AFRICAN ASSOCIATION FOR LEXICOGRAPHY

28th International Conference

1 – 4 July 2024

AFRICAN ASSOCIATION FOR LEXICOGRAPHY

28th International Conference

1 – 4 July 2024

Abstracts

Hosted by: Department of African Languages, University of Pretoria

Conference organisers: Elsabé Taljard and Michelle Goosen

Abstract reviewers: Blanche Assam, Herman Beyer, Maret Blom-de Wet, Sonja Bosch, Theo Bothma, Dané Claassen, Gilles-Maurice de Schryver, André du Plessis, Ilse Feinauer, María Teresa Fuentes Morán, Pedro Fuertes-Olivera, Wanga Gambushe, Michelle Goosen, Rufus Gouws, Marissa Griesel, Langa Khumalo, Phillip Louw, Evans Lwara, Rooweither Mabuya, Sponono Mahlangu, Paul Mavoungou, Gift Mheta, Guy Roger Mihindou, Buyiswa Mini, Lorna Morris, Steve Ndinga-Koumba- Binza, Eventhough Ndlovu, Dion Nkomo, Bulelwa Nosilela, Thalepo Otlogetswe, Danie Prinsloo, Bonny Sands, Elsabé Taljard, Michele van der Merwe, Menno van Zaanen, Zola Wababa, Daniela Waldberger, Beata Wójtowicz

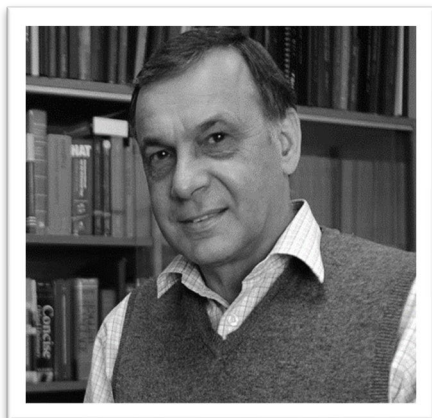**Abstract booklet editors:** Sonja Bosch, André du Plessis and Lorna Morris

Table of Contents

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
2023 – 2025

President:	Prof L. (Langa) Khumalo
Vice-President:	Prof S.E. (Sonja) Bosch
Secretary:	Prof D. (Dion) Nkomo
Treasurer:	Prof E. (Elsabé) Taljard
Editor *Lexikos*:	Mr A.H. (André) du Plessis
Members:	Dr P. (Phillip) Louw
	Dr L. (Lorna) Morris
	Dr H.S. (Steve) Ndinga-Koumba-Binza
Conference organisers:	Prof Elsabé Taljard & Ms Michelle Goosen

MESSAGE FROM THE AFRILEX PRESIDENT

On behalf of the AFRILEX Board, I am deeply honoured to officially welcome you to the 28th Annual International AFRILEX Conference. I extend a special welcome to all the present honorary AFRILEX members and our two keynote speakers Prof. Sven Tarp, who is the Director of the Centre for Lexicography at Aarhus University in Denmark, and Prof. Sonja Bosch, who is Professor Emerita and Professor Extraordinaria at the University of South Africa.

This 28th Annual International AFRILEX Conference is hosted by the University of Pretoria, a university that pioneered the development of corpora for all official languages in South Africa at the turn of the century. It also led advances in corpus lexicography and became the natural home of the Digitization Node for the South African Centre for Digital Language Resources (SADiLaR). SADiLaR is a national research infrastructure that is part of the South African Research Infrastructure Roadmap. Its strategic function is to create, manage, and distribute digital language resources as well as applicable software for all official languages in South Africa through its digitisation programme, while through its digital humanities programme, SADiLaR stimulates and supports computational research and development in the humanities and social sciences. I want to particularly thank our hosts Professor Elsabé Taljard and Ms Michelle Goosen for a sterling job in organizing this year's conference at the University of Pretoria.

I extend the Board's appreciation to SADiLaR for the pre-conference workshop titled *Towards a sustainable National Digital Term Bank for the official languages of South Africa: Collaboration vs Fragmentation*. I am happy to note that the workshop is led by Prof. Justus Roux, Dr Friedel Wolf and Mr Juan Steyn. The workshop is timely in the context of the imminent National Policy on Terminology and will cover, inter alia, critical issues on optimal

procedures and infrastructures required to facilitate and manage the technological aspects of a sustainable national digital term bank, and the role of SADiLaR as a national infrastructure on matters related to longer-term planning.

This year's conference programme displays a range of presentations covering a broad spectrum of very interesting research topics. The programme also has three special sessions, the first is a Globalex virtual session, the second is a session that has a special focus on dictionary publishing (the Publishers' Session), and the third is a session dedicated to the National Lexicography Units (the NLU Session). I therefore would like to invite every delegate to attend and listen to all these engaging sessions.

I would like to take this opportunity to invite all the presenters to develop their presentations into article manuscripts and submit them for peer review and possible publication in the AFRILEX's International Gold Open Access journal *Lexikos* (http://lexikos.journals.ac.za/pub/index). I want to further express my gratitude to the abstract reviewers who were generous with their time and expertise.

Finally, I wish to thank the AFRILEX Board members for their tenacity and hard work in organizing and making this conference a success. Thank you to our Vice President, Prof. Sonja Bosch, Dr Lorna Morris, and Mr André du Plessis for the sterling work in compiling and editing the abstracts booklet. We all thank our Treasurer, Prof. Elsabe Taljard for the stringent financial oversight that ensures our sound financial position. Thank you to our Secretary Prof. Dion Nkomo for the timely correspondence, and to Dr Phillip Louw and Dr Hugues Steve Ndinga-Koumba-Binza, for their hard work and support in the various activities leading up to this conference. Your work on workshops and growing the field of lexicography in the continent is visible and impactful.

I wish you all an engaging, stimulating, and successful 28th International AFRILEX conference.

Langa Khumalo
President: AFRILEX

AFRILEX CONFERENCE PROGRAMME

28th International Conference of the African Association for Lexicography
1 – 4 July 2024, University of Pretoria, Pretoria, South Africa

	MONDAY 1 JULY
	PRECONFERENCE WORKSHOP *Towards a sustainable National Term Bank for the official languages of South Africa: Collaboration vs Fragmentation* Plant Sciences Auditorium
09:00-09:20	**Welcome** Mr Juan Steyn **Introduction and setting of the broader scene** Prof. Justus Roux
09:20-09:40	**International and national tendencies and practices in Lexicography and Terminology** Prof. Rufus Gouws
09:40-10:00	**Current state and practices of Terminology development in South Africa** Dr Mariëtta Alberts
10:00-10:30	**Terminology development in African languages** Prof. Dion Nkomo
10:30-11:00	**TEA** Foyer of Graduate Centre
11:00-11:20	**NLU's view on terminology development** Dr Sponono Mahlangu
11:20-12:00	**Role of SADiLaR: advice on formats, backups, licensing, software, platforms; Community of Practice – Reflections** Mr Juan Steyn and Dr Friedel Wolff

12:00-12:55	**Summary; Open-floor discussions and contributions; Resolutions** Mr Juan Steyn and Prof. Justus Roux	
12:55-13:00	**Conclusion and workshop closure** Mr Juan Steyn	
13:00-14:00	**Conference registration and packed lunch** Foyer of Graduate Centre	
14:00-14:10	**Welcome by AFRILEX President** Plant Sciences Auditorium Prof. Langa Khumalo	

Parallel Sessions

	Plant Sciences Auditorium (Chair: H. Steve Ndinga-Koumba-Binza)	**Graduate centre 1-68** (Chair: Justus Roux)
14:10-14:40	*Implementation progress of the provisions of the African Union Language Plan of Action in Zimbabwe: A lexicographer's perspective* (Eventhough NDLOVU)	*Contextualisation in dictionary consultation and Generative Artificial Intelligence – a reader's perspective* (Theo BOTHMA)
14:40-15:00	**TEA** Foyer of Graduate Centre	
	Graduate centre 1-64 (Chair: Danie Prinsloo)	**Graduate centre 1-68** (Chair: Edgard Ella)
15:00-15:30	*Spelling variations in A dictionary of Setswana personal names* (Thapelo J. OTLOGETSWE & Goabilwe N. RAMAEBA)	*Treatment of Dialectal Varieties in Lumbu Lexicography* (H. Steve NDINGA-KOUMBA-BINZA, Maryse Edith MALOUGHOU MANGAMA, Paul Achille MAVOUNGOU & Ludwine MBINDI ANINGA)
15:30-16:00	*A Morphophonological Analysis of Transliterations in the UFS Sesotho Accounting Terminology List* (Malefu Justina MAHLOANE)	*Knowledge on the Margins: The Treatment of Variation in isiXhosa Dictionaries* (Dion NKOMO & Thandile NTUMBA)

16:00-16:30	*Comparison of selected Setswana monolingual dictionaries in their coverage of multiword expressions* (Keabetswe PHETOANE)	*Dictionnaire des Expressions Idiomatiques du Lumbu: The Latest Addition to Gabonese Dictionary Production* (Ludwine MBINDI ANINGA, H. Steve NDINGA-KOUMBA-BINZA, Maryse Edith MALOUGHOU MANGAMA & Paul Achille MAVOUNGOU)
18:30	**WELCOMING COCKTAIL** Foyer of Graduate Centre	

	TUESDAY 2 JULY	
09:00-09:15	**Opening by Deputy Dean Teaching and Learning, Faculty of Humanities** Prof. Sandy Africa Plant Sciences Auditorium	
09:15-10:05	**Keynote Address 1** Plant Sciences Auditorium (Chair: Prof. Sonja Bosch) *Lexicography has a bright future: Shifting the focus from dictionaries to databases* Prof. Sven Tarp	
10:05-10:30	**TEA** Foyer of Graduate Centre	
Parallel Sessions		
	Graduate centre 1-64 (Chair: Buyiswa Mini)	**Graduate centre 1-68** (Chair: Theo Bothma)
10:30-11:00	*Revisiting and re-imagining Kropf's A Kafir-English Dictionary 125 years on* (Dion NKOMO & Sanele NTSHINGANA)	*The Van Warmelo dictionary card collection – a lexicographic appraisal* (Elsabé TALJARD, Danie PRINSLOO & Michelle GOOSEN)

11:00-11:30	*An Etymological Dictionary of Khoekhoe-Nguni Loanwords* (Camilla Rose CHRISTIE)	*Using computational tools and a corpus lexicography framework in developing an isiZulu LSP Dictionary* (Mthuli BUTHELEZI & Cael MARQUARD)
11:30-12:00	*Towards a Dictionary Production Strategy for African Languages: A Note from the Lexicography Agenda of CASAS* (H. Steve NDINGA-KOUMBA-BINZA & Mantoa Rose MOTINYANE)	*Corpus-based dictionaries for low-resource languages* (Mmasibidi SETAKA & Menno VAN ZAANEN)
12:30-13:30	**LUNCH** Adlers Restaurant (on campus)	
	Graduate centre 1-64 (Chair: Eventhough NDLOVU)	**Graduate centre 1-68** (Chair: Thapelo Otlogetswe)
13:30-14:00	*Inclusion of Etymological information in Swahili lexicographic works: a challenge to be surmounted* (Norbert MTAVANGU)	*Terminology management programmes within the South African context* (Michelle GOOSEN & Elsabé TALJARD)
14:00-14:30	*Peering into Ghana from a Lexicographic Perspective* (Hasiyatu ABUBAKARI, Anthony A. MUSAH & Dion NKOMO)	*Developing a Dictionary as a Resource for Language Acquisition: The Case of isiXhosa Dictionary of collocations* (Mlondolozi PUTE, Sebolelo MOKAPELA & H. Steve NDINGA-KOUMBA-BINZA)
14:30-15:00	*Using digital tools for data collection in lexicography: Mzansi Taal dictionary as case of analysis* (Kganathi SHAKU & Mmagonkahloleng MAKUA)	*A Triangle of Relationships Among Three Major Language Practices: Translation, Terminology and Lexicography in African Languages: The Case of isiXhosa* (Buyiswa MINI)
15:00-15:30	**TEA** Foyer of Graduate Centre	
	Graduate centre 1-64 (Chair: André du Plessis)	**Graduate centre 1-68** (Chair: Phillip Louw)

15:30-16:00	*Making sense of kuningi using a corpus linguistic analysis* (Langa KHUMALO)	*Towards 200,000 Entries: How Regional Languages Contribute to the Great Dictionary of the Indonesian Language* (Chusna AMALIA & Dzien Nuen ALMISRI)
16:00-16:30	*Information retrieval structures in online lexicography* (Rufus GOUWS & Theo BOTHMA)	*Indonesianization: Standardization and Codification of Terms in Indonesian* (Dewi PUSPITA & Dira HILDAYANI)
16:30-18:00	**AFRILEX ANNUAL GENERAL MEETING** Plant Sciences Auditorium	
18:30	**CONFERENCE DINNER** Function Hall (on campus)	

	WEDNESDAY **3 JULY**
09:00-09:10	**Announcements** Plant Sciences Auditorium
09:10-10:00	**Keynote Address 2** Plant Sciences Auditorium (Chair: Prof. Langa Khumalo) *Lexicography beyond dictionaries – Exploring Semantic Relations in African Wordnets* Prof. Sonja Bosch
Plenary sessions	
	Graduate centre 1-64 (Chair: Marissa Griesel)
10:00-10:30	*IsiXhosa.click: online, open, user-friendly, and searchable isiXhosa-English dictionary software* (Cael MARQUARD)
10:30-11:00	**TEA** Foyer of Graduate Centre

	Graduate centre 1-64 (Chair: Lorna Morris)
11:00-11:30	*GREDYLEX's contribution to Dictionary Use and Culture in Education in Gabon* (Edgard Maillard ELLA, Ludwine MBINDI ANINGA, Gilles SAPHOU-BIVIGAT & Léandre Serge SOAMI)
11:30-12:00	*The presentation of information related to food in the selected African language dictionaries* (Julius TAJI & Dion NKOMO)
12:00-13:00	**SPECIAL SESSION 1:** **Publishers' Session** Plant Sciences Auditorium (Chair: Phillip Louw) OUPSA, Pharos & SANLU
13:00-14:00	**LUNCH** Adlers Restaurant
14:00-16:00	**SPECIAL SESSION 2:** **NLU Session** Plant Sciences Auditorium (Chair: Dion Nkomo) South African National Lexicography Units
16:00-16:30	**TEA** Foyer of Graduate Centre
16:30-16:45	**AFRILEX CLOSURE** Plant Sciences Auditorium Prof. Langa Khumalo

	THURSDAY **4 July**
09:00	CONFERENCE EXCURSION

6ᵀᴴ GLOBALEX WORKSHOP ON LEXICOGRAPHY AND NEOLOGY PROGRAMME

	WEDNESDAY **3 July**
09:15-09:30	**Welcome & Introduction** (Annette KLOSA-KÜCKELHAUS) Graduate Centre 1-56
Session 1 (Chair: Elsabé Taljard)	
09:30-10:00	*Designing a Digital Lexicon of Poetic Neologisms: A Case Study in Kazantzakis' Work* (Nikos MATHIOUDAKIS)
10:00-10:30	*The Role of Combining Forms in Creating Neologisms: A Data-Driven Approach* (Yongwei GAO)
10:30-11:00	**TEA** Foyer of Graduate Centre
11:00-11:30	**HIGHLIGHT PRESENTATION:** *European Network on Lexical Innovation – ENEOLI* (Giovanni TALLARICO et al.)
Session 2 (Chair: Giovanni Tallarico)	
11:30-12:00	*From Neologism Extraction to Dictionary Description: Methodological Issues in Corpus Balance, Word Unit Bias and LLM Assistance* (Kilim NAM, Soojin LEE & Hae-Yun JUNG)
12:00-12:30	*Neology and Nomenclature: Concept Clarification as a Precursor to Computational Lexicography* (Peter Juel HENRICHSEN)
12:30-13:00	*Neologisms and their Functions in Critical Discourse* (Petra STORJOHANN)
13:00-14:00	**LUNCH** Adlers Restaurant
Session 3	

(Chair: Rufus Gouws)	
14:00-14:30	*Lexicographic Representation of Verbal Semantic Neology in Spanish Dictionaries* (Victoria DE LOS ANGELES BOSCHIROLI & Marina BERRI)
14:30-15:00	*Semi-Automatic Detection of New Words in Georgian* (Tamar LALUASHVILI & Tinatin MARGALITADZE)
15:00-15:30	*Digging for -ings: A Survey of Selected (Pseudo)Anglicisms in Dictionaries of Modern Greek* (Anna VACALOPOULOU)
15:30-16:00	**TEA** Foyer of Graduate Centre
Session 4 (Chair: Sonja Bosch)	
16:00-16:30	*From taggare to blessare: Verbal Hybrid Neologisms in Italian Youth Slang* (Maria Pia DI BUONO)
16:30-17:00	*Lexical Change and Variation in the Parenting Domain: Analysis and Terminographic Proposal* (Paola CAÑETE-GONZÁLEZ & Elisabet LLOPART-SAUMELL)
17:00-17:30	**Closure and Discussion** (Ilan KERNERMAN)

KEYNOTE PRESENTATION 1

Lexicography has a bright future: Shifting the focus from dictionaries to databases

Sven TARP (st@cc.au.dk)
Centre for Lexicography, University of Aarhus, Aarhus, Denmark; Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa; Centre for Lexicographical Studies, Guangdong University of Foreign Studies, Guangzhou, China

There is little doubt that lexicography, as a time-honoured cultural, practical and academic discipline, is at a crossroads from a global perspective. A complex mix of new technological breakthroughs (especially AI), an outdated business model, new information sources, new user habits, and new types of needs has created a situation in which lexicography simply cannot continue as usual. The time between new dictionary releases has become longer and longer. The number of titles sold, both in print and digitally, has fallen dramatically, and so has publisher revenue. In many countries around the world, this has led to the closure of dictionary publishers, including some of the most prestigious, and the unfortunate redundancy of a large number of skilled and experienced lexicographers. All this has understandably led to a certain pessimism and concern among lexicographers, not only about their own personal situation, but also about the future of their discipline as a whole.

Without underestimating the many challenges that lie ahead, the paper will convey an optimistic message full of hope. It will do so by using history as the best teacher. Four main factors have driven lexicography for thousands of years: the information needs observed in society, the ever-available technology used to meet these needs, the experience and increasing skill of its practitioners, and a growing superstructure of reflection and theoretical statements. Using examples from early European lexicography, the paper will discuss how dictionaries grew out of glosses inserted into manuscript texts, and how these glosses were later ignored when the term "lexicography" was coined in the 17th century to refer exclusively to the writing of dictionaries. It will then show that, against all odds, these glosses survived the advent of printing and are now experiencing a renaissance that is placing them increasingly at the centre of lexicographic work alongside dictionaries, thus providing a holistic vision of the discipline. However, these glosses are no longer the old classical ones, but new types of glosses that are being created, stored in databases and presented to users in writing and reading aids, as well as in other types of digital software, by means of the latest technologies, in particular AI.

With this in mind, the paper recommends that lexicographers should shift their focus from dictionaries to databases containing both new and old types of lexicographic data that can serve various tools, including but not limited to digital dictionaries. In this context, it distinguishes between two different categories of lexicographic databases that can already be observed in practice, namely lemma-centred databases and problem-centred databases. As the terms imply, the former are the traditional ones, built around lemmas, while the latter are completely new types of databases that do not focus on specific words (lemmas), but on classes of grammatical and stylistic problems that appear in texts.

All of this will be illustrated with a wealth of examples, including some from an ongoing project using ChatGPT for various tasks to create a lexicography-supported writing assistant for Spanish learners. The paper will argue that this new kind of high-tech product requires a high level of lexicographic knowledge and skills on the one hand, and close interdisciplinary

collaboration with IT experts and other stakeholders on the other. With such collaboration, knowledge and skills in place, it is almost certain that wonderful new lexicographic products, adapted to the needs, expectations and habits of current and future users, will see the light of the day. On this basis, it concludes that both lexicography and its practitioners have a bright future if the latter know how to adapt, accepting that the products, tasks and methods, as well as the knowledge, culture and skills required of lexicographers, will be increasingly different from those of the past. With a clear vision of the challenges and a good dose of determination and courage to face them, there is no need for gloom and doom.

<center>*****</center>

KEYNOTE PRESENTATION 2

Lexicography beyond dictionaries – exploring semantic relations in African Wordnets

Sonja Bosch (seb@hbosch.com)
Department of African Languages, University of South Africa, Pretoria, South Africa

Central to the expansion of lexicography beyond traditional dictionaries is the advent of lexical databases known as Wordnets, initially originating in the 1980s with the Princeton English WordNet (PWN) and then progressively adapted for many global languages. Wordnets offer a comprehensive mapping of semantic relationships between words, and go beyond the linear entries found in dictionaries, by offering an enhanced understanding of how words relate to each other in terms of synonyms, hypernyms, hyponyms, antonyms and more. Nouns, verbs, adjectives, and adverbs are organised into sets of cognitive synonyms called synsets, with each set representing a unique concept. These synsets are then connected through conceptual-semantic and lexical relations.

Wordnets are fundamental language resource components for natural language processing, aiding among others, in word sense disambiguation, machine learning, artificial intelligence, and computational linguistics. They not only enhance our understanding of word meanings but also contribute to the development of more sophisticated language technologies, making language processing more context-aware and semantically accurate.

This presentation aims to explore semantic relations in the multilingual African Wordnet (AfWN) project, which deals with a large scale, freely available, lexical database of the nine official African languages of South Africa. The individual language wordnets are in varying phases of development, but they are all linked via the PWN which is used as a type of scaffold. This approach is known as the "expand" method and is not unusual for the development of wordnets for languages with limited resources. The AfWN project resorts under the Unisa African Languages Node of the South African Centre for Digital Language Resources (SADiLaR).

Despite challenges such as lexicalisation differences, the construction of AfWNs makes it possible to cover extensive information on aspects of word meaning that are not easily covered in traditional print dictionaries. The reason is that electronic or digital databases are generally not adversely affected by space, size or even time constraints as is the case with printed matter. They therefore lend themselves ideally to the inclusion of additional data, such as indigenous knowledge that is frequently underrepresented in conventional dictionaries. This information is often dispersed across a variety of sources including mono- and bilingual dictionaries, interdisciplinary sources and flat-structured online databases.

The treatment of the word category adjective in Wordnets deserves special attention. Unlike nouns and verbs, which are hierarchically organised in Wordnets, adjectives are organised in terms of antonymy. Traditionally, the semantic expression of adjectives in the PWN has been handled through the word category adjective which is an open class in English. Many other languages, however, have a limited class of adjectives, for instance languages belonging to the Bantu language family, have a small class of adjectives with a membership ranging from ten to approximately forty or fifty. This category is morphologically described as consisting of an adjective prefix and an adjective stem.

By means of examples of adjectives extracted from the SIL Comparative African Wordlist (SILCAWL), it will be demonstrated that a variety of constructions or word classes is used in African languages to express concepts marked as adjectives in the PWN. Therefore, the development of Wordnets for African languages necessitates a novel approach that focuses on semantic relations concerning adjectives rather than morphological structures. The need for new approaches in structuring adjectives within AfWNs, reflecting the diverse linguistic structures, will be discussed.

<center>*****</center>

PRECONFERENCE WORKSHOP

Towards a sustainable National Digital Term Bank for the official languages of South Africa: Collaboration vs Fragmentation

Given the imminent *National Policy on Terminology* of the Terminology Coordination Section of the National Language Services and the history of fragmentation of activities in the field, this workshop intends to focus on:

— the necessity for cooperation among all stakeholders in the field,
— an inclusive pathway addressing the general and specific needs of different sectors among other things by creating communities of practice, training, and reskilling activities,
— optimal procedures and infrastructures to facilitate and manage the technological aspects of a sustainable national digital term bank,
— the role of SADiLaR as a national infrastructure on matters related to longer-term planning at the start of terminology development projects.

Target audience: Members of government departments including DSAC and DTI; Members of National and Provincial Language Boards and NLUs; PanSALB; Members involved with the development of the *Draft National Terminology Policy*; Members of the Terminology Coordination Section of the NLS.

Outcomes:
— Participants with an understanding of the current state of terminology practice and challenges in South Africa.
— Management of expectations regarding future developments and policy implementation.
— Contribution of participants in formulating practical steps towards collaboration.

Procedures:
— Presentations of relevant topics followed by questions/comments.
— Dedicated discussions by all participants and formulation of actions for future planning and implementation.

<center>*****</center>

GLOBALEX WORKSHOP ON LEXICOGRAPHY AND NEOLOGY

The aim of the GWLN series is to bring together cross-world views combining the detection, classification, description and evaluation methods of neologisms with lexicographic processing, compilation and representation for different languages. GWLN takes place annually, with support of the Global Alliance for Lexicography (Globalex), alternating in conjunction with the international conferences of the continental associations for lexicography. The main theme of the 6th Globalex Workshop on Lexicography and Neology (GWLN-6) is Neologisms, Lexicography, and LLMs, including all types of exchange among large language models, neology, and lexicography. GWLN-6 will feature an overview of the new COST Action – European Network On Lexical Innovation – ENEOLI (CA22126), whose major challenges are to: (1) Define the core terminology of neology through the creation of a born-digital specialised multilingual glossary; (2) Adapt digital methodologies and tools to identify and account for lexical innovation; (3) Carry out comparative studies on lexical innovation in European languages; and (4) Provide specific training in neology for editors, translators, technical writers, educators, and journalists.

*****

CONFERENCE SPECIAL SESSIONS

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 overemphasised. 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. As with the 27th International AFRILEX Conference, this year's conference provides an opportunity for both practising lexicographers in commercial publishing and from the South African National Lexicography Units (NLU) to discuss not only the interaction between theory and practice, but also to reflect individually and collectively about all matters that affect their work. The two sessions, the NLU session and the Publishers' session, will further provide an opportunity for the editors and/or lexicographers to highlight some of their recently completed projects as well as to announce any future endeavours. The publishers' session features the South African National Lexicography Units (publisher for many of the NLUs), Pharos Dictionaries and Oxford University Press (Southern Africa) as major players in the field of lexicographic practice in South Africa. The NLU session features representatives from, inter alia, the Afrikaans NLU, the isiNdebele NLU, the Setswana NLU, the Sesotho NLU, the Tshivenḓa NLU and the Xitsonga NLU.

*****

PRESENTATIONS: AFRILEX CONFERENCE

Peering into Ghana from a Lexicographic Perspective

Hasiyatu ABUBAKARI (haabubakari@ug.edu.gh)
Institute of African Studies, University of Ghana, Legon, Ghana; School of Languages &
Literatures: African Language Studies, Rhodes University, Makhanda, South Africa
Anthony A. MUSAH (aamusah@uew.edu.gh)
College of Languages Education, University of Education, Winneba, Ghana; School of
Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South
Africa
Dion NKOMO (d.nkomo@ru.ac.za)
School of Languages and Literatures: African Language Studies, Rhodes University,
Makhanda, South Africa

Ghana is a multilingual country with a strong tradition of linguistic and applied linguistic research. However, compared to other language-related professional activities and academic areas of study, lexicography exhibits a dearth of scholarship when it comes to the Ghanaian context. Virtually not much of note is known about Ghanaian dictionaries or dictionary use in the country. No comprehensive survey and systematic analyses of dictionaries in Ghanaian languages exist. One also cannot easily find substantial information about individual dictionaries in different languages nor dictionary user-research in Ghana. In fact, the best one gets from a Google Scholar search[i] of "Ghana + dictionaries" are references to the *Historical dictionary of Ghana*, *A dictionary of Ghanaian English* and a book chapter (O'Brien & Sherris 2023) on the uses of bilingual dictionaries among the Safaliba people in Ghana.

Accordingly, one may justifiably ask the question, "Does lexicography exist in Ghana?" Such a question necessitated a comprehensive desktop survey of dictionaries and other lexicographic resources that have been produced for various Ghanaian languages. The survey was guided by the following central questions:

— To what extent has lexicography contributed towards the development and use of various languages that are spoken in Ghana from a historical perspective?

— What types of lexicographic resources are available in Ghana and how are they spread across the different Ghanaian languages?

— What are the factors that have shaped the present lexicographic situation in Ghana?

— What are the factors that can necessitate and inspire futurist lexicographic practice in Ghana?

— What are the prospects for the necessary lexicographic practice and a robust lexicographic infrastructure in a multilingual Ghanaian context in the 21st Century?

Through the above questions, the survey enabled us to get an insight into the resourcefulness, or lack thereof, of Ghanaian languages from a lexicographic perspective. It enabled us to uncover dictionaries that exist in various Ghanaian languages in terms of their types, users, functions and compilers, as well as the motivations behind the production of the

---

[i] https://www.google.com/search?q=%E2%80%9CGhana+%2B+dictionaries%E2%8%9D&rlz=1C1GCE A_enZA1082ZA1082&oq=%E2%80%9CGhana+%2B+dictionaies%E2%80%9D&gs_lcrp=EgZjaHJvbW UyBggAEEUYOTIHCAEQIRigATIHCAQIRigAdIBCTE4MzRqMGoxNagCCLACAQ&sourceid=chro me&ie=UTF-8 (accessed on 27 April 2024).

works. As it is generally the case in many African communities, the Ghanaian lexicographic situation is characterized by dictionaries that were produced by missionaries and colonial linguists (cf. Durand 1953) for the country's selected languages. The dictionaries are typically bilingual, pairing Ghanaian languages with English, and the overall lexicographic standards in terms of scope and depth in the treatment of Ghanaian languages are low when compared with lexicographic resources in English language, for instance.

However, some Ghanaian languages, particularly minority languages such as Kusaal, Dagaare, and Buli have survived in what Gouws and Ponelis (1992) call a pre-lexicographic environment well into the 21st Century. Such languages did not benefit from the lexicographic efforts of missionaries and colonial linguists, but contemporary language documentation efforts with support from mother-tongue speaking linguists have included lexicographic endeavours as illustrated by the publication/compilation of the *Kusaal (Agole) Dictionary* (Naden 2015), the *Comparative Dictionary of Central Mabia Languages* (Naden 2021) and *An Orthographic Dictionary of Kusaal* (Awimbila and Naden 2022). Such lexicographic works appear to be central to the future prospects of lexicography in Ghana and would need to be accompanied and guided by dictionary research for theoretical insights. Furthermore, it would be vital for current and future lexicographic projects to exploit technological advances and interdisciplinary expertise that was not available for the pioneering dictionaries in Ghanaian languages. Indeed, collaborative lexicography would be important.

Unsurprisingly, English dictionaries of all kinds feature in the Ghanaian lexicographic context since English is the country's official language that is used in education and other prestigious domains of society. Most of those dictionaries are publications of global publishing powerhouses such as Oxford University Press, Merriam-Webster and Macmillan. In the future, dictionary user research would be needed in Ghana to determine how these English dictionaries, together with dictionaries in the indigenous languages of Ghana, are valued in society in general, but more specifically within the education sector. Such knowledge would be crucial for future lexicographic practice, building on the modest findings of the present research and its potential implications for language documentation and revitalization, educational initiatives, as well as community development in the so-called knowledge economy.

References

Awimbilla, G. & Naden, T. 2022. *An Orthographic Dictionary of Kusaal*. Agole – Fullest version. Tamale: M-buk publications.

Blench, R., & Dendo, M. 2006. *A dictionary of Ghanaian English*. http://www.rogerblench.info/Language/English/Ghana%20English%20dictionary.pdf [13 July 2016].

Durand, J.B.W.E. 1953. *Dagaare-English Dictionary*. Navrongo: St John Bosco's Press.

Gouws, R.H. & Ponelis, F.A. 1992. The development of Afrikaans and the lexicographical tradition. L. Zgusta (ed.). *History, Languages and Lexicographers*: 77-104. Tübingen: Max Niemeyer Verlag.

Naden, T. 2015. *Kusaal (Agole) Dictionary*. Accra: G.I.L.L.B.T.

Naden, T. 2021. *Comparative Dictionary of Central Mabia Languages*. Unpublished Monograph.

O'Brien, C.A. & Sherris, A. 2023. Exploring the Uses of Bilingual Dictionaries among the Safaliba in Ghana and the Gorontalo in Indonesia. R-L. Valijärvi and L. Kahn (eds.). *Teaching and Learning Resources for Endangered languages*: 145-165. Leiden: Brill.

Owusu-Ansah, D. 2014. *Historical dictionary of Ghana*. Lanham: Rowman & Littlefield.

*****

Towards 200,000 Entries: How Regional Languages Contribute to the *Great Dictionary of the Indonesian Language*

Chusna AMALIA (chusna.amalia@kemdikbud.go.id; ch.amalia.ach@gmail.com)

Language Development and Cultivation Agency, Ministry of Education, Culture, Research, and Technology, Indonesia

Dzien Nuen ALMISRI (dzien.nuen@kemdikbud.go.id; inunesia@gmail.com)

Language Development and Cultivation Agency, Ministry of Education, Culture, Research, and Technology, Indonesia

The *Kamus Besar Bahasa Indonesia* (KBBI), or the *Great Dictionary of the Indonesian Language*, is the official dictionary of the Indonesian language. Compiled by the Indonesian Language Development and Cultivation Agency (Badan Bahasa), the Ministry of Education, Culture, Research, and Technology, the dictionary holds significance as the most comprehensive and precise Indonesian dictionary to date. It serves as the authoritative reference for standard as well as nonstandard Indonesian.

The KBBI undergoes biannual updates, typically occurring in April and October. The last update to the KBBI occurred in October 2023. Presently in its sixth edition, it boasts an extensive compilation of 120,557 entries. In 2024, Badan Bahasa proposes an ambitious plan to add at least 80,000 new entries to KBBI. By the October 2024 update, it is expected that KBBI will contain 200,000 entries. Anticipatedly, a minimum of 50% of the new entries in the KBBI will be sourced from regional languages (Badan Bahasa 2024). As for now, the editorial team of KBBI are working towards the aforementioned goal.

Indonesia, being an archipelagic nation, is a home to 718 regional languages, excluding dialects and subdialects (Badan Bahasa 2023), with many languages yet to be documented. These languages speakers are spread across 38 provinces. The highly diverse array of languages poses a challenge for the KBBI editorial team in compiling words adapted from regional languages. Apart from the editorial team, the KBBI also relies on contributors from Badan Bahasa's regional offices which are located in 30 provinces to collect proposed words from their respective local languages for inclusion in the KBBI. As of May 2024, the data taken from KBBI shows a total of 8,486 entries from regional languages are included in the KBBI. Meanwhile, within the same timeframe, there are 43,731 indigenous words proposed for inclusion in the KBBI. In other words, approximately 19% of the proposed words successfully made it into the KBBI, with Javanese and Sundanese having the highest representation. Some might consider this percentage insignificant given the vast number of regional languages in Indonesia.

There are five criteria for a word to be included in the KBBI: 1) Uniqueness: The proposed word has a meaning that does not yet exist in Indonesian. These words will serve to fill a lexical gap, i.e. an absence of meaning in Indonesian; 2) Euphonics: The proposed word does not contain unusual sounds in Indonesian or, in other words, conforms to the phonological rules of Indonesian. This requirement is intended to ensure that the words are easily pronounceable by Indonesian speakers with various native language backgrounds; 3) Consistency with Indonesian orthography: The word can be formed and form other words according to the rules of word formation in Indonesian, such as affixation and compounding; 4) Non-negative connotations: Words with negative connotations are not encouraged for inclusion as they may not be well-received among users; 5) Frequency: The frequency of a word's usage is measured by its occurrence frequency and range. Frequency is the rate of occurrence of a word in a corpus, while range is the distribution of the appearance of that word across several regions. (Badan Bahasa 2022).

This paper will discuss the lexicographic practice carried out by the KBBI editorial team, particularly the selection of words from regional languages for inclusion in the KBBI based on

the five criteria mentioned above, as well as the problems encountered during this process. Some issues arise either from the proposed words themselves, for example the absence of a spelling system in the source language, or from external influences, such as the objection from informants regarding the adaptation of regional language words into Indonesian. This paper will address what types of words from regional languages have successfully been included in the KBBI, whether they are cultural words, expressions, etc. The findings from this research hopefully will be useful as guidelines for the editorial works particularly concerning word inclusion in the KBBI and ultimately will contribute to achieving the target of 200,000 entries in the KBBI by the end of 2024.

References
Badan Bahasa. 2022. Bagaimana Sebuah Kata Masuk ke KBBI (How is a Word Included in the KBBI). https://badanbahasa.kemdikbud.go.id/artikel-detail/839/bagaimana-sebuah-kata-masuk-ke-kbbi. [29/12/2023].
Badan Bahasa. 2023. Data Bahasa di Indonesia (Data of Languages in Indonesia). https://petabahasa.kemdikbud.go.id/databahasa.php. [9/12/2023].
KBBI Daring (Online KBBI). 2023. https://kbbi.kemdikbud.go.id/KamusUsulanDaerah. [06/05/2024].
Badan Bahasa. 2024. *Kerangka Acuan Kerja Produk Kodifikasi Bahasa Tahun 2024* (Framework of Reference for KBBI 2024). Jakarta: Language Development and Cultivation Agency, Ministry of Education, Culture, Research, and Technology.

*****

Contextualisation in dictionary consultation and Generative Artificial Intelligence – a reader's perspective

Theo J.D. BOTHMA (theo.bothma@up.ac.za)
Department of Information Science, University of Pretoria, Pretoria, South Africa

A considerable amount of research is being done about the use of Generative Artificial Intelligence (GenAI) in creating dictionaries, dictionary databases and writing assistants. There are widely different perceptions about the usefulness of the technology for writing dictionary articles, suggesting sample sentences for dictionary articles, and the use of GenAI to support human intelligence (see, e.g. Fuertes-Olivera 2024; Huete-García & Tarp 2024). However, less has been done about the support that GenAI could give the reader / end-user in the dictionary consultation process.

Readers require the micro- or macro- access structure of a dictionary article to be clear and logical to facilitate access to the correct meaning in a specific context. Access structures have been researched in detail over many years by lexicographers, for both print-based and e-dictionaries (e.g. Gouws 2022). Furthermore, contextual metadata are provided in dictionaries in the form of labels and usage notes. It is the responsibility of the reader to match the context of the text in which an unknown word occurs to the meanings provided in the dictionary with the aid of any of the metadata to determine the meaning of the word.

Linking words in an e-text on an e-reader or in a browser to an e-dictionary entry is standard, and the reader can access the e-dictionary by clicking on the word. The link is, however, to the first lemma that corresponds to the word in the text, and not necessarily to the lemma in context, and not to the specific meaning in context. In both cases, it remains the responsibility of the reader to evaluate all options, by reading the dictionary article(s) and determining which of the provided meanings is correct in context (Bothma & Gouws 2022; Bothma & Fourie 2024). The reader needs some grammatical knowledge to determine the part of speech of the word on which they clicked, and then needs to navigate through sometimes

lengthy articles to determine the correct meaning in context, making use of the structure of the article and interpreting the metadata available.

The question is whether, and to what extent, GenAI could assist the reader to link to the correct lemma and the correct meaning of the lemma in context, i.e., whether GenAI can parse the part of speech correctly, and whether it can deduce from the context of the text what the applicable meaning of the word is. This paper will discuss this at the hand of a number of examples. The experiments were done using the chatbot GPT-4 on Microsoft Copilot; this version of the chatbot is freely available within Microsoft Copilot. (Earlier versions are not dependable, and often "hallucinate".)

Currently, if a reader would like to make use of a chatbot to obtain lexicographic information about a word in a text, they have to access their preferred chatbot and then provide one or more prompts to obtain an answer (as was done in this paper). The context that the reader has to provide can be the sentence or the paragraph in which the word occurs. If, in the latter case, the context is still not clear, the reader has to provide further prompts to guide the GenAI.

We will demonstrate that GenAI can (in general) correctly distinguish between homographs and assign the correct meaning in context. If there are insufficient keywords in the sentence, the reader may have to add further contextual prompts. The sources the chatbot used for the answer are listed, which implies that the reader can check the references for the validity of the GenAI answers (if required). Examples of a simple disambiguation of homographs in a sentence will be given, as well as complex examples where the GenAI uses contextual markers in the sentence or paragraph, and where it requires additional contextual prompts.

This does not imply that GenAI will replace dictionaries, considering that their main source of information seems to be e-dictionaries that are not behind firewalls. However, the ease and efficiency of use makes GenAI a good supplementary information source. The principle of clicking on a word to obtain the required information remains the ideal, and the current linking technology should be enhanced by the efficiency offered by GenAI, specifically the ability to determine results based on context. This implies that the linking technology should be redeveloped to include the ability to provide context-aware results.

References

Bothma, T.J.D. & Fourie, I. 2024. Contextualised dictionary literacy, information literacy, and information behaviour in the e-environment. *Library Management* (forthcoming).

Bothma, T.J.D. & Gouws, R.H. 2022. Information needs and contextualization in the consultation process of dictionaries that are linked to e-texts. *Lexikos* 32(2): 53-81.

Fuertes-Olivera, P. 2024. Making lexicography sustainable: Using ChatGPT and reusing Data for lexicographic purposes. *Lexikos* 34: 123-140.

Gouws, R.H. 2022. Nuwe raamstruktuurtipes in aanlyn woordeboeke. *Lexikos* 33(1): 382-403.

Huete-García, A. & Tarp, S. 2024. Training an AI-based writing assistant for Spanish learners: The usefulness of Chatbots and the indispensability of human-assisted intelligence. *Lexikos* 34: 21-40.

*****

Using computational tools and a corpus lexicography framework in developing an isiZulu LSP Dictionary

Mthuli BUTHELEZI (buthelezi@bfsu.edu.cn)
Department of Linguistics, University of KwaZulu-Natal, Durban, South Africa
Cael MARQUARD (cael.marquard@gmail.com)
Department of Computer Science, University of Cape Town, Cape Town, South Africa

We propose compiling an isiZulu Language for Special Purposes (LSP) dictionary to be deployed and made available through an online software interface. This study develops alongside the much-needed studies on open-source code for creating dictionaries[i] and on the work of writing online dictionaries for languages like isiZulu[ii].

The herein proposed isiZulu LSP dictionary offers distinct advantages over existing online isiZulu dictionaries. Firstly, it is tailored to specialized linguistic and literature terms. Secondly, its open-source code enables *localisation* to other African languages, fostering inclusivity and accessibility. Thirdly, its online interface is available in isiZulu. This is a novel approach in interactive interface design that has not been seen in other online isiZulu dictionaries. Finally, its development addresses the critical need for wide-ranging dictionaries in African languages. This endeavour aligns with the imperative to advance Human Language Technologies (HLT) in African languages and contributes to pedagogy and terminology development, ultimately transforming isiZulu pedagogy.

Importantly, the algorithm of the online isiZulu LSP dictionary is a *template-based* algorithm (Keet & Khumalo 2014), and it is *localised* from the isiXhosa.click algorithm.

The dictionary utilizes an LSP corpus, which can be scaled, reused, and redistributed, thus enhancing its linguistic richness. Therefore, this dictionary has the capacity to be improved in the future.

This study utilises the corpus lexicography framework to establish an isiZulu dictionary for the linguistics and literature terms. Corpus lexicography as a theoretical framework and a research methodology provides for the use of term extraction and computational techniques (Khumalo 2015). The textual datum in the isiZulu LSP dictionary is intrinsically a corpus datum. Therefore, in this study, the corpus lexicography framework guides the process of term extraction in the process whereby the potential lexical entries are identified as headwords for the possible addition to the dictionary.

The selected isiZulu LSP corpus from Kumalo (1995) provides the terms which are incorporated into the dictionary. The extracted isiZulu terms have their English translation equivalents. The proposed dictionary is intended for teachers and learners of linguistics and literature. The established algorithm for the proposed isiZulu LSP dictionary also serves as an electronic database for term compilation, analysis and storage.

The isiXhosa.click[iii] online dictionary software will be adapted to provide a usable front-end to the isiZulu LSP dictionary. Initially developed as an online dictionary for isiXhosa, it allows for live word search and community participation in the dictionary. Community participation is facilitated through tools on the site to suggest additions, modifications, and removals of words in the database. These suggestions are visible in a moderation dashboard, where they are manually edited and approved. This review-by-experts process is intended to prevent vandalism and the capturing of incorrect information.

IsiXhosa.click provides a search function which is the primary way users discover words. Words may be searched either by their isiZulu form, their isiZulu stem, or their English translation. The search is typo-tolerant and allows for even misspelled queries to return results. Therefore, the software is easy to use even for inexperienced users. Additionally, the

---

[i] https://github.com/topics/dictionary-software (open-source dictionaries code)

[ii] https://zuludict.acdh.oeaw.ac.at/zuludict/ (example of initiatives on creating online isiZulu dictionaries)

[iii] https://ched.uct.ac.za/dot4d/implementation-student-led-initiatives/isixhosaclick

site is suitable for mobile use, increasing its accessibility.

IsiXhosa.click also supports capturing of grammatical information, addition of example sentences for words, and linking related words together. This software has been selected for the following reasons. Firstly, it currently targets isiXhosa, which is similar to isiZulu, meaning that the software already fulfills many of the dictionary's requirements. Secondly, it is open-source, meaning that the knowledge acquired from the isiZulu LSP dictionary development project may be directly applicable to other efforts aiming to adapt the software to other languages. Thirdly, the online interface is suitable for mobile use, expanding its potential reach. Lastly, its community editing tools may extend the lifespan of the isiZulu LSP dictionary by allowing interested third parties to continue its maintenance and improvement. Therefore, there are plans to expand these dictionaries and make them into 'living dictionaries' with the continual appraisal and addition of lexical items and their example sentences.

Key tasks involved in the adaptation of the isiXhosa.click software to the isiZulu LSP include importing the isiZulu LSP data into the database, adapting the interface to support localisation for isiZulu, creating translations for interface text and labels, and deployment of the modified software such that it may be accessed publicly. In order to better support searching by isiZulu stems, a new isiZulu stemmer should be bootstrapped from the existing isiXhosa algorithm. Because this is only used as a search heuristic, accuracy is not important, hence a simple greedy stemming approach will suffice. Importantly, the software will be extended to support *localisation* of its interface to other languages. Thus, the language development efforts intended in the proposed study underscore the importance of integrating the lexicographic and the corpus linguistic insights with computational methodologies to address the challenges and opportunities presented by African languages in the digital age.

References
Keet, M. & Khumalo, L. 2014. Toward verbalizing ontologies in isiZulu. *Controlled Natural Language* 8625: 78-89.
Khumalo, L. 2015. Semi-automatic term extraction for an isiZulu linguistic terms dictionary. *Lexikos* 25: 495-506.
Kumalo, M. 1995. *English/IsiZulu linguistic and literary terms: Amatemu esiNgisi/esiZulu elingwistiksi nelithrisha*. Cape Town: Vivlia Publishers.

<div align="center">*****</div>

Problems of Equivalence in the Compilation of *English–ChiBarwe Bilingual Elementary Science Terms Dictionary*

Tendai CHIRIMAUNGA (chirimaungat@staff.msu.ac.zw)
Midlands State University National Language Institute, Gweru, Zimbabwe
Loice NGOMACHEZA (ngomachezal@staff.msu.ac.zw)
Midlands State University National Language Institute, Gweru, Zimbabwe

This paper discusses some theoretical and practical implications arising from the compilation of the *English–ChiBarwe Bilingual Elementary Science Terms Dictionary*. This is an ongoing lexicography project being executed by the Lexicography and Linguistics (LL) unit of the Midlands State University National Language Institute (MSUNLI) in collaboration with the Translation and Terminology Development (TTD) unit. The argument avowed in this paper is that in the compilation of this specialised bilingual dictionary in the field of techno-science, lexicographers are mostly concerned with semantic equivalence between the Source Language (SL) and the Target Language (TL). Consequently, the practice by these bilingual

dictionary compilers is usually that of giving one-word equivalents. However, given the disparities and incommensurability between the language and culture of the SL and TL, it is sometimes difficult to find equivalence. Theoretically, one cannot expect an exact match between two languages that express different cultural realities. Given this scenario, compilers of the bilingual dictionary end up bridging the gap between languages by giving translational equivalents rather than relying solely on one-word equivalents. The *English–ChiBarwe Bilingual Elementary Science Terms Dictionary*, like other bilingual dictionaries, also displays this characteristic. This paper discusses the problems of translation equivalents in the compilation of this bilingual science dictionary where lexicographers will be dealing with divergent languages and cultures, traditional practices of lexicography and the absence of a reliable ChiBarwe corpus.

In lexicography, the basic activity of semantics is the analysis of meaning. The meaning of a lemma or expression is a concept related to the thing or idea that it refers to or that it represents which can be described with other words. In monolingual dictionaries, only one language is used for both the lemma and the glosses. On the other hand, in bilingual dictionaries, the meaning of a lexical item is derived through the process of translation of the SL word into that of the TL. A bilingual dictionary, then, is one where two languages are used, one for the lemma and the other for the glosses. In the case of the English–ChiBarwe Bilingual Elementary Science Terms Dictionary, the lemmas are in English and the glosses are in ChiBarwe. As noted by Bassnett-McGuire (1988:2), a minimally acceptable translation is one where the surface meanings of the SL and TL are approximately similar and also one that preserves the structures of the SL. In other words, the ideal in translation is to achieve structural and semantic equivalence. Absolute equivalence between any two languages, however, is rare. The reason for this is that one is dealing with languages with different grammatical structures and sociocultural settings. The semantic problems involved in bilingual dictionaries are more complicated than those in monolingual ones because the latter are written for people who participate in and understand the culture being described, whereas the former describes a culture which differs from that of the users (Al-Kasimi 1983:58).

In this case, the major task of a bilingual lexicographer is to find appropriate equivalents (Al-Kasimi 1983:58) but in some cases, the entries take the form of an explanatory or a descriptive equivalent. These explanatory or descriptive equivalents are necessitated by the absence of a translational equivalent in the target language. In the *English–ChiBarwe Bilingual Elementary Science Terms Dictionary*, there are instances where it is difficult to give exact equivalents, and this is where the lexicographer is left with little choice but to use explanations and employ various term creation techniques. Data for this research were collected through analysing the ChiBarwe Language Database and analysis of existing literature. Since the researcher is the chief lexicographer in the compilation of this dictionary, the researcher also utilised extracts from the uncompleted dictionaries to draw relevant examples for this paper. It was discovered that there are numerous problems of equivalence in the compilation of the *English–ChiBarwe Bilingual Elementary Science Terms Dictionary* due to the linguistic and cultural differences between English and ChiBarwe. Research findings are presented and analysed qualitatively using Chimhundu's (1996) Scan and Balance Theory. The Scan and Balance theory considers the process of translation as involving creating rather than merely searching for equivalence which would lead to the untranslatability of some technical documents as those found in the field of techno-science.

References:

Al-Kasimi, A.M. 1983. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
Bassnett-Mcguire, S. 1988. *Translation Studies*. London: Routledge.
Chimhundu, H. 1996. Translation in African LLDs and Communication for Development: proposing the Scan and Balance Theory. *Proceedings from the XIVth World Congress of FIT, Melbourne, Australia, 9–16 February*: 448-459. Melbourne: AUSIT.

*****

An Etymological Dictionary of Khoekhoe-Nguni Loanwords

Camilla Rose CHRISTIE (imbriferumver@gmail.com)
Department of Linguistics, Rhodes University, Makhanda, RSA

Click loan, a contact-induced change in which lexical items containing clicks are borrowed into a historically clickless language, was and is a widespread phenomenon across southern Africa. (Gunnink et al. 2015; Sands & Gunnink 2019). Multiple languages in the S40 Nguni group of the broader BANTU language family have gained click words via contact with extinct and undocumented Khoekhoe- and !Ui-branch languages. Today, S41 isiXhosa and S42 isiZulu are the largest and most widely spoken click languages on earth.

Despite the ubiquity of Khoekhoe loanwords in the Nguni languages, there is no consolidated resource dedicated to their etymological origins. The available discussions treat only a fraction of the total Nguni click loanwords. The longest list ever compiled (Meinhof 1905) contains only 166 click etymologies, drawn from the limited Nama and isiXhosa lexicographic resources of the late 19th century. It has never been translated from German, and so is inaccessible to most South Africans. Subsequent discussions (Anders 1937; Bourquin 1951; Lanham 1964; Louw 1977, 1979; Ownby 1985; Argyle 1986; Herbert 1990) are numerous but scattered, and none etymologises more than 50 items.

Furthermore, all available resources are now outdated, and so inevitably error-prone. For one thing, they predate the modern consensus (Güldemann 2014) that the 'Khoisan' grouping comprises three genetically distinct lineages. As a result, authors such as Louw (1977:138-144) and Argyle (1986:59-64) erroneously draw etymological comparisons with click languages in the KX'A language family. For another, they predate several major new developments including the provision of expanded lexicographic resources for Namibian Khoekhoe (Haacke & Eiseb 2002); the redocumentation of the Kora language (Du Plessis 2019); and the documentation of the Nǀuuki language (Collins & Namaseb 2011; Sands & Jones 2022). These developments have broadened our knowledge of click languages historically spoken in South Africa, and, in turn, have the potential to improve our understanding of click loan into Nguni.

This paper discusses the compilation of a modern lexicographic resource dedicated to Khoekhoe-Nguni loanwords. Using comparative methods, it compiles all existing Khoekhoe-Nguni etymologies into a look-up dictionary, while also adding over 300 new etymologies as sourced from modern standardised references. Although it focuses on Khoekhoe-Nguni loanwords, it also collates a smaller set of around 100 !Ui-Nguni loanwords. In addition to providing sociohistorical discussion of the Khoekhoe-Nguni contact event itself, it also offers a linguistic analysis of the rephonologisation strategies employed during loan.

The anticipated lexicographic output is a sociohistorical linguistic print resource, ideally accessible to both linguists and lay readers, that will facilitate research both into the phonological behaviour of click consonants during loan, and into the precolonial contact dynamics that shaped the southern African linguistic landscape.

References

Anders, H.D. 1937. The clicks. *The South African Journal of Science* 33: 926-939.

Argyle, W.J. 1986. The extent and nature of Khoisan influence on Zulu. *Sprache und Geschichte in Afrika* 7(2): 43-71.

Bourquin, W. 1951. Click-words which Xhosa, Zulu and Sotho have in common. *African Studies* 10 (2): 59-81.

Collins, C. & Namaseb, L. 2011. *A Grammatical Sketch of Nǀuuki with Stories*. Köln: Rüdiger-Koppe Verlag.

Du Plessis, M. 2019. *Kora: A lost Khoisan language of the early Cape and Gariep*. Pretoria: UNISA University Press.

Güldemann, T. 2014. 'Khoisan' linguistic classification today. Tom Güldemann and Anne-Marie Fehn (eds.), *Beyond 'Khoisan': Historical relations in the Kalahari Basin*: 1-41. Amsterdam: John Benjamins.

Gunnink, H., Sands, B., Pakendorf, B. & Bostoen, K. 2015. Prehistoric language contact in the Kavango-Zambezi transfrontier area: Khoisan influence on southwestern Bantu languages. *Journal of African Languages and Linguistics* 36(2): 193-232.

Haacke, W.H.G., & Eiseb, E. 2002. *Khoekhoegowab-English dictionary.* Windhoek: Gansberg Macmillan.

Herbert, R. K. 1990. The sociohistory of clicks in Southern Bantu. *Anthropological Linguistics* 32(4): 295-315.

Lanham, L.W. 1964. The proliferation and extension of Bantu phonemic systems influenced by Bushman and Hottentot. H. Lunt (ed.). *Proceedings of the 9th international congress of linguists*: 382-391. The Hague: Mouton.

Louw, J.A. 1977. Clicks as loans in Xhosa. Jan Snyman (ed.). *Bushman and Hottentot linguistic studies*: 82-100. Pretoria: UNISA University Press.

Louw, J.A. 1979. A preliminary survey of Khoi and San influence in Zulu. *Khoisan Linguistic Studies* 5: 8-21.

Meinhof, C. 1905. *Hottentottische Laute und Lehnworte im Kafir*. Leipzig: Brockhaus.

Ownby, C. P. 1985. Early Nguni History: The Linguistic Evidence and its Correlation with Archaeology and Oral Tradition. PhD dissertation. Los Angeles: University of Los Angeles.

Sands, B., & Gunnink, H. 2019. Clicks on the fringes of the Kalahari Basin Area. Emily Clem, Peter Jenks, Hannah Sande (eds.), *Theory and description in African linguistics: Selected papers from the 47th annual conference on African linguistics*: 703-24. Berlin: Language Science Press.

Sands, B., & Jones, S. 2022. *Nǀuuki Namagowab Afrikaans English: ǂXoaki ǂxanisi / Mîdi di ǂKhanis / Woordeboek / Dictionary. African Sun Media*: Cape Town.

*****

GREDYLEX's contribution to Dictionary Use and Culture in Education in Gabon

Edgard Maillard ELLA (ellamaillard@yahoo.fr)

Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques, Institut de Recherche en Sciences Humaines, Libreville, Gabon

Ludwine MBINDI ANINGA (sunchine9208@gmail.com)

Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques, Institut de Recherche en Sciences Humaines; Centre de Recherches en Etudes Germaniques et Interculturelles, Université Omar Bongo, Libreville, Gabon

Gilles SAPHOU-BIVIGAT, (saphoubivigat@yahoo.fr)

Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques, Institut de Recherche en Sciences Humaines, Libreville, Gabon

Léandre Serge SOAMI (sergesoami@gmail.com)

Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques, Institut de Recherche en Sciences Humaines, Libreville, Gabon

In Gabon, dictionaries are still used fairly sparingly in the education system and by some barely in the practice of their profession. This is even truer for works in local languages. As Kalonji (1993:46) points out, in sub-Saharan African countries which still have a strong oral tradition, the dictionary is regarded as an imported object of the Western culture. However,

as Dru (1998:1) puts it, the dictionary is one of the essential tools for mastering a language for both text production and oral communication performance. Moreover, local lexicographic production is becoming increasingly important in Africa against the backdrop of the desire to preserve and promote native languages and cultures.

The aim of this paper is to report on the Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques' (Research Group on Linguistic and Lexicographic Dynamics – GREDYLEX) contribution to dictionary use and culture in education in Gabon through *Kabi*, GREDYLEX Practitioners' Journal. GREDYLEX is a research unit of the Institute for Research in Human Sciences within the National Centre for Scientific and Technological Research. The paper focuses on two volumes of *Kabi*, which resulted from the study of dictionary use and culture in Gabonese primary and secondary schools.

The research problem of the study can be stated in three questions. First, should GREDYLEX promote dictionary culture in Gabon through French dictionaries produced in France for use in the Gabonese education system? Second, how can dictionary culture from French works help to promote dictionary culture in native languages in Gabonese schools? Third, what are the measures to be put in place by GREDYLEX so that dictionary culture is relevant for both French dictionaries and native language dictionaries?

To answer the abovementioned questions, GREDYLEX conducted a fieldwork approach combining both practice and theory. The study was designed following the framework of the general theory of lexicography set out by Wiegand (1989), namely:

a) the first unit of Theoretical Constituent A which is the general theory relating to dictionary aims, and

b) the first component of Theoretical Constituent D, which is the theory of lexicographic research on language consisting of a typology of dictionaries and its principles and the second in a category of all the presentations of the results of linguistic lexicography as texts about language.

Additionally, a complementary methodological approach was adopted from Saporito (1997) and Dru (1998). Both suggest educational approaches for learning the dictionary use at school. Saporito (1997:8) argues that dictionary learning should make the use of dictionaries more effective and that children must have this instrument constantly at hand, in the same way as the calculator or the atlas of geography. According to Dru (1998:12) dictionary learning must be carried out over the long term and on a daily basis so that children acquire reflexes about using the dictionary.

The study has resulted in the two volumes of *Kabi*. In reference to the first volume aimed at primary school, preliminary results report on the possibility of strengthening the use of the dictionary or introducing it into the curricula. Then, with the second volume aimed at high school, results show that dictionary use can be also introduced into school curricula. In this regard, Mabika Mbokou (2001) and Ella (2015) both suggest that dictionary can help to reinforce or improve not only the curricula but also language production, notably writing, reading and speaking.

The current paper is organized into two sections, each one consisting of the presentation of the different parts of each volume of *Kabi*. For the volume aimed at primary school, it is about the theoretical approach of the dictionary, the practical approach of the dictionary, the activity sheets and the corrections of the activity sheets. For the one aimed at secondary school is about the activity protocol, the activity schedule, the questionnaires, the activity sheets, the progress reports of each class and the teachers' progress report. These two volumes of *Kabi* can contribute to the cultural practice of using dictionaries in Gabon.

References

Dru, A. 1998. La pratique du dictionnaire. Crinon, J. et al. (eds.). *Extrait des Mémoires Professionnels des Stagiaires de Deuxième Année.*
http://www.francais.creteil.iufm.fr/mémoires/dru.htm

Ella, E.M. 2015. Plaidoyer pour l'introduction de la pratique du dictionnaire comme enseignement dans l'éducation primaire au Gabon. *Afrique, Langues et Cultures* 5: 73-97.

Kalonji, Z. 1993. *La Lexicographie Bilingue en Afrique francophone: L'Exemple Français-Cilubà.* Paris: L'Harmattan.

Mabika Mbokou, L. 2001. Le rôle du dictionnaire dans le système éducatif. Emejulu, J.D. (ed.). *Éléments de Lexicographie Gabonaise*: 206-222. New York: Jimacs-Hillman.

Saporito, J-C. 1997. Fichier d'utilisation du dictionnaire. *Le Nouvel Educateur* 97: 13-20.

Wiegand, H.E. 1989. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J et al. (eds.). *Wörterbücher. Dictionaries. Dictionnaires. Ein Internationales Handbuch zur Lexikographie*: 246-280. Berlin: De Gruyter.

<div align="center">*****</div>

Terminology management programmes within the South African context

Michelle GOOSEN (michelle.goosen@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

The Constitution of South Africa from 1996 makes provision for the equal use of the 11 official languages of the country and the promotion and development of previously historically marginalised languages. However, it is not always feasible – even within the national government departments – to provide users with multilingual (standardised) terminology in the 11 official languages. According to Antia (2015:470), the national departments are expected to publish documents in all 11 official languages, but – in cases where this is not possible – documents are published in six languages. However, this statement contradicts the Use of Official Languages policy of 2012 (https://www.gov.za/documents/use-official-languages-act). According to this policy, the national departments are expected to make all official documents communicated to the public available in at least three official languages.

In relation to terminology management and terminology management programmes in South Africa, Ferreira found in 2002 that the National Language Services' (NLS) Terminology Coordination Department was at that stage the only organisation that focused on terminology management. More recently, Antia (2015:473) states that terminology is currently developed by both state institutions and private institutions on a systematic or ad hoc basis within formal and informal frameworks. Some terminology development initiatives are stand-alone projects, while other projects are part of integrated terminology development initiatives. He claims that the fragmented nature of these terminology management programmes makes it difficult to give an accurate overview of the work that has already been done. Khumalo (2017:260) also adds that to date there are several institutions that develop terminology - he refers here to South African universities, the NLS within the Department of Sport, Arts and Culture (DSAC) and the South Africa Norway Tertiary Education Development Program (SANTED). However, Khumalo (ibid.) does not give any further information regarding which universities are involved or in which subject fields and/or languages terminology is being developed.

Therefore, the need exists to give a critical account of the main terminology management programmes within the South African context. Through this, it will be possible to determine

which terminology has already been developed in which subject fields and languages. This will also prevent duplication of subject fields and the proliferation of terms. Resources can then be used much more effectively to develop terminology for other subject fields. Furthermore, it should also be determined whether these terminology management programmes make use of the latest technological development available for terminology management.

This paper firstly considers using terminological knowledge bases as a means to make terminological information available to users. The profile of the target user of a terminological product forms part of this discussion. Secondly, the different aspects of terminology management are discussed. It is argued that terminology management is carried out within different contexts for different purposes. Examples within the South African context are given to illustrate this statement. Thirdly, a critical account is given of the main terminology management programmes within the South African context. It is indicated that before South African terminology management programmes can implement the latest technological developments for terminology management, i.e. terminological knowledge bases, the lack of terminological databases – specifically open-source terminological databases – must be addressed. It is also pointed out that a lack of expertise and resources within the South African context, remains a challenge. Lastly, this paper determines to what extent the South African Centre of Digital Resources' (SADiLaR) (www.sadilar.org) infrastructure can be used as a national platform for the coordination of terminological resources in South Africa. Given the limited resources, this developed national platform is the ideal platform to host a terminological database or terminology portal.

Multilingual terminology programmes are a strategic resource which can be utilised in the functional development of a country's languages. Thus, the importance of such programmes within the South African context, must not be overlooked or underestimated.

References

Antia, B.E. 2015. Language policy and terminology in South Africa. Kockaert, H.J. & Steurs, F. (eds.). *Handbook of Terminology* 1: 467-488. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ferreira, D.M. 2002. Terminologiebestuur in Suid-Afrika met spesifieke verwysing na die posisie van histories ingekorte tale. PhD dissertation. Bloemfontein: University of the Free State.

Khumalo, L. 2017. Intellectualization through Terminology Development. *Lexikos* 27: 252-264.

South African Centre of Digital Language Resources (SADiLaR). 2024. www.sadilar.org. [01/05/2024].

Use of Official Languages Act 12 of 2012. 2012. https://www.gov.za/documents/use-official-languages-act. [18/09-2018].

*****

Information retrieval structures in online lexicography

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

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

Theo J.D. BOTHMA (theo.bothma@up.ac.za)

Department of Information Science, University of Pretoria, Pretoria, South Africa

In the planning and compilation of both printed and online dictionaries lexicographers need to devise data distribution structures that can be applied to allocate the different lexicographic data to their respective search positions. These positions might be search zones

within the dictionary article, outer texts in the search region, or, in the case of online dictionaries, a search position in the search domain, i.e. a dictionary portal, or the search universe, i.e. the internet.

To achieve a successful dictionary consultation procedure the intended target user must be familiar with the data on offer and must be able to access the data in order to retrieve the required information. In printed dictionaries with a static article and word book structure the relation between inserting data and retrieving information is not complex and most dictionaries follow more or less the same ordering systems. Online dictionaries display dynamic structures with data often presented in various levels and layers that compel the user to drill down into articles or into specific comments of articles, and to embark on search procedures that must negotiate new structures like extended comments and even new search positions like the search tunnel. It is of paramount importance that users should be aware of the full spectrum of data included in a lexicographic product and should be able to select the data types of interest to them.

The data distribution structure of a dictionary must be complemented by an access structure that determines different search routes a user can follow to reach the required item or item text in a dictionary article or an outer text. However, the users of lexicographic tools need assistance to ensure an optimal retrieval of information from the data. Consequently, lexicographers need to embark on dedicated ways to establish an information retrieval structure that can enable users to achieve an adequate dictionary consultation.

Dictionaries and other lexicographic instruments can and should play a vital role in the provision of knowledge. The success of the language services provided by lexicographic products depends on the user successfully accessing the lexicographic data and retrieving the required information to find solutions to the problems that initiated the consultation process. In this regard, lexicography must take cognizance of the real needs of real users of real dictionaries and other lexicographic instruments. Presenting data is not enough. Structures must be in place so that the envisaged user can interpret and use the data in an appropriate way. This demands the implementation of an information retrieval structure. An information retrieval structure is a mirror image of the data distribution structure. It is characterised by an ordered sequence of steps, indicated by appropriate data indicators, that can be followed by the user of a dictionary or other lexicographic product to achieve an optimal retrieval of information from different search positions in the specific product. An information retrieval structure reflects the search venues, and the respective information retrieval routes a user follows to reach the required data from which information can be retrieved. These routes can either follow all the steps or utilise a restricted sequence of steps to ensure a rapid information retrieval.

This paper focuses on different aspects of information retrieval structures in different lexicographic instruments. These instruments include a variety of dictionaries but also reading and writing assistants and other lexicographic products. Proposals are made for a structured and rapid access to data and the use of data indicators to guide users to items occurring in complex dictionary articles. The lexicographic needs and reference skills of the users are at the heart of the proposals being made. New forms of access structures and search routes are suggested to assist the user in the best possible way. The emphasis is also on user-friendliness of lexicographic instruments that can enhance the quality of intuitive dictionary use.

<p align="center">*****</p>

Making sense of *kuningi* using a corpus linguistic analysis

Langa KHUMALO (langa.khumalo@nwu.ac.za)
SADiLaR, North-West University, Potchefstroom, South Africa.

In 2023 the Pan South African Languages Board (henceforth PanSALB) pronounced the word *kuningi* as the inaugural South African word of the year for social media. In its statement of 16 October 2023, PanSALB stated in this pronouncement that the word is reflective of "[…] real language use" (Huluhulu 2023). *Kuningi* is an isiZulu word literally meaning "it's a lot". This study analyses the meaning of *kuningi* using a corpus linguistic approach. Two analysis corpora (AC) were created from well-known social media platforms, namely X (formerly Twitter) and Instagram. The other two analysis corpora (AC) were created from two prominent isiZulu newspapers, respectively *Isolezwe[i]* and *Ilanga[ii]*. The corpus data that was collected for this study was limited to the data produced between 1 January 2023 and 30 November 2023.

The study uses the corpus linguistic approach as a basis for the theoretical analysis. According to Sinclair (2005), a corpus is "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research." The advantage of such a theoretical approach is that "[…] a corpus [is] stored in a computer, it is easy to find, sort and count items, either as a basis for linguistic description or for addressing language-related issues and problems" (Kennedy 1998: 11). A corpus is thus a collection of naturally occurring texts derived from real life language use in either written or spoken form, which is then processed, stored, and accessed by means of computers (Khumalo 2015: 497). Such a corpus is then useful as a basis for, inter alia, investigating language use.

The analysis corpora from X and Instagram (henceforth ACX and ACI) form the basis for the analysis of the word *kuningi* from a social media perspective, and therefore is viewed as projecting the most understood public meaning of *kuningi*. Because of the nature of the text types that are derived from these platforms, the corpus data is small. ACX has 870 tokens, while ACI has 4 218 tokens. *Kuningi* is the most frequent word in ACX as shown in Figure 1. It is the third most frequent word in ACI.

---

[i] https://www.isolezwe.co.za/
[ii] https://ilanganews.co.za/

Figure 1. Frequency List for the ACX.

The study uses the two corpora derived for the most popular isiZulu newspapers Isolezwe (henceforth ACISO) and Ilanga (henceforth ACILA), to comparatively analyse the standard use, and therefore, standard meaning of *kuningi*. The two corpora are also limited to the period 1 January 2023 - 30 November 2023. ACISO and ACILA respectively have 4 061 tokens and 2 915 tokens. To enable the comparison, the study uses Voyant Tools to query and analyse the two sets of corpora. Voyant Tools is used a lot in digital humanities to facilitate the analysis of textual data using computational tools. Voyant Tools is an open-source, web-based application for performing data analysis. It enables close text reading and interpretation using sophisticated functionalities for analysing the key-word-in-context. Its use enabled the swift creation of the wordlist, concordance analysis and other (syntactic) correlations. The study will show that in the social media context the word *kuningi* has extended its semantic range to include "[…] is hard", "[…] is difficult" and "[…] is complicated". While its use in the formal public media has retained its general meaning "[…] many" and "[…] a lot".

Section 1 briefly discusses the PanSALB and its role in the promotion of the 12 official languages in South Africa. The context of the South African Word of the Year is also discussed in the section. Section 2 discusses crowdsourcing and the corpus. The section discusses in detail the four corpora and how these were carefully collected. We analyse the keyword in Section 3 using a corpus linguistics as a framework, we discuss the findings in Section 4 and conclude in Section 5.

References

Huluhulu, N. 2023. Announcement of SA Word of the Year 2023. https://www.pansalb.org/announcement-of-sa-word-of-the-year-2023/ [23 April 2024].

Khumalo, L. 2015. Semi-automatic term extraction for an isiZulu linguistic terms dictionary. *Lexikos* 25: 495-506.

Kennedy, G.D. 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.

Sinclair, J. 2005. Corpus and Text: Basic Principles. Wynne, M. (ed.). *Developing Linguistic Corpora: A Guide to Good Practice*: 1-16. Oxford: Oxbow Books.

*****

Toward Shona Techno-Science Terms Development

Allan T. MAGANGA (maganagaa@staff.msu.ac.zw)

Midlands State University National Languages Institute, Gweru, Zimbabwe

The thrust of this paper is on analysing the mainstreaming of indigenous languages in the teaching and learning of scientific disciplines in the early stages of the Zimbabwean education system. Currently, the teaching and learning process is done in English language – which is a second language to the majority of the learners. Therefore, this paper is an effort to understand how the glossary compilation project being carried out at the Midlands State University National Languages Institute (henceforth MSUNLI) and also sponsored by Zimbabwe's Ministry of Higher and Tertiary Education, Innovation Science and Technology Development (hereafter MHTEISTD) is useful in unlocking the teaching and learning of techno-scientific knowledge through Shona (an indigenous language) in Zimbabwe's Primary Education system. It also holds that the ongoing elementary glossary of science terms making in specialised fields, science-techno as in this case, relies heavily on term creation strategies in the local language and thus becomes a sure way toward decolonising the Anglophone oriented curricula on issues related to the medium of instruction. As part of the codification principle in the indigenous language empowerment or development process, the project seems to underscore that, terminology development is in essence corpus planning in praxis since there are no language boards or any other board that deals with corpus planning in Zimbabwe. In its effort towards this codification related exercise, MSUNLI can be credited for spearheading the MHTEISTD's endeavour in path-finding the teaching and learning of Science Technology Engineering Mathematics (henceforth STEM) learning areas in the 'constitutionally recognised languages'. This, as the paper further argues, is also in tandem with Sustainable Development Goal 4 on Quality Education which "ensure[s] inclusive and equitable quality education and promote lifelong learning opportunities for all" especially L1 speakers of languages of limited diffusion (LLD) or the so-called previously marginalised languages. This is against the backdrop that the ongoing specialised term creation for the enriched glossary, now at production stage and successfully carried out in all "the officially recognised languages of Zimbabwe" (Chapter 1 Section 6), has compiled techno-scientific nomenclature for primary school level learners as a clear demonstration "that there are many functions which the local languages can perform without necessarily competing with English," Chimhundu (1996:151).

Guided and informed by Chimhundu's (1996) scan and balance theory, the paper unravels the different ways of developing terms that were created during the indigenisation exercise of Western scientism. This approach is qualitative in nature since it is hinged on intuitive knowledge and deductive syllogism. The theory became the lens that was utilised for translanguaging 'harvested terms' from the textbooks that are used as primary sources by learners in schools. Creating terms from the English source to the respective indigenous language was influenced by the scan and balance theory which underscores minimising issues to do with language loss and gaining during the search for equivalences. This is because during the enriched glossary of terms making process, it was found handy for its emphasis on balancing, "senses or ranges of meaning, usage…and impact and then mak[ing] selections accordingly" (Chimhundu 1996:155) and accuracy whereas in this paper it was co-opted for its innate quest for sameness. The primary focus of this paper is on the ChiShona language specialised glossary terms for science and technology with particular interest on the strategies utilised – approximation, re-phonologising, borrowing, coining, onomatopoeia – for terminological expansion in Shona as one of the languages in search of being granted the medium of instruction for STEM education. In terms of English terminology source, the terms were harvested from STEM related learners' textbooks and ascribed meanings using artificial intelligence applications. The paper purposively samples three articles for each strategy that

was deployed or co-opted during the compilation process. The rationale of zeroing in on three articles per each term creation strategy is guided and informed by the fact that an article has meta-data for each central list entry. Therefore, the utilised sample is quite informative and enriching.

The paper preliminarily concludes that with this enriching techno-scientific glossary of terms as a teaching and learning aid for both the educator and the learner, teaching techno-scientific concepts in the Ministry of Primary and Secondary Education's Heritage Based Curriculum Framework implementation will be functional and valid as teaching and learning material since at this 'kick off stage' there is no proper resource in place for such endeavours. Thus, the creation and expansion of science nomenclature for the scientific disciplines will definitely reform and capacitate the teaching and learning process from elementary level, which is a critical base in the learning process. This paper is of the view that the possibility of reaching the peak of technological advancement just like the Asian tigers – Japan, South Korea and China – starts with an educational reform, precisely capacitating the medium of instruction. In a nutshell, the paper concludes that without the glossary of terms for elementary learners, any effort towards innovation and industrialisation remains elusive. What the MSUNLI and the MHTEISTD have initiated will not only empower the techno-scientific landscape but also contribute towards an effective and efficient human capital base, rooted in its linguistic heritage, through such an initiative.

References
Chimhundu, H. 1996. Translation in African LLDs and Communication for Development. *Proceedings from the XIVth World Congress of FIT, Melbourne, Australia, 9–16 February*: 448-459. Melbourne: AUSIT.
Elementary School Level Glossary of Terms: Creation of Science Terms Based for ECD to Grade 7. 2024. MSUNLI Publications (forthcoming).
The Constitution of Zimbabwe Amendment (No. 20) Act, 2013 (Act 1 of 2013).

*****

A Morphophonological Analysis of Transliterations in the UFS Sesotho Accounting Terminology List

Malefu Justina MAHLOANE (MahloaneMJ@ufs.ac.za)
Academy for Multilingualism, University of the Free State, Qwaqwa, South Africa

Curriculum planning in South Africa involves careful consideration of language policies to ensure inclusivity, accessibility, and linguistic diversity. The Language Policy for Higher Education (2020), issued by the Department of Higher Education and Training (DHET) emphasises the importance of promoting multilingualism and the use of indigenous languages in various academic contexts, including teaching, learning, research, and official communication.

Terminology development plays a vital role in supporting the objectives outlined in the language policy framework. By developing terminology in indigenous languages, higher education institutions can enhance the accessibility of academic content to students from diverse linguistic backgrounds. This not only promotes linguistic diversity, but also contributes to the preservation and revitalisation of indigenous languages.

Terminology, serving as one of the tools of effective communication within specialised fields, embodies various definitions across linguistic and scholarly contexts. Amongst others, terminology is characterised as a concise concept adhering to native language rules, encapsulating elements of the external or internal human world. Therefore, terms are

delineated as linguistic units that unambiguously correspond to concepts in social, political, scientific, technological, and artistic realms, distinguished by its precise semasiological boundaries (Kaguera 2015:45). Further elaborating on the concept, it is defined as a linguistic unit used to accurately express field-specific concepts in knowledge, production, or culture, fulfilling communicative needs within human activity domains. There are multiple processes utilised during terminology development, one such process being transliteration wherein terms are adapted to fit the phonological, morphological, and syntactic patterns of the target language.

This research has focused on the adaptation of transliteration with a particular interest in Sesotho accounting terminology that was recently developed by the University of the Free State (UFS). One of the objectives of the UFS Language Policy that was approved in 2023 is to intellectualise Sesotho as an academic language by, amongst other things, developing Sesotho terminology for the various subjects offered at the university. Therefore, this paper explores the implications of transliterating the evolution of Sesotho terminology. Drawing on the functional (descriptive) approach to terminology (Vakulenko 2014:15), which speaks to presenting terms to enhance contextual understanding, this study delves into potential contradictions with theories of Sesotho morphology, particularly looking at the usage of the language in the academic context.

It argues that achieving successful academic use of Sesotho necessitates a consistent and logical method for presenting transliterated terms alongside effective data management tools and comprehensive guidelines. Additionally, this study recommends that institutions collaborate on terminology projects for similar subjects to ensure consistency and reliability of the term lists. Uniformity is crucial for maintaining the integrity of terminology lists and avoiding inconsistencies.

The research employs a systematic morphophonological analysis of consonant clusters in anglicized Sesotho accounting terms. Preliminary findings suggest potential discrepancies in adhering to Sesotho morphological rules, highlighting the need for systematic approaches to terminology development.

Moreover, the research compares the morphological adaptations of English terms used in Sesotho financial management terminologies across various institutions, aiming to discern patterns and inconsistencies in loanword adaptation within the financial sector terminology in particular. These findings contribute to the discourse on terminology evolution in Sesotho, emphasizing the significance of consistency in transliterating. The implications extend to language policy formulation, academic curriculum development, and the broader promotion of Sesotho in scholarly discourse, fostering linguistic diversity and academic excellence. Through these efforts, the study aims to advance the integration of Sesotho into academic spheres while preserving its linguistic integrity and cultural heritage. Further research will be carried out on how to improve terminology management methods for maintaining consistent and standardised term lists.

References

The Language Policy Framework for Public Higher Education Institutions. 2020. https://www.gov.za/sites/default/files/gcis_document/202011/43860gon1160.pdf. [11 May 2024].

Kageura, K., 2015. Terminology and lexicography. Kockaert, H. and Steurs, F. (eds.). *Handbook of terminology* 1: 45-59. Amsterdam: John Benjamins.

Vakulenko, M.O. 2014. Term and terminology: basic approaches, definitions, and investigation methods (Eastern-European perspective). *Terminology Science & Research/Terminologie: Science et Recherche* 24: 13-28.

<div align="center">*****</div>

IsiXhosa.click: online, open, user-friendly, and searchable isiXhosa-English dictionary software

Cael MARQUARD (cael.marquard@gmail.com)

Department of Computer Science, University of Cape Town, Cape Town, South Africa

IsiXhosa.click (https://isixhosa.click) is an open-source, online, easy-to-use isiXhosa-English dictionary. It supports typo-tolerant live search, allowing users to find words quickly by typing their first few letters. Word pages provide example sentences, related words, and linguistic information. The software allows users to submit corrections and new vocabulary, which are published after review. IsiXhosa.click is a community-driven, crowd-sourced dictionary project that still enforces quality standards. The website's database and source code are freely available under open-source licenses.

IsiXhosa.click started in 2021 as an independent, student-led project by a team of home-language English speakers learning isiXhosa. Existing isiXhosa dictionaries were found to be inconvenient and costly. They use proprietary software; do not allow re-use, redistribution and modification; and do not easily permit users to suggest edits. Thus, the software for IsiXhosa.click was built expressly for the project. A key design goal was ease of use for those unfamiliar with dictionaries. The interface is simple and uses minimal jargon, preferring to use terms familiar to learners of isiXhosa. Development effort concentrated on commonly used features, particularly the ability to discover words in the dictionary.

Instead of reading an ordered list of words, users primarily find words through the site's search feature. Live results are returned immediately as they type. Words may be searched by their full isiXhosa form, their isiXhosa stem, or their English translation. Search is provided by the Tantivy[i] library. A fuzzy term query allows for typo-tolerance within the search. Results are ranked by how well they match the search term using the Jaro-Winkler distance (Winkler 1990). Because the search is typo-tolerant, bilingual, and stemming-enabled, searching is intuitive to users of any lexicographic skill level. Even words which users misspell may be found, which is particularly useful to learners who are new to transcribing isiXhosa.

Words may also be discovered through the "related words" sections. Once a word is found through search, clicking it brings up a page describing it in full. Here, applicable grammatical information such as part of speech, infinitive, and noun class is presented. Example sentences using the word are displayed, if available. Lastly, there is a section listing related words, hyperlinked to their entries in the dictionary. The linking method is also listed (e.g., whether they have related meanings, or whether one word is the plural form of the other). This assists users to find the most relevant entry in the dictionary.

Since IsiXhosa.click is a fully open-source project, it constitutes an Open Education Resource (OER) (Wiley & Hilton 2018). The software is released under the GNU AGPL 3.0 license and available on GitHub[ii]. It may be used and adapted freely for research or other projects. The dictionary database is released under CC BY SA 4.0. Site archives in CSV format are uploaded to a public mirror[iii]. This allows for interoperability between other tools, such as Microsoft Excel.

---

[i] https://github.com/quickwit-oss/tantivy
[ii] https://github.com/IsiXhosa-click/isixhosa_click
[iii] https://github.com/IsiXhosa-click/database

Project members initially created IsiXhosa.click's word list by adapting resources with permission. In 2022, IsiXhosa.click received a grant through the DH OER Champions programme[i] to expand its word list. Additional entries have been added and updated through a crowd-sourcing approach.

The software was created using an iterative, exploratory approach. When features were developed, they were published to the site for user feedback. This process guided the selection and implementation of future features of the software.

IsiXhosa.click is a website and hence accessible via all major platforms. The site is adapted for mobile use. No special software is required, so the dictionary may be accessed at any time via the Internet.

IsiXhosa.click allows users to suggest modifications to the word list. Users may click an edit button to capture missing information or correct errors and may also submit new words. Suggestions are queued for moderator review. This prevents vandalism and allows the implementation of a quality assurance process.

IsiXhosa.click is an isiXhosa-English dictionary with focus on usability. Words are easily discovered through a typo-tolerant, bilingual, live search. It is crowd-sourced, yet resistant to vandalism, allowing users to improve its database over time. Finally, it is open source, and may be freely adapted for future projects.

References

Wiley, D. & Hilton J. 2018. Defining OER-Enabled Pedagogy. *International Review of Research in Open and Distributed Learning* 19(4): 132-138.

Winkler, W. E. 1990. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. Proceedings of the Survey Research Methods Section, American Statistical Association (1990), *Washington DC, United States of America*: 354-359. http://www.asasrms.org/Proceedings/papers/1990_056.pdf. [21/04/2024].

<div align="center">*****</div>

The *Dictionnaire des Expressions Idiomatiques du Lumbu*: The Latest Addition to Gabonese Dictionary Production

Ludwine MBINDI ANINGA (sunchine9208@gmail.com)
Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques, Institut de Recherche en Sciences Humaines, CENAREST; Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo, Libreville, Gabon
Hugues Steve NDINGA-KOUMBA-BINZA (nkbinza@uwc.ac.za)
Centre for Advanced Studies of African Society, University of the Western Cape, Cape Town, South Africa
Maryse Edith MALOUGHOU MANGAMA (ndongme@yahoo.fr)
Département d' Etudes Germaniques, Université Omar Bongo; Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo, Libreville, Gabon
Paul Achille MAVOUNGOU (moudika2@yahoo.fr)
Département des Sciences du Langage, Université Omar Bongo; Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo, Libreville, Gabon

Lumbu lexicography was recently enriched with a second dictionary (Mbindi Aninga et al.,

---

2023), in addition to the first one published almost fifteen years ago (Mavoungou & Plumel 2010). Entitled *Dictionnaire des Expressions Idiomatiques du Lumbu* (*Dictionary of Lumbu Idiomatic Expressions*, henceforth DEIL), the dictionary is both the latest addition to dictionary production in Gabon and the latest addition to the sententious literature of the African languages of Gabon.

The fruit of a long collaboration between Gabonese lexicographers and linguists who are not necessarily speakers of the language, the compilation of this dictionary followed the methodological and metalexicographic principles in use in modern dictionary production, namely Wiegand's general theory of lexicography (Wiegand 1983) and the user-driven approach (Mpofu 2007; Tarp 2009; Frankenberg-Garcia 2020) with necessary adaptations due to the specific context and issues of Gabonese lexicography (Mavoungouet al. 2014; Mavoungou 2010; Ndinga-Koumba-Binza 2005; Emejulu 2002; Gouws 2001).

The DEIL focuses on collocations and locutions in Lumbu (Bantu, B44). However, the distinction between collocation and locution is not easy (Siepmann 2005; Sbisà 2013). What distinguishes collocation from locution is the absence or presence of a base. The phrase has no base, while the collocation always has one. Phrases are polylexical units coded in language. They consist in sequences of multiword expressions, collocations, idiomatic expressions, to name a few. Those sequences of lexemes can be considered as independent lexical units. As such, they signify and are selected as a whole by the speaker.

In addition to collocations and locutions, the DEIL includes proverbial phrases and sentences, sayings and paroemias. This explains the use of "idiomatic expressions" concept to gather under one term different discursive items that have approximately similar discursive functions. In this monodirectional Lumbu-French dictionary, each idiomatic expression appears under the lemmatised keyword of the phrase. The dictionary also contains the translation and definition of the idiomatic expression, as well as the contexts of use and encyclopaedic notes.

Following insights proposed by Bergenholtz and Gouws (2016) on a metalexicographic genre of dictionary reviews, this paper primarily intends to introduce the DEIL. In addition to a general presentation of the scope of the dictionary, the paper covers the following topics:
    i.   the lexicographic processes in the compilation of the DEIL,
    ii. macro- and microstructural components of the dictionary, and
    iii. the insightful contributions to the Gabonese lexicography research and dictionary production.

References

Bergenholtz, H. & Gouws, R.H. 2016. On the metalexicographic genre of dictionary reviews, with specific reference to *LexicoNordica* and *Lexikos*. *Lexikos* 26: 60-81.

Emejulu, J.D. 2003. Challenges and Promises of a Comprehensive Lexicography in the Developing World: The Case of Gabon. Botha, W.F. (ed.). *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 195-212. Stellenbosch: Bureau of the WAT.

Frankenberg-Garcia, A. 2020. Combining user needs, lexicographic data and digital writing environments. *Language Teaching* 53(1): 29–43.

Gouws, R.H. 2001. Lexicographic training: approaches and topics. Emejulu, J.D. (ed.). *Eléments de Lexicographie Gabonaise. Tome 1*: 58-94. New York: Jimacs-Hillaman Publishers.

Mavoungou, P.A. 2010. *Lexicographie et Confection des Dictionnaires au Gabon*. Stellenbosch: SUN Press.

Mavoungou, P.A. & B. Plumel. 2010. *Dictionnaire Yilumbu-Français*. Libreville: Editions Raponda Walker.

Mavoungou, P.A. Moussounda Ibouanga, F. & Pambou, J.A. 2014. *Le Dico des Makaya et des Mamadou. Contribution à l'Etude du Français du Gabon*. Libreville: Editions Odette Maganga (ODEM).

Mbindi Aninga, L., Maloughou Mangama, M.E.N., Mavoungou, P.A. & H.S. Ndinga-Koumba-Binza. 2023. *Dictionnaire des Expressions Idiomatiques Lumbu*. Libreville: Éditions Malugu.

Mpofu, N. 2007. The user perspective in lexicography: the lemmatisation of fixed expressions in Duramazwi Guru reChiShona. *Lexikos* 17: 360-371.

Ndinga-Koumba-Binza, H.S. 2005. Considering a lexicographic plan for Gabon within the Gabonese language landscape. *Lexikos* 15: 132-150.

Sbisà, M. 2013. Locution, illocution, perlocution. Sbisà, M. & K. Turner (eds.). *Pragmatics of Speech Actions*. Berlin/Boston: De Gruyter Mouton.

Siepmann, D. 2005. Collocation, colligation and encoding dictionaries. Part I: lexicological aspects. *International Journal of Lexicography* 18(4): 409–443.

Tarp, S. 2009. Reflections on lexicographical user research. *Lexikos* 19: 275-296.

Wiegand, H.E. 1983. On the structure and content of a general theory of lexicography. Hartmann, R.R.K. (ed.). *LEXeter'83 Proceedings*: 13-30. Tübingen: Max Niemeyer

*****

A Triangle of Relationships Among Three Major Language Practices: Translation, Terminology, and Lexicography in African Languages: The Case of isiXhosa

Buyiswa M. MINI (buyiswa.mini@ru.ac.za)

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

This is a reflective article on an identified need for foregrounding lexicography as part of the translation and term creation processes in African languages, focusing on isiXhosa. The reflective thoughts and ideas presented here have been evolving in the author's mind over past and current experiences in, and involvement with, lexicography, translation in various fields and capacities, and the creation of new terms.

The purpose of this paper is to address translation inadequacy in isiXhosa (and, by extension, other African official languages) that has been observed in many translated materials, in the form of concerning issues such as mistranslations, literal translations, language inauthenticity, ambiguity, vagueness and other inaccuracies. These issues result in difficult-to-understand translated materials. Some users reject such materials outrightly, while others may use them only because they have to. This situation results in diminished value and undermined status of isiXhosa, reducing its meaning of 'official language'. The paper seeks to address this situation through a process of auto-ethnographic reflection on the author's endeavours to promote comprehensibility, and thus easy accessibility in translated products in selected translated works she has done in the recent past. Such endeavours were undertaken by constantly pausing and reflecting on the readability and comprehensibility of her translation, and its compatibility with the language's natural structure. Thus, the conceptual framework chosen for this paper is based mainly on concepts expounded in a collection of articles (cf. Pietrzak 2019), in an online journal of translation. These concepts revolve around reflection and self-examination. They include translator self-concept and self-constructed competence, resulting in authenticity, and other real-life related skills, adequate to meet both language-related standards and market demands.

This paper argues that there is a need, in many isiXhosa translation and term creation works, for the promotion of clarity of meaning, which is in turn promoted by simplicity, accuracy, term consistency in all contexts, and language and terminology authenticity. It further argues that one of the most significant ways of effecting such comprehensibility is the creation of dictionaries, simultaneously with, and to accompany, every translation product. Referring to a different lexicographical focus, Wang and Chen (2024:42) speak of the inclusion of newly formed translation terms, which they aptly call neologisms, in a dictionary. Words

developed or adapted for accurate and easy flowing translation diction, as well as terms newly created to meet gaps in the case of new concepts, must be recorded in a systematic, accurate, context-consistent, easy-to-find way for the target users of a translation product. Collaborative translation-lexicographic mutual support among isiXhosa translators and lexicographers, as well as the possibility of promotion of dictionary culture (cf. Nkomo & Wababa 2013:348) could in this way become a reality. The proposed dictionary will be subject- or field-specific, thus becoming a step towards meeting the need for "other small dictionaries ... [in] these [African] languages" (Nkomo & Wababa 2013:362). The paper reports and demonstrates, quoting from translated works, how the dictionary development part has been an innovative thought on the author's part, resulting from moments of pausing and reflecting significantly, creating, evaluating, and recreating suitable translation-diction and easy-to-understand terminology.

Most of the literature reviewed on translation-related reflective writing (e.g., Imane & Said 2023) focuses mainly on the training and education of future translators. The current paper's author argues for the recognition and application by already practising translators, of reflection and self-examination, with view to constantly assessing one's translation and improving its quality. The paper demonstrates how a translator can use this meta-cognitive strategy to enhance their practice, and how lexicographical knowledge application can take the process beyond the usual reflective journaling activity and raise the standard of their translation work.

For a detailed exposition of the observed relationship between translation, term creation and lexicography, this paper relies heavily on selected translation works the author has engaged in especially during the current decade of the 21st century.

The paper concludes by giving a summary of a number of other reflection-resultant translation guidelines and comprehensibility principles and strategies for improved IsiXhosa translation quality.

References
Imane, T. and Said, B. 2023. Investigating the Presence of Reflective Practice in Translator and Interpreter Training. *In Translation* 10(2): 266-283.
Nkomo, D. and Wababa, Z. 2013. IsiXhosa Lexicography: Past, Present and Future. Lexikos 23: 348-370.
Pietrzak, P. 2019. (ed.). *inTRAlinea Special Issue: New Insights into Translator Training*. https://www.intralinea.org/specials/article/2431. [18 May 2024]
Wang, A., & Chen, X. 2024. On the Inclusion of Neologisms in Oxford Advanced Learner's Dictionary (10th edition). *Lexikos* 34: 41-50.

*****

Inclusion of Etymological information in Swahili lexicographic works: a challenge to be surmounted

Norbert MTAVANGU (nmtavangu@udsm.ac.tz; mtavangu@yahoo.com)
Department of Foreign languages and linguistics, University of Dar es Salaam, Dar es Salaam, Tanzania

Beyond describing the meaning and linguistic information of lexical items, a dictionary narrates the history of lexemes. In tracing etymological information, a comprehensive dictionary would indicate the most primitive form of the defined word and its core meaning. From there, the lexicographer would trace evolutive stages up to the most updated form and meaning while elaborating various intermediary structural and semantic transformations.

From the etymological data, the reader can establish principles of meaning formation in the language as well as adaptation as well as accommodation of foreign concepts and linguistic forms which are essential processes in language development. Thus, for fast-growing languages like Swahili, these etymological records are essential not only for retrospective needs but also in determining the prospect and readiness of the language to accommodate new concepts.

Swahili is the most widely spoken African language. It is considered to be the most apt to unite Africa and well-equipped to create lexical terms in science and technology. However, publications on Swahili lexicography seems to display a misunderstanding of the importance of etymology in language corpus development. In Swahili lexicography, a strong foundation in etymology can be traced to early lexicographic works (see Sacleux 1939; Johnson 1939). In these works, the reader can find original roots and complex semantic networks between native lexemes while at the same time, establish the original meaning of foreign words and channels through which the words entered Swahili language. Regrettably, lexicographic, and academic works published from 1970 onwards seem to favour synchronic lexicographic studies at the expense of diachronic ones. In the latter perspective, anything related to the background of the lexical items is ignored and sometimes discouraged (see Zawawi 1979; Mdee 1986, 1993, 1998, 1999). In our view, these attitudes seem to constrain the whole process of word formation in the language obliging language planners to face and accommodate a growing influx of new terms and concepts into the language.

The present study is aimed at tracing the evolution of etymological thoughts within the Swahili lexicography and collecting stakeholders' views on the place of etymology in word formation. The author draws insights from two data collection methods. In the first place, the author studied etymological information from monolingual and bilingual Swahili dictionaries published from the 19th to date. In the second, the researcher analyses responses from a questionnaire distributed to experts in language and linguistics in Tanzania enquiring about the place of etymology in Swahili lexicographic studies.

The findings demonstrate low apprehension of the importance of etymological information in dictionaries. A similar, picture is evidenced in contemporary Swahili lexicographic works. In these works, a few dictionaries indicate etymological information. Even these dictionaries which include the information, do not go beyond marking the language of origin of foreign words. This low awareness of the importance of the history of word forms and meanings is partly explained by the overall conceptualization of language which eventually led to the removal of etymological content in educational curricula at all levels. Considering this as a back sliding, the author therefore recommends awareness raising on the importance of trace word history especially for the adoption of new words and incorporation of emerging concepts into the Swahili language.

References

Johnson, F 1939. *A Standard Swahili-English Dictionary*. Dar es Salaam: Oxford University Press.

Mdee, J. S. 1986. Mtunga Kamusi ya lugha sanifu na usanifu wa vidahizo vyake: Mifano kutoka Kiswahili. *Kiswahili Jarida la Taasisi ya Uchunguzi wa Kiswahili* 53(1-2): 84-91.

Mdee, J. S. 1993. Lexicographic needs of Swahili Learners: Implications for Swahili Learner's Dictionary, sources of vocabulary for Swahili Dictionary. Dar es Salaam: Institute of Kiswahili Research

Mdee, J. S. 1997 *Nadharia na Historia ya Leksikografia*. Dar es Salaam: TUKI

Mdee, J. S. 1999. Dictionaries and the Standardization of Spelling in Swahili. *Lexikos* 9: 119-134.

Sacleux, C. 1939. *Dictionnaire swahili-français*. Paris: Travaux et mémoires de l'Institut d'ethnologie.
Zawawi, S, M. 1979. Loan words and their effect on the Classification of Swahili nominals. Leiden: Brill.

<div align="center">*****</div>

The Treatment of Dialectal Varieties in the Lumbu Lexicography

Hugues Steve Ndinga-Koumba-Binza (nkbinza@uwc.ac.za)
Centre for Advanced Studies of African Society, University of the Western Cape, South Africa
Maryse Edith Maloughou Mangama (ndongme@yahoo.fr)
Département d' Etudes Germaniques, Université Omar Bongo, Libreville, Gabon
Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo
Paul Achille Mavoungou (moudika2@yahoo.fr)
Département des Sciences du Langage, Université Omar Bongo, Libreville, Gabon
Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo
Ludwine Mbindi Aninga (sunchine9208@gmail.com)
Groupe de Recherche sur les Dynamiques Linguistiques et Lexicographiques, Institut de Recherche en Sciences Humaines, CENAREST, Libreville, Gabon
Centre de Recherche en Etudes Germaniques et Interculturelles, Université Omar Bongo

Lumbu or Yilumbu is a cross-border language spoken in Gabon and Congo (Brazzaville). In Gabon, although many Lumbu speakers live in the country's two main cities (Libreville and Port-Gentil), the majority of the Lumbu speakers reside in Nyanga Province where three dialectal varieties of the language are found. These varieties are known as Ghangu Lumbu (in Mayumba and its surroundings), Ossoye Lumbu (near Tchibanga towards Mayumba) and Mongo Lumbu (in Moulengui-Binza and its surroundings). In Ogooué-Maritime Province, where only one dialectal variety of the language is found, Lumbu speakers are found in the towns of Gamba and Sette-Cama and their surroundings. The variety found in Gamba and Sette-Cama is called Menaane Lumbu.

The literature review on Yilumbu discusses the linguistic research and lexicography of the Yilumbu language, particularly focusing on the treatment of dialectal variation in Lumbu lexicography. As far as the linguistic research activity is concerned, despite being a minority language, Yilumbu has been the subject of sustained linguistic research over the past three decades. Gamille (2013) is a major systematic description of the language, focusing on the Ghangu dialect spoken in Mayumba. Emejulu and Pambo Loueya (1990) provide a phonetic and phonological sketch of the Menaane variety. Regarding lexicographic outputs, Yilumbu has seen significant lexicographic outputs, including two doctoral theses, one Master's thesis, two dictionaries, and numerous research articles and book chapters. Mavoungou (2002) laid the foundation for metalexicographical criteria and introduced the grammar of the language. Mavoungou and Plumel (2010) produced the first Yilumbu-French bilingual dictionary, while Saphou-Bivigat (2010) initiated the encyclopaedic branch of Lumbu lexicography. Mbindi Aninga et al. (2023) proposed the second Lumbu dictionary, focusing on idiomatic expressions and including bilingual translation, definition, contexts of use, and encyclopaedic notes. As far as the treatment of dialectal variation is concerned, the lexicographic works mentioned above deal with dialectal variation by providing descriptions and dictionaries that cover different dialects and varieties of the Yilumbu language. While some works focus on specific dialects like Ghangu or Menaane, others aim to encompass a broader range of dialectal

variation within the language.

As part of the planning of the macrostructure of the Yilumbu-English-French dictionary, Mavoungou (2002) has dealt with all possible dialectal forms. Moreover, dialectal forms are the result of various phenomena, allophony and allomorphy in particular. Allophones will be understood here as the various members of the sounds which are grouped together to form phonemes. For example, the [nts] in ntsalë (hunger) and [ndz] in ndzoondzi (judge) form two allophones of the phoneme /ndz/. These are free variants because they are not determined by phonetic conditions. They rather occur in free variation in quite a number of utterances in Yilumbu. In terms of standardisation, decision has been taken to use the trigraph "ndz" to represent both the sounds [ndz] and [nts]. Nevertheless, where necessary, the user will still be provided with the form in "nts".

The aim of the current study is to examine and determine the situation of each dialectal variety within Lumbu lexicography. Hence, the research question underpinning this study reads as follows: How does lexicography research deal with dialectal issues? The subsequent questions should lead to determine whether there is a dialectal variety that is the most promoted above the others in Lumbu lexicography, and to determine the extent of negligence of the less-promoted dialectal varieties of Yilumbu.

Following a qualitative research methodology, this study adopts systematic literature review, text analysis, short questionnaires and semi-structured interviews as investigation methods. A thematic analysis and an interpretative method are used to analysed collected data. Preliminaries results of this analysis show that Lumbu lexicography, in general, follows a more inclusive approach in terms of the different dialects of the language with regard to dictionary production. However, lemmatisation processes and speech data collection for lexicographic corpora tend to favour a standardisation-motivated approach, which provides a platform for a one-dialect domination. Final recommendations of this study contribute to the standardisation and intellectualisation programmes of the Yilumbu language.

References
Emejulu, J.D. & F. Pambo Loueya. 1990. ◎ilumbu. *Revue Gabonaise des Sciences de l'Homme* 2: 197-201.
Gamille, L.G. 2013. Éléments de Description Phonologique et Morphologique du Lumbu, Langue Bantu (B44) du Gabon Parlée à Mayumba. Doctoral thesis. Paris: Université Sorbonne Nouvelle.
Mavoungou, P.A. 2002. Metalexicographical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu-English-French. Doctoral thesis. Stellenbosch: Stellenbosch University.
Mavoungou, P.A. & B. Plumel. 2010. *Dictionnaire Yilumbu-Français*. Libreville : Editions Raponda Walker.
Mbindi Aninga, L., Maloughou Mangama, M.E.N., Mavoungou, P.A. & H.S. Ndinga-Koumba-Binza. 2023. *Dictionnaire des Expressions Idiomatiques Lumbu*. Libreville : Editions Malugu.
Saphou-Bivigat, G. 2010. A Theoretical Model for an Encyclopaedic Dictionary for the Gabonese Languages with Reference to Yilumbu. Doctoral thesis. Stellenbosch: Stellenbosch University.

*****

Towards a Dictionary Production Strategy for African Languages: A Note from the Lexicography Agenda of CASAS

Hugues Steve NDINGA-KOUMBA-BINZA (nkbinza@uwc.ac.za)
Centre for Advanced Studies of African Society, University of the Western Cape, South Africa

Mantoa Rose Motinyane (mmotinyane@uwc.ac.za)
Centre for Advanced Studies of African Society, University of the Western Cape, South Africa

This paper is to initiate a discussion on a pan-African strategy for dictionary production. In certain countries, the production of dictionaries of African languages is part of a national language planning framework, but not in the vast majority of African countries. After consulting the pretexts of fifty dictionaries and a dozen of metalexicographic studies, the heartrending observation is that very few of these dictionaries and dictionary projects explicitly indicate to be motivated by the desire to intellectualize or to equip these languages with resources. It is therefore no surprise that these dictionaries end up simply as trophies on the shelves of authors and their sponsors.

Initiatives such as the creation of the African Academy of Languages (ACALAN) in 2001 raised a lot of hope as ACALAN immediately participated in a number of projects across Africa from the development of a linguistic atlas of Africa to the harmonization and standardization of the writing systems of vehicular cross-border languages (Bamgbose 2011). Despite some notable successes, it can still be said that ACALAN has considerably failed in its promises regarding the intellectualization of languages. Amongst the multiple reasons, the absence of a strategy or an agenda to produce dictionaries and reference works is noted.

Another notable initiative is the creation of national lexicography units (NLU's) within the Pan South African Language Board (PanSALB). Amongst successes of this initiative is undeniably the move from a Euro-centric to an Afro-centric approach to dictionary compilation for African languages (Prinsloo 2015). Nowadays, South African languages dictionaries are compiled by mother-tongue speakers through the NLU's (Kumalo 1999; Mongwe 2006; Alberts 2011). Nevertheless, some challenges and disappointments have been noted in the operation and achievements of the NLU's (Alberts 2022). To these we must add the insignificant impact in the educational system which continues to favor dictionaries produced by commercial publishers, whose strict quality control measures guarantee the good quality of dictionaries at the expense of user affordability. Likewise, the South African experience with the creation of PanSALB and the NLU's has not resonated and served as a model at the continental level as expected, owing to South Africa's political and economic leadership in Africa.

Taking cognizance of these facts, reflection is therefore necessary for the usefulness of currently published dictionaries and for a strategic dictionary production for African languages in the future. The Centre for Advanced Studies of African Society (CASAS) has chosen to initiate this discussion within the framework of this paper.

Over the years CASAS has been doing work that is critical for African languages. In terms of lexicographic work, CASAS has produced wordlists, dictionaries and translations of scientific texts from French, English and Portuguese into African languages. Monolingual dictionaries have been produced in several languages. The experience of CASAS, which is fully committed to the development, intellectualization and revitalization of African languages through a network of fellows in several African countries, allows us to observe the extent of the problem. What has been done so far is not enough and the advancement of new technologies such as those of artificial intelligence renders all this work even unusable. Hence the question: What strategy is needed for the production of usable dictionaries of African languages in the 3rd millennium?

Adopting the approach by Bowers & Creamer (2021), which consists of incorporating principles of grounded theory approach into a systematic review process, this paper outlines the lexicography agenda of CASAS as it suggests a strategic plan that can be adopted for all African languages. As Emejulu (2003:195) puts it: "Language is one of the most vital factors of self-realization for a people. Lexicography is a major linguistic cornerstone, and should not be neglected."

References
Alberts, M. 2011. National Lexicography Units: past, present, future. *Lexikos* 21: 23-52.
Alberts, M. 2022. The South African National Lexicography Units – two decades later. *Lexikos* 32(3): 1-24.
Bamgbose, A. 2011. African languages today: the challenge of and prospects for empowerment under globalization. E.G. Bokamba, R.K. Shosted & B.T. Ayalew (eds.). *Selected Proceedings of the 40th Annual Conference on African Linguistics*: 1-14. Somerville, MA: Cascadilla Proceedings Project.
Bowers, A.W. & E.G. Creamer. 2021. Core principles of grounded theory in a systematic review of environmental education for secondary students. *International Journal of Social Research Methodology* 24(6): 713-726.
Emejulu, J.D. 2003. Challenges and promises of a comprehensive lexicography in the developing world: the case of Gabon. W. Botha (ed.). *'n Man wat beur. Huldingingsbundel vir Dirk van Schalkwyk*: 195-212. Stellenbosch: Bureau of the WAT.
Kumalo, M.B. 1999. The National Lexicography Units – existing and prospective. *Lexikos* 9: 211-216.
Mongwe, M.J. 2006. The role of the South African National Lexicography Units in the Planning and Compilation of Multifunctional Bilingual Dictionaries. MPhil Thesis. Stellenbosch: Stellenbosch University.
Prinsloo, D.J. 2015. Analyzing words as a social enterprise: lexicography in Africa with specific reference to South Africa. J. Miller (ed.). *Analysing Words as a Social Enterprise: Celebrating 40 Years of the 1975 Helsinki Declaration on Lexicography*: 42-59. Adelaide, Australia/Auckland, New Zealand: AustraLex.

*****

Implementation progress of the provisions of the African Union Language Plan of Action in Zimbabwe: A lexicographer's perspective

Eventhough NDLOVU (evennthough@yahoo.co.uk)
Department of African Languages, University of the Free State, Bloemfontein, South Africa

The Language Plan of Action for Africa mandates each African Union Member State to ensure that universities, research institutes and other institutions concerned with the study and promotion of African languages have a unique role to play in strengthening the role these languages play in the daily lives of the African peoples. These institutions in the Member States are expected to strike a proper balance between the scientific study of the African languages and their actual use and practical promotion. Each Member State is mandated to render its national universities and other research and related institutions as primary instruments for the practical promotion of African languages, as regards such critical promotional activities, as the compilation of technical and general dictionaries, the training of teachers of language, translators, interpreters, broadcasters and journalists and the updating of vocabulary in African languages.

Against this background, this paper examines the implementation progress of these provisions in Zimbabwe focusing on the compilation of technical and general dictionaries as well as the updating of vocabulary in African languages. The study employed observations to see how many technical and general dictionaries have been compiled and in which language(s) and the role of government in their compilation. Observations were also carried

out to identify and ascertain terminology development works in African languages in the country and the role of government in these initiatives. This was corroborated using semi-structured interviews with language lecturers and researchers in universities and research institutes. Findings of this study show that not much has been done in Zimbabwe in terms of the compilation of technical and general dictionaries and the updating of vocabulary in African languages. It emerged that technical and general dictionaries only exist in Shona and Ndebele. It was observed that there are only two Shona technical dictionaries compiled by the former African Languages Research Institute (ALRI), namely *Duramazwi ReUtano Neurapi* (a Shona medical terms dictionary) and *Durazwi Remimanzi* (a Shona musical terms dictionary). There are two monolingual general-purpose dictionaries, *Duramazwi ReChiShona* and *Duramazwi Guru ReChiShona*. There is only one technical dictionary in isiNdebele, *Isichazamazwi SezoMculo*, a Ndebele musical terms dictionary and only one monolingual general-purpose dictionary, *Isichazamazwi SesiNdebele*. The number of both dictionary types and the languages in which these lexicographic activities take place is disheartening in a multilingual country which in 2013 accorded 16 languages, officially recognised language status. It was also observed that there are no dedicated and coordinated terminology development bodies. Terminology development is haphazard and largely uncoordinated. The recent closure of ALRI in the face of the binding and forceful constitutional language provisions reflects the country's reluctance to prioritise language issues in its development trajectory and failure to adhere to the provisions of the Language Plan of Action for Africa. This is worrisome given that the country adopted the 2013 Constitution which has dedicated Sections 6 and 63 on languages, which in my view should have created an impetus to finance lexicographic and terminology development activities. It is also a serious cause for concern that such lack of meaningful progress is happening in a country which hosted the 1997 Intergovernmental Conference of Ministers on Language Policy in Africa, which culminated in the adoption of the Harare Declaration. The Harare Declaration further affirmed the African Union Member States' commitment to the resolutions of the Organisation of African Unity Language Plan of Action for Africa.

In light of these findings, the study makes a case for adequate and sustained allocation of the necessary financial and material resources needed to support the compilation of technical and general dictionaries as well as the updating of vocabulary in African languages. The study recommends the adoption of a principle of financial allocation in the National Budget for the compilation of technical and general dictionaries as well as the updating of vocabulary in African languages. There is need for a financial provision in the form of revenue allocation formula, which will be linked to the National Budget. There must be a certain percentage of the National Budget reserved for the compilation of technical and general dictionaries as well as the updating of vocabulary in African languages. There is need for dedicated and well-resourced institutional support structures to support lexicographic and terminography initiatives which will spearhead the compilation of technical and general dictionaries and updating of vocabulary in all the officially recognised languages. There is need for human capital development and professional skills upgrading of lexicographers and terminographers to align with the global best practices and trends in these disciples.

References
Organisation of African Unity/African Union. 1986/2006. Language Plan of Action for Africa. http://www.bisharat.net/Documents/OAU-LPA-86.htm [17 January 2024].

Revisiting and re-imagining Kropf's *A Kafir-English Dictionary* 125 years on

Dion NKOMO (d.nkomo@ru.ac.za)
School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa
Sanele NTSHINGANA (sanele.ntshingana@uct.ac.za)
School of Languages and Literatures: African Languages, University of Cape Town, Cape Town, South Africa

The publication of Rev. Johann Heinrich Albert Kropf's *A Kafir English Dictionary* a century and a quarter ago was a landmark in the history of isiXhosa. The lexicographer was persuaded to publish the dictionary after compiling an isiXhosa vocabulary list for personal use, having struggled to get a comprehensive language learning reference upon his arrival in the Eastern Cape where he was posted to serve as a missionary in 1845. Pahl (1989:xxxviii) labelled the dictionary a "masterly and scholarly work" and regarded Kropf as "the first lexicographer to have fathomed and thoroughly mastered the intricacies" of the language. Pahl (1989:xxxviii) went further to recommend the dictionary, advising that "all future lexicographical work must be based on it." Indeed, the Lovedale Mission Press assigned Rev. Robert Godfrey to produce a revised and enlarged edition of Kropf's dictionary, publishing it in 1915. In spite of his historic accomplishment, Kropf had already acknowledged some omissions which resulted in the dictionary being published with an appendix of words that were collected after the commencement of printing in 1895 and could not be included and properly treated in its main text. Although a manuscript of the third edition of the dictionary was finalised by Godfrey by 1942, it was never published due to some contestations around the revised orthography that Godfrey had adopted. Insights into the third edition are provided by Opland (2019) in an edited volume on Godfrey's contributions to isiXhosa lexicography. Opland's (2019) volume includes extracts from the second edition, published in 1915 and the unpublished third edition, as well as Godfrey's reflective notes on his lexicographic methods. The second edition of Kropf's dictionary remains one of the major references for isiXhosa translators and other language scholars, alongside the *Greater Dictionary of (isi)Xhosa*. In fact, the strong influence of *A Kafir English Dictionary* manifests itself in the *Greater Dictionary of (isi)Xhosa* through abbreviations 'Kr' and 'Kr-G' which acknowledge the first and second editions by Kropf and Godfrey respectively as sources of material used in the latter dictionary.

This presentation seeks to revisit Kropf's dictionary and re-evaluate its special place in isiXhosa lexicography, language and scholarship. The dictionary has been a subject of critical scrutiny, with Kropf's critics pointing out serious short-comings such as mistranslation of culturally specific items, errors, and omissions (Moropa and Kruger 2000, Nkomo 2020). However, no comprehensive study of the dictionary has ever been published apart from Moropa and Kuger's (2000) analysis of cultural mistranslations. This presentation demonstrates that the problems pointed out by these critiques also persist beyond the cultural domain. For example, the concept *umthetho*, which can be argued to exist in the political authority domain, is derived from the verb *-thetha* "to speak". Words such as *isithethi* [a speaker], *isithethe* [tradition] carry the sense of speaking or discoursing. Kropf and Godfrey define *umthetho* as a "commandment, law, and ordinance" (Kropf and Godfrey, 1915: 409). The sense of "speaking" or discourse is omitted from this definition and a new epistemology with respect to law and political authority is introduced with many implications in terms of

understanding of this concept. In this contribution, we will analyse how Kropf and later on Godfrey handled terms that fall into the political authority domain such as *umbuso*, *umthetho*, *iphakathi*, *ikomkhulu*, and *inkosi* amongst others and imagine how these terms would be handled outside of the limits of these early missionary lexicographers. Our contribution does not only take into account the political, cultural and hence epistemic context in which Kropf compiled the dictionary, but it also considers the methods that he used to collect, process and compile his lexicographic data, noting in them an interesting interface between vernacular lexicography and professional lexicography. How Godfrey built on Kropf's legacy and how that legacy lives on in the *Greater Dictionary of (isi)Xhosa* and *Isichazi-magama SesiXhosa*, the first monolingual dictionary in the language, receives due attention. Accordingly, Kropf's dictionary is not only revisited, but also re-imagined in order to highlight how modern isiXhosa lexicographers and dictionary users can continue to use this minefield but with the benefits of hindsight.

References
Kropf, A. 1899. *A Kafir–English Dictionary*. Alice: Lovedale Press.
Kropf, A. and Godfrey, R. 1915. *A Kafir–English Dictionary*. 2nd Ed.. Alice: Lovedale Press.
Moropa, K., and Kruger, A. 2000. Mistranslation of culture-specific terms in Kropf's *Kafir–English Dictionary*. *South African Journal of African Languages* 20(1): 70–79.
Nkomo, D. (2020). Vernacular lexicography in African languages: From early days to the digital age. Dictionaries: Journal of the Dictionary Society of North America, 41(2): 213-243.
Opland, J. 2019. Lexicography: Notes on Xhosa Lore and Language (1909–1934) – Robert Godfrey. Durban: University of KwaZulu-Natal Press.
Pahl, W.H. et al. 1989–2006. *The Greater Dictionary of (isi)Xhosa*. Alice: University of Fort Hare.
Tshabe, S. et al. 2008. *Isichazi-magama SesiXhosa*. Pietermaritzburg: Nutrend Publishers.

<div align="center">*****</div>

Knowledge on the Margins: The Treatment of Linguistic Variation in isiXhosa Dictionaries

Dion NKOMO (d.nkomo@ru.ac.za)
School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa
Thandile NTUMBA (thandile.ntumba@gmail.com)
School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa

The recognition of nine historically marginalized African languages as official languages alongside English and Afrikaans in the 1996 Constitution of the Republic of South Africa can been hailed as the boldest language planning decision in recognition of nation's linguistic diversity and the linguistic rights of the majority of South Africans. One of the official languages, isiXhosa, which is spoken by 16,3% of the population according to the 2022 census results (Census 2022), is characterized by a complex internal variation. However, controversy remains around the classification of isiXhosa and the number of varieties that constitute the language (see for example, Nyamende 1994). Nyamende (1994:203) argues that some of the dialects "could have been regarded as independent languages of the Nguni cluster, but, perhaps due to the missionary influence which now carried the Ngqika, Ndlambe and Thembu variants, the converted speakers of the aforementioned variants were then subjected to the use of standard Xhosa at the mission stations and seminary schools." Over two centuries since isiXhosa was written for the first time, there persists a sense of injustice among speakers

of some regional varieties about the standardization of isiXhosa which subsumed the varieties albeit placing them on the peripheries of one of the widely spoken South African official languages. For example, Maqam (2014) posited that mother-tongue education that intervened by replacing English with isiXhosa was not effective among isiMpondo speakers because isiXhosa is a language that the learners only encounter at schools.

Considering the critical role played by dictionaries in the standardization of language, this presentation considers the lexicographical treatment of regional variation in isiXhosa. Three major isiXhosa dictionaries, namely, *A Kafir English Dictionary*, *The Greater Dictionary of (isi)Xhosa* and *Isichazi-magama SesiXhosa* are studied. *A Kafir English Dictionary* is a bilingual dictionary that was compiled by Rev. Albert Kropf and first published in 1899. It was revised and published as a second edition in 1915 under the editorship of Rev. Robert Godfrey. Kropf is hailed as "the first lexicographer to have fathomed and thoroughly mastered the intricacies" of the language (Pahl 1989: xxxviii). *The Greater Dictionary of (isi)Xhosa* is a trilingual and tri-volume dictionary that culminated from the Fort Hare Xhosa Dictionary Project which commenced in 1968 and ended with the publication of Volume 1 in 2006, having started with Volume 3 in 1989, followed by Volume 2 in 1991. *Isichazi-magama SesiXhosa*, the first monolingual dictionary in the language, was then published in 2006 under the auspices of the isiXhosa National Lexicography Unit. The dictionaries were studied by reading front matter texts (forewords/introductions) that outline the lexicographers' stance regarding regional varieties, by studying abbreviations used in the dictionaries, some of which highlight the representation of data on regional varieties, as well as dictionary articles that include entries related to regional variation.

Kropf's *A Kafir English Dictionary* asserted the stance that sought to standardize isiXhosa around the Tshiwo Xhosa (isiRharhabe and isiGcaleka) which had served as the literary standard until the 1950s. However, *The Greater Dictionary of (isi)Xhosa* set out to become an up-to-date, scientific and comprehensive work that would not confine itself to Tshiwo Xhosa by accommodating the varieties of the abaMbo, abaThembu, amaBhaca, amaHlubi, amaBomvana, amaMfengu and amaMpondomise. However, Tshiwo Xhosa lexical items are the ones that are entered as lemmata, while the other varieties are presented in brackets towards the end of dictionary articles. This approach is carried over onto *Isichazi-magama SesiXhosa*. While the inclusion of regional varieties in the latter two dictionaries enables users to access rich vocabulary that was excluded in Kropf's dictionary, it is noted that the varieties are presented and treated marginally. Their vocabularies are not treated as the main lemmata and cannot be accessed independent of the standard Tshiwo variety. At best, the treatment of regional variation in isiXhosa lexicography therefore exhibits a proscriptive approach whereby included varieties are marked and not recognized in their own right. For users who are speakers of the marginalized varieties, particularly those like isiMpondo which are significantly different from the standard language, dictionaries are another site of exclusion, just like the education system which compels the learners to leave their spoken forms outside the school. Perhaps this demonstrates the need for dictionaries of regional varieties in isiXhosa in order to complement the standard language dictionaries.

References
Kropf, A. 1899. *A Kafir–English Dictionary*. Alice: Lovedale Press.
Kropf, A. and Godfrey, R. 1915. *A Kafir–English Dictionary*. 2nd Ed. Alice: Lovedale Press.
Census 2022. https://census.statssa.gov.za/#/ [28 April 2024].
Pahl, W.H. et al. 1989–2006. *The Greater Dictionary of (Isi)Xhosa*. Alice: University of Fort Hare.

Tshabe, S. et al. 2008. *Isichazi-magama SesiXhosa*. Pietermaritzburg: Nutrend Publishers.
Nyamende, A. Regional variation in Xhosa. *Stellenbosch Papers in Linguistics Plus* 26: 202-217.
*****

Spelling variations in *A dictionary of Setswana personal names*

Thapelo J. OTLOGETSWE (otlogets@ub.ac.bw)
Faculty of Humanities, University of Botswana, Gaborone Botswana
Goabilwe N. RAMAEBA (ramaebagn@ub.ac.bw)
Department of English, University of Botswana, Gaborone, Botswana

This paper looks at spelling variations in *A dictionary of Setswana personal names* (Otlogetswe & Ramaeba 2024). Setswana spelling variations have been studied before in monolingual dictionaries (Otlogetswe 2015) but never in Setswana personal names. Spelling rules which guide Setswana orthography are outlined in *Standard Orthography of 1981* (Ministry of Education 1981) and *Mokwalo o o lolameng wa Setswana* (Chebanne et al. 2008) though the historical roots of Setswana orthography may be traced to Lichtenstein (1928-1930 and 1973) and Livingstone (1875). This study uses a textual analysis method to critically assess the dictionary headwords for spelling variants. Preliminary results of our study show that there are numerous inconsistencies in the spelling of Setswana personal names. The inconsistencies fall into different categories.

Many Setswana names still use some elements of old Setswana orthography. In the old orthography the voiceless postalveolar affricate [ʧ] was rendered as "ch" while in recent orthography it is spelt as "tšh". This has led to some Setswana names still being spelt with "ch" as in the examples, *Chaba* old spelling of *Tšhaba*; (nation), *Chabadiile* old spelling of *Tšhabadiile*; (nations have gone) and *Chabaditsile* old spelling of *Tšhabaditsile*; (nations have arrived). For many years Setswana orthography spelt the voiceless postalveolar fricative as [sh] until it was replaced by [š] in the 1980s. There are, therefore, many names that reflect this old orthographical practice. These names include *Bashi* instead of *Baši*, *Boshafatso* instead of *Bošafatso* and *Keshupile* instead of *Kešupile*.

Northern Setswana has historically lost the voiceless alveolar lateral unaspirated affricate [tɬ], /tl/ and voiceless alveolar lateral aspirated affricate [tɬʰ], /tlh/ consonantal sounds. In their places the voiceless alveolar unaspirated plosive /t/ and the voiceless alveolar aspirated plosive /tʰ/ are used. This historical loss has had an enduring influence on how speakers of the Northern Setswana dialect spell personal names. Some of their spelling conventions approximate their pronunciation and not that of Setswana standard orthography. Therefore, a name such as *Tlhalefang* is spelt *Thalefang* and *Atlholang* is spelt as *Atholang*. The Northern Setswana name spellings are captured in the dictionary as they are so spelt by their owners. This is in line with the recommendations of *Standard Setswana Orthography of 1981* which notes that, "As a general rule, proper names of people should be spelled as used by the persons themselves" (p. 19). The dictionary, however, gives the standard spelling at each entry of Northern Setswana, for the benefit of the user.

Other names do not follow standard orthography where the vowels are concerned and use [i] instead of [e], as in *Baaitsi* instead of *Baaitse*, *Matshidiso* instead of *Matshediso*, *Balibi* instead of *Balebi*. This is probably because the name givers gave it by ear and in their hearing, the [e] sounded like an [i]. In other cases, the situation is reversed, name givers write [e]instead of [i], as in, *Gabaiphewe* instead of *Gabaiphiwe* or *Molapise* instead of *Molapisi*.

In other cases, instead of [o] the name givers use [u]. This can also be attributed to the sound of [o] which is heard by name givers as [u]. Speakers, therefore, spell *Babotsi* as *Babutsi*,

Moalosi as *Moalusi* and *Monnosi* as *Monnusi*. Some, instead of using the voiceless, labio-dental fricative [f] use the glottal fricative [h]. Using the sound [h] instead of [f] is a common feature of Setswana speech and there is much confusion in written Setswana where [h] and [f] should be used. There is, however, a Setswana morphological rule which illustrates that [h] verbs give [kh] nouns and [f] verbs give [ph] nouns in strengthening. For instance, the verb *hupela* generates the noun *khupelo* while the verb *fenya* generates the noun *phenyo*. The [h] and [f] spelling confusion is pervasive in names. For example, speakers write *Bahedile* instead of *Bafedile*, *Bahemedi* instead of *Bafemedi* and *Gaehole* instead of *Gaefole*. In many old Setswana names, instead of choosing either [h] or [f], name givers have used both consonants spelt [fh]. There are therefore many names which use [fh] instead of [f], such as *Gadifhele* instead of *Gadifele*, *Kefhilwe* instead of *Kefilwe*, *Molefhe* instead of *Molefe*, and *Molefhi* instead of *Molefi*.

This paper discusses spelling variations in Setswana personal names as observed in Otlogetswe and Ramaeba (2024) to highlight the varying spellings for the same name. These variations are a result of several factors such as the influence of the standardised old and new Setswana orthography, various Setswana dialects and some morphological and phonological processes.

References

Chebanne A., Otlogetswe, T.J., Seloma, P., Molosiwa A., Moatlhodi B., Kgolo, N. & David, M. 2008. *Mokwalo o o lolameng wa Setswana.* Gaborone: Pentagon Publishers.

Lichtenstein, H. 1928-30. *Travels in Southern Africa in the years 1803, 1804, 1805 and 1806* (a reprint of the translation from the original German by Anne Plumptre). Translated by Plumptre, A. Cape Town: The Van Riebeeck Society.

Lichtenstein, H. 1973. *Foundation of the Cape (1811) and About the Bechuanas (1807).* O.H Spohr, (ed. and trans.) Cape Town: A. A. Balkema.

Livingstone, D. 1875-85. Missionary Travels and Researches in South Africa. London.

Ministry of Education. 1981. *Setswana Standard Orthography of 1981.* Gaborone: Ministry of Education.

Otlogetswe, T.J., Ramaeba, G.N. 2024. *A dictionary of Setswana personal names.* Gaborone: Mokala Publishing.

Otlogetswe, T.J. 2015. Treatment of spelling variants in Setswana monolingual dictionaries. *Lexikos* 25: 262-284.

<center>*****</center>

Comparison of selected Setswana monolingual dictionaries in their coverage of multiword expressions

Keabetswe PHETOANE (phetoanekt@tut.ac.za)

Registrar's Office: Secretariat and Legal Services, Tshwane University of Technology, Pretoria, South Africa

Traditional dictionaries primarily focus on individual words and their definitions. However, language is full of meaningful units that go beyond single words, i.e. multiword expressions (MWEs). Identifying and including MWEs in dictionary entries is becoming increasingly important because they are abundant in natural language. They can range from simple verb-particle combinations "kick the bucket" to complex idioms "spill the beans" and phrasal verbs "make up" (Moon 1998). These expressions often have a non-compositional meaning, i.e. their meaning cannot be simply derived from the sum of their individual words.

Jackson (2003) suggests three ways to define a "word", namely the written word (orthographic word) that refers to a sequence of letters with spaces around them, like "cat" or "run"; spoken word (phonological word) which focuses on the sounds that make up a word

in speech. Rules like stress and syllable structure determine the boundaries of spoken words (e.g. "water" vs. "wa-ter"); and lexeme, a word in the vocabulary of a language. A lexeme is what is usually captured as a dictionary entry or as a "headword". A lexeme can be made up of several written words, like "warble fly," "war chest," or "ward of court".

African languages dictionaries, often fail to capture the richness and complexity of MWEs. They might list individual words within an MWE, but they don't provide information about their combined meaning, usage patterns, or idiomaticity (Otlogetswe 2009). This can lead to confusion and difficulty for dictionary users, especially learners of a new language since dictionaries play a crucial role in understanding meaning.

The paper argues that focusing solely on individual words does not provide the full picture since MWEs often hold specific meanings within particular contexts. This paper focuses on Setswana monolingual paper dictionaries, namely *Thanodi ya Setswana* (Kgasa and Tsonope 1998), *Thanodi ya Setswana* (Mareme 2008), and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012b). Extracting and including MWEs in dictionaries helps users understand how MWEs are used in real-world situations. Dictionaries are no longer just for finding definitions or translation equivalents; learners, translators, and professionals rely on them for understanding nuances of language and appropriate usage of larger structures like MWEs. Including MWEs therefore provides comprehensive information about how language actually functions. By incorporating MWEs, dictionaries can become more informative, user-friendly, and reflective of the way language is used in real-world communication.

The paper focuses on how the identification and inclusion of MWEs can enrich dictionary content, enhance meaning clarity and capture figurative language, thus improving dictionary usability. This will be achieved by comparing selected Setswana dictionaries in their coverage of multiword expressions. Otlogetswe (2012a) indicates that since a dictionary is often seen as a list of simple words, a dictionary may be enriched by including idioms, multiword units and collocates.

In conclusion, the paper suggests that the quality and comprehensiveness of dictionaries can be improved by incorporating MWEs, which are essential units of meaning in natural languages. By looking at the words that a particular word frequently appears with, i.e. collocates, we can discover different types of multiword expressions (MWEs) like proverbs, idioms, and phrasal verbs. These expressions can then be added to dictionaries as separate entries. The selected Setswana monolingual dictionaries will be analysed and compared to see how well they cover multiword expressions. The comparison may identify gaps in the coverage of MWEs in existing Setswana dictionaries and further provide valuable insights into the current state of MWE lexicography in Setswana. This information could be used to inform the development of new dictionaries or the revision of existing ones. By including MWEs with clear explanations and usage examples, dictionaries become more informative and user-friendly for learners, translators and other professionals.

References

Jackson, H. 2003. *Lexicography: An introduction*. USA and Canada: Routledge.

Moon, R. 1998. Fixed Expressions and Idioms in English: A Corpus-Based Approach. Clarendon Press Oxford Inc: New York.

Kgasa, M.L.A. and Tsonope, J. 1998. *Thanodi ya Setswana*. Gaborone: Longman.

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

Otlogetswe, T.J. 2009. Populating Sub-entries in Dictionaries with Multi-word Units from Concordance Lines. *Lexikos* 19: 446-45.

Otlogetswe, T. 2012a. Enriching a Dictionary Database with Multiword Expressions. H.S. Ndinga-Koumba-Binza and S. Bosch (eds.). *Language Science and Language Technology in*

Africa: A Festschrift for JC Roux: 215 – 236. SUN MeDIA Stellenbosch.
Otlogetswe, T.J. 2012b. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.

<div align="center">*****</div>

Indonesianization: Standardization and Codification of Terms in Indonesian

Dewi PUSPITA (dewi.puspita18@gmail.com)
Language Development and Cultivation Agency, Jakarta, Indonesia
Dira HILDAYANI (dirahildayani@gmail.com)
Language Development and Cultivation Agency, Jakarta, Indonesia

Being a developing nation, Indonesia persists in its efforts to advance science and technology to avoid falling behind other nations. In this endeavour, Indonesians continue to study various fields of science and specialties from foreign literature, which contains many technical terms. In order to disseminate the new knowledge to students and the wider community, foreign literature, especially the technical terms, needs to be translated.

There are certain procedures that must be followed when translating foreign terms into Indonesian. This process is not merely translation, but rather term formation or term equalization, as translation is just one of the processes. Although the formation of terms can be done by anyone, if it is not in accordance with established standards, it cannot be called a standard term and cannot be used in formal writing. The resulting terms should therefore be standardized and codified nationally.

The standardization and codification of terms in Indonesia began with the cooperation of three neighbouring countries: Brunei Darussalam, Indonesia, and Malaysia in an assembly called MABBIM (Majelis Bahasa Brunei Darussalam-Indonesia-Malaysia, "Language Council of Brunei Darussalam-Indonesia-Malaysia"). Initially, the three countries agreed to standardize foreign terms in such a way that they could be shared across the three countries. However, over the years, language policies and language usage in the three countries have diverged. Due to this circumstance, the three countries opted to terminate collaboration in the area of equivalencing foreign technical terms.

After MABBIM, the standardization and codification of term formation or equalization in Indonesia are regulated in the Regulation of the Indonesian Minister of Education and Culture. The ministerial regulation states that standardization and codification are conducted by the Language Development and Cultivation Agency. The Language Development and Cultivation Agency has compiled guidelines for this terminology work, entitled *Pedoman Umum Pembentukan Istilah* (General Guidelines for the Formation of Terms, Sugono 2006). The guidelines were developed with reference to ISO TC/37 (1969), which provides standardization of terminology work. The guidelines state that equalizing foreign terms, also called the Indonesianization of foreign terms, is done through translation, absorption, or a combination of translation and absorption. Equalization is done by experts in the field together with linguists (which are mostly lexicographers of the Language Agency) so that the concept contained by the term remains the same and in accordance with Indonesian language rules. Foreign terms that have been Indonesianized are then published in a special dictionary as well as an application that can be accessed by the wider community. Technical terms resulting from the Indonesianization process that are already widely used by Indonesian speakers will be incorporated into the Indonesian Comprehensive Dictionary, *Kamus Besar Bahasa Indonesia*.

Problems arise in every process, both in the process of Indonesianization and after the new terms are published. Problems that arise include difficulties in finding appropriate

equivalents, concepts that do not exist in Indonesia, and public rejection of foreign terms that have been Indonesianized. These issues will be further discussed in the paper along with other emerging issues. This paper will show how the process of Indonesianizing foreign technical terms is conducted: where technical terms are derived from, what the Indonesianization procedure is, and how the resulting new terms are stored and published. This paper will also discuss how terminology work in Indonesia differs from other countries, both those with good terminology management and those without, and to compare the process of how English technical terms are translated into other languages as described by Drozd and Roudny (1980) in the case of old Czechoslovakia, Moghadam and Far (2015) in the case of Iran, and Polyakova et al (2019) in the case of Russia. The results of this comparison will be immensely helpful for the improvement and development of the process of standardization and codification of terms in Indonesia.

References

Drozd, L. and Roudny, M. 1980. Language Planning and Standardization of Terminology in Czechoslovakia. *International Journal of the Sociology of Language* 23: 29-41.

International Organization for Standardization. 1969. Language and Terminology (ISO/TC 37). https://www.iso.org/committee/48104.html

Moghadam, M.Z. and Far, M.D. 2015. Translation of Technical Terms: A Case of Law Terms. *Journal of Language Teaching and Research* 6(4): 830-835.

Polyakova, L.S. et al. 2019. Peculiarities of translation of English technical terms. *IOP Conference Series: Materials Science and Engineering* 483(2019): 012085. doi:10.1088/1757-899X/483/1/012085.

Regulation of the Indonesian Minister of Education and Culture Number 18 Year 2021 concerning Standardization and Codification of Language Rules.

Sugono, D. (ed.). 2006. *Pedoman Umum Pembentukan Istilah* (General Guidelines for the Formation of Terms). Jakarta: Pusat Bahasa, Departemen Pendidikan Nasional.

<div align="center">*****</div>

Developing a Dictionary as a Resource for Language Acquisition: The Case of isiXhosa Dictionary of collocations

Mlondolozi PUTE (mpute@uwc.ac.za)
Department of African Language Studies, University of the Western Cape
Sebolelo MOKAPELA (smokapela@uwc.ac.za)
Department of African Language Studies, University of the Western Cape
Hugues Steve NDINGA-KOUMBA-BINZA (nkbinza@uwc.ac.za)
Centre for Advanced Studies of African Society (CASAS), University of the Western Cape, South Africa

Many scholars such as Robins (1967), Namvar et al. (2012) and Baker (2016) have acknowledged the importance of the meaning relationships that exist between vocabulary items and have used the concept of collocation to study it. Collocations are word combinations that express a particular meaning, whose co-occurrence in communication is greater than chance (Lewis 1997). Collocations are very significant in language. For instance, Hill (2000) contends that collocations make up 70 per cent of the English language that is used daily. According to Nation (2001:318), "all fluent and appropriate language use requires collocational knowledge". However, it is common knowledge that general-purpose dictionaries "provide only some grammatical collocations and a very few lexical collocations" (Benson 1990:32). This sparked an increased attention of lexicographers and dictionary makers, especially in the African context (Gouws 2015; Taljard & Prinsloo 2023; Mbindi

Aninga et al. 2023).

Dictionaries have been playing a significant role in language acquisition since the 8th century (Mustafai 2018). They try to provide solutions to language-related challenges that their target users come across daily. Following the importance that the subject of collocations has gained in linguistics, there has been an emergence of collocation dictionaries. Of course, collocation dictionaries are important because they offer collocational knowledge in a very direct manner. Unfortunately, there seems to be a limited number of collocation dictionaries (especially bilingual collocation dictionaries) available compared to the need for them. A large number of the available collocations' dictionaries are written in the English language (Galal 2015). It is known that there has not been a single collocation dictionary published in isiXhosa.

In light of the benefits of collocation dictionaries and their possible benefits, the following question has to be asked: How can an isiXhosa dictionary of collocations contribute to isiXhosa learning and enhance proficiency? The current paper comes within the project to develop a metalexicographic model for an isiXhosa dictionary of collocations. Within the current study, it is hypothesised that the use of the proposed isiXhosa dictionary of collocations will enhance the understanding and production of isiXhosa texts for those that are learning isiXhosa as a second language and thus achieve collocational competence as well as some fluency in speaking (and writing) isiXhosa, and also help isiXhosa home language speakers improve their language abilities.

The current paper will review the challenges with achieving collocational competence in isiXhosa and fluency in speaking and writing the language, as well as reviewing literature on the term collocation and classifications that the dictionary will focus on such as lexical, semantic and/or syntactic. In terms of purpose, this paper mainly intends to provide an account of a research project which is concerned with the compilation of a collocation dictionary of isiXhosa. For this purpose, the study will adopt the methodological principles of the systematic literature review within a qualitative approach.

Subsequently, the paper will outline the conceptualisation plan for the intended dictionary in view of gathering inputs from experts. The results of the projected study include among others to build a corpus of texts in isiXhosa and to compile a learner's dictionary of isiXhosa collocations. These results should contribute to the process of developing language resources for isiXhosa acquisition.

References

Baker, P. 2016. The shapes of collocation. *International Journal of Corpus Linguistics* 21(2): 139-164.

Benson, M. 1990. Collocations and general-purpose dictionaries? *International Journal of Lexicography* 3(1): 23-34.

Galal, M. 2015. Lost in collocation: when Arabic collocation dictionaries lack collocations. *English Language and Literature Studies* 5(2): 18-38.

Gouws, R.H. 2015. The presentation and treatment of collocations as secondary guiding elements in dictionaries. *Lexikos* 25: 170-190.

Hill, J. 2000. Revising priorities: From grammatical failure to collocational success. M. Lewis (ed.). *Teaching Collocation: Further Development in the Lexical Approach*: 47-69. Oxford: Oxford University Press.

Lewis, M. 1997. Implementing the Lexical Approach: Putting Theory into Practice. Hove: Language Teaching Publications.

Mbindi Aninga, L., Maloughou Mangama, M.E.N, Mavoungou, P.A. & H.S. Ndinga-Koumba-Binza. 2023. *Dictionnaire des Expressions Idiomatiques Lumbu*. Libreville: Editions Malugu.

Mustafai, J. 2018. Types of Dictionaries. *Knowledge-International Journal* 22(6): 1661-1668.

Namvar, F., Mohdnur, F., Ibrahim, N. & J. Mustafa. 2012. Analysis of collocations in the

Iranian postgraduate students' writings. *The Southeast Asian Journal of English Language Studies* 18(1): 11-22.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Robins, R. 1967. *A Short History of Linguistics*. London: Longman.

Schmitt, N. (ed.). 2004. *Formulaic Sequences*. Amsterdam: John Benjamins Publishing Company.

Taljard, E. & D.J. Prinsloo. 2023. Collocations and African language lexicography. S.E. Bosch, L. Morris & A.H. du Plessis (eds.). *African Association for Lexicography. 27th International Conference, 26-29 September 2023*: 29-30. Bloemfontein: AFRILEX.

Wiktorsson, M. 2003. Learning Idiomaticity: A Corpus-Based Study of Idiomatic Expressions in Learners' Written Production. Doctoral Thesis. Lund, Sweden: Lund University.

*****

Corpus-based dictionaries for low-resource languages

Mmasibidi SETAKA (mmasibidi.setaka@nwu.ac.za)

South African Centre for Digital Language Resources, Potchefstroom, South Africa

Menno VAN ZAANEN (menno.vanzaanen@nwu.ac.za)

South African Centre for Digital Language Resources, Potchefstroom, South Africa

Dictionaries are helpful resources that, among other things, document, standardize, and preserve languages while also providing them significance and authority. They can also help learners become more proficient readers and writers. However, lexicographic resources are often either absent or scarce in many low-resource languages.

Before the availability of and access to large amounts of language samples like corpora, lexicographers used to rely on intuition to choose lemmas, senses, meanings, etc., and the chances of a dictionary compiler gathering all the essential information without the help of corpora and just intuition were questionable (Williams 2003; Gouws and Prinsloo 2005). As such, the arrival of computers had an effect on how dictionaries would ultimately be compiled. As a result, large language corpora, stored and analyzed on a computer, became more popular and greatly impacted the dictionary making processes as they provide valuable information to the lexicographer, who needs to access a wide range of linguistic and cultural knowledge. Consequently, Krishnamurthy and Nicolas (2000) argue that corpora have become the preferred source of evidence in the creation of dictionaries.

Current trends in dictionary compilation indicate that corpus-based approaches are highly suitable for 'big' languages such as English, French, Dutch, etc. that have large amounts of digital language resources available. Also, it cannot be denied that dictionaries of the future will rely heavily on electronic corpus-based studies to assist in dictionary making. Oxford dictionaries (2024) argue that "It seems likely that by the middle of this century, if not before, all dictionaries will be in electronic form." This means that the high resource languages are expected to reach that level. This paper will argue that though such 'big' languages greatly benefit from using corpora, it is unclear how suitable corpus-based approaches are for the compilation of dictionaries for 'small' languages that are low-resourced and have little or no corpora. Preliminary results indicate that creating a dictionary for a low-resource language without corpora is difficult (perhaps like in the past) because lexicographers do not have sufficient examples of the use of the language to allow them to make conclusions regarding this language. Motjope-Mokhali et al. (2020) argue that many words that are currently in use do not appear in these dictionaries, resulting in a limited vocabulary and the presence of several obsolete words.

This paper will examine traditional versus corpus-based approaches, differences between

high and low-resource languages, and measures of corpus-based approaches through experiments on aspects of correctness, completeness, usability, and time-efficiency.

While the powerful nature of corpus-based approaches cannot be denied, it is equally important to realize that low-resourced languages also need to be brought into the picture to form part of this way of compiling dictionaries.

This research will give us an understanding of the extent to which corpus-based approaches can be used to create, improve quality of resources and, most importantly, advance low-resource languages.

In conclusion, this paper is part of a larger project that is currently in progress. The aim of the project is to get a better understanding of how dictionaries of low-resource languages can be compiled with the aim to improve the process with respect to the use of modern-day corpora and related techniques. The results obtained will indicate the core strengths and weaknesses (or possibilities and impossibilities) of using corpora for low-resource languages. This will provide more insight on how suitable this approach is in such cases. It may be the case that there are situations where a non-corpus-based approach is more suitable, but that intermediate approaches that combine corpus and non-corpus-based approaches can be considered as well.

References
Gouws, R.H. & Prinsloo, D.J. 2005. *Principles and practice of South African lexicography*. Stellenbosch: AFRICAN SUN MeDIA.
Krishnamurthy, R. and Nicholls, D. 2000. Peeling an onion: The lexicographer's experience of manual sense-tagging. *Computers and the Humanities* 34(1-2): 85–97.8
Motjope-Mokhali, T., Kosch, I., and Mafela, M. J. 2020. Sethantso sa sesotho and Sesuto–English dictionary: A comparative analysis of their designs and entries. *Lexikos* 30: 258–274.
Oxford Dictionaries. 2024. How have dictionaries changed over the years? https://premium.oxforddictionaries.com/words/how-have-dictionaries-changed-over-the-years#:~:text=In%20the%20nineteenth%20century%20dictionaries,regional%20words%2C%20or%20technical%20jargon. [14 May 2024].
Williams, G. 2003. From meaning to words and back: Corpus linguistics and specialised lexicography. *ASp. la revue du GERAS* (39-40): 91–106. Groupe d'étude et de recherche en Anglais de spécialité. https://doi.org/10.4000/asp.1320 [15 May 2024}.

<div align="center">*****</div>

Using digital platforms for data collection in lexicography: *Mzansi Taal* as a case of analysis

Kganathi SHAKU (shakukj@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa
Mmagonkahloleng Brudence MAKUA (mmakua@ufh.ac.za)
Department of English and Comparative Literature, University of Fort Hare, East London, South Africa

This paper explores the use of social media as an effective source of data collection in lexicography. It focuses on the data collection procedure for *Mzansi Taal*, an online dictionary which records words and phrases from the colloquial languages used by South Africans. The following colloquial languages are included, so far, Tsotsi Taal, Spitori, Ringas, Scamtho, Afrikaaps and Gayle. Such languages are spoken in South African Township such as Soweto, Soshanguve, Mamelodi in Gauteng, and District Six, Khayelisha and the Cape Flats in the Western Cape. Just like other dictionaries, the lexicographic process of *Mzansi Taal* follows a

robust step by step procedure for dictionary compilation (SIL International 2024). The *Mzansi Taal* dictionary procedure has five steps which are followed for a word to end up in a dictionary. These steps are, research and collection, term selection, term computerization, quality assurance and term publishing. For instance, the following diagram illustrated procedure for term selection criteria step:
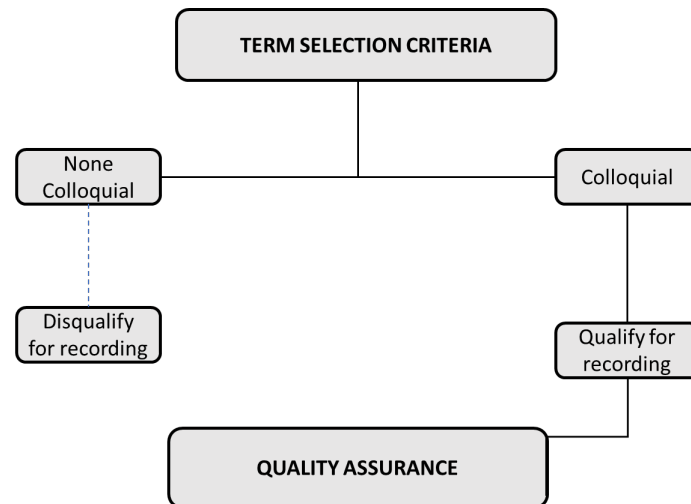
```
                    ┌─────────────────────────────┐
                    │   TERM SELECTION CRITERIA   │
                    └─────────────────────────────┘
                                 │
         ┌──────────────┐                      ┌──────────────┐
         │     None     │──────────────────────│  Colloquial  │
         │  Colloquial  │                      └──────────────┘
         └──────────────┘                              │
                │                                       │
         ┌──────────────┐                      ┌──────────────┐
         │  Disqualify  │                      │  Qualify for │
         │ for recording│                      │   recording  │
         └──────────────┘                      └──────────────┘
                                                       │
                    ┌─────────────────────────────┐
                    │     QUALITY ASSURANCE       │
                    └─────────────────────────────┘
```

Figure 1: *Mzansi Taal* term selection criterion

Traditionally, lexicographers would engage in reading of printed material like books and newspapers, reliance on self-experience and oral literature, to identify popularly used terms and frequency lists that have been used for years. Such process informs the lemmatization of words in the compilation process of a dictionary. Although technology brought more advanced digital platforms such as social media, such platforms are not fully exhausted and considered as effective methods of data collection in lexicography. It is without doubt that social media users use language, coin new words and phrases on daily basis as they are members of a digital linguistic community. As a result, social media streets have become a virtual village where language evolves (Zanzotto & Pennacchiotti 2012). Therefore, if platforms such as social media are explored effectively, they have the potential to be used as convenient sources of data collection in lexicography. This makes it easy for modern lexicographers like the *Mzansi Taal* team, to collect words and phrases — with flexibility of time and cost effectively. The collection of words and phrases for *Mzansi Taal* dictionary demonstrated the effectiveness of using digital platforms in the field of lexicography.

When words or phrases are collected for a dictionary, secondary information sources such as printed books, online books, newspapers etc. and primary sources like live interactions, become central to the process of term collection. In a traditional approach to dictionary building, material collected for the purpose of a dictionary came in a form of oral and print material. When a dictionary is compiled, the collection is the first step, and Gouws & Prinsloo (2005:17) mentioned that: "During this phase the lexicographers have to prepare the collected material for the next steps of the lexicographic process."
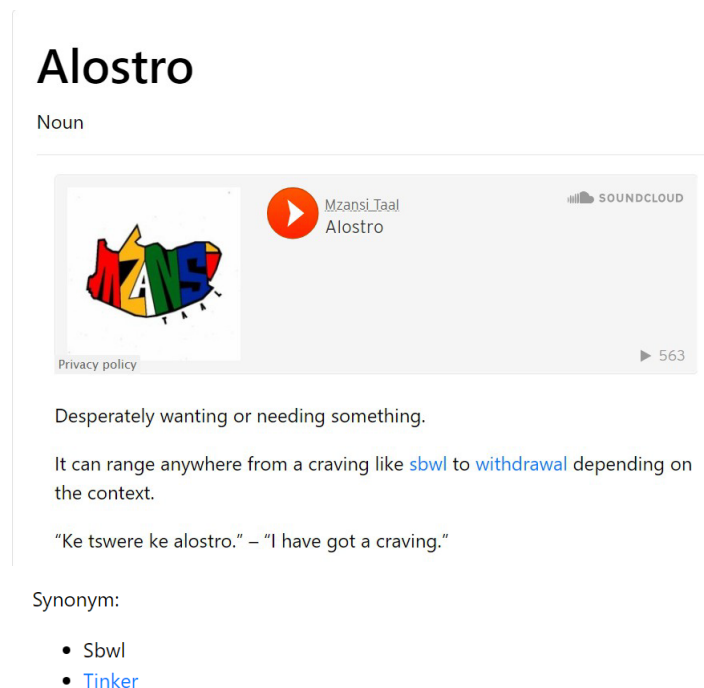
The *Mzansi Taal's* data collection strategy is advanced since it mostly uses digital platforms like social media; this became a paradigm shift from the traditional way of collecting data, where lexicographic data were collected from either oral or written (print) material as explained by Gouws & Prinsloo (2005). *Mzansi Taal* uses X, Facebook, Instagram and TikTok to collect lexicographic data. With the dominance of social media in the 21st century, it is

necessary for digital platforms to be considered in the data collection phase of lexicography. However, platforms such as social media have not yet been explored as effective collection material in lexicography.

This paper follows a qualitative research approach, as it is guided by an open and unstructured approach to the analysis of an online dictionary. It also explores and narrates the researchers' experiences as the members of *Mzansi Taal* dictionary; and its discussion is communicated narratively rather than analytically (Kumar 2019).

The paper uses digital ethnography (also known as virtual ethnography, online ethnography, cyber-ethnography and netnography) to narrative researchers' experiences of compiling an online dictionary. The Program for Ethnographic Research and Community Studies (PERCS) (Accessed 15th March 2024 Digital Ethnography, PERCS, Elon University), explains digital ethnography as a new subfield in social sciences, that studies the cultural and social domains of human interactions through internet technologies.

The *Mzansi Taal* dictionary is established by young South Africans (non-academics from Mzansi Taal organisation and academics from the University of South Africa, Northwest University and Stellenbosch University). They collect and record popularly used terms, these are the terms that are frequently used on social media, hence there are words that are not frequently or popular used on social media. 10% of data is collected from face-to-face interactions, whilst 90% is collected on social media. The documented words are assigned word classes, definitions, synonyms (if applicable), and usage examples. A case in point is the word 'Alostro' presented below:

Alostro

Noun

Mzansi_Taal
Alostro

SOUNDCLOUD

Privacy policy

▶ 563

Desperately wanting or needing something.

It can range anywhere from a craving like sbwl to withdrawal depending on the context.

"Ke tswere ke alostro." – "I have got a craving."

Synonym:

- Sbwl
- Tinker

Figure 2: Mzansi Taal Word Entry (*Mzansi Taal* Dictionary, Accessed 15th March 2024)

Digital tools such as Facebook, X (previously known as Twitter), YouTube, Instagram, LinkedIn, TikTok and Snapchats are recognised as possible sources to be used in lexicographic research.

The purpose of this paper is to demonstrate how digital platforms such as social media can be effectively used for the collection of corpora for a dictionary. Nonetheless, the use of social

media in lexicography should not be considered as a replacement of the already existing practices, it should be regarded as an alternative. The collection of data for lexicography would also be determined by the type of the dictionary and its purpose.

References
Digital Ethnography. 2024. https://www.elon.edu/u/academics/percs/digital-ethnography/. [25 March 2024].
Gouws, R.H. and Prinsloo, D.J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
Kumar, R. 2019. Research Methodology: A Step-by-Step Guide for Beginners Vol. 4. Sage Publications Limited.
Mzansi Taal. 2024. https://www.mzansitaal.co.za/. [15 February 2024].
SIL International. 2024. Dictionary Development Process. https://www.sil.org/dictionaries-lexicography/dictionary-development-process. [16 May 2024].
Zanzotto, F.M. and Pennacchiotti, M., 2012. Language evolution in social media: a preliminary study. *Linguistica Zero*: 209-230.

<center>*****</center>

The presentation of information related to food in the selected African language dictionaries

Julius Taji (juliustaji@gmail.com)
Department of foreign languages & linguistics, University of Dar es Salaam, Tanzania
Dion Nkomo (d.nkomo@ru.ac.za)
School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa

This study seeks to investigate the presentation of information related to various types of food that are most common and typical to certain communities in selected dictionaries in African languages. The study specifically aims to determine the extent of details presented in the lemmas denoting food, discuss the positioning of the information related to food within the dictionary, and examine the relationship between the food information and the relevant speech communities from cultural and environmental perspectives. The study used data from several African-language dictionaries that were selected mainly because of their familiarity to the researchers. The dictionaries are:

— *Ruhaya-English-Swahili Dictionary* (Muzale 2006)
— *Yao-English-Swahili Dictionary* (Taji 2017)
— *Ngoni-English-Swahili Dictionary* (Mapunda 2016)
— *Kamusi la Kiswahili Fasaha* (BAKIZA 2010)
— *Isichazamazwi SesiNdebele* (Hadebe et al 2001)
— *The Greater Dictionary of (isi)Xhosa* (Pahl et al. 1989-2006)

In our analysis, we adopt the methodology employed by Nkabinde (2003) who classified cultural information in Zulu into several aspects, one of which being food and beverages. An analysis of the dictionaries indicates that there are both similarities and variations in the presentation of food information. The findings show that in terms of positioning of the cultural information related to food, while most dictionaries present food information in the main text just like other entries, some dictionaries present further details related to food in the back matter. This includes a back-matter text that merely lists some typically Ndebele traditional food items in *Isichazamazwi SesiNdebele,* a monolingual dictionary for one major indigenous language in Zimbabwe. While this back-matter text does not provide definitional or descriptive information, it facilitates a quick reference and identification of the food types

and guides the user to the main text where the information is provided. Another monolingual dictionary, *Kamusi la Kiswahili Fasaha*, which is based on the Zanzibar dialect, even does it better. For example, the lemma mchele 'rice' is presented in the main text and given the relevant Swahili definition, but the lexicographer further presents information on different types of food that are prepared by using mchele 'rice' and their mode of preparation in a back-matter text. The lexicographer thus explains, "rice flour is used to prepare various types of food such as chila, mikate ya mchele au ya kumimina, vitumbua, vijoya, vibibi, mabobwe, vipopoo and tambi. A related approach is notable in *The Greater Dictionary of isiXhosa* which, in addition to defining utywala (alcohol) in the main text, presents brief essays under the headings Ukusilwa kotywala besiXhosa (the brewing of Xhosa beer) and Utywala bomzi (the beer-drink of the home) in ways that offer cultural information about traditional beer among amaXhosa people of South Africa.

It is notable that it is difficult to translate some names of food items identified in the studied dictionaries into English. For example, the various foodstuffs derived from rice flour are culturally specific to the Swahili speaking community, particularly Zanzibar. In terms of the relationship between the lemmas depicting food and the environment, we found a close relationship between the food terminologies and the environment or economic mainstay of the communities. For example, in the *Ruhaya-English Swahili Dictionary*, the lemmas denoting food mainly depict various meals made from banana, which is the stapple food in the Ruhaya speaking community. A few examples include entotomya 'cooked ripe banana', embiire 'banana for beer', ekitooke 'banana for roasting', and ekyai 'banana beer'. Similarly, in the *Yao-English-Swahili Dictionary* (which is based on the Tanzanian variety), a significant number of food lemmas are related to legumes and root plants, particularly cassava – since these are the most common types of crops grown in the Yao community. A few examples include matamba 'ugali from cassava flour', makondowole 'ugali from fermented dried cassava', chipalasya 'peas which have been husked and cooked afterwards', chinyanya 'pounded cassava', chipwenje 'cooked mixture of maize and beans', chisambula 'cooked cassava leaves, chimbilimbisya 'stew made from pounded ground nuts.' The food from cassava is also largely attested in the *Ngoni-English-Swahili Dictionary*. Examples include linindi 'cassava', chimbwinya 'soaked cassava for grilling', myakaya 'liquor from cassava.' These findings indicate that the lemmas related to food presented in African language dictionaries largely reflect the types of crops cultivated in the speech communities whose language is the focus of lexicographic description. Indeed, the findings suggest that food is not only related to culture, but it is also a key ingredient for the formation of culture and identity. This in turn confirms a close relationship between language, culture, and environment, which the lexicographers appear to be cognisant to as they go transcend regular lexicographic treatment of the relevant lemmas in the main texts in order to provide comprehensive cultural information.

References
BAKIZA. 2010. *Kamusi la Kiswahili Fasaha*. Nairobi: Oxford University Press.
Hadebe, S. et al. 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
Mapunda, G. 2016. *Ngoni-English-Swahili Dictionary*. Dar es Salaam: Languages of Tanzania Project.
Muzale, H. 2006. *Ruhaya-English-Kiswahili & English-Ruhaya-Kiswahili Dictionary*. Dar es Salaam: Languages of Tanzania Project.
Nkabinde, A.C. 2003. The implications of culture for dictionaries of the African languages. *Lexikos* 13: 168-182.
Pahl, H.W. et al. 1989-2006. *The Greater Dictionary of isiXhosa*. Alice: University of Fort Hare.
Taji, J.J. 2017. *Yao-English-Swahili Dictionary*. Dar es Salaam: Languages of Tanzania Project.

<center>*****</center>

The Van Warmelo dictionary card collection – a lexicographic appraisal

Elsabé TALJARD (elsabe.taljard@up.ac.za)
Department of African languages, University of Pretoria, Pretoria, South Africa
Danie Prinsloo (danie.prinsloo@up.ac.za)
Department of African languages, University of Pretoria, Pretoria, South Africa
Michelle Goosen (michelle.goosen@up.ac.za)
Department of African languages, University of Pretoria, Pretoria, South Africa

N J van Warmelo (1904-1989) was a linguist and state anthropologist (1930-1969), who left a collection of approximately 70 000 lexicographic index cards to the Department of African languages (University of Pretoria). Lemmas are in Sepedi, with most translation equivalents, paraphrases and translation of usage examples and usage notes provided mainly in English, although some Afrikaans translations, Latin nomenclature for plants and animals, and comments in German are also found. The aim of this presentation is to (a) give a detailed description of the Van Warmelo card collection and the contents of typical dictionary cards from this collection of approximately 70 000 cards, and (b) provide a critical analysis of the data with specific reference to its possible lexicographic application, including its use for the compilation of paper and electronic bilingual Sepedi-English dictionaries.

As far as (a) is concerned, it will be shown which data categories are present and how they are structured. The identification of data categories is an iterative process and is done by manual analysis of a number of cards until a saturation point is reached. The huge percentage of occurrence of lexical gaps in the data will be highlighted and problematic aspects of the distinction between translation equivalents and glosses will be indicated. Issues such as inconsistency in the presentation and organisation of the data will be discussed and orthographic differences will be noted.

As far as (b) is concerned, three options for lexicographic application of the Van Warmelo data are discussed and illustrated. In the first instance, cards can be digitised for inter alia, archival purposes. In this regard, we indicate that digitisation goes beyond making a mere copy of the cards; it already involves some processing of the data such as the identification of data categories represented on the cards, compilation of a metadata list, and the development of a living protocol document for the analysis and processing of the cards. The data categories that have thus far been identified on the cards are the following: lemma, part of speech, (possible) tonal indication, indication of circumflexes, noun class indication, morphological / syntactic information, translation equivalents, glosses, usage examples, translation of usage examples, stem indication, dialectal information, cross references, synonyms, usage notes, reference to sources and certain unexplained codes. The ultimate aim of this level of processing of the data would be the compilation of a data set consisting of each lemma in the metadata list, which can be linked to the pdf version of the card. Compare Figure 1 for examples of the layout and data presentation on the lexicographic index cards.
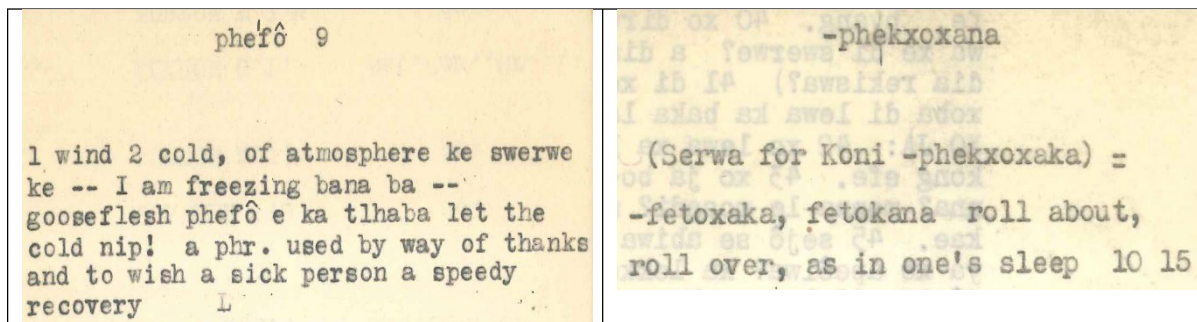
Figure 1: Examples of layout and data presentation of Van Warmelo index cards: *phefo* and *phekxoxana*.

Secondly, the data can be used for the compilation of a bilingual dictionary with conservation of the data as presented on the cards. For example, if certain data categories for a specific lemma are empty, the lexicographer will not attempt to fill them. Each dictionary article will therefore be an exact copy of the data as presented on each card. The data will however be structured according to the data categories as identified by manual inspection of a number of cards as explained above. The formulation of model entries for paper and electronic dictionaries will be attempted. For a paper dictionary the focus will be on maximum utilisation of available dictionary space for the dictionary articles, as well as front and back matter and the medio structure. For an electronic dictionary, true electronic features which are relevant for the purpose of this particular lexicographic product, such as clickable options for access to data categories such as usage examples and usage notes will be utilised.

In the third instance, a fully-fledged electronic dictionary can be compiled, using the data on the cards as a basis. For this option, the lexicographer can decide to what extent the original data can be expanded. The proposed model entails maximal utilisation of true electronic features of which hovering and clicking enabled by hyperlinking and pop-up boxes are major elements. So, for example, all words in translation equivalent paradigms and example sentences can be hyperlinked to the articles of lemmas in the dictionary. Singular and plural forms of nouns can be hyperlinked to each other as well as to pop-up boxes reflecting the full scope of noun classes, etc., cf. Prinsloo and Van Graan (2021).

It will be concluded that analysis and processing of the Van Warmelo data makes an invaluable contribution to Sepedi lexicography.

References

Prinsloo, D.J. and N.D. van Graan. 2021. Principles and Practice of cross-referencing in paper and electronic dictionaries with specific reference to African languages. *Lexicography: Journal of Asialex: Lexi* 8.1: 32-58.

Van Warmelo Lexicographic Index Card Collection, currently housed in the Resource centre of the Department of African Languages, University of Pretoria.

*****

Language Empowerment through Science Term Creation (SSTC) Processes and Procedures: A Mirage?

Beatrice TARINGA (taringab@staff.msu.ac.zw)

Midlands State University National Language Institute, Gweru, Zimbabwe

This article interrogates the process and procedures that were employed in the SSTC project at a language institute in Zimbabwe. Furthermore, it seeks to ascertain the implications on

Shona language and users in the teaching and learning process. The primary source for this study is volume one of *Shona Science Terms*. Of particular interest is how the basic principles of lexicography were applied, the promises, opportunities and challenges that came with the project to the end users, policy makers and language researchers. The significance of the study is that it appraises the custodians, user system and policymakers on the experiences of the researchers on the ground. The study is informed by Swift and Levine's 1987 theory of empowerment which states that both processes and procedures in any social endeavour may be empowering or disempowering depending on context in which they are used. The paper rests on the assumption that lexicographic activities and corpora building enrich languages and capacitate them into media of instruction. This is against the backdrop that, Zimbabwe, like most of formerly colonised African countries, inherited colonially designed language policies that revered Western languages while pushing National Indigenous Languages (NILs) to the periphery. Forty-four years after independence, linguistic neo-colonialism is still intact with stakeholders making frantic efforts to capacitate NILs into serving in teaching and learning and other spheres.

In post-independence Zimbabwe, 1980 to be precise, English language hegemony continued since it was recognised as the official language while Shona and Ndebele were national languages. This longstanding language situation as the paper holds, has of late been revisited since the current Zimbabwean constitution now recognizes sixteen (16) official languages. This has necessitated the call for use of indigenous languages as media of instruction in the infant teaching and learning process. Thus, the official recognition of the sixteen (16) NILs in Zimbabwe is a landmark development in that it gave the green light to stakeholders to use the NILs in education. In a celebratory mood to the constitutional proclamation, the Ministry of Primary and Secondary Education further buttress the position by making the officially recognised national languages media of instruction for infants which is the first 4 years of schooling from ECD-A to Grade 2. In related developments, teacher training institutions established programmes and enrolled pre-service teachers for especially those NILs that had been marginalised before the 2013 constitutional proclamation. The linguistic developments necessitated the Indigenous Languages Term Creation projects. The move was meant to avail the glossaries of terms and dictionaries that serve as frames of reference in facilitation and learning in various learning areas with the NILs as Languages of Teaching and Learning. The corpora building and lexicographic works have been going on since 1992 when the African Languages Research Institute (ALRI), that was housed under the University of Zimbabwe, was established. ALRI had seen the production of various dictionary outputs. Midlands State University National Language Institute (MSUNLI) which was established in 2018 and is housed under Midlands State University is furthering the projects that have been initiated at ALRI.

The article is set to report on the process and procedures that are employed in the SSTC as well as uncover the implications on Shona Language Status. There are lexicographic principles that guide term creation, glossary and dictionary making that should guide the processes and procedures. The article also attempts to spell out the difference in these lexicographic activities MSUNLI from the previous ones by ALRI. The article is premised on the assumption that lemmatization algorithms often rely on linguistic patterns and syntactic structure of the target language. The findings reveal that rephonologising is twice as much as semantic extension and coinage combined. The rephonologised terms though enriching does taint Shona language with English language remodeled vocabulary. The vocabulary brings unnaturalness in their use in discourse which makes their use artificial and another form of codeswitching.

The lion's share that rephonologising commands in the term creation projects is accounted for as resulting from rarely one-for-one translation equivalence and the need to avoiding massive coinage that is likely to create another foreign language from either target or source languages. The article recommends similar studies in other NILs apart from Shona.

<div align="center">*****</div>

Dictionaries as pedagogic tools in South Africa: A Case study at selected schools in Makhanda

Zola WABABA (zola.wababa@ecdoe.gov.za)
Eastern Cape Department of Education, Bisho; School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa
Dion NKOMO (d.nkomo@ru.ac.za
School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa

South Africa's Curriculum and Assessment Policy Statement (CAPS 2011) prescribes dictionaries as 'core resources' that should be used alongside textbooks for different language and content subjects. This is in line with metalexicographical literature, which regards dictionaries as problem-solving tools which assist users to address their cognitive and communicative needs. In language subjects, CAPS also guides educators and learners towards a variety of practical activities that seek to develop dictionary skills and integrate them in all pedagogical activities. However, Nkomo (2015) observed that notwithstanding the educational policy on dictionary use encapsulated in CAPS, dictionary culture remained almost non-existent in South African schools, especially in communities who speak indigenous African languages. Nkomo's study was based on the content analysis of CAPS documents and a questionnaire survey among educators in Makhanda.

The present paper, also based on research conducted in selected Makhanda schools, went further in its quest to get an empirical perspective of the pedagogical use of dictionaries in real teaching and learning contexts. The aim of the study was two-fold. Firstly, it sought to investigate aspects of dictionary culture in the schools, focusing on both Intermediate Phase (IP) educators and learners in both language and selected content subjects. Secondly, the study sought to identify the types of dictionaries that the educators and learners had access to determine their relevance in terms of curriculum alignment for IP, as well as their user-friendliness includes the way dictionary are used. Respectively, the two-fold-aim was to conduct dictionary-user research and dictionary criticism in the context of typical township schools attended by mainly black and African-language speaking children. In addition to questionnaire surveys among educators and learners, this study conducted classroom observations, content analysis of observed lessons and dictionary criticism of dictionaries that were found in the selected schools.

The questionnaire survey confirmed a poor dictionary culture among educators. While some of them showed an awareness of some relevant information that could be obtained from dictionaries for use of daily teaching and learning, there was little evidence of the integration of dictionaries in lesson planning or in actual teaching and learning. The latter point was vividly demonstrated in the observation of selected Language (English), Technology, Mathematics, and Social Science lessons. During the lessons some challenging words were noted by the educator who would draw learners' attention to such words and try to unpack them for or with the class. Other challenging words were identified by the first author of this

paper. In some instances, educators asked learners to consult available dictionaries for the meaning of challenging words, which appeared to be an equally challenging task for the learners who seemed to have limited or no abilities to find words from dictionaries. Most learners indicated that they had never used dictionaries before. In other instances, dictionaries were not available in the classrooms.

The dictionaries that were found in some of the classrooms were not relevant for the IP and for isiXhosa speaking learners. The dictionaries were mostly monolingual, which did not offer the isiXhosa learners whose English language was still at an early stage of development an opportunity to use their mother-tongue. Interestingly, the only bilingual dictionary found in circulation in the two schools was an Afrikaans-English bilingual pocket dictionary. In terms of dictionary structures, the available dictionaries were not user-friendly for Grade 4 learners. For example, some of the English dictionaries were advanced and provided multiple senses for words, leaving learners with a difficult task of identifying the correct meaning, while pocket dictionaries used small and condensed text that is not user-friendly. In fact, no South African school dictionaries were found in the two schools. When interviewed, educators noted that although they were aware of the pedagogical value of dictionaries, using dictionaries in their teaching and learning activities would be a difficult task because, while CAPS underscores their importance, dictionaries were of secondary importance as the Department of Education did not prioritize them when it comes to supplying teaching and learning materials. Accordingly, this study concluded that while children in other communities are now exposed to advanced lexicographical resources, the majority of South African learners still lack exposure to dictionaries, basic awareness of their importance and skills of consulting them. This confirms results of earlier dictionary user research (see Gouws 2013; Taljard et al. 2011).

References

CAPS. 2011. *Curriculum and assessment policy statement Grades 7-9*. Department of Basic Education, South Africa.

Gouws, R. H. 2013. Establishing and developing a dictionary culture for specialised lexicography. V. Jesensek (ed.). *Specialised lexicography: Print and digital, specialised dictionaries, databases. Lexicographica. Series Maior*: 51-62. Berlin: De Gruyter.

Nkomo, D. 2015. Developing a dictionary culture through integrated dictionary pedagogy in the outer texts of South African school dictionaries: The case of Oxford Bilingual School Dictionary: IsiXhosa and English. *Lexicography* 2(1): 71-99.

Taljard, E., Prinsloo, D. J., & Fricke, I. 2011. Use of LSP dictionaries in secondary schools – A South African case study. *South African Journal of African Languages* 31(1): 87-109.

PRESENTATIONS: GLOBALEX WORKSHOP

European Network on Lexical Innovation – ENEOLI

Giovanni TALLARICO (giovanni.tallarico@univr.it)
Dipartimento di Lingue e Letterature Straniere, Università di Verona, Verona, Italy
Rute COSTA (rute.costa@fcsh.unl.pt)
NOVA CLUNL – Centro de Linguística da Universidade NOVA de Lisboa, Lisboa, Portugal
Špela ARHAR HOLDT (spela.arharholdt@ff.uni-lj.si)
Faculty of Arts and Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia
Vincent BALNAT (balnat@unistra.fr)
Département d'études allemandes/LiLPa (UR 1339), Université de Strasbourg, Strasbourg, France
Petar BOŽOVIĆ (petarb@ucg.ac.me)
English Language and Literature Department, Faculty of Philology, University of Montenegro, Nikšić, Montenegro
Judit FREIXA (judit.freixa@upf.edu)
Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Spain
Kris HEYLEN (kris.heylen@ivdnt.org)
Katholieke Universiteit, Leuven, Belgium; Instituut voor de Nederlandse Taal, Leiden, The Netherlands
John HUMBLEY (humbley.john@orange.fr)
CLILLAC-ARP, Université Paris Cité, Paris, France
Ilan KERNERMAN (ilan@lexicala.com)
Lexicala by K Dictionaries, Israel
Annette KLOSA-KÜCKELHAUS (klosa@ids-mannheim.de)
Leibniz-Institut für Deutsche Sprache, Mannheim, Germany
Ana OSTROŠKI ANIĆ (aostrosk@ihjj.hr)
Institute for the Croatian Language, Zagreb, Croatia
Ana SALGADO (anasalgado@fcsh.unl.pt)
NOVA CLUNL – Centro de Linguística da Universidade NOVA de Lisboa, Lisboa, Portugal
Federica VEZZANI (federica.vezzani@unipd.it)
Dipartimento di Studi Linguistici e Letterari (DiSLL), Università degli Studi di Padova, Padova, Italy
Weronika WOŹNIAK-ŻAK (weronika.wozniak.zak@edu.uni.lodz.pl)
Doctoral School of Humanities, Faculty of Philology, University of Lodz, Poland

In recent years neology has gained new international insights thanks to the Globalex Workshop on Lexicography and Neology (GWLN) series of events and publications, but the study of lexical innovation as the most dynamic aspect of language change is yet to find a more prominent position in general linguistic studies.

Approaching neology from a multilingual or global point of view, however, is the goal of the recently launched COST Action (European Cooperation in Science and Technology) "European Network on Lexical Innovation – ENEOLI" (CA22126 – https://www.cost.eu/actions/CA22126/, October 2023 – October 2027). As of March 2024, 218 members from 44 different countries are participating in the Action. The primary objective of

ENEOLI is to establish a network of researchers across Europe and beyond, fostering the exchange of best practices and methodologies in research on lexical innovation, also known as neology, a central though all too frequently neglected dimension of natural language study. From a linguistic point of view, the study of neology "contributes to a better understanding of the lexical system of a given language and its evolution" (Sablayrolles 2019:7), while from an extralinguistic point of view, "the inventory of neologisms also gives much information about language communities in their material lives and social representations" (ibid.).

The key challenges addressed by the Action can be summarised as follows: (1) Defining and explaining key concepts in the field of lexical innovation by creating an open-access multilingual glossary where none existed hitherto; (2) Reviewing and sharing methodological implementations, digital resources and tools for identifying and tracking neologisms over time, using both natural language processing approaches and sociolinguistic/psycholinguistic methods; (3) Conducting comparative studies of neologisms in European languages, exploring factors such as the impact of technology on lexical innovation and contact-induced language change in the digital era; and (4) Providing specific neology training for professionals (e.g. translators, teachers, technical communicators, terminologists).

The Action aims to advance the state of the art in neology through a four-fold approach: (1) Compiling a born-digital open-access multilingual glossary defining and illustrating the terminology of lexical innovation; (2) Investigating and developing methods, digital resources, and tools for the study of lexical innovation; (3) Conducting diachronic and synchronic comparative studies of neology; and (4) Developing innovative forms of neology training.

ENEOLI is dedicated to advancing research on neology, with a specific emphasis on enhancing the lexicographic dimension. The core deliverable, a multilingual glossary collecting the terminology of lexical innovation, is an innovative digital lexical resource whose architecture is designed to accommodate a large number of languages and to fulfil various user needs. The glossary will serve both didactic and pedagogical functions, as well as facilitate intercultural comprehension and cooperation across different languages worldwide.

To achieve these aims, the Action is organised into four work packages: WG1: Multilingual glossary of neology; WG2: methods, digital resources and tools for neology; WG3: Diachronic and synchronic comparative studies of neology, and WG4: Training in Neology.

This paper provides an overview of the project, outlining the goals of each work package and describing ongoing work, and highlighting its lexicographic aspects.

References
Sablayrolles, J.-F. 2019. *Comprendre la néologie. Conceptions, analyses, emplois*. Limoges: Lambert-Lucas.

*****

Neologisms and their Functions in Critical Discourse

Petra STORJOHANN (storjohann@ids-mannheim.de)
Leibniz-Institut für Deutsche Sprache, Mannheim, Germany

New lexical items can be key elements in the constitution of discourse in the sense of Critical Discourse Analysis (Teubert 2010, Partington et al. 2013). A neologism may have discursive and communicative weight during a specific moment or within a defined timeframe. For example, amid the pandemic German witnessed an influx of new terms related to Covid, many carrying immediate significance in public conversations and contributing to

the depiction of key knowledge structures (Storjohann & Cimander 2022). The contributions of new words in the construction of a discourse are understood as discourse functions. Neologisms introduce new ways of thinking by embodying novel conceptual knowledge and influencing collective thought processes. They play a crucial role in distinguishing emerging social practices by clarifying vague concepts and laying the groundwork for further lexical innovation, e.g. through word formation. Within crises-related discourse, neologisms become fundamental in shaping linguistic reality, acting as pivotal nodes within a structured network. They have the potential to expose ideological content and cultural values, as these terms can highlight key issues in debates. Terms such as Cancel-Culture, Wokeness and Greenflation capture a range of views and beliefs in societal dialogues across ideological, cultural, and political spectrums. However, when neologisms are documented lexicographically, their roles in discourse are not specified.

This paper aims to demonstrate how the key aspects highlighted by new vocabulary are systematically documented in a recently developed dictionary. The novel approach adopted integrates specific elements of discourse lexicography, capturing aspects such as "die Gebundenheit des Wortschatzes in topikalischen, sprecherbezogenen, textlichen, zeitlichen und funktionalen Hinsichten" (translates as "the relevance of vocabulary in topical, speaker-related, textual, temporal, and functional respects"), as highlighted by Kämper (2006:350). Some content also prompts users to delve into complex connections within the lexicon, moving beyond single headwords. This new resource allows users to access discourses by using a reverse approach where one can find all new terms associated to a particular discourse. For instance, journalists frequently inquire, "What are the latest terms related to the energy crisis?" Conversely, the paper will also show the capacity to amalgamate these individual lexeme-specific insights into summarised discourse depictions. This involves providing an outlined overview of the discourse associated with various neological keywords, such as the Bundestag election campaign in 2021, the U.S. election in 2020, climate crisis, energy crisis, and the Ukraine war. This interplay between information specifically attributed to single lexemes and broader linguistic knowledge spanning multiple lexemes highlights the relationship between neologisms and (critical) discourse. Clearly, the novel neologisms resource is not a conventional discourse dictionary; rather, it seeks to transcend the constraints of previous lexicographical methods.

Currently under examination is the lexicographic integration of discourse and the evolution of neologisms. The shifts in discourse patterns are intricately connected to alterations in paradigmatic structures between neologisms and other pivotal terms. Identifying the central themes and key terms that are paradigmatically similar within a public discourse has become more accessible through vector-based word embedding models (Kupietz et al. 2018, Fankhauser & Kupietz 2019). For instance, the term Energiekrise (energy crisis) has become a prominent focal point in critical discourse intertwined with the Ukraine war since 2021. It frequently intersects with other recent crises, forming clusters of terms that encompass various facets of the Covid-Krise (Covid crisis), Chipkrise (chip crisis), Gaskrise (gas crisis), Versorgungskrise (supply crisis) and Klimakrise (climate crisis). Indeed, many new words that have emerged since 2020 can be sorted into these interconnected groups, showcasing specific additions to the vocabulary. While Bubenhofer (2022) employed the word embedding approach to identify central themes, recurring motifs, cognitive frameworks, narratives, and arguments, this paper explores whether new methods can be utilised for lexicographic objectives.

References

Bubenhofer, N. 2022. Exploration semantischer Räume im Corona-Diskurs. Kämper, H. & Plewnia, A. (eds.). *Sprache in Politik und Gesellschaft: Perspektiven und Zugänge*: 197-216. Berlin/Boston: De Gruyter.

Fankhauser, P. & Kupietz, M. 2019. Analyzing domain specific word embeddings for a large corpus of contemporary German. *Proceedings of ICL, Cardiff*. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/9117/file/Fankhauser_Kupietz_Analyzing_domain_specific_word_embeddings_2019.pdf. [15.11.2023].

Kämper, H. 2006. Diskurs und Diskurslexikographie. Zur Konzeption eines Wörterbuchs der Nachkriegsdiskurse. *Deutsche Sprache* 4/2006: 334-353.

Kupietz, M., H. Lüngen, H., P. Kamocki & Witt, A. 2018. German Reference Corpus DeReKo: New Developments – New Opportunities. Calzolari, N. et al. (eds.), *Proceedings of the 11th LREC 2018*: 4353-4360. Miyazaki: ELRA.

Partington, A., Duguid, A. & Taylor, Ch.. 2013. Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS). Amsterdam: John Benjamins.

Storjohann, P. & Cimander, L. 2022. Annäherung an den Coronadiskurs in der öffentlichen Kommunikation mithilfe von Neologismen. Jakosz, M. & Kałasznik, M. (eds.). *Corona-Virus-Pandemie: Diverse Zugänge zu einem aktuellen Superdiskurs*:25-50. Göttingen: V&R Unipress.

Teubert, W. 2010. *Meaning, Discourse and Society*. Cambridge: Cambridge University Press.

*****

From Neologism Extraction to Dictionary Description: Methodological Issues in Corpus Balance, Word Unit Bias and LLM Assistance

Kilim NAM (nki@yonsei.ac.kr)
Korean Language and Literature, Yonsei University, Seoul, South Korea
Soojin LEE (sjmano27@naver.com)
Korean Language and Literature, Kyungpook National University, Daegu, South Korea
Hae-Yun JUNG (haeyun.jung.22@gmail.com)
Korean Language and Literature, Kyungpook National University, Daegu, South Korea

Corpus research, which was first pioneered in lexicography, has since developed into the methodology of corpus linguistics and has been expanded, refined, and eventually adopted by neological studies. News corpora in particular have been used as authentic language data and a strong basis for the identification of institutionalised neologisms (as opposed to nonce words), the dating of their first appearances, and the investigation of their usage trends. Their status as the greatest resource for neologism collection and study has been abundantly discussed (Renouf 2013; Boussidan 2013; Nam et al. 2020; Freixa & Adelstein 2013; Klosa & Lüngen 2018). However, the spread of web languages and the emergence of large language models (LLMs) today have a considerable impact on the creation and diffusion of neologisms, as well as on the application of language resources. In that sense, it has become crucial to re-examine not only the bias of [+formal] and [+written] news corpus, but also the neologism extraction methods centred on single-word units. This study critically reviews the methodology for Korean neologism research, which has consisted in the semi-automated extraction from news corpora from 2005 to date, and explores ways to improve dictionary compilation as to reflect the dynamics of language from the cognitive perspective of discourse communities and individual speakers.

Chapter 2 examines the news media language bias in collecting neologisms[i] by using a

---

[i] The list of unregistered words was monitored by using the list of headwords of the biggest Korean dictionary *Urimalsaem* as well as neologism lists collected from 1994 to present by the National

Python programme to analyse a news corpus of 500 million words, a 14-million-word corpus of online posts from forums and social media, and a 6-million-word instant messages (IM) corpus, which roughly span from 2020 to 2022, in order to compare the appearances, frequencies, domains and distributions of neologisms in different communication contexts. The analysis shows a significant bias of the news corpus toward public domains, such as politics, economics, and society, and scantily accounts for everyday language, neglecting expressions related to food or emotions for instance. A case in point is the particularly productive derivational suffix *-sulep-* 'be like', which is used to form adjectives in Korean. Adjectives constitute a part of speech that well expresses one's evaluations, attitudes, and emotions. The analysis of unregistered '*-sulep-*' derivational adjectives shows that there are 531 unregistered derivatives (35% of the total) across all web genres, 79% of which, however, are found only in online posts and/or IMs. While the '*-sulep-*' derivatives found in the news corpus are often related to politics or social issues (e.g. *yunsekyelsulepta* 'be very Yoon-Suk-Yeol-like'; *pheymisulepta* 'be feminist-like'), those from online posts and IM are often formed from bases denoting aspects of the daily life, such as food (e.g. *hansiksulepta* 'have a traditional Korean vibe'; *kokwumasulepta* 'be stifling [just as when eating sweet potato]'). In addition, the chapter discusses the value of such everyday language products as headword candidates.

Chapter 3 discusses the issues of the single-word and formal neologism bias. A comprehensive account of the lexicon of native speakers for a given era relies not merely on the identification of new forms but also on the analysis of their frequencies, distributions, and the discourse context in which they appear. This means that it has become crucial for the advancement of neologism research to develop a methodology for extracting semantic units such as phrases and collocations. Neological phrases are harder to identify than single-word neologisms and are often related to semantic neologisms, thereby being often dismissed from neologism extractions and headword selections. Instead, this study is to provide a closer look at phrase unit neologisms such as *kkwul ppalta* 'idle time away [instead of working] (literally, 'suck honey')' that are deemed worth including in the dictionary.

Lastly, this study tests the recommendation of headwords and the compilation of dictionary microstructures for the neologisms presented here when prompting major Korean LLMs such as CLOVA by Naver. For now, it seems that CLOVA struggles not only to identify but also to define neologisms. For example, when prompted to recommend verbal neologisms or define a given neologism, it could only provide nominal forms and example sentences containing the lemma to define. In contrast, foreign LLMs such as Chat GPT could give an explanation of the lemmas it provided, along with pragmatic information such as 'slang' or 'informal' labels. Moreover, ChatGPT could provide verbal forms, although their quality and qualification as neologisms were somewhat questionable. This seems to point towards the inadequacy of the Korean data and existing dictionaries learning by LLMs for the identification and description of neologisms. Twenty years ago, Sinclair (2004:188-192) emphasized that the linguistics community needed to prepare for larger corpora in order to contribute to the future information society. In line with this, this study argues that Korean neologism research, to contribute to society in the modern age of LLMs, needs to think its methodology anew, turning to larger, balanced corpora and the contextualized extraction of semantic units for the dictionary to depart from the prescription of an idealised language and instead, be in tune with the dynamicity of actual language as spoken by native Koreans.

Institute of Korean language and the Centre for Korean Language Information at Kyungpook National University.

References

Boussidan, A. 2013. Dynamics of semantic change: Detecting, analyzing and modeling semantic change in corpus in short diachrony. Doctoral dissertation. Lyon: Université de Lyon.

Freixa, J. & Adelstein, A. 2013. Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. *Comunicación presentada en Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación, 28-30 October 2013, Mexico City, Mexico.*

Klosa, A. & Lüngen, H. 2018. New German words Detection and description. Čibej, J., Gorjanc, V., Kosem, I., & Krek, S. (eds.). *Proceedings of the XVIII EURALEX International Congress Lexicography in Global Contexts 17-21 July 2018, Ljubljana, Slovenia.*

Nam, K., Lee, S. & Jung, H. Y. 2020. The Korean Neologism Investigation Project: Current Status and Key Issues. *Dictionaries: Journal of the Dictionary Society of North America* 41(1):105-129.

Renouf, A. 2013. A finer definition of neology in English. Hasselgård, H., Ebeling, J. & Oksefjell Ebeling, S. (eds.). *Corpus perspectives on patterns of lexis*: 177-207.

Sinclair, J. 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.

*****

Lexicographic representation of verbal semantic neology in Spanish dictionaries

Marina BERRI (mberri@campus.ungs.edu.ar)

Instituto del Desarrollo Humano, CONICET, Universidad Nacional de General Sarmiento, Los Polvorines, Argentina

Victoria de los Ángeles BOSCHIROLI (vboschir@campus.ungs.edu.ar)

Instituto del Desarrollo Humano, Universidad Nacional de General Sarmiento, Los Polvorines, Argentina

Semantic neologisms (i.e. new uses and meanings of existing forms) have proved hard both to detect and to characterize in terms of their neologicity (i.e a new item's property of being new, cf. Sablayrolles 2003, Estopà 2015, Bouzidi 2010, Adelstein 2019, among others), given the lack of formal identifying features -which have mostly made them all but invisible for automatic tracking-, and the difficulty involved in differentiating meanings, senses and variations of a sense (Adelstein 2019). Furthermore, they tend to be less frequently represented in dictionaries than other neological items, even in those devoted exclusively to neologisms: in the *NEOMA. Diccionario de neologismos del español actual*, with 2400 headwords, less than 5% are marked as semantic neologisms, while in the *Antenario*, the ratio is even lower (about 2,5%). This degree of under-representation is unsurprising, since, as Adelstein (2022) points out, most semantic neologisms tend to go unperceived as neological by speakers -in other words, they often fail the "neological feeling" (Sablayrolles 2003) test.

This work aims to study aspects of the microstructural lexicographic representation of verbal semantic neologisms in Spanish dictionaries, in particular, how the new, emerging meanings are acknowledged and explained lexicographically. There are two main issues involved in lexicographic representation: the inclusion of the new item in the macrostructure of the dictionary, an aspect that is often focused on especially with regards to dictionary updating, and its treatment in the microstructure. If a semantic neologism is represented in a dictionary of neologisms, these are independent processes, which usually involve the creation of a new entry where references to previous senses may or may not be made. On the other hand, in the case of general language dictionaries, these processes only affect the microstructural level: unlike other types of neologisms, whose meaning is explained through a definition under a new headword, the neological meaning is incorporated into an existing

entry which, as a result, undergoes changes. Thus, semantic neologisms may be included in ways other than as new senses with conventional explanatory definitions –namely, through example sentences, through the label "fig." ('figurative') or by expanding a preexisting definition. Our aim is to assess how these strategies relate to the different stages a neologism goes through in the neologic process (Freixa 2022), in particular, the linguistic, cognitive and social subprocesses.

In previous works (Berri & Boschiroli 2022, 2023), we suggested that the concepts of norm and exploitation, as introduced by Patrick Hanks (2013), can be used as criteria to measure the degree of neologicity of a verbal semantic neologism since both the frequency of occurrence of a collocational pattern and its stability may be used as evidence to determine what stage of the process it is in, especially at linguistic level, but also cognitively. We have thus proposed analysis for the following semantic neologisms identified and recorded in BOBNEO: "anestesiar" ('to anesthetize'), "coreografiar" ('to choreograph'), "mapear" ('to map'), "catapultar" ('to catapult'), "centrifugar" ('to centrifuge') and "dilapidar" ('to squander'). In this presentation we would like to analyze and assess how and to what extent the new senses of these verbs that are undergoing different stages of the neologic process are represented lexicographically in regularly and recently updated general Spanish dictionaries, such as the *Diccionario de la lengua española* and *Diccionario del español actual*, and reflect about the problems of inclusion and representation in dictionaries of neologisms.

References
Adelstein, A. 2019. Hacia una definición de neologicidad en semántica. M.L. Perassi & M. Tapia (eds.). *Palabras como puentes. Estudios lexicológicos, lexicográficos y terminológicos desde el Cono Sur*: 209-230. Córdoba: Editorial Buena Vista.
Adelstein, A. 2022. Neología y semántica: grados de neologicidad en el ámbito nominal. E. Bernal, J. Freixa & S. Torner (eds.). *La neología del español. Del uso al diccionario*: 361-371. Madrid-Frankfurt: Editorial Vervuert.
Adelstein, A. & Freixa, J. (eds.) (2018-). *Antenario. Diccionario en línea de neologismos de las variedades del español (Antenario)*. http: //www.antenario.wordpress.com. [13/03/2024].
Berri, M. & Boschiroli, V. 2022. El contexto oracional como parámetro de neologicidad en la neología semántica verbal. Unpublished Manuscript.
Berri, M. & Boschiroli, V. 2023. Neologicidad metafórica y grados de inestabilidad. P. C. Hernández, & Galvani Gelusini M. G. (eds.). *Estudios SAEL 2023*: 75-88. Córdoba: Editorial Humanitas.
Bouzidi, B. 2010. Néologicité et temporalité dans le processus néologique, *Algérie* 9: 27–36.
Observatori de Neologia de la Universitat Pompeu Fabra Banco de neologismos lexicográficos (BOBNEO). http://bobneo.upf.edu/inicio.html. [13/03/2024].
Estopà, R. 2015. Sobre neologismos y neologicidad: reflexiones teóricas con repercusiones metodológicas. I. M. Alves & Simões Pereira, E. (eds.). *Neologia das Línguas Românicas*: 111-150. São Paulo: Humanitas.
Freixa, J. 2022 El proceso neológico. E. Bernal, J. Freixa & S. Torner (eds.). *La neología del español. Del uso al diccionario*: 51-72. Madrid-Frankfurt: Editorial Vervuert.
Hanks, Patrick. 2013. *Lexical Analysis. Norms and exploitations.* Cambridge-London: The MIT Press.
Real Academia Española. 2014. *Diccionario de la lengua española* 23ª ed. (DLE) https://dle.rae.es/. [13/03/2024].
Sablayrolles, J. F. 2003. Le sentiment néologique. J.F. Sablayrolles (ed.). *L'innovation lexicale*: 156-167. Paris: Honoré Champion.
Sánchez Manzanares, C. (Dir.). 2016. *NEOMA. Diccionario de neologismos del español actual*. https://www.um.es/neologismos/index.php/ [13/03/2024].
Seco, M., Andrés, O. & Ramos, G. 2023. *Diccionario del español actual* (3a edición) (DEA). https://www.fbbva.es/diccionario/ [13/03/2024].
*****

Digging for -ings: a survey of selected (pseudo)anglicisms in dictionaries of modern Greek

Anna VACALOPOULOU (avacalop@athenarc.gr)
ILSP, Athena RC, Athens, Greece

One of the many ways for a language's vocabulary to extend through neology is through lexical borrowing. A very specific type of this phenomenon is the anglicism, which is created when a language borrows a word or phrase from the English language and incorporates it into its vocabulary. This process can be reflected at various levels in the borrowing language, such as its phonology, morphology, grammar, or vocabulary. For historical reasons, the prominence of the English language has caused anglicisms to be the most widely spread lexical items that are borrowed at a global level. This trend has been further enhanced by the introduction of vocabulary from terms related to international business, as well as new technologies, most notably computers and the internet, as these are originally coined in English and very quickly make their way into foreign lexicons. For these reasons, among others, English is also the most popular foreign language to learn, which makes anglicisms spread more easily. In some cases, these items are adopted verbatim in the new languages, usually when these languages share the same alphabet as English, that is, Latin. In the case of languages such as Greek, that use a different alphabet, these items normally follow a process of transliteration or other types of anglicization that help words and phrases assimilate into the new language. In some cases, the influence of English is so prominent that cases of pseudoanglicism also occur, resulting in lexical items that appear to be English because of their form but, in fact, are not.

This paper explores the inclusion and treatment of a special category of anglicisms and pseudoanglicisms as they occur in dictionaries of modern Greek. This category consists of nouns ending with -ing, a suffix that corresponds to the present participle and gerund of a verb, which is typical of English morphology and has been incorporated in other languages through the process of borrowing. The methodology is based on a selection of five Greek dictionaries available online: HLNG, LKN, MILNEG, SLANG.gr, and Wiktionary. These are compared with each other in terms of their macrostructure to find out how inclusive they are in their treatment of this phenomenon. Dictionaries that are available exclusively on paper were not considered taking space constraints into account. The decisive criterion for this selection was that the dictionaries are currently in use. Among the selected dictionaries, three are the products of professional lexicographers, while the other two are works of crowdsourcing. At the second stage of the process, the resulting entries are examined in terms of their orthography and cross-checked against a corpus of modern Greek (elTenTen19) to test whether their treatment reflects actual language use according to the data. Then, moving further into the lexicographic microstructure, the possible existence of Greek equivalents is researched in all five dictionaries. These findings are also cross-checked against the corpus for frequency to see which of the existing alternatives native users of Greek tend to use and which are possibly more alien to them. Finally, the list of lexical items in the findings is explored as to which domain or semantic field they belong to as well as their register, style, and attitude; this exploration aims to unveil possible trends in their coinage and usage.

References

elTenTen19. Corpus of the Greek Web. 2020. https://app.sketchengine.eu/. [20/04/2024].

HLNG. *Practical Dictionary of Modern Greek*. 2014.
https://christikolexiko.academyofathens.gr/. [20/04/2024].
LKN. *Modern Greek Dictionary*. 1998. https://www.greek-
language.gr/greekLang/modern_greek/tools/lexica/. [20/04/2024].
MILNEG. *Big Electronic Dictionary of Modern Greek*. n.d. http://lexicon-
neohel.patakis.gr/el/lexikon. [20/04/2024].
SLANG.gr. n.d. https://www.slang.gr/. [20/04/2024].
Wiktionary. n.d. https://en.wiktionary.org/wiki/Wiktionary:Main_Page. [20/04/2024].

<div align="center">*****</div>

The role of combining forms in creating neologisms — A data-driven approach

Yongwei GAO (ywgao@fudan.edu.cn)
College of Foreign Languages & Literature, Fudan University, Shanghai, China

Combining forms (CFs), a lexical concept first appeared in the English language as early as 1884 and defined by the *Oxford Dictionary of English* as "a form of a word normally used in compounds in combination with another element to form a word", has traditionally played an important role in forming new words in the English language. Technical CFs like *immuno-* and *-plasty* have spawned at least one hundred technical terms. Like affixes, CFs can be used to attach to existing words, word elements and even other CFs to form new compounds. Traditionally, the productivity of most CFs is paled by comparison with that of affixes. Unlike affixes, the number of CFs has always been on the rise while there are seldom new affixes emerging in the English language. *Cyber-*, for instance, a relatively new CF, has been very productive in forming new English compounds whose number may have amounted to dozens or even over a hundred, as are recorded by major English dictionaries such as *Collins English Dictionary* and *Oxford Dictionary of English*. Frequently used *cyber-* compounds include *cyberattack, cyberbullying, cybercafé, cybercrime, cyber-espionage, cybershopping, cyberspace,* and *cyberterrorism*, to mention a few. Other newly emerged CFs like *e-* (e.g. *e-book*), *-flation* (e.g. *shrinkflation*), *-fluencer* (e.g. *petfluencer*), *-geddon* (e.g. *snowgeddon*), *-licious* (e.g. *babelicious*), and *-preneur* (e.g. *dadpreneur*) have so far formed scores of new words that have been widely used in the English language. There are also several CFs-to-be that have formed many new compounds, some of which, however, have been wrongly assumed to be blends. *-cation* is a case in point. As a splinter from *vacation*, it has been seen in use in at least ten neologisms such as *bizcation, coronacation, daycation, haycation, mancation,* and *nakation*. As a result of their vitality, the current number of CFs has apparently far surpassed that of affixes, as is shown in the *Oxford English Dictionary* in which there are over two thousand CFs and only several hundred affixes (including both prefixes and suffixes).

This paper takes a data-driven approach to the study of new CFs and their productivity in forming new compound words. The new CFs to be studied number at about twenty, including neoclassical elements and clipped forms of existing English words, such as *perma-, -pocalypse, -razzi, robo-, -tacular, -tastic,* and the above-mentioned CFs. The new compounds formed by these CFs are selected from major English dictionaries like the OED and several neologism dictionaries that the author compiled in the past few years, such as *A 21st Century English-Chinese Dictionary of Neologisms* (2021), *An English-Chinese Dictionary of Neologisms in the New Era* (2023), and *An English-Chinese Dictionary of Neologisms in Present-day English* (2023). The data to be used are taken from corpora like News on the Web and news archives such as Google News. The productivity of all the CFs in question will be examined with the help of corpus data, and the mechanism in forming compounds with CFs will also be investigated. The status of neologisms as compounds or blends will be scrutinized as well.

References
Collins English Dictionary: https://www.collinsdictionary.com/dictionary/english.
Gao, Y. W. 2021. *A 21st Century English-Chinese Dictionary of Neologisms*. Shanghai: Fudan
    University Press.
Gao, Y. W. 2023. *A Dictionary of Blends in Contemporary English*. Beijing: The Commercial
    Press.
Gao, Y. W. 2023. *An English-Chinese Dictionary of Neologisms in Present-day English*. Beijing:
    The Commercial Press.
Oxford Dictionary of English: https://premium.oxforddictionaries.com.
 Oxford English Dictionary: https://www.oed.com/.
*****

From *taggare* to *blessare*: verbal hybrid neologisms in Italian youth slang

Maria Pia DI BUONO (mpdibuono@unior.it)
University of Naples "L'Orientale", Naples, Italy

Hybrid neologisms, formed directly within the language through various operations of word derivation and formation (Aprile 2005) and characterised by the use of non-native structures, constitute a diverse system of phenomena and a significant part of contemporary Italian (Dardano et al. 2000), with a growing influence on this linguistic system (Frenguelli 2005; Giovanardi et al. 2003).

This contribution aims to investigate the phenomenon of verbal hybrid neologisms formed from a non-native morpheme, derived from an English verb, through the analysis of their diffusion in youth slang on social media, as the creation and dissemination of Anglo-Italian neologisms and pseudo-anglicisms can be the result of a bottom-up Anglicization (Lubello 2014:73).

Due to the fact that documenting the language evolution brought by the introduction of neologisms can be challenging in identifying and classifying new formations, but also in representing their meanings, we propose a methodology to support the discovery and the semantic description of neologisms.

As the first step, using the NLTK Python library, we extract the complete list of single-word English verbs from WordNet and apply a set of manually defined derivative rules to create the hybrid patterns. In Italian, hybrid verb neologisms are usually formed applying the 1st pattern conjugation, represented by the highly productive inflectional morpheme *-are* in the infinitive form. The derivative rules, accounting for the process of morphological blending, include any adjustment rules pertaining to Italian word formation that depend on the type of lexical morpheme used as root. Thus, for instance, monosyllabic verbs ending with a consonant and a vowel are adjusted by an elision of the final vowel and the doubling of the remaining consonant, like *lovvare* from *love* (it is worth stressing that this adjustment for monosyllabic verbs is not consistent, as it might be influenced by an interference with native words, e.g., *zonare* from *zone*).

From this procedure, we obtain a list of hybrid formations which present different statuses according to some dictionaries, e.g., Treccani dictionary: i. The formation is already recorded into dictionaries (e.g., *taggare*); ii. The word has been already identified but not included in dictionaries (e.g., *shippare* described in the Treccani Web portal in 2019); iii. The word is not present in dictionaries and has not been discussed in the Treccani Website (e.g., *blessare* and

lovvare[i]). Besides these types, we also identify another group that presents an overlap with native forms, e.g., *zappare* that can be the results of a neologism formation from *zap* a native form which means *hoe the ground* (for the sake of this work, we do not consider the overlapping formations, as we would need a deeper semantic analysis to distinguish native forms from neologism usages).

The list of hybrid verbs is employed in combination with some of the hashtags related to the most popular reality shows and TV programs among young people to extract tweets and create our corpus. Indeed, the practice of commenting on reality shows and TV programs through social networks is widespread, especially among the youth, as evidenced by the daily trending topics published by the social media formerly known as Twitter (now X).

We refine the collected data on the basis of the number of occurrences of each candidate and apply a multidimensional approach to describe the candidate semantics. For the most widespread occurrence, e.g., *shippare*, we use an unsupervised approach, i.e., Latent Dirichlet Allocation - LDA (Blei et al. 2003), to obtain natural clusters for such occurrences. LDA allows for observing the data based on spontaneous aggregations, not decided a priori and it can be applied to detect neologisms and novel word senses (Kim 2022; Matsumoto et al. 2019; Lau et al. 2012). It represents the underlying information in a text by grouping words in a way that each group is representative of a specific topic. We set the number of topics to 3 and 5 and observe the obtained results, namely clusters of keywords that represent each topic with numerical values indicating the relevance of each word within the topic (for example, for *shippare* the most relevant keywords in Topic 0 are "gregorelli", "romeraci", "giulia", "fan", and so on. This suggests that this topic might concern discussions about shipping different couples and their respective fandoms). Furthermore, we can account for the context usage as the similarity of the topics might indicate that tweets have a very uniform linguistic structure. Finally, being derived from English verbs, our candidates can be evaluated and enriched through the use of English external resources such as VerbNet and WordNet to describe their frame and compare and derive the sentence frame. The obtained results allow for an initial assessment of the diffusion of certain hybridisms and their contexts of use to outline their meaning and semantic fields.

References
Aprile, M. 2005. *Dalle parole ai dizionari.* Bologna: Il Mulino.
Blei, D.M., Ng, A.Y. & Jordan, M.I. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan.): 993-1022.
Dardano, M., Frenguelli, G., & Perna, T. 2000. L'italiano di fronte all'inglese alle soglie del terzo millennio. L'italiano oltre frontiera. *Atti del V convegno internazionale dell'AISLLI, 22-25 April, 1998, Leuven, Belgium*: 31-55.
Frenguelli, G. 2005. La composizione con elementi inglesi. Giovanardi, G. (ed.). *Lessico e formazione delle parole. Studi offerti a Maurizio Dardano per il suo 70° compleanno*: 159-176. Firenze: Cesati.
Giovanardi, C., Gualdo, R. & Coco, A. 2003. *Inglese-italiano 1 a 1: Tradurre o non tradurre le parole inglesi.* Lecce: Manni Editori.
Kim, W. 2022. Distributional Corpus Analysis of Korean Neologisms Using Artificial Intelligence. PhD Thesis. University of Georgia.
Lubello, S. 2014. L'itangliano è ancora lontano? Qualche riflessione sull'influsso dell'inglese. *Lezioni d'italiano. Riflessioni sulla lingua del nuovo millennio.* 63-84. Bologna: il Mulino.
Matsumoto, K., Ren, F., Matsuoka, M., Yoshida, M. & Kita, K. 2019. Slang feature extraction by analysing topic change on social media. *CAAI Transactions on Intelligence Technology*

---

[i] Some of these are reported in other web portals, like Slengo (https://www.slengo.it/define/lovvare) and user-made web portals on Italian neologisms.

4(1): 64-71.

Lau, J.H., Cook, P., McCarthy, D., Newman, D. & Baldwin, T. 2012. Word sense induction for novel sense detection. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, April 2012, Avignon, France*: 591-601.

Treccani Web Portal. 2024. https://www.treccani.it/magazine/lingua_italiana/domande_e_risposte/lessico/lessico_727 .html. [12/01/2024].

<p style="text-align:center">*****</p>

Neology and Nomenclature – Concept clarification as a precursor to computational lexicography

Peter Juel HENRICHSEN (pjh@dsn.dk)
Danish Language Council, Bogense, Denmark

The Danish Language Council (DSN) is currently developing new concepts and procedures for elicitation of neologisms, with a keen focus on computability. We present some of our newest developments for sharing and discussion.

DSN is responsible for maintaining Retskrivningsordbogen (RO), the dictionary defining the Danish orthographic norm (Schack 2012). RO develops rather slowly, adopting a few hundred new lexemes each year at most, in contrast to the enormous daily production of neologisms in the media and everywhere else. This calls for effective selection criteria. DSN have traditionally used qualitative judgments for verification (Andersen et al. 2014, Jensen et al. 2024), but today's intense media stream has made computational methods necessary. Such methods require formally tightened definitions of certain concepts. In the following, we briefly introduce a few central terms.

An 'exologism' is, basically, a word form appearing in a corpus *C*, though absent in a dictionary *D*. Thus, each pair of language resources, for example,

C_G = DAGW (Danish GigaWord Corpus, Derczynski et al. 2021)

D_R = Retskrivningsordbogen (63k lemmas),

generates a set of candidate tokens. For (C_G,D_R) the characteristic set includes: 'toogfyrre' (forty-two)

'Pretoria'   (Pretoria)

'kiwitærte' (kiwi-tart)

'øjebæ'   ('øje'+'bæ', eye+poo, ≈ugly-building)

Some of these tokens are not really foreign to D_R, merely absent for reasons of parsimony; numerals (e.g. 'toogfyrre') and proper names ('Pretoria') are examples of domains deliberately restricted in D_R. We define an exologism as a word form *W* appearing in a corpus *C* while unsupported in a dictionary *D* (formal/computational definitions are in the paper) and a neologism as an exologism in (*C″*,*D″*) where *C″* faithfully represents contemporary language production, and *D″* faithfully represents the shared vocabulary at an earlier state in time (e.g. Δt=10Y). Claiming a neologism *N* thus comes with an obligation to (*i*) establish a corpus *C″*, (*ii*) quantify *N* in *C″*, (*iii*) define Δt, (*iv*) compile a dictionary *D″*, and (*v*) prove that *N* is unsupported in *D″*. While these procedures are by no means trivial, most challenges are inherent to neology as such and must be addressed independent of methodology. We find that the formal approach supports division of labour (procedures *i-v*) and recycling of resources (*D″* and *C″*). It also allows a stricter classification of neologisms. Last but not least, it paves the way for software development.

A current example is DSN's application NeoClink (based on CLINK (Henrichsen 2024), a morphological text parser using categorial grammar and type logic). NeoClink is used for

unsupervised extraction of neologisms from text streams. Each input token is broken down into material components ('Morphs'), then analysed for morphological function ('Sequent') and semantic relations ('Semantics'). See table 1 for examples; further details are in the paper.

Token	Morphs	Sequent	Semantics	Class
"øjebæ"	[øje][][bæ]	N $X \setminus Y / Y$ N ==> N	bæ(øje)	DAN|DANderog
"antiwoke"	[anti][woke]	X/X A ==> A	¬(woke)	DANpre|ENG
"tjak"	[tjak]	X ==> X	?tjak	OOV

Table 1. NeoClink lexemes (reduced CLINK templates). OOV=*Out-of-vocabulary*.
X,Y,Z=category variables. N,A=*category constants* (in casu noun and adjective).

DSN's daily text feed from Infomedia (www.infomedia.dk, ≈800M tokens/year) and other text sources are screened on a regular basis using NeoClink, the resulting suggestion list evaluated against Nyordslisten (DSN's manually compiled list of neologisms, cf. Jensen et al 2023a, 2023b). NeoClink typically scores very high for recall (>0.9), meaning that most hand-picked candidates are also in NeoClink's output; however much lower for precision (0.2-0.4), NeoClink often over- accepting (*a*) exologisms as neologisms and (*b*) lexical redundancies as exologisms (*a:b*≈1:3). This profile makes NeoClink useful as a source of supply while the final decision about inclusion in RO of course remains with the responsible editor.

DSN's traditional classification of neologisms (Jarvad 1996) is mainly example-based and thus hard to implement. NeoClink's template-based analysis provides a computationally feasible alternative (cf. table 1, 'Class').

Apart from supporting DSN's software development, the semi-formal take on neology has also facilitated our communications, not only internally (lexicographers, computational linguists, assistants), but also across institutional boundaries (e.g. in cooperation with university departments). This is not to say that neology has suddenly become easy; but at least some of the problems now have names.

References

Andersen, M., et al. (eds.). 2014. *Neologismer. Dansk Sprognævns 2. seminar om nye ord. København 5-6 November 2013*. Bogense: DSN-Publications.

Derczynski, L., et al. 2021. The Danish gigaword corpus. Dobnik, S. & L. Øvrelid (eds.). *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), May 31–2 June, 2021, Reykjavik, Iceland.*

Henrichsen, P.J. 2022. Det Centrale Ordregister. Et indeks over det danske sprog. *Proceedings Nordiska Forening i Lexikografi 16, 27- 29 April 2022, Lund, Sweden*: 113-126.

Jarvad, P. 1996. Nye ord - Hvorfor og hvordan? *Nydanske Sprogstudier* 21: 83–91.

Jensen, E.S. et al. 2023a. (eds.) *Nyordslisten. Nye ord i dansk 1955 til i dag.* Bogense: DSN-Publications. https://dsn.dk/ordboeger/nye-ord-i-dansk.

Jensen, E.S. et al. 2023b. Nye ord i dansk. *Proceedings Nordiska Forening i Lexikografi 17, 24–26 May 2023, Bergen, Norway.*

Schack, J. et al. 2012. *Retskrivningsordbogen.* Bogense: DSN-Publications. https://dsn.dk/ordboeger/retskrivningsordbogen

*****

Semi-Automatic Detection of New Words in Georgian

Tamar LALUASHVILI (tamar.laluashvili.1@iliauni.edu.ge)

Ilia State University, Tbilisi, Georgia

Tinatin MARGALITADZE (tinatin.margalitadze@iliauni.edu.ge)

Centre for Lexicography and Language Technologies, Ilia State University, Tbilisi, Georgia

The present study is a part of the three-year project, dedicated to the comprehensive study of neologisms in the Modern Georgian language. The work is supported by Shota Rustaveli National Science Foundation of Georgia (FR-23-4304). This paper aims to present the methodology worked out at Ilia State University for the semi-automatic detection of new words in Modern Georgian.

Changes, taking place in a language are usually very slow and difficult to notice. These changes occur over long periods before they become perceptible on a synchronic level. But there are exceptions from this general tendency and the contemporary Georgian language is a very good example of this. Currently, Georgia and the Georgian language are in an interesting era from a historical point of view. After the collapse of the former Soviet Union, our country saw the emergence and rapid development of a free market economy, multiparty political system, private banking sector, the Georgian national armed forces, etc. There developed intense relations with foreign countries on diplomatic, as well as on educational, economic and merely personal levels. These processes are only intensified by the increasing availability of computer, telecommunication and mobile technologies. Consequently, all prerequisites are in place which cause considerable changes in the Georgian lexis. The detection and study of latent diachronic processes taking place in the Georgian language becomes especially interesting at this linguistic-historical moment.

The study of neologisms is particularly relevant for lexicography, which, in addition to studying the theoretical aspects of the issue, also serves purely practical purposes. It involves updating existing dictionaries, adding new words, and assigning new meanings to the already existing ones. Contemporary users evaluate the quality of dictionaries by their ability to keep pace with the latest vocabulary and meanings. Dictionaries that fail to capture modern vocabulary tend to lose their appeal and popularity. The study of neologisms is also important as it reveals the lexical creativity of a language at a certain stage of its development.

Neologisms introduced into the Georgian language have been the focus of many Georgian researchers. However, there was no established methodology for identifying new words in modern Georgian. Therefore, we decided to work out such a methodology and for this purpose, we studied existing methods of detecting new words for other languages (Cabré & Nazar 2012; Janssen 2009; Kernerman & Klosa-Kückelhaus 2021) and formulated our approach.

For this project, a special corpus was composed including textual material from Georgian-language online newspapers and magazines, news websites, social media sites, websites of various governmental and non-governmental organizations, Georgian Wikipedia and some other sites, covering the last 15 years. Our methodology for the identification of Georgian new words is based on the combination of several approaches. We have applied the lemmatization tool for the Georgian language developed at Ilia State University. The lemmatizer is available on the website of the University at the URL https://qartnlp.iliauni.edu.ge (Lobzhanidze 2021). The lemmatizer relies upon the dictionaries, integrated with it, which are mainly composed of the word lists of the 8-volume *Explanatory Dictionary of the Georgian Language* and various other Georgian normative dictionaries. The lemmatization tool is capable of lemmatizing only those lexical units, which are included in these resources and serves as a kind of exclusion source allowing us to look for neologism candidates in unlemmatized, out-of-vocabulary (OOV) lexis. After the tokenization and lemmatization of the corpus material, the OOV lexical units are subjected to analysis and the potential neologisms are sampled. In order to reduce the number of OOV lexis, we applied to it another lemmatizer, developed by Paul Meurer (2014) for the

Georgian National Corpus (GNC, Gippert & Tandashvili 2015). The GNC contains over 200 mln tokens and P. Meurer's lemmatizer can recognize much more words than the one applied by us for the first phase as an exclusion source. As a result, the number of unlemmatized vocabulary was reduced by 45 %.

Georgian word embedding software proved to be very instrumental in finding additional neologism candidates (https://wordembedding.spellchecker.ge).

As a result of the study, we have identified 1000 lexical neologisms in Modern Georgian. The study of semantic development of existing words is planned for the next phase within the framework of the grant project. We mostly selected words that belong to the common vocabulary of Georgian, including colloquial words and slang. We also selected some general terms from the fields of social media, online media, tourism. New terms from specialized fields were excluded from the study.

Based on the research carried out: (a) an online dictionary of neologisms will be composed and published; (b) the methodology and tools for the detection of neologisms will be perfected; (c) a website will be set up for the monitoring of neologisms in future which will be published on the website of Ilia State University.

References
Cabré, M.T. & Nazar, R. 2012. Towards a New Approach to the Study of Neology. *Neologica* 6: 63-80.
Gippert, J. & Tandashvili, M. 2015. Structuring a diachronic corpus. The Georgian national corpus project. Gippert, J. & Gehrke, R. (eds). *Historical Corpora. Challenges and Perspectives. Corpus Linguistics and Interdisciplinary Perspectives on Language,* v. 5. Tübingen: Narr.
Janssen, M. 2009. Detección de Neologismos: una perspectiva computacional. *Debate Terminológico* 5: 68-75.
Kernerman, I. & Klosa-Kückelhaus, A. (eds.) 2021. Lexicographic Neology and Neological Lexicography. Special issue of *International Journal of Lexicography,* 34(3).
Lobzhanidze, I. 2021. *Principles of Morpho-Syntactic Annotation and Morphological Analysis of the Finite State*. Tbilisi: Ilia State University Publishing (in Georgian).
Meurer, P. 2014. The Morphosyntactic Analysis of Georgian. *Georgian National Corpus.*
http://gnc.gov.ge/gnc/page.

<div align="center">*****</div>

Lexical change and variation in the parenting domain: analysis and terminographic proposal

Paola CAÑETE-GONZÁLEZ (pcanete@udec.cl)
Departament of Foreign Languages, Universidad de Concepción, Concepción, Chile
Elisabet LLOPART-SAUMELL (elisabet.llopart@upf.edu)
Institute for Applied Linguistics, Universitat Pompeu Fabra, Barcelona, Spain

As a result of development, society has experienced vertiginous changes in different domains that, undoubtedly, have an impact on our daily lives. Individuals have been challenged by new traditions, ideas, and ways to communicate with each other, among other changes. Because of these new trends, speakers are exposed to new words that have appeared, precisely, to denominate all these new realities. An area that has experienced relevant innovations in the last decades is parenting (Molina 2006). For this reason, the emergence of new words in this sphere is only the tip of the iceberg, since it shows how parents and, consequently, society are increasingly more interested in learning skills and methods on how to take care of and educate their children (Wileczek 2020). Nevertheless, no specific studies devoted to neology of this domain can be found, unlike what happens in the case of other

social phenomena, such as the coronavirus pandemic (Klosa-Kückelhaus and Kernerman 2022, Luna 2022, Ponce de León 2021, Rodríguez 2023) or the feminist movement (Novoa and Rebolledo 2019, Gandarias 2017 or Di Paolo 2022).

Nowadays, there are plenty of magazines, blogs, and books that offer guidelines on this broad topic, but also social networks are filled with this type of content. These publications are a valuable source of information to study the new lexicon related to parenting. That said, the main objectives of this paper are, in the first place, to observe the changes experienced in this field by analyzing the terms in Spanish language that have appeared in the last few years. For example, regarding the positive upbringing methods, we can find neonyms such as *crianza positiva* 'positive parenting' (CDC 2024), *crianza respetuosa* 'respectful parenting' (El País 2018, Eres Mamá 2023, Suavinex 2022), *crianza natural* 'natural parenting' (Criar con Sentido Común 2022), *crianza afectiva* 'affective parenting' (UNICEF n.d.), *crianza con apego* 'attachment parenting' (Bebés y más 2021), and so on. Therefore, an analysis of the social context is conducted to correlate the emergence of different groups of neonyms and the changes experienced by our society in this domain. In the second place, there is also an applied objective which consists in the design of a glossary proposal devoted to parenting.

The theoretical framework in which this work is based takes into account both neological and terminological studies (Cabré 1992, L'Homme 2004, Sager 1990, Termcat 2010), since the object of study is mainly neonyms, that is, specialized neologisms. Besides, we also follow the typical steps of any terminographic work. To do that, we will collect a corpus of texts from different textual genres (books, academic articles, specialized magazines, and blogs) written in the XXI century. Once the corpus is created, neonyms will be extracted both in a semiautomatic way (using a terminological management tool) and manually, to detect formal and semantic neonyms, from these sources. The neologicity of the units will be determined by psycholinguistic and lexicographical criteria (Rey 1976, Cabré 1992). Later, all the new terms collected will be classified in different thematic domains, such as feeding, sleeping, hygiene, behavior, among others, with the aim of building up a conceptual map of this areaof knowledge. Afterwards, the terminological variation, both conceptual and denominative, will be analyzed in more detail due to the different theories and schools from which the terms related to parenting come from. In this preliminary proposal we show that, on the one hand, there is a gap in this field, since none of the glossaries available (Bebés y más 2019, Educapeques 2017, Infancia en positivo n.d., Madresfera n.d., Mamma Bra 2023, Puigdemasa y Villanueva 2023, Suavinex n.d.) follow the terminological standards, and that many times the dissemination of this vocabulary is made by the so-called experts that, in fact, are not specialists in the field. For this reason, one of the contributions of the glossary will be to transfer quality knowledge, that is, based on specialists and institutions on parenting.

References

Bebés y más. 2021. *Las ocho "B" de la crianza con apego.* https://www.bebesymas.com/ser-padres/ocho-b-crianza-apego-1. [15/02/2024].

Bebés y más. 2019. *¿Padres primerizos? Este es el glosario de términos que utilizarás con frecuencia en los primeros meses de tu bebé.* https://www.bebesymas.com/ser-padres/padres-primerizos-este-glosario- terminos-que-utilizaras-frecuencia-primeros-meses-tu-bebe. [15/02/2024].

Cabré, M. Teresa. 1992. *La terminologia. La teoria, els mètodes, les aplicacions*. Barcelona: Empúries.

Criar con Sentido Común. 2022. *Crianza natural del bebé: qué es y cómo aplicarla.* https://www.criarconsentidocomun.com/crianza-natural-del-bebe/. [15/02/2024].

Di Paolo, Marilicia. 2022. El feminismo y sus implicaciones en el lenguaje español: el caso de los "man-neologisms". *Investigación y género*: 51-64.

Educapeques. 2017. *Diccionario para padres y madres.*https://www.educapeques.com/escuela-

de-padres/diccionario-para-padres.html. [15/02/2024].

El País. 2018. *La crianza respetuosa: ¿qué significa realmente?*
https://elpais.com/elpais/2018/06/01/mamas_papas/1527850952_523895.html. [15/02/2024].

Eres Mamá. 2023. *¿Qué significa la crianza respetuosa?* https://eresmama.com/significa-la-crianza-respetuosa. [15/02/2024].

Gandarias Goikoetxea, Itziar. 2017. ¿Un neologismo a la moda?: Repensar la interseccionalidad como herramienta para la articulación política feminista. *Investigaciones Feministas* 8(1):73-93.

Infancia en positivo. n.d. *Términos básicos en disciplina positiva.*
https://infanciaenpositivo.com/terminos-basicos-en-la-disciplina-positiva- destaparse-conectar. [15/02/2024].

Klosa-Kückelhaus, Annette and Kernerman, Ilan. 2022. *Lexicography of Coronavirus- related Neologisms*, Berlin/Boston: De Gruyter.

L'Homme, Marie-Claude. 2004. *La Terminologie: principes et techniques*. Montréal: Les Presses de l'Université de Montréal.

Luna, Rosa. 2022. Comportamiento de la neología en tiempos de coronavirus. *Boletín de la Academia Peruana de la Lengua* 71: 301-330.

Madresfera. n.d. *Glosario Madresférico.* https://www.madresfera.com/glosario/.

Mamma Bra. 2023. *Glosario de términos sobre la lactancia materna.*
https://mammabra.com/blog/glosario-de-terminos-sobre-lactancia- materna.html. [15/02/2024].

Molina, María Elisa. 2006. Transformaciones Histórico Culturales del Concepto de Maternidad y sus Repercusiones en la Identidad de la Mujer. *Psykhe* 15(2): 93-103.

Novoa, Sergio and Rebolledo, Valentina. 2019. *Feminismo e innovación léxica: ¿qué hay de nuevo en la prensa chilena?* Tesina para optar al grado de Licenciado en Traductología, Universidad de Concepción.
http://repositorio.udec.cl/xmlui/bitstream/handle/11594/10140/rebolledo_w_v.p df?sequence=1&isAllowed=y. [15/02/2024].

Ponce de León, Ramón. 2021. Neologismos relacionados con la pandemia de COVID-19 en México. *Lingüística Mexicana. Nueva Época* 4(1): 7-29.

Puigdemasa, Marta and Villanueva, Paola. 2023. *Madrebulario*. Madrid: Planetadelibros.

Rey, Alain. 1976. Le néologisme: un pseudoconcept? *Cahiers de Lexicologie* 28: 3-7.

Sager, J. C. 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.

Suavinex. 2022. C*rianza respetuosa: qué es y sus beneficios.*
https://www.suavinex.com/livingsuavinex/crianza-respetuosa. [15/02/2024].

Suavinex. n.d. *Glosario de maternidad: embarazo y posparto.*
https://www.suavinex.com/livingsuavinex/glosario-maternidad/. [15/02/2024].

Termcat. 2010. *El diccionari terminològic*. Vic: Eumo Editorial; Barcelona: Termcat. Recuperado de https://arxiu.termcat.cat/enprimerterme/el-diccionari- terminologic.pdf. [15/02/2024].

UNICEF. n.d. *Preguntas frecuentes sobre crianza positiva.*
https://www.unicef.org/mexico/preguntas-frecuentes-sobre-crianza-positiva. [15/02/2024].

Wileczek, Anna. 2020. Linguistic images of modern parenthood on parenting websites. Kwartalnik Pedagogiczny 254(4): 197-215.

<center>*****</center>

Designing a Digital Lexicon of Poetic Neologisms: A Case Study in Kazantzakis' Work

Nikos MATHIOUDAKIS (nikosmathious@gmail.com)
Department of Modern Greek Studies, University of Belgrade, Belgrade, Serbia

In the present study we deal with the design, creation and development of a digital dictionary of poetic neologisms in the modern Greek, having taken as a case study the poetry of Nikos Kazantzakis, especially his Odyssey. Our aim is to describe the construction of a specialized digital dictionary, in the absence of specialized dictionaries of Greek authors, but also as a model for the development of similar literary dictionaries.

For Modern Greek, several scholars, among them A. Christofidou, E. Kollia, Ch. Charalampakis and N. Mathioudakis, are engaged in the search for neologisms in literary texts, characterizing specific poetic neologisms as an element of the author's style. These

researchers also try to identify cases in which readers can understand the importance of new formations or existing ones with other meanings.

Regarding the creation of a specialized dictionary, we rely on modern theories of specialized lexicography (Bergenholtz & Tarp 1995; Sterkenburg 2003; Nielsen & Tarp 2009; Fuertes-Olivera & Tarp 2014) – with the use of new technologies in the case of dictionary development in a digital environment. According to Tarp & Gouws (2019), contextualization and personalization are a major issue in modern lexicography.

The Digital Lexicon of Neologisms of Kazantzakis (LeNEKAZ) is a pioneering tool as it is the first digital dictionary for modern Greek authors. Also, it is set to become an easy-to-use and functional digital reference hub for both the academic and educational community, since it will be online and with open access. Kazantzakis was revealed as a wordhunter, an amateur linguist, who has created thousands of words that are not yet recorded in dictionaries (undictionaried words, Mathioudakis [2012]2020) – mostly compounds and multi-compounds words (Mathioudakis & Karasimos 2014, 2023). The LeNEKAZ was created based on the Kazantzakian epic Odyssey, but it can be extended to the rest of his work. After all, this specific poem is a work of reference, regarding his language and his idiosyncratic vocabulary.

The platform is implemented using open-source software, based on modern technologies for the construction of word treasures and other technologies of the semantic web, with integrated functions of annotation, documentation (in-context) and search for words. According to the standard mark-up and information encoding support infrastructure, the content of the dictionary is prepared by encoding the entries and linking each entry to its context. The design and construction work are divided into three sections: (a) Information Architecture, (b) Design, and (c) Development.

The macrostructure of LeNEKAZ contains approximately 4,500 neologisms, as lemmas, that were collected from the Odyssey of Nikos Kazantzakis and are morphological formations, which were not found in any of the seven basic dictionaries of modern Greek. The microstructure includes a detailed description of the lexicographic information of the lemma, which are divided into ten parts: (1) the lemma, (2) the part of speech (noun, verb, adjective, proverb), (3) the interpretation, (4) the morphological analysis, (5) the frequency in the poem, (6) the word type in the poem, (7) the position in the poem, (8) the context from Odyssey, (9) the code of the verse (in digital form by crosslinking), and (10) the English translation of the lemma by Kimon Friar.

The phenomenon of poetic neology is a timeless issue which develops and transforms according to the needs of each era. New words are constantly synthesized morphologically (or even existing words change meaning) in order to satisfy the linguistic sense of each author, poet or novelist. In direct contrast, there is a need to create specialized dictionaries and/or glossaries to record idiosyncratic literary formations, as they are part of our daily lives, while they can be an unexplored field of study, revealing linguistic and stylistic elements for a literary text, but also highlighting elements of our own language and its mechanisms, especially at the level of derivation and compounding.

References
Bergenholtz, H. & Tarp, S. (eds.). 1995. *Manual of Specialised Lexicography: The preparation of specialised dictionaries*. Amsterdam: John Benjamins Publishing Company.
Fuertes-Olivera, P. A. & Tarp, S. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin: De Gruyter.
Mathioudakis, N. [2012]2020. *The Odys(s)ey of Words. Neological Undictionaried Words in Nikos Kazantzakis' epic* [in Greek]. Athens: Kapa Publishing House.

Mathioudakis, N. & Karasimos, A. 2014. Dialectic and Idiomatic Aspects in Odyssey by Nikos Kazantzakis. *Selected Papers of the 11th International Conference on Greek Linguistics*: , 1070-1088. Rodes, Greece: University of Aegean.

Mathioudakis, N. & Karasimos, A. 2023. The dialectical and neological "Odys(s)eia" of Nikos Kazantzakis: a morphological and stylistic approach [in Greek]. *Glossologia* 30 (2022-2023): 141-163.

Nielsen, S. & Tarp, S. (eds.). 2009. *Terminology and Lexicography Research and Practice*. Amsterdam: John Benjamins Publishing Company.

Van Sterkenburg, P. (ed.). 2003. *A Practical Guide to Lexicography*. Amsterdam: John Benjamins Publishing Company.

Tarp, S. & Gouws, R. 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.