*. Malibu: Undena.
- Borg, A. 2021. *Rewriting Dialectal Arabic Prehistory*. Leiden: Brill.
- Bulakh, M., Kogan, L. 2013. More on Genealogical Classification of Ethiopian Semitic. Kogan, L. (ed.) *Babel und Bibel* 7, 599–608. Winona Lake: Eisenbrauns.
- Faulkner, R. 1981. *A Concise Dictionary of Middle Egyptian*. Oxford: Ashmolean Museum.
- Goddard, C. 2018. *Ten Lectures on Natural Semantic MetaLanguage*. Leiden: Brill.
- Kogan, L. 2015. *Genealogical Classification of Semitic: The Lexical Isoglosses*. Berlin: DeGruyter.
- Leslau, W. 1976. *Concise Amharic Dictionary*. Wiesbaden: Harrassowitz.
- Leslau, W. 2006. *Comparative Dictionary of Ge‘ez*. Wiesbaden: Harrassowitz.
- Muchiki, Y. 1999. *Egyptian Proper Names and Loanwords in North-West Semitic*. Atlanta: SBL.
- Noonan, B. 2016. Egyptian Loanwords as Evidence for the Authenticity of the Exodus and Wilderness Traditions. Hoffmeyer, J. et al. (eds.) *Did I Not Bring Israel Out of Egypt?* 49–68. Pittsburgh: PSUP.
- Phillips, J. 1997. Punt and Aksum. *Egypt and the Horn of Africa*, JAH 38: 423–457.
- Takács, G. 2012. Semitic–Egyptian Relations. Weninger, S. et al. (eds.) *The Semitic Languages*, 7–17. Mouton: DeGruyter.
- Weninger, S. 2005. Der Wortschatz des klassischen Äthiopisch. Younansardaroud, H. et al. (eds.) *Studia Semitica et Semitohamitica*, 465–488. Münster: Ugarit-Verlag.

Search and you will find: innovative access structures in mobile dictionaries

André H. DU PLESSIS (ahdp@sun.ac.za)

Bureau of the WAT, Stellenbosch, South Africa; Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa.

The access structure of any dictionary is an important one. It provides the user with a means to obtain the required data or to extract valuable information. This dictionary structure mostly overlaps with other structures such as the macro-, micro- and mediostructure, which in turn creates different access routes a dictionary user can follow to try and satisfy his or her lexicographic needs. In printed dictionaries the access structure can already be found on the front page, while in electronic or online dictionaries this structure is usually visible on the home or landing page (Du Plessis, 2015:51–52). Gouws and Prinsloo (2005:64) differentiate between two types of access structures: the outer access structure, which relates to the access or search routes that lead a user to the lemma sign, and the inner access structure, which, among other things, involves the search routes and zones within a dictionary article. With online and mobile lexicographic resources, the traditional functions and uses of dictionary structures are adapted for the online and mobile medium in order to better satisfy user needs. The disadvantage of this, however, is that it becomes increasingly difficult to distinguish between different dictionary structures (Müller-Spitzer, 2013:368–369). Fortunately, e-dictionaries have the freedom to provide new and innovative ways for users to access lexicographic material. For

example, the departure from a traditional macrostructure leads electronic dictionaries to provide a central outer access structure through the search engine of the dictionary (Gouws, 2014). Other features of the outer access structure that have now become standard practice in mobile and online dictionaries are *recent searches*, *favourite searches* and *voice searches*. 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. The integration of other device functions or tools, are also being employed in mobile dictionaries in new and exciting ways. Although the features of a traditional access structure are prevalent in many electronic dictionaries, there are some that have taken advantage of the advancement of technology or by integrating standard or new web/mobile practices to the access structure.

This paper will highlight and evaluate some of these innovative access routes in mobile dictionaries. These dictionaries include, among others, the *Muiswerk Woordenboek*, *Oxford Dictionary of English*, *Collins English Dictionary Complete and Unabridged*, *Merriam Webster* and *Dictionary.com*. This is done to determine which access routes deserve further investigation or should be considered as a standard practice in mobile and possibly other e-dictionaries. The evaluated dictionaries all presented one or more innovative access structure, while some, like *Dictionary.com* and *Oxford* are designed to really utilise the mobile phone's features in order to provide and present data in ways a mobile dictionary user would need or expect it. A key finding is that the success and innovativeness of the access structures could be linked to the way in which the dictionary applications not only satisfy the users' lexicographic needs, but also their technological expectations i.e., how well the dictionary is suited to the medium and exploits the standard or expected practices of said medium. The way in which the dictionary involves the user in the lexicographic process is also another interesting component that can improve the usability and innovativeness of the application.

Lastly, it is also important to take note of the ever-changing landscape of web and mobile development. Many of these access routes may have been around for many years as part of existing online data structures. Practical lexicographers are becoming more versed in web and app development 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

- Du Plessis, A.H. 2015. 'n Analise van die selfoon-WAT: 'n grondslag vir die verbetering van selfoonwoordeboeke. Ongepubliseerde MA-tesis. Universiteit Stellenbosch: Stellenbosch.
- Gouws, R.H. en Prinsloo, D.J. 2005. Principles and Practice of South African Lexicography. Stellenbosch: AFRICAN SUN MeDIA.
- 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. Berlyn: De Gruyter. 367-380

Gouws, R.H. 2014. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3): 481-504.

Frequency or keyness?

Zorica ĐUROVIĆ (zoricag@ucg.ac.me)

Faculty of Maritime Studies Kotor, University of Montenegro, Kotor, Montenegro

The possibility of compiling electronic corpora as from the second part of the previous century has brought new opportunities into vocabulary research. This also resulted in the development of a series of computer software solutions for lexical analysis of texts and building of vocabulary lists for language learners. In this paper, we are discussing the differences in building technical vocabulary lists according to their frequency and keyness in corpora of English for Specific Purposes.

The relevant methodology for frequency count is based upon the use of the AntWordProfiler software (Anthony, 2014) whilst the keyness count will be carried out through the AntConc software of the same author (Anthony, 2014). The advantage of these programmes is that they are free of charge, they have been regularly updated (thus the most recent versions are readily available) and can be used for any language.

The main research question of the paper regards the criterion upon which a word list for English for Specific Purposes (ESP) is to be built to provide the language learners with sufficient and core technical vocabulary. The frequency count would anticipate words appearing most frequently across the corpora. Keyness as a corpus linguistics method, on the other hand, refers to the frequency of the words in a special purpose corpus, compared to their frequency in a reference General English (GE) corpus. The two methods are tested and discussed with reference to a professional corpus of marine engineering instruction books, with English for Marine Engineering Purposes generally considered as extremely demanding, vocabulary wise (Hsu, 2014; Đurović, 2021).

Building a word list upon a frequency count using AntWordProfiler (Anthony, 2014) surpasses the AntConc software in a very important aspect. It provides us with the opportunity to eliminate available lists from further analysis, but also measures the coverage of each list, as well as their cumulative coverage in the corpora. In building technical vocabulary lists this tool especially comes in handy since it provides us with the opportunity to exclude the lists of most frequent GE words (or any other type of word list) from further processing, thus focusing the frequency count on technical vocabulary to be mastered in order to reach the adequate comprehension of the professional genre(s).

Keywords, however, are the words of a “special status” (Stubbs, 2010:21). They are “key” because they capture the essence of particular types of discourses (Culpeper & Demmen, 2015: 1). Therefore, we examined the possibility of combining the two methods for building word lists, glossaries or dictionaries upon both criteria. In particular, the meticulously presented methodology, the results of which are comparable to numerous other word frequency lists, has been supplemented and upgraded by including the keywords extracted from the corpus. The two methods can be applied to any ESP branch either individually or combined, as proposed in the paper.

References

- Anthony, L, 2014. AntWordProfiler, Build 1.4.1.0., [Computer Software], Tokyo: Waseda University. Available from <https://www.laurenceanthony.net/software/antwordprofiler/>
- Anthony, L, 2012. AntConc [Computer Software], Tokyo: Waseda University. Available from <https://www.laurenceanthony.net/software/antconc/>
- Culpeper, J, Demmen, J, 2015. Keywords. In Biber, D. & Reppen, R. (eds.) *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Đurović, Z, 2021. Using corpus linguistics methods for the purpose of enhancing the teaching and learning of Marine Engineering English vocabulary (Upotreba metoda korpusne lingvistike u svrhu unapređenja nastave i učenja leksike engleskog jezika brodomašinske struke), unpublished doctoral dissertation. Nikšić: University of Montenegro, Faculty of Philology.
- Hsu, W, 2014. Measuring the Vocabulary Load of engineering Textbooks for EFL Undergraduates, *English for Specific Purposes* 33, pp. 54-65.
- Stubbs, M, 2010. Three Concepts of Keyness, in M. Bondi and M. Scott (eds.) *Keyness in Texts*, Amsterdam: John Benjamins.

The concept of *lexicographic condensation*: Review and perspectives in digital lexicography

María Teresa FUENTES MORÁN (tfuentes@usal.es)

Faculty of Translation and Documentation, University of Salamanca, Salamanca, Spain.

The framework of the *general theory of the form of dictionaries* (e.g. Wiegand & Fuentes Morán, 2009) describes the concept of *lexicographic textual condensation* (Wiegand, 1996). This concept attempts to explain the particular, albeit not unique, form in which data are presented in dictionaries. As part of this presentation, essentially linguistic data are expressed throughout the dictionary with forms of expression that are not part of natural language. The methods of textual condensation include the following procedures that are applied consecutively and systematically: shortening, abbreviating, omitting, shifting, substituting, summarizing, and embedding.

One of the notable weaknesses in the design of this concept is the fact that it is based on a principle that can only be applied from a theoretical perspective (e.g. Rascon Caballero, 2021:104-106). That is, lexicographers in a real context do not generally formulate the contents in natural language and then *transform* them through condensation procedures in order to include the data in the dictionary according to the required format. Therefore, an actual *condensation* is not taking place, although the lexicographic *text* does present different levels of *density*. In spite of this fact, the strengths in the application of this concept include its contribution to the detailed description of the differences between natural language and the forms of expression used in the dictionary and, with it, to an accurate description of the different levels of *lexicographic textualization* (e.g. Wiegand, 1996). This type of description makes it possible to establish the bases of a partial methodology for the planning, design, comparison and assessment of dictionaries on a formal plane - which is intrinsically connected to their contents (Wiegand, 1998).

The concept of *lexicographic textual condensation* was coined for printed dictionaries, but it may be applied, with some caveats, to digital dictionaries, which also follow these procedures. In fact, textual density in dictionaries is not only motivated by the need to save space - which is more evident in printed dictionaries - but also, as many other current means of communication, by the search for greater dynamism and effectiveness in the transmission of the message (Torres del Rey & Fuentes Morán, 2013).

We apply this approach to the study of some examples of digital dictionaries for human use, and more specifically to Spanish-English bilingual dictionaries of different kinds. To do so, we will use the methodology that was implemented in previous studies on printed dictionaries (e.g. Fuentes Morán & Pradas Macías, 2009; Gouws, 1999; Nielsen, 2002), with some alterations that will allow us to work with digital dictionaries. From a descriptive perspective, the following questions are presented:

- What *condensation* procedures are used in the digital dictionaries in our sample?
- For what types of data?
- What differences can be observed between them and printed dictionaries?
- Do general patterns emerge in the dictionaries in the sample?

In parallel with these questions, and from an epistemological perspective, we will discuss the extent to which the concept of *lexicographic textual condensation* is useful and applicable to current digital dictionaries and, where appropriate, what modifications or adjustments are required.

We will refer once again to the concept of *lexicographic textual condensation* and apply it to current lexicography, in an attempt to establish the bases to redefine *informative density* and *lexicographic textual density*, and to establish some of the parameters that may allow us, in the future, to study legibility and readability in digital dictionaries for human use with a scientific approach.

References

- Fuentes Morán, M. T., Pradas Macías, E. M. 2009. Zur Textverdichtung in spanischen Lernerwörterbüchern. *Lexicographica: International annual for lexicography* (25): 3-10.
- Gouws, Rufus H. 1999. Textual Condensation in the *duden*. *Das Grosse Wörterbuch der Deutschen Sprache* in Zehn Bänden. In: *Duden. Das große Wörterbuch der deutschen Sprache* in zehn Bänden. Band II. Berlin, Boston: Max Niemeyer Verlag, 279-294.
- Nielsen, S. 2002. Textual Condensation in the Articles of *de gruyter wörterbuch deutsch als fremdsprache*. In: *Perspektiven der pädagogischen Lexikographie des Deutschen* II: Untersuchungen anhand des »*de Gruyter Wörterbuchs Deutsch als Fremdsprache*«. Tübingen: Max Niemeyer Verlag, 587-608.
- Rascon Caballero, A. 2021. *Teoría y práctica del ejemplo lexicográfico bilingüe*. Barcelona: Universidad Autónoma de Barcelona.
- Torres de Rey, J, Fuentes Morán, M. T. 2013. Diccionarios electrónicos: (re) estructuras de acceso. In: *Documentos electrónicos y textualidades digitales*. Salamanca: Universidad de Salamanca, 237-245.
- Wiegand, H. E. 1996. Textual Condensation in Printed Dictionaries. A Theoretical Draft. *Lexicos* (6): 133-158.
- Wiegand, H. E. 1998. Lexikographische Textverdichtung. Entwurf zu einer vollständigen Konzeption. In: *Symposium on Lexicography VIII. Proceedings of the Eighth Symposium*

- on Lexicography May 2-4, 1996. Tübingen: Niemeyer (Lexicographica. Series Maior 90), 1-35.
- Wiegand, H. E., Fuentes Morán, M. T. 2009. Estructuras lexicográficas: aspectos centrales de una teoría de la forma del diccionario. Granada: Tragacanto.

The Concept of Lexicographic Data: Some Lexicographic, Economic and Technical Reflections

Pedro A. FUERTES-OLIVERA (pedro@emp.uva.es)

Department of Afrikaans and Dutch, University of Stellenbosch, South Africa; International Centre for Lexicography, Universidad de Valladolid, Spain

According to Fuertes-Olivera & Tarp (2020), lexicography is currently being shaped by three related tendencies: (a) the increasing use of disruptive technologies; (b) the necessity of finding new business models that can finance new lexicographic projects; and (c) the existence of growing competition from other information sources, e.g. Google. These trends are also influencing the practice of lexicography, for example, the selection of lexicographic data. They are any data that have been prepared or accepted by lexicographers and stored in a DWS with the aim of helping humans and/or machines convert them into information *in a straightforward manner*.

Lexicographers typically assume that they must use corpora for data selection and recommend this practice uncritically, i.e. without offering any cost analysis of the process, e.g. how much time and money the process takes. For instance, Kilgarriff (2013) indicates that “word sketches” offer a glimpse of the existing meanings of a particular lemma in a corpus. This is true. It is also true that existing dictionaries, Google searches and minitexts also offer a glimpse of the meanings of a particular lemma (Tarp & Fuertes-Olivera, 2016). The question is: Can we *always and only use* corpora for our new lexicographic project, or can we also use existing tools, e.g. dictionaries and Google searches and minitexts for it? To the best of my knowledge, no scientific research and practice starts from scratch; all create something new starting from and using existing knowledge and practice. Hence, in my view, we can and should use any method — corpora, Google searches and minitexts, grammar books, existing dictionaries, etc. — for selecting our lexicographic data provided that they meet our lexicographic purposes and budget constraints, including time. This means that we must initiate the process of selection by asking questions related with lexicographic, economic and technical issues. Regarding lexicographic questions, ingrained practice such as the use of recursive definitions must be discarded because they are not lexicographic data as they are useless, especially for machine users. For instance, the Spanish defining formula for defining “action nouns”, i.e. deverbal nouns that refer to an action or an event (example 1), is pure nonsense:

dilución
acción y efecto de diluir (action and effect of diluting)

Example 1: definition of *dilución* in Spanish dictionaries, e.g. in DLE

Similarly, out-of-context data do not qualify because they do not help users to convert them

into information in a quick and easy way. For instance, dictionaries do not typically indicate that almost 50% of the meanings of most words are metaphorical extension of their literal meaning (my research on 90,000 Spanish meanings currently stored in the DWS of the *Diccionarios Valladolid-UVa* (Fuertes-Olivera, 2019). For example, the Spanish word *locomotora* (locomotive) is very much used in its metaphorical meaning of driving force of, e.g. economic projects. Dictionaries such as the *Diccionario de la Lengua Española* and *Lexico* (Spanish) do not mention the metaphorical extension at all.

Finally, lexicographic data must be prepared for dynamic use, i.e. being adapted to different user and situation types. This means that we must also ask questions regarding the technical distribution of data in the DWS. I will illustrate this problem with a technical document created for preparing several dictionaries of construction Spanish dictionaries.

References

- Fuertes-Olivera, Pedro A. (2019): “Designing and Making Commercially Driven Integrated Dictionary Portals: The *Diccionarios Valladolid-UVa*”. *Lexicography* 5: 21-41.
- Fuertes-Olivera, Pedro A. & Tarp, Sven (2020): A Window to the Feature: Proposal for a Lexicographically-assisted Writing Assistant. In: *Lexicographica* 36, 257-286.
- Kilgarriff, Adam (2013): “Using Corpora as Data Sources for Dictionaries”, in Howard Jackson (ed.), *The Bloomsbury Companion to Lexicography*, 77-96. London: Bloomsbury.
- Tarp, Sven and Fuertes-Olivera, Pedro A. (2016): “Advantages and Disadvantages in the use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-UVa”. *Lexikos* 26: 273-295.

Aspects of Vernacular Dictionaries: A Case Study of a Hakka Dictionary

Audrey HEIJNS (ajheijns@gmail.com)

School of Foreign Languages, Shenzhen University, Shenzhen, China

The aim of this study is to explore aspects of vernacular dictionaries based on the case study of a Hakka dictionary compiled by Peter Adriaan van de Stadt (1876–1940). Van de Stadt had been trained and appointed as interpreter of Chinese for service in the Dutch East Indies (now Indonesia). Since the mid-nineteenth century the Dutch colonial government’s purpose for employing interpreters was to improve communication with Chinese immigrants, who had settled there long before the Dutch arrived. Besides their work as interpreters, some also published research, literary translations and learning resources for language learning. An important contribution was Van de Stadt’s bilingual learners’ dictionary for the spoken language of Hakka published in 1912.

First, I will establish the basics of making “a profile of the intended user group and a typology of user situations where problems or needs may pop up that can be solved by providing lexicographic data in a dictionary” (Bergenholtz & Tarp, 2003:173). The dictionary under investigation was aimed at improving oral communication between Dutch officials and the temporarily employed immigrants in the tin mines on Bangka and Belitung. I will examine the historical background and specific problems that could be solved by learning the vernacular through the dictionary. All miners were Chinese, mostly China born Hakkas who were recruited from China (Somers Heidhues, 1991:5). The number of new mine workers fluctuated but almost every year a few thousand arrived from and returned to China (Mollema, 1927). To

establish the users' needs, I will investigate the organization of the tin mines to better understand the situation that could lead to misunderstandings, conflicts or problems related to contractual issues, abuse or exploitation, living conditions and other legal or practical matters.

The second step is to find out whether the dictionary meets the features and functions of a Bilingual Learners' Dictionary, including: 1) usefulness for all proficiency levels; 2) inclusion of the L1>L2 part; 3) help to users to better understand the workings of L2; 4) inclusion of grammatical, collocational, stylistic information about the headword; 5) function of being bilingual, usable for both productive and receptive purposes; and 6) function of being monodirectional (Adamska-Sałaciak & Kernerman, 2016). Preliminary findings show that the Hakka dictionary meets these criteria. The dictionary starts with an introduction to the pronunciation and intonation of Hakka, its sentence structure, and a list of references that the compiler consulted and recommends to learners. In part 1 (L1>L2) the entries are listed alphabetically according to Dutch words with their Hakka equivalents in characters and romanization. Some give different usage, some give explanations, and yet others give back-translations of Chinese phrases. Hakka as a spoken rather than written language uses Chinese characters for common words, but for some words or syllables that are typical Hakka, no characters exist. For the latter, Van de Stadt gives a circle with the pronunciation in romanization. Earlier research already concluded that a lot of colloquial material was included given the large number of circles (Sybesma, 2013:135). In part 2 (L2>L1), the compiler lists Chinese characters according to pronunciation with the basic meaning and refers to examples in part 1.

Based on my findings, I will analyse how the dictionary meets the needs of the users, by focusing on aspects of the vernacular and how the information on the workings of Hakka can help the learner build sentences. By categorizing the different types of aspects of the vernacular illustrated with relevant examples, I will show how the dictionary can help learners communicate with the Chinese workers, and establish a relationship and understanding with them.

The outcome of this study will hopefully shed light on aspects of vernacular dictionaries more in general. It aims not only to contribute to historical research on vernacular dictionaries but also to serve as reference material for future compilation of vernacular dictionaries.

References

- Adamska-Sałaciak, A., Kernerman, I. 2016. Introduction: Towards Better Dictionaries for Learners. *International Journal of Lexicography* (29:3): 271-278.
- Bergenholtz, H., Tarp, S. 2003. Two opposing theories: On H. E. Wiegand's recent discovery of lexicographic functions. *Hermes, Journal of Linguistics* (31): 171-196.
- Mollema, J. C. 1927. *Gedenkboek Billiton, 1852-1927*. The Hague: Nijhoff.
- Somers Heidhues, M. F. 1991. A Note on the History of Belitung. *Indonesia* (51): 1-20.
- Sybesma, R. 2013. A History of Chinese Linguistics in the Netherlands. *Chinese Studies in the Netherlands: Past, Present and Future*. Leiden: Brill.
- Van de Stadt, P. A. 1912. *Hakka Woordenboek*. Batavia: Landsdrukkerij.

In progress: the Survey of South African Sign Language Place Names

Chrismi-Rinda LOTH (lothC@ufs.ac.za)*

Jani DE LANGE (delangeJC@ufs.ac.za)*
Susan LOMBAARD (lombaSC@ufs.ac.za)*
Emily MATABANE (matabE@ufs.ac.za)*
Gideon KOTZÉ (KotzéG@ufs.ac.za)*
Annalene VAN STADEN (vanstadenA@ufs.ac.za)*

*Department of South African Sign Language and Deaf Studies, University of the Free State, Bloemfontein, South Africa

This paper reports on a project to collect, document and distribute South African Sign Language (SASL) place names. The importance of documenting place names has been well established (from various perspectives). Place names in sign languages, or signed place names, however, is an under-researched area of study.

As with spoken languages, signed place names carry a specific lexicographic as well as symbolic load, in that they constitute artefacts of cultural heritage and function as socio-political markers. However, these names are generally not standardised, codified or even officially acknowledged. Instead, local Deaf communities assign names to places relevant to them through a process of conventionalisation. They employ at least two different naming methods to allocate place names (see for example Lombaard, 2020). Additionally, sign languages are also subject to sociolinguistic variables, such as dialectical variation, (intra)generational differences, language contact, and language acquisition methods. This dynamism results in a high level of variation. Subsequently, signed place names form a unique parallel toponymy that is not only unwritten but also dynamic. For these conventionalised names to be recognised, whether officially or informally, they must first be documented.

In South Africa there are some initiatives to develop word lists and online dictionaries. For example, Real SASL (Real SASL, 2022) is a crowd-sourced online dictionary, and the National Institute for the Deaf developed an online dictionary (NID, 2021). In all of these resources, place names are only included incidentally. The existing lack of data and recognition means that these names, and with them their inherent heritage and insight into Deaf Culture, are neglected. It also means that there currently exists a corpus gap in SASL. Our project, the Survey of South African Sign Language Place Names (henceforth ‘the Survey’), addresses the imperative to document SASL place names.

The study of signed place names, in South Africa in particular, constitutes a novel research area. For this reason, the first phase of the Survey aims to establish local research protocols that are cognisant of the ethical considerations. During this initial phase, data collection will be limited in scope. The pilot study will focus on a selection of informants from the Bloemfontein area. Data will be collected through means of structured interviews conducted by a well-known and respected person in the Deaf community. The interviews will elicit names relevant to the informant (such as where they grew up), ask their signs for specific places (such as the provinces and major cities), and include a brief profile section to determine sociolinguistic factors (such as age and school attended). The interviews will be recorded by video. In order to protect the privacy of the informants, all signs will be rerecorded by a native SASL user before being entered into the database. SASL place names are not standardised, therefore the database will include all collected variations.

A further consideration is the distribution of this data. It should be made available in a way that serves the Deaf community, and that is readily accessible. Internationally there is some precedent to use online platforms to promote and make sign languages accessible, such as the

New Zealand Sign Language Dictionary (NZSL Dictionary, n.d.). We plan to eventually make our database available freely, in collaboration with other partners. In the meantime, we are launching a mobile app of South African place names, which will include these collected signed names and their variations. The database will also be shared with stakeholders in the Deaf community, for terminology development, language planning activities and any other purposes they might deem necessary.

We will report in further detail on the efficacy of our research protocols, share some preliminary results, and explore the potential of this database as a lexicographic resource. Finally, we will reflect on the contribution this project could make to the promotion and corpus development of SASL, and the associated socio-political benefits to the Deaf community.

References

- Lombaard, S. 2020. Signed place names in South Africa: the Deaf way. In: C Loth (ed). Recognition, Regulation and Revitalisation: Place names and indigenous languages: Proceedings of the 5th international symposium on place names 2019. 18-20 September 2019. Bloemfontein, South Africa: Sunmedia:113-129.
- NID (National Institute for the Deaf). 2021. Resource Hub. Available: .Accessed on 22 March 2022.
- NZSL Dictionary. n.d. New Zealand Sign Language Dictionary. Available: <https://www.nzsl.nz/>. Accessed on 22 March 2022.
- Real SASL. 2022. Real South African Sign Language. Available: <https://www.realsasl.com/>. Accessed on 22 March 2022.

Towards an Integrated Approach to IsiZulu Terminography and Lexicography

Njabulo MANYONI (Manyonin@ukzn.ac.za)

University of KwaZulu-Natal, Durban, South Africa

The first isiZulu bilingual dictionary called the *Zulu-English Dictionary* was compiled by Wits University lecturers CM Doke and BW Vilakazi and published in 1949 (the work commenced in 1940). This was followed by another bilingual English-IsiZulu dictionary in 1958 called the *English-isiZulu/isiZulu-English Dictionary* that was compiled by Doke, Malcom, Sikakana and Vilakazi. The Zulu Dictionary Project commenced on 1 December 1977, headed by AC Nkabinde at the University of Zululand's African Languages Department and yielded two publications, *Isichazamazwi 1* and *Isichazamazwi 2* in 1981 and 1985 respectively, with a more comprehensive version printed in 1998. Another bilingual dictionary (isiZulu/English), the *Scholar's ZULU Dictionary* was compiled by G.R. Dent and CLS Nyembezi (Nkabinde, 1999). Subsequently, *Afrikaans/Zulu Woordeboek met Engelse Vertalings* by Ernst Kotze and Patrick Wela and a monolingual Zulu dictionary, *Isichazimazwi Sanamuhla Nangomuso* by Sibusiso Nyembezi, were produced respectively in 1991 and 1992. Whilst there are a host of other isiZulu dictionaries like the Oxford series and others, there is still a shortage of Language for Specific Purpose (LSP) dictionaries that have been compiled using the official terminology that has been created by *inter alia* the University of KwaZulu-Natal, and the Department of Sport, Arts and Culture (national and provincial).

The structure that is responsible for all non-commercial lexicographic projects for isiZulu is the National Lexicographic Unit (NLU) which was established in keeping with the

PANSALB Amendment Act of 1999 which stipulates that the Board shall establish "national lexicographic units for each of the official languages of South Africa". Currently, isiZulu dictionaries compiled through the NLU processes are *Isichazamazwi SesiZulu* compiled by M Mbatha and The Official Foundation Phase CAPS Dictionary which was published by the NLU's in 2018. Gouws & Prinsloo (2005: 10-11) highlight the importance of clear primary and secondary comprehensive lexicographic processes in the compilation of dictionaries. This paper highlights the shortcomings or total lack of these processes in the current setting for isiZulu.

The University of KwaZulu-Natal, the Department of Sport, Arts and Culture (national and provincial), The South African Weather Service, University of Zululand and other institutions have been involved in terminology development for isiZulu. Terminology has been developed in the disciplines such as Law, Information Technology, Physics, Anatomy, Psychology, Social Work and Environmental Sciences among others. All these terms have been developed with the involvement of the official legislated body for the development of isiZulu Terminology, *uMzukazwe* which is the isiZulu National Language Body. Khumalo (2017) states that the development of discipline specific terminology is the university's response to the assertion that African languages cannot be used as languages of teaching, learning and research because of their lack of discipline specific terminology. This therefore, emphasizes the need to not only create terminology but to further compile LSP dictionaries to allow for the terms to be accessible in a more user-friendly manner.

Nkomo & Madiba (2011) state that glossaries result from the collection, description and presentation of terms, where terms are words that are specific to a particular discipline. The University of KwaZulu-Natal's Language Planning and Development Office has developed terminology for over 10 disciplines and in the process also published term lists for Architecture and Law. The terms are uploaded into various repositories developed by the ULPDO which marks the end of the terminology development project. However, there is a clear shortage of specialized dictionaries that document and describe these terms.

This paper highlights the paucity of specialized isiZulu dictionaries in the presence of a vast amount of isiZulu terminology. By integrated approach in this study, we refer to a careful process that will link the terminology development processes to lexicography in the compilation of specialized isiZulu dictionaries. Furthermore, the paper makes recommendations on the enhancement of the liaison and co-operation between the isiZulu NLB and NLU in maximizing outputs aimed at developing isiZulu.

The presence of published glossaries in isiZulu for the Architecture and Law disciplines presents an opportunity to create a multidisciplinary isiZulu dictionary from the existing glossaries. It is also important to highlight that the process will only require the addition of lexicographic information which can be in this instance the grammatical information about the headword (is it a noun, verb etc.), examples of usage and synonyms (and/or antonyms).

References

- Gouws, R.H. & Prinsloo, D.J. 2005. Principles and practice of South African lexicography. Stellenbosch.: SUN PReSS, African Sun MeDIA
- Khumalo, L. 2017. Intellectualization through Terminology Development. *Lexikos*. (27): 252-264
- Nkabinde, A. C. 1999. The Zulu Dictionary Project. *Lexikos*, (9): 239 – 24.
- Nkomo, D. & Madiba, M. 2011. The Compilation of Multilingual Concept Literacy Glossaries

at the University of Cape Town: A Lexicographical Function Theoretical Approach. *Lexikos* (21): 144 – 168.

The Inclusion of Lexical Particularities in the Comprehensive Dictionary of Gabonese French

Paul A. MAVOUNGOU (moudika@yahoo.fr)

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

Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo

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

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

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

Virginie Ompoussa (vompoussa@yahoo.com)

Département des Sciences de l'Information et de la Communication, Université Omar Bongo

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

Groupe de Recherche en Langues et Cultures Orales, Université Omar Bongo

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

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

The debate on the definition of Gabonese French has resulted in the launch of the project on the *Dictionnaire Général du Français Gabonais*, i.e., a comprehensive dictionary of Gabonese French (DGFG). The project was initiated and is hosted by the *Centre de Recherche en Etudes Germaniques et Interculturelles* (CREGI) at Omar Bongo University.

The main point in the debate was that Gabonese French is not merely the set of lexical particularities and basilectal forms of speech found in the Gabonese utterances of the French language. It is rather the full lexicon set comprising all varieties of the French language as spoken in Gabon as well as French standard language found in Gabon in various domains of life such as government, academia, medias, etc.

This study focuses on lexical particularities that would have to be included in the lemmata and the microstructural component of the DGFG. Lexical particularities are known as words (nouns, verbs, etc.) that bear the quality of being particular and pertaining to a specific place. Several studies of these types of lexical items have been conducted for French in parts of the Francophone world (Dolatabadi, 2016).

For instance, the particularity of French items in Belgium are called Belgicisms (Jacquet, 2014; De Surmont, 2009) and those of Quebec are Quebecisms (Boulanger, 1999). For both French and English in Canada, lexical particularities are known to be Canadianisms (Dawley, 2022; Hien & Reguigui, 2020; Dollinger, 2015; Dollinger *et al.* 2012). In Gabon, the term “Gabonism” was coined (Ngou-Milama & Mve-Ondo, 2016; Moussounda Iboanga, 2008; Dodo Bounguendza, 2008 & 2016). A general understanding from a linguistic perspective is that a Gabonism refers to a term or a phrase specific to the French spoken in Gabon. A few reference works have been compiled and published on these lexical particularities of Gabonese French (Mavoungou, 2013). Nyangone *et al.* (2016) argue that these dictionaries are not an actual reflection of Gabonese French.

The purpose of this paper is to highlight the processes of the inclusion of these lexical particularities within the DGFG. The paper will focus on two processes that would determine the lemmatization process. The detailed outline of the abovementioned processes will be the

core structure of the current paper. Each process will be presented in a section of its own. The organisation and presentation of the data will align with the principles common to functional perspective methods in theoretical lexicography (Tarp, 2014).

The first process is corpus building. One of the lexicographic processes in the compilation of the DGFG is the building of the *Corpus de la Langue Française du Gabon* (CLFG). The work on the building of CLFG is still ongoing, and the compilation of DGFG is intended to be based on the CLFG.

The second process is the determination of Gabonisms. This process has three sub-processes. The first sub-process is to determine between slang and popular language that may range between basilect and upper mesolect. Slang in Gabonese French is called Toli-Bangando. Not all Toli-Bangando terms deserve to be included in the lemmata of the DGFG. There are two reasons that motivated the exclusion of these Toli-Bangando terms: the English basis of these terms and the phonological, grammatical, or semantic detorsion that may feature in Toli-Bangando items. This sub-process is called the sociolectal categorisation.

The database of lexical particularities of Gabonese French is made of Gabonisms. However, a few lexical items accepted as particularities in the public opinion are not of Gabon's origin. One example of such items is the word “*folon*”, which designates a very popular vegetable (*amaranthus viridis*) in Gabon. Although a few Gabonese native languages have a term for this vegetable, the item in Gabonese French originated from Ewondo, a language from Cameroon. The word came via Fang, a cognate language to Ewondo. A determination between items of native origin and items of foreign origin will be the second sub-process for the determination of Gabonisms. Items of foreign origin may be included in the lemmata of the DGFG based on their frequency in the CLFG.

The third sub-process in the determination of Gabonisms is at the semantic level. One of the features of several items in the Gabonisms database is the phenomenon of semantic shift. Two types of words go through this phenomenon. The first type of words are items from Gabonese native languages that undergo a semantic shift when becoming French Gabonisms. The second type of words are lexical items of standard French that are made local with a different or extended or reduced meaning. Both types of lexical items may be considered for the lemmata of the DGFG. However, the lexicographic treatment will certainly be different. This discussion on the inclusion of lexical particularities in the comprehensive dictionary of Gabonese French should contribute to the theoretical and practical grounding of Gabonese French lexicography.

References

- Boulanger, JC. 1999. Images de la norme du français québécois. Les perspectives lexicographiques contemporaines. *Cahiers de Lexicologie* 75(2): 13-127.
- Dawley, E. 2022. Lexical cleansing: Anglicisms and Francophone Québécois identity. *The French Review* 95(3): 35-55.
- De Surmont, JN. 2009. L'apport de la sociolinguistique à la lexicographie: l'exemple récent des belgicismes dans l'édition du Nouveau Petit Robert 2008. *Lexikos* 19: 55-71.
- Dodo Bounguendza, E. 2008. Dictionnaire des Gabonismes. Paris: L'Harmattan.
- Dodo Bounguendza, E. 2016. Le français dans les corps de défense au Gabon: esquisse d'analyse d'un inventaire lexical. *Revue Afrique, Langue et Cultures* 6: 141-174.
- Dolatabadi, H. 2016. Typologie des particularités lexicales des français parlés dans l'espace francophone. *Mélange Francophone. Annales de l'Universitate Dunărea de Jos de Galați*

- Fascicule XXIII, volume X(13): 97–111.
- Dollinger, S. 2015. How to write a historical dictionary: a sketch of the Dictionary of Canadianisms on Historical Principles, Second Edition. Ozwords 24(2): 2-6.
- Dollinger, S., L.J. Brinton & M. Fee. 2012. Revising The Dictionary of Canadianisms on Historical Principles: A Progress Report, 2006—(April) 2012. Dictionaries: Journal of the Dictionary Society of North America 33: 164-178.
- Hien, A. & A. Reguigui. 2020. Sémantique des locutions anatomiques dans les canadianismes. Reguigui, A., J. Boissonneault, L. Messaoudi, H. El Amrani & H. Bendahmane (eds.), Langues en Contexte. Languages in Context. Gatineau: Editions OKAD. 219-246.
- Jacquet, A. 2014. Les journalistes en belgique causent-ils « belge », une fois? Des belgicismes sur les sites d'information. Le Discours et la Langue 6(1): 177-193.
- Mavoungou, P.A. 2013. Gabonese French dictionaries: survey and perspectives. Lexikos 23: 255-272.
- Moussounda Iboanga, F. 2008. Pour une lecture des particularismes gabonais: cas de la ville de Mouila. Faits de langues et identitaires. Le Français en Afrique 23: 113-133.
- Ngou-Milama, E. & B. Mve-Ondo. 2016. La Refondation de l'Université du Gabon. Enjeux Sociétaux et Ethique de la Co-responsabilité. Paris: Editions des Archives Contemporaines.
- Nyangone Assam, B., H.S. Ndinga-Koumba-Binza & V. Ompoussa. 2016. What French for Gabonese French dictionaries? Lexikos 26: 162-192.
- Tarp, S. Theory-based lexicographical methods in a functional perspective: an overview. Lexicographica 30: 58-76.

The historical account and contribution of the *Greater Dictionary of IsiXhosa* to African Language Lexicography

Buyiswa M. MINI (buyi.mini@gmail.com)

This paper seeks to reflect on the vision, goal, and practical processes and methods that were involved in the making of the trilingual, tri-volume *Greater Dictionary of IsiXhosa* (The GDX) as well as on its lexical contents and its significance in the lexicographical work of IsiXhosa and the other African languages of South Africa. It seeks to reveal, from its historical and metalexicographical narratives, how the work was conceived, initiated, and carried forward, initially with vision, commitment and precision in almost all its aspects. In many ways as the processes were progressively improved and fine-tuned, an aspect of the field, lexicography, was unconsciously introduced to, or extended and expanded in, the field of IsiXhosa lexicography in particular, and of African languages lexicography, by extension.

There are some behind the scenes stories that need to be brought to light, not only for their being interesting and sometimes adventurous, but also for the benefit of current and future researchers in the field of lexicography in these languages. These include accounts of how an old existing dictionary of missionary times (by Albert Kropf, 1899, 1915) served as the basic source for lemmas, and how the lemmas were verified and authenticated in the light of evidence of the occurrence of the process of language change in the IsiXhosa language – language change which had occurred, by that time, both naturally (as exemplified, in today's IsiXhosa lexicon, by the following neologistic transformations: imvu => igusha (a sheep); ixhwane => itakane (lamb); bawo => tata (father/dad), and the gradually waning use to a point of almost total disappearance of some words like inkazana (woman), or semantic adaptations and some

natural neologisms.) Other language change instances were the result of linguistic ecology (cf. Connell et al, 2021).

Knowledge of the ways in which field research was undertaken for the contents of the GDX, and of the ways in which the field research teams connected and built rapport with IsiXhosa speaking communities for information gathering, are also of significant value to researchers. Although such research was far from strictly being the scientific research process, some of the activities of the field research teams can be defined as aspects of the scientific research process. Such aspects include the common sense of building rapport with the communities and establishing gateways, and the focus group interviews techniques of information gathering which followed.

The paper will reflect also on the various decisions that had to be made with regards to the most suitable entry type and mode, every time a different type of a vocabulary item was encountered. These include the decision for stem-alphabetisation of lemmas, and creation of necessary abbreviations not previously known in the language. An appraisal of the successes with regard to abbreviation formation, and of the influence of stem-alphabetisation on the language's orthographical changes today, also forms part of this paper. The history of the development of the GDX includes the change from the truncated name of the language, namely *Xhosa* in the third volume's title to the full form of the language's name, namely **isiXhosa**, in the titles of the second and first volumes of the GDX. The second and the first volumes were the second and third or last of the volumes of the GDX to be completed and published.

The paper also considers how the GDX continues to be a basis for further lexicographical work at the present IsiXhosa National Lexicography Unit, as for instance the production of a monolingual dictionary and bilingual dictionaries involving, at least, the three languages represented in the bigger work. All these were envisaged from the early times of the compilation of the GDX. Lastly, the process of how computerization caught up, and became integrated, with the other lexicographical processes during the production of the GDX, is part of the GDX's interesting story and life time.

The research techniques involved are document study, personal narratives, and in-depth individual interviews for triangulation.

References

- Connell, B.A.; Griffiths, S.S., Martin, M., and Hayward, L. (2021). Language ecology, language endangerment, and relict languages: Case studies from Adamawa (Cameroon-Nigeria). *Open Linguistics* 7(1):244-300
- Kropf, A. 1899. Kafir-English Dictionary. Alice: Lovedale.
- Mini, B.M., Tshabe, S.L., Shoba, F. and van der Westhuizen (Eds.). 2003. The Greater Dictionary of isiXhosa. Vol 2. Alice: University of Fort Hare.
- Pahl, H.W., Pienaar, A.M., and Ndungane, T.A. (Eds.) 1989. The Greater Dictionary of [isi]Xhosa. Vol 3. Alice: University of Fort Hare.
- Tshabe, S.L. and Shoba, F. (Eds.). 2006. The Greater Dictionary of isiXhosa. Vol 1. Alice: University of Fort Hare.

Help! A user guide to user guides in electronic dictionaries

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

Department of Afrikaans and Dutch, Stellenbosch University, South Africa

User guides are one of the compulsory texts in a printed dictionary as they “ensure a successful use of the dictionary” (Gouws, 2004:74). The question is whether the same should be said about user guides in electronic dictionaries. Intuitive use of dictionaries “may have been a dream a few decades ago, but it is now achievable due to new technologies” and “it is the lexicographers’ sole responsibility that their products can be used successfully by the target group” (Tarp & Gouws, 2020:25).

This paper examines this dichotomy and asks whether user guides are necessary in electronic dictionaries, and if so, what should such guidance contain? In printed dictionaries, user guides can be found in the extra matter, and they either contain brief notes, or annotated diagrams of dictionary entries or dictionary pages.

Electronic dictionaries have more options for user guides: they could be accessed by a menu item or a help button on every page. They could be presented as a pop-up screen, with help relevant to only the pertinent part of the entry, or the user could be taken to a generic user guide or FAQ page and find the information they are looking for.

Are user guides even necessary in electronic dictionaries? Tarp and Gouws state that “users should be able to find what they need in as few clicks as possible” (Tarp & Gouws, 2020:25). According to Tarp and Gouws (2020) “data should be presented to the users in a way that guarantees intuitive use and smooth consultation by means of appropriate techniques” (Tarp & Gouws, 2020:25). However, Griškevičienė and Berg-Olsen say that the intention to make a dictionary so user-friendly as to make a consultation of a Help section unnecessary is “unattainable ... because any dictionary, no matter how specialised, will be used by a wide range of different people, with different backgrounds and different motivations for looking up information in the dictionary in a specific setting” (Griškevičienė & Berg-Olsen, 2012:648). The aim of this paper is to establish whether user guidance should be a compulsory text in electronic dictionaries, as in their print counterparts, or whether the lexicographers have met their responsibility to design dictionaries that can be used successfully without such a section. In this study, nine electronic dictionaries are examined to establish A) whether they have a user guide or help section, B) how easy it is to find, and C) how comprehensive it is. The results vary a great deal.

The nine dictionaries included in this survey are: Macmillan Dictionary, which is based on the Macmillan English Dictionary; Cambridge Learner’s Dictionary; Collins Dictionary, which is the online Collins Unabridged English Dictionary; The Merriam-Webster.com Dictionary; Merriam-Webster Learner’s Dictionary; Lexico, which is a collaboration between Dictionary.com and Oxford University Press; LDOCE online, which is the online version of the Longman Dictionary of Contemporary English; the online Dictionary of South African English; and Wordsmyth Children’s Dictionary, which is a children’s online dictionary with no print counterpart.

Four print dictionaries are included to show examples of user guides in print dictionaries. These dictionaries are the Cambridge Advanced Learner’s Dictionary, the Oxford South African Concise Dictionary, the Longman South African School Dictionary, and the Collins New School Dictionary.

This paper will show what user support is given in print dictionaries and compare that to the user support offered in online dictionaries. Literature on user guides and user friendliness in electronic dictionaries will be presented. The question of whether user support is necessary in electronic dictionaries will be discussed. Recommendations for better user support will also

be given.

References

Printed dictionaries

- Bullon, S. 2007. Longman South African School Dictionary. England: Pearson Education Limited.
- Cullen, K., Higgleton, E., Collins, C. 2002. Collins New School Dictionary (2e). Great Britain: HarperCollins Publishers.
- Van Niekerk, T. and Wolvaardt, J., 2010. Oxford South African Concise Dictionary. South Africa: Oxford University Press (Southern Africa).
- Walter, E. ed., 2008. Cambridge Advanced Learner's Dictionary. United Kingdom: Cambridge University Press.

Online Dictionaries

- Cambridge University Press. 2018. Cambridge Learner's Dictionary. [Online] Available at: <https://dictionary.cambridge.org/>. [10/04/2022]
- Dictionary Unit for South African English. 2018. Dictionary of South African English. [Online] Available at: <https://dsae.co.za/>. [10/04/2022]
- HarperCollins. 2020. Collins Dictionary. [Online] Available at: <https://www.collinsdictionary.com/dictionary/english>. [10/04/2022]
- Macmillan Education Limited. 2020. Macmillan Dictionary. [Online] Available at: <https://www.macmillandictionary.com/>. [10/04/2022]
- Merriam-Webster. 2015. Merriam-Webster.Com Dictionary. [Online] Available at: <https://www.merriam-webster.com/>. [10/04/2022]
- Merriam-Webster. 2020. Merriam-Webster's Learner's Dictionary. [Online] Available at: <https://learnersdictionary.com/>. [10/04/2022]
- Oxford University Press. 2020. Lexico.com. [Online] Available at: <https://www.lexico.com/>. [10/04/2022]
- Pearson Education Limited. 2020. LDOCE Online. [Online] Available at: <https://www.ldoceonline.com/>. [10/04/2022]
- Wordsmyth. 2018. Word Explorer Children's Dictionary. [Online] Available at: <https://kids.wordsmyth.net/we/>. [10/04/2022]

Other sources

- Gouws, R.H. 2004. Outer texts in bilingual dictionaries. *Lexikos*, 14:67-88.
- Griškevičienė, A. & Berg-Olsen, S. 2012. A golden mean? Compromises between quantity of information and user-friendliness in the bidirectional Norwegian-Lithuanian Dictionary. In Proceedings of the 15th EURALEX International Congress (pp. 648-653). Oslo: University of Oslo, Department of Linguistics and Scandinavian Studies.
- Tarp, S. & Gouws, R.H. 2020. Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture. *Lexikos*, 30:470-498.

Dictionary Base and Lemmata for a Trilingual Dictionary of Fauna

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

The present paper is an excerpt of a dictionary conceptualization plan. The project aims at compiling and producing a trilingual dictionary of fauna. The languages involved in the project are English, French and Civili with the latter as the source language. The main project puts an emphasis on the animals' categories and naming, the hunting methods, the reproduction methods, their traditional value in the Vili culture and their use. The projected dictionary is the first of its kind in Civili and may contribute to a general dictionary in the language.

The purpose of the current paper is to present the “dictionary base” (Wiegand, 1984: 14) as well as the lemma selection method within the process of compiling the intended dictionary. According to Wiegand (1984:14), the dictionary base refers to “the complete linguistic material forming the empirical basis for the production of a language dictionary”. It comprises “at least the lexicographical corpus as the set of all the primary sources: primary sources may be defined as all sources not themselves language dictionaries, the secondary sources as the set of all language dictionaries consulted, and other linguistic material” (Wiegand, 1984: 14). In other words, the dictionary base is the complete dataset for a projected dictionary. In the field of lexicography, Svensén (2009:6) refers to the term “data” as “suitable materials to form the basis of the description”.

Based on Wiegand’s theoretical grounds, the current paper intends to provide answers to the three main matters of this study, i.e., nature and characteristics of the data to be used for the planned dictionary, the collection methodology of these data, and the lemmatization process as well as the core motives of the lemma selection type.

As for the difference between data collection and data selection, the lemmata of a planned dictionary are formed from the selected data through the process of lemmatization (Kanis & Skorkovská, 2010; Ndinga-Kouumba-Binza & Saphou-Bivigat, 2012). In fact, according to Svensén (2009:6), the lemma has a very special position in the dictionary, and its establishment, form and function are subject to theoretical criteria aimed at “meeting the needs of practical lexicography” (Svensén, 2009:93).

Regarding the research methodology, the data collection method required a pre-survey. The pre-survey allowed to proceed above all with the consultation of preliminary works on the subject. In addition to identifying the study areas, this literature review shed light on existing research and studies on the subject. This required a rigorous examination of the documentation. This helped the investigation in developing a better understanding of the study object and a good knowledge of the field of investigation. It also gave support in the formulation of the research problem and hypotheses. Material organization was included, because the researcher must know the means at his/her disposal to carry out the research. Finally, this pre-investigation also resulted in the selection of research participants, data sampling, and the development of the research instruments. The latter included a questionnaire and an interview schedule.

It should be indicated that this paper adopts the conceptualization by Bergenholz et al. (1998) for data distribution structures in macro- and microstructures of special-field dictionaries. Focusing on macrostructural underpinnings, this paper argues that the lemma selection method used in the planned dictionary comes out of this relationship in the sense that (i) data are selected (not collected) to form part of the lemmata according to the typology of the planned dictionary, (ii) the planned dictionary typology is in turn determined by the genuine purpose of the planned dictionary, i.e. user-oriented typology, “focusing not on the content of

the dictionaries but on their so-called genuine purpose” (Svensén, 2009:21), and (iii) the genuine purpose is determined by the users’ needs.

Subsequently, the paper will also outline the process of data collection as well as data selection for the planned dictionary. The paper will provide a full description of the dictionary base, i.e., sources used in a dictionary compilation. In fact, lexicographic data are always inseparable from their sources. In the understanding of Wiegand’s theory, lexicographic data are gathered from three types of sources, i.e., primary sources, secondary sources, and tertiary sources. This is opposed to Svensén’s view, which indicates that “[t]he sources used for the compilation of a dictionary ... are of two kinds: primary sources and secondary sources” (Svensén 2009:39). The current paper is of the view that the determination of source types also depends on the planned dictionary typology.

References

- Bergenholtz, H., Tarp, S. & H.E. Wiegand. 1998. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: Hoffmann, L., Kalverkamper, H. & Wiegand, H. E. (eds.), *Fachsprachen. Languages for Special Purposes. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / An International Handbook of Special-Language and Terminology Research*. Vol. 2. Berlin / New York: de Gruyter, 1762-1832.
- Kanis, J. & L. Skorkovská. 2010. Comparison of different lemmatization approaches through the means of information retrieval performance. In: Sojka, P., A. Horák, I. Kopeček & K. Pala (eds.), *Text, Speech and Dialogue. TSD 2010*. Berlin/Heidelberg: Springer.
- Ndinga-Koumba-Binza, H.S. & G. Saphou-Bivigat. 2012. Dictionary basis and lemmatisation issues for an encyclopedic dictionary of Yilumbu. In: Ndinga-Koumba-Binza, H.S. & S.E. Bosch (eds.), *Language Science and Language Technology in Africa: A Festschrift for Justus C. Roux*. Stellenbosch: SUN Press. 237-250.
- Svensén, B. 2009. *A handbook of lexicography: the theory and the practice of dictionary-making*. Cambridge University Press.
- Wiegand, H.E. 1984. On the structure and contents of a general theory of lexicography. On the structure and contents of a general theory of lexicography, in R.R.K. Hartmann (ed.). *LEX'eter '83* proceedings. Tübingen: Max Niemeyer. 13–30.

An examination of the reception of the dictionaries compiled by the University of Zimbabwe's African Languages Research Institute (ALRI)

Eventhough NDLOVU (evennthough@yahoo.co.uk)

Department of African Languages/University of the Free State, Bloemfontein, South Africa

Since its inception in 2000, the University of Zimbabwe's African Languages Research Institute (henceforth: ALRI) has produced a total of 2 Ndebele dictionaries and 5 Shona dictionaries. The first Ndebele general-purpose dictionary, *Isichazamazwi SesiNdebele* (Hadebe, et al., 2001) turns 21 this year and the only Ndebele specialised dictionary, *Isichazamazwi SezoMcilo* (Nkomo and Moyo, 2006) turns 16 this year. The first Shona general-purpose dictionary, *Duramazwi reChiShona* (Chimhundu, et al., 1996) turns 26 this year. *Duramazwi Guru reChiShona* (Chimhundu, et al., 2001) turns 21 this year. The first and

only, Shona specialised medical terms dictionary; *Duramazwi reUrapi neUtano* (Mpfu et al. 2004) turns 18 this year and the specialised Shona musical terms dictionary, *Duramazwi reMimhanzi* (Mheta et al., 2005) turns 17 this year. *Duramazwi reDudziramutauro noUvaranomwe* (2007) turns 15 this year. Given these milestones, this study examines the reception of the afore-mentioned dictionaries by the target users in order to gauge the dictionary culture of Ndebele and Shona speakers. The study seeks to assess the target users' awareness, knowledge and use of these dictionaries. The primary sources of data were sales records from the publishers of these dictionaries. These records provided data on the number of copies sold and the people who bought the dictionaries. The researcher also used observations in which strategic institutions which offer language mediation services and training as well as teach these languages were selected to see whether they have and use copies of these dictionaries. Document analysis was also used to ascertain whether these dictionaries are part of the school, college and university curriculum in teaching these languages and related subjects. Document analysis was also used to ascertain whether there is a curriculum for imparting dictionary skills. This involved an analysis of the Shona and Ndebele syllabus for primary and secondary education. It also involved an analysis of the course/ module outlines of translation, interpretation, Shona and Ndebele majors at college and university level. The Music and Heritage Studies syllabus and course/module outlines and prescribed reading/reference materials were also analysed to see whether the music terms dictionaries are prescribed as key reading materials in these courses/modules. Semi-structured interviews with purposively sampled language mediators, teachers, lecturers and majors in these languages at primary, secondary, college and university level were used to complement data from the other sources.

A semi-structured questionnaire was also used to collect data on the target users' awareness, knowledge and use of the dictionaries under study. Preliminary research findings of this study paint a somewhat bleak picture which shows that since the publication of these dictionaries, the dictionary culture of the target users has not changed. It emerged that the majority of the target users do not own these dictionaries, are not aware of them and do not use them and they are not prescribed as key reference materials or tools. It was also observed that there is no curriculum for the teaching of dictionary skills. This study further confirms the prevailing myth that in African speech communities, dictionary assistance is not required in one's mother tongue. The findings of this study make clear the dire need for Zimbabwe to invest in a curriculum that will cultivate, ingrain and enhance the culture of using dictionaries in local languages. They make clear the need for a dedicated curriculum across all levels of education coordinated by ALRI to improve dictionary skills. There is need to incorporate the teaching of dictionary skills into the tertiary, primary and secondary schools syllabi and dedicated teaching of dictionary use, an area of education that is somehow taken for granted, and either neglected or, at best, skimpily treated by ALRI and education systems across all levels in their training mandate. The findings of the study also make clear that the closure of ALRI was a mistake since there is still a lot of work that needs to be done to cultivate, ingrain and enhance the culture of using dictionaries in local languages. They also make clear the need for a re-branded ALRI which will assume a coordinating role at national level in the teaching of dictionary use in order to improve dictionary culture, especially in view of dictionaries in local languages.

References

Chimhundu, H, Mberi, NE, Batidzirai, RB, Chitauro, MB, Mangoya, E, Masasire, A,

- Matambirofa, F. 1996. Duramazwi ReChiShona. Harare: College Press.
- Chimhundu, H, Mangoya, E, Chitauro-Maweme, MB, Chabata E, Mpofu, N, Mavhu, W. 2001. Duramazwi Guru ReChiShona: Rinezvirungamutauro. Harare: College Press.
- Chimhundu, H, Chabata, E. 2007. Duramazwi reDudziramutauro noUvaranomwe. Gweru: Mambo Press.
- Hadebe, S, Dube, K, Ndlovu, TM, Mhlabi, SJ, Khumalo, L, Maphosa, M, Ndlovu, V, Masuku, J, Dube, T. 2001. Isichazamazwi SesiNdebele. Harare: College Press.
- Mheta, G, Chinouriri, B, Mavhu, W, Zivenge, W. 2005. Duramazwi reMimhanzi. Gweru: Mambo Press.
- Mpofu, N, Chimhundu, H, Mangoya, E, Chabata, E. 2004. Duramazwi reUrapi neUtano. Gweru: Mambo Press.
- Nkomo, D, Moyo, N. 2006. Isichazamazwi SezoMculo. Gweru: Mambo Press.

How great is thy dictionary? Cross-referencing conditions and practices in *The Greater Dictionary of (isi)Xhosa*

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

School of Languages and Literatures: African Language Studies, Rhodes University

Bulelwa NOSILELA (b.nosilela@ru.ac.za)

School of Languages and Literatures: African Language Studies, Rhodes University

Wanga GAMBUSHE (w.gambushe@gmail.com)

School of Languages and Literature: University of Cape Town

The Greater Dictionary of (isi)Xhosa (GDX), is a tri-volume trilingual dictionary with isiXhosa as the target language of lexicographic treatment while Afrikaans and English comments, including translation equivalents are provided. Conceptualised in 1967 and commencing in the following year, the GDX was completed in 2006, with its volumes appearing separately, starting with Volume 3 (Q-Z) in 1989 (Pahl et al. 1989), followed by Volume 2 (K-P) in 2003 (Mini et al. 2003) and Volume 1 (A-J) in 2006 (Tshabe & Shoba 2006). The dictionary was conceptualised to become a rich lexicographic project attempting a comprehensive coverage and treatment of isiXhosa vocabulary from linguistic and extra-linguistic perspectives (Pahl et al., 1989: xiii). Against this brief overview of the GDX, this paper attempts a critical analysis of cross-reference conditions (Nielsen 1999) and procedures employed by the evolving editorial teams over almost four decades. Cross-reference conditions are understood as “the reference prerequisites” and “mediostructurally relevant aspects” (Wiegand 2004:207) that necessitate cross-referencing. The vital role of cross-referencing as a lexicographic device is reiterated by Bolinger (1985:69) who notes that “half of the lexicographer’s labor is spent” on re-establishing “an infinitude of natural connections that every word in any language contracts with every other word” which are, however, decontextualized by unnatural dictionary structures. This paper analyses cross-referencing in the GDX in the following instances:

1. wordlist-internal cross-referencing connecting two or more lexical items in the GDX,
2. cross-referencing between lexical items and outer-texts, and
3. external-source cross-referencing.

In the first instance, lexical items with semantic relationships and formal relationships such as orthographic and geographic variations are randomly identified as they exhibit cross-reference conditions before evaluating if and how cross-referencing was deployed to establish the relevant relationships that are usually camouflaged by alphabetical lemmatisation. The use of synonym definitions in the GDX and the prevalence of *isihlonipho* (euphemism) in isiXhosa constitute major cross-reference conditions that oblige lexicographers to link different lexical items and dictionary users to move from one lemma to another for meaning information. Orthographic variation linked to isiXhosa dialects make formal relationships necessary between words such as *ninji*, *nintshi*, *nintsi* and *ninzi* (many).

In the second instance, the inclusion of comprehensive grammatical and sociolinguistic descriptions of the language as well as anthropological texts on various practices and beliefs of *amaXhosa* in the GDX front and back matter present cross-reference conditions from the wordlist to the outer texts. In such instances, explicit cross-refencing is employed to refer the user from an article in the main text to the relevant GDX outer texts. For example, entries *kgl Add 40* (khangela iAddendum 40) advise the user to consult Addendum 40 which describes *ukushwama* (ritualistic eating of sacrificial meat) more comprehensively. Other relevant examples include cross-referencing from lemmata representing grammatical terms and formatives to grammatical and user-guide texts in the front-matter.

In the final instance, cross-reference conditions are occasioned by the use of other publications which are explicitly acknowledged as constituting the GDX dictionary basis, especially as sources of lemmata and usage examples. Beyond general acknowledgements, explicit cross-referencing is used in some of those cases. The treatment of the lemma *ubuqati* (indifference) includes the entry (Kr-G), which refers the user to the 2nd volume of the *Kafir-English Dictionary* edited by Albert Kropf (Kr) and revised by Robert Godfrey (G). Another interesting example is the lemma *imvu* (sheep) from where a cross-referencing is made to the Bible to indicate that while this word exists in isiXhosa, it is more used in Christian context, with its synonym *igusha* being used in general contexts.

As highlighted above, diverse cross-reference conditions exist in the GDX because of the nature of the language and the dictionary. Consequently, cross-referencing is extensively used to establish various types of relationships in the language and in the dictionary. While successful cross-referencing procedures are identified, instances of circular and dead cross-references also prevail as a major shortcoming that diminish the greatness of the GDX as a lexicographic resource.

References

- Bolinger, D. 1985. Defining the indefinable. In Ilson, R. (Ed.). 1985:69-73. *Dictionaries, Lexicography and Language Learning*. Oxford: Pergamon Press.
- Godfrey, R. 1915. *Kafir-English Dictionary*. 2nd Ed. Alice: Lovedale.
- Kropf, A. 1899. *Kafir-English Dictionary*. Alice: Lovedale.
- Mini, B.M., Tshabe, S.L., Shoba, F. and van der Weshuizen (Eds.). 2003. *The Greater Dictionary of isiXhosa*. Alice: University of Fort Hare.
- Nielsen, S. 1999. Mediostructures in Bilingual LSP Dictionaries. *Lexicographica* 15:90-113.
- Pahl, H.W., Pienaar, A.M., (Eds.) Tshabe, S.L. and Mini, B.M. 1989. *The Greater Dictionary of Xhosa*. Alice: University of Fort Hare.
- Tshabe, S.L. and Shoba, F. (Eds.). 2006. *The Greater Dictionary of isiXhosa*. Alice: University of Fort Hare.

Wiegand, H.E. 2004. Reflections on the Mediostructure in Special-Field Dictionaries. Also According to the Example of the Dictionary for Lexicography and Dictionary Research. *Lexikos* 14:195-221.

The lemmatisation and treatment of verbal suffixes in Sepedi

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

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

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

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

Words represent the majority of lexical items in the lexicon of a language but the lexicon also includes items smaller than words such as stems and affixes and lexical items consisting of more than one word. Good dictionaries lemmatise affixes e.g. the prefix *on-* to indicate negation in HAT such as *suiker/onsuiker* ‘pure/impure’ or the suffix *-ing* to form principle participles in COBUILD such as *sing/singing*. Sepedi examples of multi-word lexical items are *ela hloko* ‘pay attention’ *ipona molato* ‘admit guilt’, *ka gobane* ‘because’, etc. As far as items smaller than words such as nominal affixes are concerned, the class prefixes *mo-* (class 1), *ba-* (class 2) etc. and suffixes *-ng* (locative), *-gadi* (augmentative) and *-ana* (diminutive) should be lemmatised. Current Sepedi paper dictionaries generally do not lemmatise these affixes, and the ones that do, do so inconsistently. Likewise, the full range of individual verbal suffixes should be lemmatised with special attention to the treatment of combinations of verbal suffixes, cf. Bosch and Pretorius (2017:173).

This paper focuses on the lemmatisation and treatment of Sepedi verbal suffixes in paper and electronic dictionaries. African language dictionaries and Sepedi dictionaries in particular fail to lemmatise verbal suffixes. Given the space restrictions in paper dictionaries lexicographers should attempt to maximally utilize available space presenting the information about the form and meaning of the suffix in a user-friendly way to answer the questions most likely to be asked by the target users. In electronic dictionaries the lexicographer should also utilize all electronic features enabled by the computer such as hyperlinking to relevant information mainly obtained through hovering and clicking activities by the users in the process of information retrieval.

Gouws and Tarp (2017:391) list a number of important features applicable to all e-dictionaries such as improved search methods and access routes, use of hidden data, that is data that are not always on display but can be called up when needed, use of pop-up windows and hypermedia to present additional data, inclusion of video and audio options, new forms of internal and external linking, etc. Model dictionary entries for paper and electronic dictionaries will be presented. It will be argued that the lexicographer should firstly compile the full set of verbal suffixes and study the occurrences of combination of suffixes. Access to electronic text corpora and sophisticated corpus query tools will ensure that the data utilized in the dictionary is based on real language use, and will also provide the lexicographer with additional information such as frequency of use and frequent collocations. Van Wyk (1995:87) calculated that a single verb can occur with up to 4104 suffixes and combination of suffixes.

-iš- (causative verbal suffix) [iʃ] cause, let/force something to happen. **Ba mo ngwadiša lengwalo** They let/force him to write a letter

In an electronic dictionary this basic entry could be substantially enhanced through hyperlinking manifesting as hovering and clicking in the following way:

- Audible pronunciation through clicking
- A pop-up box of a selection of the most frequently used verbal suffixes through hovering and a full table of all verbal suffixes through clicking
- A cross-reference to the articles of *cause*, *let*, *force*, *something* and *happen* as well as all of the English words in the translation of the example sentences through clicking.
- An explanation of phonological changes caused by affixation of the relevant suffix through clicking.

It will be concluded that failure to lemmatise suffixes constitutes a major shortcoming in dictionaries for African languages and Sepedi in particular. Concise but informative separate entries for verbal suffixes should be presented in paper dictionaries and for electronic dictionaries entries featuring true electronic features.

References

- (COBUILD) Sinclair, J. (ed.) 1995. Collins COBUILD English Dictionary. London: HarperCollins.
- (HAT) Odendal, F.F. and Gouws R.H. 2005. Fifth edition. Verklarende Handwoordeboek van die Afrikaanse Taal. Cape Town: Pearson Education South Africa.
- Bosch, S. E. and Pretorius, L. 2017. A computational approach to Zulu verb morphology within the context of lexical semantics. *Lexikos* 27:152-182.
- Gouws, R.H. and Prinsloo D.J. 2005. Principles and Practice of South African Lexicography. Stellenbosch: African Sun Media.
- Gouws, R.H. and Tarp, S. 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography*, 30(4), pp. 389-415.
- Van Wyk, E.B. 1995. Linguistic assumptions and lexicographical traditions in the African languages. *Lexikos* 5, pp. 82-96.

Turning Bilingual Lexicography Upside Down: Realization of an old dream

Sven TARP (st@cc.au.dk)

Centre of Excellence in Language Technology, Ordbogen A/S, Denmark, Department of Afrikaans and Dutch, Stellenbosch University, South Africa; International Centre for Lexicography, University of Valladolid, Spain; Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; and Centre for Lexicography, University of Aarhus, Denmark

This paper focuses on the automatic inversion of bilingual dictionaries, which is an old dream among lexicographers. Over the years, different methods have been tested to make this dream come true without requiring too much manual work. The success rate is depressing. Only a few specialized dictionaries dealing with non-culture-specific topics have passed the test. However,

a months-long research stay at the Danish company Ordbogen A/S in 2021 gave me a unique opportunity to experiment with new technologies and methods that could crack the hard nut. In this respect, the paper reports on the second part of a project conducted during the research stay (see Tarp 2022). It is based on the positive outcome of the first part of the project, where machine translation was used to bilingualize a monolingual L2 database into an L2-L1 database. The second part of the project, which is discussed here, aimed at developing and testing a new method that could effectively invert this database into an L1-L2 one without losing semantic content or requiring too much manual work from the lexicographer. The new method developed breaks completely with the traditional approach to bilingual dictionaries. Nonetheless, it yielded promising results.

The paper will first discuss the theoretical basis for the new approach. It concerns the concept of a bilingual dictionary and the introduction of the terms *object language* and *auxiliary language* with a meaning completely different from the traditional terms source language and target language (see Marello 2003, Tarp 2005, Adamska-Sałaciak and Kerner 2016, Fuertes-Olivera et al. 2018). In this respect, L2 is always the object language while L1 is the auxiliary language. As L2 is the language to be learnt, the semantic and syntactic description of this language is taken as the basis for both an L2-L1 and an L1-L2 database. In this respect, a short L2 definition of each sense of L2 words or expression plays a central role in the inversion.

During the research stay, an ad hoc program was designed that could upload relevant English lemmata, definitions, and example phrases from existing L2 databases; translate the two latter into any L1; automatically present candidates for L1 equivalents; and subsequently invert all of it starting from the L1 equivalents (which now become L1 lemmata), while the former L2 lemmata now become L2 equivalents with the translated L2 definitions in a new role as meaning differentiators – all of it without losing semantics!

After creating a monolingual lexicographic L2 database and bilingualizing it into an L2-L1 database using machine translation, the main idea was to turn the L2-L1 database into an L1-L2 database, with the original and now translated definitions of L2 words – together with the corresponding L2 example phrases and their L1 translations – following these L2 words throughout, including during their metamorphosis from lemmata to equivalents. Since much of the process can be done mechanically, this approach increases productivity. But it also improves the quality of the overall product, since L2 words – whether they appear as lemmata or equivalents – are never detached from their original semantic and syntactic L2 universe, as frequently happens in traditional L1-L2 dictionaries.

There is also an unexpected additional benefit: Once a monolingual base database is created, it can be bilingualized and turned into various languages, thereby achieving significant economies of scale.

The paper will discuss the whole inversion process in great detail and provide various examples of how it works in practice. Finally, it will briefly discuss some remaining challenges as well as the lexicographers' new role and tasks. It will conclude that – despite the ad hoc program working automatically and performing surprisingly well – the human lexicographers and their specialized skills and creativity are definitely still required to guarantee a high-quality product.

References

Adamska-Sałaciak, A. and I. Kerner. 2016. Introduction: Towards Better Dictionaries for

- Learners. International Journal of Lexicography 29(3): 271-278.
- Fuertes-Olivera, P.A., S. Tarp and P. Sepstrup. 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Diccionarios Valladolid-UVa. Lexikos 28: 152-176.
- Marello, C. 2003. The Bilingual Dictionary: Definition, History, Bidirectionality. Hartmann, R.R.K. (Ed.). 2003. Lexicography. Critical Concepts II: 325-342. London/New York: Routledge.
- Tarp, S. 2005. The Concept of a Bilingual Dictionary. Barz, I., H. Bergenholz and J. Korhonen (Eds.). 2005. Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch: 27-41. Frankfurt am Main: Peter Lang.
- Tarp, S. 2022: Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology. Lexikos, 32: 66-87.

Using machine translation to compile bilingual lexicographical databases: An unexpected but highly welcome research outcome

Sven TARP (st@cc.au.dk)

Centre of Excellence in Language Technology, Ordbogen A/S, Denmark, Department of Afrikaans and Dutch, Stellenbosch University, South Africa; International Centre for Lexicography, University of Valladolid, Spain; Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; and Centre for Lexicography, University of Aarhus, Denmark

During the summer and early autumn of 2021, the author had a months-long research stay at the Danish company Ordbogen A/S. He was assigned to the company's ODIN team, where most of its research and development (R&D) takes place. During the sojourn, a number of pre-determined projects were conducted to see how far technology has come in terms of improving and completely changing the methods to compile bilingual dictionaries. The research stay represented a unique opportunity to experiment with lexicographical data by means of new techniques and ad hoc programs created by the company's programmers.

This paper will briefly inform about the research stay and focus on one of the projects that came out with a highly unexpected and promising result. The first part of this project was to examine how machine translation can be integrated into the compilation of bilingual lexicographical databases and to which degree it makes sense to use it today. After decades of struggle, the technology has improved considerably during the past few years and it is probably only a question of time before it will be adopted at a broad scale within lexicography. It is thus necessary to be at the forefront of this development.

Initial expectations were that machine translation is not yet of immediate relevance to lexicography. The view was shared by the head of the ODIN Team, according to whom the technology only becomes really attractive and relevant for production purposes when over 70 percent of the translations are correct and can be inserted directly into the database. Both parts judged that it still takes some time before this success rate can be achieved.

The project started by testing two machine translators (Google and DeepL) using the monolingual Spanish database, which Pedro Fuertes-Olivera is compiling in Valladolid (at that moment, he had completed approximately 80,000 senses). 200 random definitions from this

database were selected and then translated into English, after which source and target definitions were systematically compared using an ad hoc classification based on lexicographical criteria. The two translators yielded very different results. The performance of Google Translate was disappointing, but as expected. About 30 percent of the translations were acceptable. Another 40 percent had major or minor errors that were disruptive to understanding. And the last 30 percent were straight-out incomprehensible. DeepL performed much better. 78% of the translated definitions were completely correct, while 22% had major or minor errors. Hence, DeepL seemed to comply with the 70% success rate that makes it relevant for production purposes.

The result must be understood correctly. Although translating 200 definitions out of 80,000 does not give statistically precise results, the tendency is convincing. However, the result cannot be generalized and directly applied to translation in the reverse language direction, i.e., from English to Spanish, or to translation between other language pairs, each of which has its own personality. It was, therefore, necessary to conduct further tests with DeepL. The target users of the definitions contained in the monolingual Spanish database are native Spanish speakers. However, since the translation of these items only makes sense if the target users are non-native speakers, it was now decided to extract empirical data from a monolingual dictionary designed for the latter. For this purpose, 200 random definitions were taken from the *Oxford Dictionary* on the Lexico platform. These definitions are characterized by a more straightforward language with less complex syntax than those from the Valladolid-UVa database. DeepL was now used to translate them into Spanish and Danish, respectively. The result was even more convincing than previously.

Of the 200 definitions translated into Spanish, 93,5% were now completely correct, 6,5% had minor problems, and only 2% were unusable. Of the 200 definitions translated into Danish, 71% were correct, 24% had minor problems, and the remaining 5% were unusable. In both tests, the unusable translations required a total rewrite, whereas the ones with minor problems only needed a slight reworking which, in most cases, could be done using another of DeepL's functionalities. Out of curiosity, a few samples of the same English definitions translated into Chinese, French, and German, respectively, were revised by colleagues, and in all three cases, the DeepL translator showed a surprisingly high accuracy rate.

The paper will provide examples from the various tests and show how to use DeepL to facilitate the lexicographers' job. Finally, it will discuss the implications for the compilation of future bilingual lexicographical databases.

Reference

Tarp, S. 2022: Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology. *Lexikos*, 32: 66-87.

To see and understand: visual elements in Foundation Phase Dictionaries

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

Department of Curriculum studies, University of Stellenbosch, Stellenbosch, South Africa.

Dictionaries used in an educational setting for teaching and learning undoubtedly have an educational purpose and function (if a dictionary function is understood according to the

lexicographic theory of functions, Tarp (2009:275)), where such an educational function is concerned with the active development of literacy skills. In the case of dictionaries for the Foundation Phase, target users have emergent literacy skills and specifically developing reading skills. For this reason, lexicographers make use of a variety of lexicographic devices, for example visual elements to render language support to the target user.

Learners in the Foundation Phase form the target users of this category of dictionaries and the focus of the article is situated within the lexicographic theory of the user perspective, and the field of pedagogical lexicography. Concepts known in reading research, language acquisition and second language teaching come into play and will be applied to visual elements in dictionaries.

According to Stein (1991: 99) sight plays a dominant part in the cognitive and linguistic development of the human brain. The way in which visual perception is fundamental to comprehension, manifests when we revert to pictures, drawings, and gestures when people do not understand our communication in everyday life. Visual illustrations (Stein, 1991: 101) in lexicographical works have been available since the Middle Ages and specifically in French, Spanish, Italian, English and German bilingual glossaries. Visual illustrations are also common in South African dictionaries and the dictionaries focused on in this article make extensive use of colour illustrations as well.

In recent research the use of visual illustrations has been described by researchers such as Lew & Doroszweska, 2009; Kemmer, 2014; Liu, 2015; Klosa, 2016; Biesaga, 2017; Taljard & Prinsloo, 2019; as well as Akram, Sajid, Nawazish & Rizwan, 2020. Klosa (2016: 516) defines an illustration as a particular kind of image, which is used in conjunction with a text and which decorates, illustrates, or explains the text. According to Klosa (2016: 516) there is always a particular relationship between the (printed or electronic) text and the image. When text and image are combined in a dictionary, both are visible at the same time, and they refer to each other. Klosa (2016: 516) considers the relationship between the definition of a headword and the illustration as complementary to each other, so that the whole meaning could be ascertained from the definition and the illustration, with the illustration completing the text and vice versa. It is generally effective when a written description of the external form of an object reaches its limits, and the visualization of the object in an illustration contributes to a better understanding. In the case of Foundation Phase learner's context of emergent literacy, illustrations tend to play an even bigger and more important role in the conveyance of meaning. The use, value and types of illustrations in dictionaries are investigated by means of a literature study and then applied within the literacy context of the target user. Use of visual illustrations in accordance with the literacy needs of the target users forms the focus of the paper

The order in which literacy skills are learned and acquired by humans, namely *listening, speaking, reading, and writing*, is significant for the description of the emergent literacy context. The assumption is made that a learner when school-ready in the Foundation Phase, should be able to listen and speak with ease in her home language, but is also ready to learn reading and writing skills in the home language.

Four bilingual dictionaries are researched in which Afrikaans functions as one member of the language pair. The *Afrikaans English Xhosa WAT/WHAT/NTONI* dictionary contains full-colour illustrations against a white background. The *Longman Grondslagfasewoordeboek Afrikaans/English* is a partially full-colour dictionary with a thematic section in colour, against a colour and white background. The *Official Foundation Phase CAPS English-Afrikaans dictionary* is a full-colour dictionary with glossy pages, against a colour and white background.

The *Oxford First Bilingual Dictionary. English and Afrikaans* also contains full-colour pictures.

Some preliminary findings include that literacy skills, like productive and receptive skills, as well as cognitive skills are developed by context-rich illustrations with familiar frameworks of meaning and reference, and with opportunity for interactivity. Illustrations providing the opportunity for interaction by means of descriptive details and story characters conveying emotions are functional. Dictionaries using probing questions in illustrations support the development of productive literacy skills.

References

Dictionaries

- Botha, W. (Ed). 2018. *Picture Dictionary: English–Afrikaans Gr R–3*. South Africa: NLU for Afrikaans.
- Gouws, R., L.Potgieter & S.Burgess (Eds.). 2010. *Grondslagfasewoordeboek Afrikaans/English*. Cape Town: Maskew Miller Longman.
- Simpson, K. (Ed.). 2021. *Oxford First Bilingual Dictionary. English+Afrikaans*. Cape Town: Oxford University Press Southern Africa.
- Van Wyk, C., I. Lusted & S. Zotswana. 2004. *Wat/What/Ntoni? Afrikaans-English-Xhosa*. Durbanville: Garamond Publishers.

Other

- Akram, M., A. Sajid, M.K. Nawazish & S. Rizwan. 2020. An evaluation of Pictorial Illustrations in Urdu Dictionaries. Available: <http://jorurdu.bzu.edu.pk>. [2022, February 22].
- Biesaga, M. 2017. Pictorial illustrations in Encyclopaedias and Dictionaries – a Comparison. Available: <http://creativecommons.org/licenses/by-sa/4.0/> [2022, February 22].
- Durkin, P. (Red.) 2016. The Oxford Handbook of Lexicography. Oxford: Oxford University Press.
- Kemmer, K. 2014. Illustrationen im Onlinewörterbuch. Text-Bild-Relationen im Wörterbuch und ihre Emperische Untersuchung. Mannheim: Amades.
- Klosa, A. 2016. Illustrations in Dictionaries; Encyclopaedic and Cultural Information in Dictionaries, in Durkin, P. (red.), The Oxford Handbook of Lexicography. Oxford: Oxford University Press: 515-531.
- Lew, R. & J. Doroszweska. 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. International Journal of Lexicography 22(3): 239-257.
- Louw, P. 2005. Ostensiewe illustrasies as mikrostrukturele items in verklarende skoolwoordeboeke gerig op leerders in die junior sekondêre fase. Lexikos 15: 90-101.
- Liu, X. 2015. Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. Lexikos 25: 210-232.
- Stein, G. 1991. Illustrations in Dictionaries. International Journal of Lexicography 4(2): 99-127.
- Taljard, E. & D.J. Prinsloo. 2019. African Language Dictionaries for Children-A Neglected Genre. Lexikos 29: 199-223.
- Tarp, S. 2009. Reflections on lexicographical user research. Lexikos, 19:275-296.

Towards a corpus of South African Sign Language: Sociolinguistic and Phonological Challenges

Andries VAN NIEKERK (andriesvanniekerk@nid.org.za)

National Institute for the Deaf; Stellenbosch University, Stellenbosch. South Africa

This presentation will discuss the challenges faced by the National Institute for the Deaf with the development of a video dictionary for South African Sign Language (SASL).

Variation occurs in sign languages as in spoken languages. Lexical variation is very common and has been related to individual schools for the deaf, so-called “schoolization”. This variation tends to extend to the surrounding deaf community as well. South African Sign Language (SASL) has lexical variation; 3.9 signs per lemma (Van Niekerk, Baker & Huddlestone, 2020). This variation was found within signers of the same generation; if all users of SASL were to be included, the variation would likely be greater.

The National Institute for the Deaf created a video dictionary to document variation and found challenges predominantly within the sociolinguistic and phonological domains.

There is great variation in the profile of the SASL user. In most cases the native signer has hearing loss. Factors that are relevant are the degree of hearing loss, age of onset, and cause of hearing loss. These are relevant because they determine the extent to which a spoken language was acquired, which possibly influences the fluency achieved in SASL. This raises the question of what defines a native user.

In spoken languages a native user is quite easy to define – a person who grew up in a given language community would typically be viewed as a native speaker of that language. Deaf people tend to not acquire language from their primary caregivers.

The age the deaf person first attended a school for the deaf is another crucial factor. This age would typically be the first time the person was exposed to SASL. What age then determines whether a person qualifies as a native user of that language? These questions arose during our dictionary compilation. How we define a native SASL user is critical to the integrity of our dictionary.

In our view, a native signer is defined as a deaf person who attended a school for the deaf for their entire school career and who prefers to communicate in SASL. Ideally the definition of a native signer would be a deaf person born to deaf parents where SASL is the preferred language within the family. However, the vast majority of deaf people are born to hearing parents and it would severely limit our access to informants and not be representative of the language community if they were to be excluded from our definition.

Phonologically, what differentiates one sign from another? There is no SASL corpus or body of research that lists all phonemes. The dictionary followed mainstream phonological comparison methodology (McKee & Kennedy, 2000). Phonological parameters were compared to determine which signs differed phonetically; these were then grouped under the same entry. Given that there is insufficient research in the phonology of SASL, it would be premature for us to decide what is phonetic variation and what is phonological variation.

There is no established terminology for the phonological parameters. One of these parameters is movement. One can imagine the multitude of movements that are possible within the signing space. It would be premature for us to create labels for these without a corpus from which conclusive data could be drawn.

Current documenting in the NID SASL Video Dictionary covers two phonological parameters: the location of the sign and the handshape used. To date we have found 70 handshapes for SASL and anticipate that more will be found. Several of these handshapes are only used in a small number of signs, and are likely to eventually disappear due to natural standardisation. Our responsibility is to make sure that a record remains of these signs in order to create a dictionary and eventually a corpus that are truly representative of this language.

The challenges that we encounter as we progress with our dictionary – and which we will be reporting here – can benefit from the inputs of the wider lexicographic community.

References

- McKee, D., & Kennedy, G. 2000. Lexical comparison of signs from American, Australian, British and New Zealand sign languages. *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, 49-76.
- Van Niekerk, A. 2020. A lexical comparison of South African Sign Language and potential lexifier languages. Unpublished Masters dissertation. Stellenbosch University.

An analysis of a corpus-based approach for purposes of developing a bilingual multidisciplinary e-dictionary for students at the University of KwaZulu-Natal

Tholakele Glenrose ZUNGU (ZunguT@ukzn.ac.za)

University of KwaZulu-Natal, Durban, South Africa

In this paper, the development of a bilingual multidisciplinary electronic dictionary using the Language for Specific Purposes corpus-based approach is explained. This is strictly for pedagogical and empowerment purposes for isiZulu-speaking students in various fields. This paper will provide an analysis of the University of KwaZuluNatal parallel corpus building project for the purposes of developing a bilingual e-dictionary. It will do so by presenting, examining, and discussing the UKZN LSP corpus building project and looking at its state of readiness for the development of a bilingual multidisciplinary e-dictionary. Otlogetswe (2006) argues that the quality of retrieved information for lexicographic purposes largely depends on the information input during corpus building. It is therefore important to ensure that the information input during the corpus building phase is representative of the terminology that is used in the respective fields. Also, if the e-dictionary is bilingual, it is important to have both languages represented during corpus building.

UKZN prides itself on being a bilingual institution, not only on paper but also in practice. The university recognizes English and isiZulu as the languages of the university. Whilst the university recognizes English as the primary language of its academic program, it also commits itself to the development and intellectualization of isiZulu to be the language of administration, teaching and learning, research, and innovation (Khumalo, 2015). There has been several challenges and milestones at UKZN over the past 10 years (Ndimande-Hlongwa & Ndebele, 2017; Kamwendo, Hlongwa & Mkhize, 2014; Ndimande-Hlongwa, Balfour, Mkhize & Engelbrecht, 2010). These milestones include but are not limited to the development and standardization of isiZulu terminology in scientific fields such as Anatomy, Law, IT, Computer Science, Architecture, Criminology, Environmental Science, Mathematics, Research, Nursing, Social Work, and Physics. The total number of standardized isiZulu terms

to date is over 9 000 and each year 1000 more terms are standardized. UKZN has ensured that all the standardized isiZulu terminology is made available through the university's Human Language Technologies (HLT's), isiZulu Term Bank and Zulu Lexicon (Khumalo, 2017). The UKZN HLTs include IsiZulu Term Bank which is available on <https://ukzntermbank.ukzn.ac.za/PublicSearch.aspx>, the Zulu Lexicon which is available for free on the App Store for iPhone users and the Google App store for android users as well as the spellchecker that is available for download on <https://jmp.sh/CTi3fbB>. Moreover, the UKZN has published bilingual glossaries (English-isiZulu) for the field of Architecture and Law. These publications are available at the UKZN press <https://www.ukznpress.co.za>. As it stands, the IsiZulu National Corpus <https://iznc.ukzn.ac.za/>, contains over 20 million tokens. It is worth mentioning that this corpus is yet to be updated online. The current statistic is over 30 million tokens. UKZN also translates all PhD graduate abstracts and other university documents into isiZulu. This is achieved through the DR9 rule which compels all PhD graduates to submit their abstracts to the University Language Planning and Development Office for translation. This is to ensure that all PhD abstracts have both an English and an isiZulu version. The university also offers bilingual tutorials for isiZulu-speaking students. These processes also play a big role in building LSP corpora.

All the abovementioned processes and procedures have been the building blocks towards the intellectualization of isiZulu at UKZN. Despite all the achievements as far as isiZulu is concerned, language continues to be a barrier to accessing knowledge (Khumalo, 2017; Ndimande-Hlongwa & Ndebele, 2017; Kamwendo, Hlongwa & Mkhize, 2014; Ndimande-Hlongwa, Balfour, Mkhize & Engelbrecht, 2010). Also, isiZulu continues to be under-resourced which makes it even harder to access and share knowledge in isiZulu (Keet and Khumalo, 2014, 2017). There are no translated field-specific textbooks in isiZulu and there are no field-specific dictionaries for isiZulu.

LSP corpus building at UKZN currently depends on the translation of PhD graduate abstracts, oral data collected during tutorials, and other university documents that are translated by the University Language Planning and Development Office. These documents are screened, scanned, and saved for purposes of corpus building. As explained above, this paper will discuss the development of a bilingual multidisciplinary e-dictionary using the LSP corpus-based approach. UKZN has done very well to develop isiZulu at UKZN. This paper will prove that the production of scientific texts in English and isiZulu at UKZN forms a solid basis for the development of a corpus-based e-dictionary for various disciplines within the university for the benefit of students at UKZN.

References

- Keet, C. M., & Khumalo, L. 2017. Toward a knowledge-to-text controlled natural language of IsiZulu. *Language Resources and Evaluation* 51(1): 131-157.
- Khumalo, L. 2015. Semi-automatic term extraction for an IsiZulu linguistic terms dictionary. *Lexikos*, 25, 495-506.
- Khumalo, L. 2017. Intellectualization through terminology development. *Lexikos* 27: 252-264.
- Otlogetswe, T. 2006. Challenges to issues of balance and representativeness in African lexicography. *Lexikos*, 16(1).
- Ndimande-Hlongwa, N., & Ndebele, H., 2017. Embracing African languages as indispensable resources through the promotion of multilingualism. *Per Linguam: a Journal of Language Learning* 33(1): 67-82.

Ndimande-Hlongwa, N., Balfour, R. J., Mkhize, N., & Engelbrecht, C., 2010. Progress and challenges for language policy implementation at the University of KwaZulu-Natal. *Language Learning Journal* 38(3): 347-357.

