AFRICAN ASSOCIATION FOR LEXICOGRAPHY

30th International Conference
1 – 4 July 2026

AFRICAN ASSOCIATION FOR LEXICOGRAPHY

30th International Conference

1 – 4 July 2026

Book of Abstracts

Hosted by: The Tshivenda National Lexicography Unit, University of Venda

Conference coordinator: Mulalo Takalani

Abstract committee: Blanche Assam, André du Plessis and Lorna Morris

Abstract reviewers: Blanche Assam, Herman Beyer, Maret Blom De Wet, Sonja Bosch, Theo Bothma, Ruth De Oliveira, André du Plessis, Wanga Gambushe, Michelle Goosen, Rufus Gouws, Paul Mavoungou, Lorna Morris, Eventhough Ndlovu, Thapelo Otlogetswe, Danie Prinsloo, Justus Roux, Elsabé Taljard

Abstract booklet editors: André du Plessis and Lorna Morris

ISBN 978-1-0483-3254-4
© 2026 AFRILEX, African Association for Lexicography

Table of Contents

AFRILEX HONORARY MEMBERS 1

AFRILEX BOARD 2

MESSAGE FROM THE AFRILEX PRESIDENT 2

AFRILEX CONFERENCE PROGRAMME ..... **Error! Bookmark not defined.**

KEYNOTE PRESENTATION 1 10

AFRILEX at Thirty
Gilles-Maurice DE SCHRYVER

KEYNOTE PRESENTATION 2 11

On AI and the Future of African Lexicography
Langa KHUMALO

PRECONFERENCE WORKSHOP 12

CONFERENCE SPECIAL SESSIONS 12

PRESENTATIONS 13

Transition to Online Dictionaries
Thato BALOYI, Karabo MAFETSA & Kganathi SHAKU 13

Data Mining the *Greater Dictionary of IsiXhosa*: Lexical Stratification and Digital Lexicography in IsiXhosa
William BENNET, Camilla CHRISTIE, Dion NKOMO & Bulelwa NOSISLELA 14

When Lexicographical Communication Malfunctions: The Case of Fraction Names in the *Afrikaanse Woordelys en Spelreëls*
Herman BEYER 16

Lexicographic Treatment and Discursive Legacy: The French Lexeme *nègre*
Ruth DE OLIVEIRA 17

Restructuring Semantic Information for Cognitive Accessibility: The Case of *e-Le Petit Robert*
Emmanuel FOMAT 19

Teaching Lexicography at Undergraduate Level: Student Perceptions and Dictionary Culture in an isiXhosa Classroom
Wanga GAMBUSHE 21

The Role of Dictionaries in the Intellectualisation of isiNdebele
Pearl KABINI & Nontokozo MDHLULI 22

Generative AI and Lemma Treatment in Sepedi: Testing M365 Copilot (GPT-5), ChatGPT (GPT-5.5) and DeepSeek-v3
Napjadi LETSOALO 24

Data Management for an e-Lexicography Project
Jimmy MABINA & Makwarela MUDAU 25

Comprehensive Research on Terminology Development, Verification and Authentication in Sesotho sa Leboa
Mapula MADIBA 27

Digital Dictionaries as Inclusive Knowledge Infrastructures: Reimagining Lexicographic Accessibility in the Global South
Muimeleli MAEBA 29

Maximising the Development of Tshivenda Terminology as a Tool for Sustainable Development Goal 4 (Quality Education) in Rural Limpopo
Mashudu MAKHADO 30

Evaluating the Status of Translation Equivalents in Selected Northern Sotho-English Bilingual Dictionaries
Mmagonkahloleng MAKUA 32

Examining How Technological, Scientific and Cultural terms are Treated in Terms of Translation Equivalents in Bilingual Dictionaries
Matlala MAKWALA & Mmagonkahloleng MAKUA 33

Thirty Years of African Lexicography: The Contribution of AFRILEX to Indigenous Language Development in South Africa
Masila MASIPA & Nontokozo MDHLULI 35

Towards Inclusive Lexicography: The Need for User-Friendly Tshivenḓa Dictionaries in Education and Society
Mashudu MAṰHABI 36

An Approach to Digital African Lexicography That Can Help Bridge the Accessibility Gap
Nontokozo MDHLULI & Pearl KABINI 37

The *Greater Dictionary of IsiXhosa* and Standardisation in IsiXhosa Writing
Buyiswa MINI 38

User Studies in the South African Lexicographic Laboratory: Methodologies and Case Studies
Lorna MORRIS 40

Digital Indigenisation of Terminology: How Community-Based Media Platforms in Limpopo Generate and Disseminate Economic Neologisms in Tshivenda and Xitsonga
Mutshidzi MUNYA 41

Exploring Female Colloquial Reference in South Africa: A case of *Mzansi Taal* Online Dictionary
Neliswa MZIMBA & Napjadi LETSOALO 43

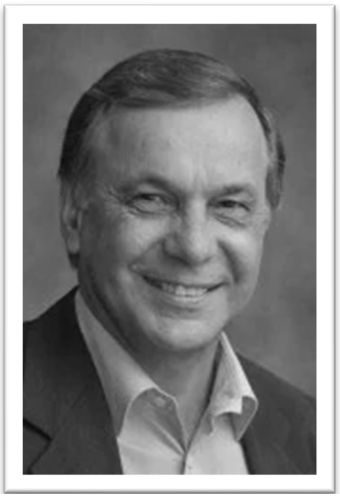
Implementation Progress of the Language Plan of Action for Africa in Namibia: A Lexicographer's Perspective
Eventhough NDLOVU 44

The Van Warmelo Dictionary Card Collection: The Use of Artificial Intelligence for Lexicographic Data Enrichment
Danie PRINSLOO & Elsabé TALJARD 46

Proverbs and Idioms as Knowledge Systems: How Should They be Treated Lexicographically?
Tshifhiwa RAMURUNZI & Shumani TSHIKOTA 47

A Corpus-Based Approach to the Lemmatization of Anthroponyms in a Xitsonga Monolingual Dictionary
Ndzalama VUMA & Madala HLUNGWANI 49

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

Prof. E. Taljard

AFRILEX BOARD

2025 – 2027

President:	Prof. D. (Dion) Nkomo
Vice-President:	Prof. H.S. (Steve) Ndinga-Koumba-Binza
Secretary:	Dr L.H. (Lorna) Morris
Treasurer:	Prof. E. (Elsabé) Taljard
Editor *Lexikos*:	Dr A.H. (André) du Plessis
Members:	Prof. H.L. (Herman) Beyer
	Dr B. (Blanche) Assam
	Dr W. (Wanga) Gambushe
Conference organiser:	Dr Mulalo Takalani

MESSAGE FROM THE AFRILEX PRESIDENT

This 30th International Annual Conference of the African Association for Lexicography (AFRILEX) is hosted at the University of Venda (UNIVEN) by the Tshivenḓa National Lexicography Unit. This is the second time that the Tshivenḓa National Lexicography Unit is hosting AFRILEX, having done so in 2006. On behalf of the AFRILEX Board, I am deeply honoured and enthused to welcome all the delegates and thank the leadership of our host institution.

This edition of the AFRILEX annual conference is historic as our association celebrates thirty-one years of existence, and for the 30th time, brings together its members and friends from across the globe to celebrate lexicographic accomplishments, share challenges and ponder on lexicographic prospects at a time when technological advancements, particularly the artificial intelligence (AI) of the 4th Industrial Revolution, challenge us to re-imagine the dictionary, dictionary-making, dictionary use and dictionary research. Hence the befitting conference theme "30 Years of African lexicography: perspectives and engagement".

Leading our celebrations, reflections and engagements with various issues of pertinence to our profession and scholarship will be our keynote speakers, Prof. Gilles-Maurice de Schryver and Prof. Langa Khumalo. Both are scholars of international repute and, most significantly, former presidents of AFRILEX, who led our association with great distinction. The AFRILEX Board is grateful to our keynote speakers for graciously accepting our invitation. Prof. De Schryver will also facilitate our pre-conference workshop, drawing insights from the legendary English lexicographer, Patrick Hanks (1940 – 2024) to help us re-imagine the lexicographic profession in the age of AI.

AFRILEX always deliver rich conference programmes that reflect our linguistic diversity, the complexities of our discipline and our association's international pool. This year is no

exception. Our programmes have also become more special because of their special sessions. Through special sessions, AFRILEX continues to pursue its founding objective to be at the forefront of African lexicography. In South Africa, the establishment of National Lexicographic Units (NLUs) represented an intentional positioning of lexicography at the heart of post-apartheid nation building through the promotion multilingualism, particularly the historically marginalised Indigenous African languages. At its formative years, AFRILEX provided NLUs with the scholarly brainpower in terms of the necessary training, guidance and reflections on various practical tasks confronting the NLUs as they seek to fulfil their mandate. These engagements have continued through the regular NLUs special sessions and joint celebrations of the International Dictionary Day, with the Tshivenḓa National Lexicography Unit hosting the most recent one in October 2025, before hosting this year's conference. As we converge in Thohoyandou, anticipate our more structured and dialogic NLUs session that seeks to promote a more flexible and richer engagement around critical issues in South African lexicographic practice.

Another major significant feature of this year's conference is a roundtable discussion featuring AFRILEX veterans comprising former presidents and founding members of the organisation. These are the giants on whose shoulders we stand tall today. Three decades after the inaugural conference that was held at Stellenbosch University, we certainly need a reminder of how far we have come, especially as we contemplate the future of African lexicography and lexicography as a profession. In that future, we need our publishers, and the publishers' session will hopefully shed some light on the dynamics that characterise the dictionary market. Lastly, AFRILEX's first conference workshop for local teachers will recognise the pivotal role of teachers in dictionary use, the development of vibrant dictionary culture and the overall lexicographic enterprise.

In appreciation of all this, the Board would like singularly to thank our keynote speakers, every conference delegate, including our AFRILEX veterans and the colleagues from the NLUs and publishers, as well as the schoolteachers, who availed themselves for the special sessions, for what would be a historic conference. We congratulate the NLUs and publishers for the dictionaries they produced over the past year, noting in particular the achievements of *Woordeboek van die Afrikaanse Taal* as they celebrated 100 years of success.

On another note, the AFRILEX Board continues to mourn the passing of Honorary Member Dr J.C.M.D. du Plessis on 8 April 2026. Dr Du Plessis remains the longest serving editor of AFRILEX's mouthpiece, *Lexikos*, having assumed the role between 1997 and 2010. Remarkably, he served the WAT for 63 years. The AFRILEX Board signed a Memorandum of Agreement with the WAT to enhance the sustainability of *Lexikos* and hope that this will preserve the legacy of Dr Du Plessis and its several founding members. Conference delegates are invited to develop their presentation and submit it for possible publication in *Lexikos*.

A quick word about our host city Thohoyandou. This is the land of King Thohoyandou who reigned over the Vhavenḓa monarchy in the late 17th and early 18th centuries. In Tshivenḓa, Thohoyandou literally means 'head of an elephant'. King Thohoyandou was a great warrior and unifier who expanded the precolonial Vhavenḓa nation. Onomastically, the city thus remains organically rooted in Vhavenḓa history. The city also enjoys geographical proximity with major tourist attractions. During our historic conference, we will be surrounded by mountains, valleys, indigenous forests and South Africa's most iconic natural and cultural heritage sites that include the mystical Lake Fundudzi, the sacred Thathe Vondo Forest and the Mapungubwe heritage site. We will be close to the Mapungubwe Heritage Site, our destination for the post-conference excursion. Overall, our conference is set to be a great

learning experience.

Accordingly, the Board express its gratitude to the leadership of UNIVEN for offering to host our conference, and Dr Mulalo Esther Takalani for being our ambassador and advocate. In turn, I would like to end by thanking all the members of the AFRILEX Board for their tireless commitment to individual tasks leading to this historic conference. From Thohoyandou, may we derive the wisdom, leadership and strength that will sustain our association and the discipline long into the future. *Ndo livhuwa*!

Dion Nkomo
President: AFRILEX

AFRILEX CONFERENCE PROGRAMME

30th International Conference of the African Association for Lexicography
1 – 4 July 2026, Tshivenḓa National Lexicography Unit, University of Venda, Thohoyandou, South Africa

	WEDNESDAY 1 JULY	
09:00-10:30	**PRECONFERENCE WORKSHOP** ***Generative AI & Lexicography: From the Insights of Patrick Hanks (1940-2024) to Chatbot Lexicography*** Gilles-Maurice de Schryver Venue: School of Education, Lecture Hall 01	
10:30-11:00	**TEA**	
11:00-13:00	**PRECONFERENCE WORKSHOP CONTINUED** Venue: Lecture Hall 01	
13:00-14:00	**CONFERENCE REGISTRATION** Venue: Foyer of Lecture Hall 01	
14:00-14:30	**OFFICIAL OPENING CEREMONY** (Chair: Dr T.D. Raphalalani) **Welcome**: Prof. N.B. Nthambeleni, Vice Chancellor & Principal: Univen **A Word from AFRILEX**: Prof. D. Nkomo, President: AFRILEX Venue: Lecture Hall 01	
Parallel Sessions, School of Education		
	Lecture Hall 02 (Chair: Steve Ndinga-Koumba-Binza)	**Lecture Hall 03** (Chair: Eventhough Ndlovu)

14:30-15:00	***Lexicographic Treatment and Discursive Legacy: The French Lexeme nègre*** (Ruth DE OLIVEIRA)	***Evaluating the Status of Translation Equivalents in Selected Northern Sotho-English Bilingual Dictionaries*** (Mmagonkahloleng B. MAKUA)
15:00-15:30	***Restructuring Semantic Information for Cognitive Accessibility: The Case of the Petit Robert*** (Emmanuel S. DOMAT)	***The Role of Dictionaries in the Intellectualisation of isiNdebele*** (Pearl L. KABINI & Nontokozo G. MDHLULI)
15:30-16:00	**TEA**	
	Lecture Hall 02 (Chair: Rufus Gouws)	**Lecture Hall 03** (Chair: Elsabé Taljard)
16:00-16:30	***Data Mining the Greater Dictionary of IsiXhosa: Lexical Stratification and Digital Lexicography in IsiXhosa*** (William G. BENNET, Camilla R. CHRISTIE, Dion NKOMO & Bulelwa NOSISLELA)	***Maximising the Development of Tshivenda Terminology as a Tool for Sustainable Development Goal 4 (Quality Education) in Rural Limpopo*** (Mashudu P. MAKHADO)
16:30-17:00	***Teaching Lexicography at Undergraduate Level: Student Perceptions and Dictionary Culture in an isiXhosa Classroom*** (Wanga GAMBUSHE)	***Digital Indigenisation of Terminology: How Community-Based Media Platforms in Limpopo Generate and Disseminate Economic Neologisms in Tshivenda and Xitsonga*** (Mutshidzi P. MUNYAI)
17:00-17:30	***The Greater Dictionary of IsiXhosa and Aspects of Standardisation in IsiXhosa Writing*** (Buyiswa MINI)	***Comprehensive Research on Terminology Development, Verification and Authorisation in Sesotho sa Leboa*** (Maphuta J. MADIBA)
18:30	**WELCOMING COCKTAIL** Venue: University Hall	

	THURSDAY **2 JULY**
09:00-09:05	**ANNOUNCEMENTS** School of Education, Lecture Hall 01
09:05-10:00	**KEYNOTE ADDRESS 1**

	Lecture Hall 01 (Chair: Prof. Dion Nkomo) ***AFRILEX at 30*** Prof. Gilles-Maurice de Schryver	
10:00-10:30	**TEA**	
Parallel Sessions, School of Education		
	Lecture Hall 02 (Chair: Wanga Gambushe)	**Lecture Hall 03** (Chair: André du Plessis)
10:30-11:00	***Examining How Technological, Scientific and Cultural Terms are Treated in Terms of Translation Equivalents in Bilingual Dictionaries*** (Matlala R. MAKWALA & Mmagonkahloleng B. MAKUA)	***The Van Warmelo Dictionary Card Collection: The Use of Artificial Intelligence for Lexicographic Data Enrichment*** (Danie J. PRINSLOO & Elsabé TALJARD)
11:00-11:30	***Generative AI and Lemm[illegible] in Sepedi: Testing M365 Copilot (GPT-5), Chat[illegible]T-5.5) and DeepSeek-v3*** (Napj[illegible]i LETSOALO)	***Transition[illegible] Onli[illegible] Dictionaries*** (M. Thato BALOYI, Kara[illegible]ETSA & Kganathi SHAKU)
11:30-12:00	***An Approach to Digital African Lexicography That Can Help Bridge the Accessibility Gap*** (Nontokozo G. MDHLULI & Pearl L. KABINI)	***User Studies in the South African Lexicographic Laboratory: Methodologies and Case Studies*** (Lorna MORRIS)
12:00-12:30	***Data Management [illegible]an e[illegible]xicography Project*** (Jimmy MABINA[illegible]warela MUDAU)	***When Lexicographical [illegible]ommun[illegible]ation Malfunctions: The Case of Fraction names[illegible]frikaanse woordelys en [illegible]pelre[illegible]*** (Herman L. BEYER)
12:30-13:30	**LUNCH** Foyer, School of Education	
	Lecture Hall 02 (Chair: Blanche Assam)	**Lecture Hall 03** (Chair: Herman Beyer)
13:30-14:00	***A Corpus-Based Ap[illegible]ach t[illegible]e Lemmatization of Anthroponyms in Xits[illegible]onolingual Dictionary*** (Ndzalama VUMA & [illegible]adala [illegible]ous HLUNGWANI)	***Proverbs and Idioms as Knowledge Systems: How Should They Be Treated Lexicographically?*** (Tshifhiwa RAMURUNZI & Shumani L. TSHIKOTA)

14:00-14:30	***Digital dictionaries as inclusive knowledge infrastructures: reimagining lexicographic accessibility in the global south*** (Mahmeleli NAEBA)	***Exploring Female Collocational Reference in South Africa: A Case of Mzansi Online Dictionary*** (Neliswa MZIMBA & Napyadi LETSOALO)
14:30-15:00	***Towards Inclusive Lexicography: The Need for User-Friendly Tshivenḓa Dictionary in Education and Society*** (Mahudu MATHABI)	***Implementation Progress of the Language Plan of Action for Africa in Namibia: A Lexicographer's Perspective*** (Eventhough NDLOVU)
15:00-15:30	**TEA**	
	Lecture Hall 02 (Chair: Ruth De Oliveira)	
15:30-16:00	***Thirty Years of African Lexicography: The Contribution of AFRILEX to Indigenous Language Development in South Africa*** (Masila J. MASIPA & Nobukozo G. MDHLULI)	
End of parallel sessions		
16:00-18:00	**AFRILEX ANNUAL GENERAL MEETING** Venue: Lecture Hall 01	
18:30	**CONFERENCE DINNER** Venue: University Hall	

	FRIDAY **3 JULY**
09:00-09:10	**ANNOUNCEMENTS** Venue: School of Education, Lecture Hall 01
09:10-10:00	**KEYNOTE ADDRESS 2** School of Education, Lecture Hall 01 (Chair: Prof. H. Steve Ndinga-Koumba-Binza) ***On AI and the Future of African Lexicography*** Prof. Langa Khumalo
10:00-10:30	**TEA**
Special sessions, School of Education	

10:30-11:30	**SPECIAL SESSION 1:** **Publishers' Session** Lecture Hall 01 (Chair: Lorna Morris) CASAS, OUPSA, SAK & SANLU
11:30-13:00	**SPECIAL SESSION 2:** **NLU Session** Lecture Hall 01 (Chair: Wanga Gambushe) South African National Lexicography Units
13:00-14:00	**LUNCH**
14:00-15:30	**SPECIAL SESSION 3:** **30 Years of AFRILEX Conferences: A Veteran's Perspective** Lecture Hall 01 (Chair: Dion Nkomo) Mariëtta Alberts, Herman Beyer, Gilles-Maurice de Schryver, Rufus Gouws, Danie Prinsloo & Elsabé Taljard
15:30-16:00	**TEA**
16:00-16:15	**AFRILEX CLOSURE** Lecture Hall 01 Prof. Dion Nkomo

	SATURDAY **4 July**
05:00	CONFERENCE EXCURSION Meeting Point: African Languages Building, Parking

***Papers marked with red cross were accepted but not presented. Presenters did not provide valid excuse for not attending the conference.**
****Papers marked with a blue cross were accepted but not presented as the presenters had a medical excuse for not attending the conference.**

KEYNOTE PRESENTATION 1

AFRILEX at Thirty

Gilles-Maurice DE SCHRYVER (GillesMaurice.DeSchryver@UGent.be)
BantUGent (the UGent Centre for Bantu Studies), Ghent University, Belgium

The title of this keynote lecture is short, to mimic my earlier research article known as '*Lexikos* at 18', which analysed the journal that is the mouthpiece of AFRILEX. In short, this presentation will be a variation of quite a number of my earlier talks and publications describing the outcomes of quantitative and qualitative studies and systematic investigations to map, measure, and analyse the literature, research trends, and scholarly impact within the field of lexicography. In the past, I worked on 'Bibliometrics in Lexicography' (De Schryver 2009a), 'Lexikos at 18' (De Schryver 2009b), 'Euralex at 30' (De Schryver 2012a), 'The International Journal of Lexicography at 25' (De Schryver 2012b), 'Asialex at 20' (De Schryver 2019), 'Asian lexicography at 30' (De Schryver 2026), and now with this keynote 'Afrilex at 30'. All these studies can be represented as follows in table format:

	ALL	AFRILEX	EURALEX	ASIALEX
Conference	'Bibliometrics in Lexicography' – *invited* (De Schryver 2009a)	'AFRILEX at 30' – *this keynote*	'EURALEX at 30' – *keynote* (De Schryver 2012a)	'ASIALEX at 20' – *keynote* (De Schryver 2019)
Journal		'*Lexikos* at 18' – *invited* (De Schryver 2009b)	'*Int. Jnl. of Lex.* at 25' – *invited* (De Schryver 2012b)	'Asian Lex. at 30' – *invited* (De Schryver 2026)

As one can notice, over the years I started by proving beyond any reasonable doubt that lexicography is a scientific discipline in and of itself, snaking my way through three continents with their respective journals and conferences, to finally come full circle — from African text to African speech. Once, I also ventured into metalexicography, covering the globe (De Schryver 2023). More specifically, during this keynote I will focus especially on the intersection of metalexicographers, lexicographers and software developers under 30 years of the AFRILEX umbrella.

References

De Schryver, G.-M. 2009a. Bibliometrics in lexicography. *International Journal of Lexicography* 22(4): 423–465.
De Schryver, G.M. 2012a. Lexicography in the crystal ball: Facts, trends and outlook. Fjeld, R.V. & Torjusen, J.M. (eds). *Proceedings of the 15th EURALEX International Congress, 7-11 August 2012, Oslo*: 93–163. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
De Schryver, G.-M. 2009b. Lexikos at eighteen: An analysis. *Lexikos* 19: 372–403.
De Schryver, G.-M. 2012b. Trends in twenty-five years of academic lexicography. *International Journal of Lexicography* 25(4): 464–506 + online supplementary material.
De Schryver, G.-M. 2023. The future of metalexicography: Reaching for the mesosphere. *Lexikos* 33(1): 136–165.
De Schryver, G.-M. 2026. Three decades of Asian academic lexicography. *Lexicography: Journal of ASIALEX* 13 (forthcoming).

*****

KEYNOTE PRESENTATION 2

On AI and the Future of African Lexicography

Langa KHUMALO (Langa.Khumalo@nwu.ac.za)
South African Center for Digital Language Resources (SADiLaR), North-West University, South Africa

Artificial Intelligence (AI) is increasingly influencing the development of lexicography and offers significant potential for transforming dictionary production and use in African languages. In multilingual contexts such as South Africa, where eleven official languages coexist within the education system, AI technologies present new opportunities for the creation of accessible, dynamic, and user-oriented lexicographic resources. This article examines the opportunities, capabilities, and challenges associated with the integration of AI in African lexicography, with particular attention to its potential role in supporting multilingual education.

AI technologies, including machine learning and natural language processing, enable faster corpus analysis, automated identification of lexical patterns, and the efficient updating of lexical databases. These capabilities can support the development of lexicographic tools that move beyond static reference works to become interactive digital tools. Such dictionaries can incorporate features such as intelligent search, pronunciation support, contextual usage examples, and multilingual linking between African languages and English. In multilingual educational environments, these tools can assist learners in accessing disciplinary vocabulary across languages, thereby supporting reading comprehension and facilitating multidisciplinary learning.

However, the application of AI in African lexicography is constrained by several structural challenges. A major limitation is the scarcity of large, balanced, and machine-readable corpora for many African languages. Much linguistic data remains undocumented, fragmented, or unavailable in digital form, particularly in relation to spoken language and emerging terminology. These data gaps can limit the accuracy and representativeness of AI-driven lexicographic systems. Addressing these challenges requires sustained efforts in corpus development, digitisation of existing resources, and the creation of accessible linguistic datasets.

The article argues that while AI cannot replace the expertise and interpretive role of lexicographers, it offers powerful tools that can enhance lexicographic practice and expand access to dictionaries. When combined with ongoing linguistic research and data development, AI has the potential to support the creation of innovative, user-centred lexicographic resources that strengthen multilingual education, improve reading development, and promote the use and status of African languages.

*****

PRECONFERENCE WORKSHOP

Generative AI & Lexicography: From the Insights of Patrick Hanks (1940-2024) to Chatbot Lexicography

Gilles-Maurice DE SCHRYVER (GillesMaurice.DeSchryver@UGent.be)
BantUGent (the UGent Centre for Bantu Studies), Ghent University, Belgium

The workshop sessions are organised in such a way that all participants will end up being convinced that Patrick Hanks was the forefather of generative AI (GenAI) in lexicography. To do so, his approach to and theories about dictionary-making are explained in such a way to show exactly what the generative pre-trained transformers (GPTs) do and use today. That of course means that dictionary-making for (British) English will serve as the starting point, whereafter the focus will shift to other languages from Europe, America, Asia, and Africa. Of the many hundreds of available large language models (LLMs), the participants will experiment with various chatbots, especially the ones known as ChatGPT, Gemini, and DeepSeek. To top it all, the participants will use the one and only lexicographic software into which GenAI has already been imbedded, and this for over three years, being the South African created TLex, which will convince everyone that in the present day and age, the GenAI tools are far more than (traditional) 'dictionary writing systems', as they are also both 'dictionaries' in their own right and 'lexicographers' plain and simple.

*****

CONFERENCE SPECIAL SESSIONS

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 can 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 has become tradition over the last decade, this year's conference again provides an opportunity for both practising lexicographers in commercial/academic 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 comment on the challenges they face as well as to highlight any recent contributions and achievements. The publishers' session features the South African National Lexicography Units (publisher for many of the NLUs), Centre for Advanced Studies of African Society, ~~Societie virrie Advancement van Kaaps~~, and Oxford University Press (Southern Africa) as major players in the field of lexicographic practice in South Africa. The NLU session features representatives from the Afrikaans NLU, the Sesiu sa Sesotho NLU, the Sesotho sa Leboa NLU, the Setswana NLU, the Siswati NLU, the Tshivenda NLU and the isiXhosa NLU. 2026 also marks a distinct privilege for AFRILEX: 30 years of conferences. In commemoration of this special milestone, this year's conference will feature a special panel discussion involving honorary members and past presidents. The session will be one of experienced reflection, with insights and brainstorming from AFRILEX veterans such Herman Beyer, Gilles-Maurice de Schryver, Rufus Gouws, and Elsabé Taljard.

PRESENTATIONS

Transition to Online Dictionaries

M. Thato BALOYI (baloymt@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa
Karabo MAFETSA (mafetks@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa
Kganathi SHAKU (shakukj@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa

The shift from print to online dictionaries has fundamentally transformed the lexicographic field, marking a significant departure from traditional print dictionaries to the dynamic realm of online resources, aligning with ongoing technological advancements (Tarp and Gouws, 2023). This transition is not merely a change in dictionary format; it represents a fundamental rethinking of how lexicographic information is created, disseminated, and consumed.

Historically, print dictionaries were used as authoritative sources, methodically produced and published in tangible form. Although robust, their structure frequently hampered the speed of changes and the amount of information available to consumers. Many of these limits have been lifted by the switch to online dictionaries, which now provide users with unparalleled access to a wide range of linguistic materials. Online dictionaries shift lexicography from post-hoc archiving to real-time language monitoring. Making frequent updates as usage shifts, colloquial language dictionaries like Mzansi Taal and Urban Dictionary can provide users with access to the most recent definitions, usage examples, and linguistic trends, as Leuckert (2024) suggests. This transition positions dictionaries as active, real information repositories.

Ferrett and Dollinger (2021) argue that the incorporation of multimedia features such as audio pronunciations, video demonstrations, and interactive quizzes has altered how users interact with language and increased their understanding of lemmas. These qualities not only allow for a better comprehension of word meanings and usage but also accommodate a variety of learning styles, making language acquisition more inclusive. Effective online dictionaries also have key elements, such as multimedia audio pronunciations. These elements set them apart from traditional print versions. Lew (2015) supports this by emphasising that online dictionaries provide new methods for obtaining lexicographic data, such as comment sections and advanced search tools, to enable users to find information quickly and efficiently. The ability to search for words, phrases, or contextual examples streamlines the user experience and corresponds to the fast-paced demands of contemporary life. Nesi (2024) posits that everyday dictionary users have become active participants through their ability to edit entries, fostering a sense of community and collaboration. This participation encourages users to contribute to the ongoing evolution of the dictionary itself. Tarp and Gouws (2020) reinforce this by arguing that dictionaries should accommodate human needs by paying attention to the quality of interaction.

This paper examines this transition by highlighting the disconnect between the average user and the static nature of print dictionaries, which necessitates the move toward online dictionaries that prioritise user needs and evolve with language. The paper outlines and

Accepted but not presented

contrasts the constraints associated with print dictionaries, such as cost and limitations, with the advantages of online dictionaries, which function as real-time repositories that incorporate interactivity, accessibility, searchability, and user-generated lexicographic data.

The inquiry is grounded in lexicographical literature to analyse how the transformation of dictionaries can bridge the disconnect between users and traditional lexicographic tools. Successful online dictionaries, like Cambridge Dictionary Online, Collins English Dictionary Online, Linguee Online Dictionary, WordReference, and Merriam-Webster Online, are analysed alongside unconventional dictionaries, such as Mzansi Taal Online Dictionary and Urban Dictionary, to validate this literature-based paper. The paper argues that while print dictionaries remain valuable for language studies, digital resources are becoming increasingly important in modern linguistic environments and in meeting user demands. Although user-centred design is a strength of online dictionaries, it also presents challenges, such as information overload and quality assurance concerns, particularly when user-generated content may produce inaccurate or biased definitions.

The paper concludes that although lexicographers have made significant progress in integrating digital resources in response to the demands of the 4IR era, there remains a need for improved representation of African and multilingual contexts in online dictionaries. Dictionaries such as Mzansi Taal and Urban Dictionary have begun to address this gap, but further research is required to overcome existing limitations.

Accepted but not Presented.

References

Ferrett, E., & Dollinger, S. 2021. Is digital always better? Comparing two English print dictionaries with their online counterparts. *International Journal of Lexicography* 34(1): 6–91.

Leuckert, S. 2024. Moving Lexicographical Mountains for the Community: A Comparison of Print and Online Resources of Mountaineering English. *International Journal of Lexicography* 37(1): 29–49.

Lew, R. 2015. Research into the use of online dictionaries. *International Journal of Lexicography* 28(2): 232–253.

Nesi, H. 2024. Are we witnessing the death of dictionaries? *Ibérica* 47(1): 7–14.

Tarp, S and Gouws, R. 2020. Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture. *Lexikos* 30: 1–29.

Tarp, S. and Gouws, R. 2023. A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and Implications for the Future. *Lexikos* 33(1): 425–447.

*****

Data Mining the *Greater Dictionary of IsiXhosa*: Lexical Stratification and Digital Lexicography in IsiXhosa

William G. BENNET (w.bennett@ru.ac.za)
Departmentof Linguistics & Applied Language Studies, Rhodes University, Makhanda, South Africa
Camilla R. CHRISTIE (imbriferumvir@gmail.com)
Departmentof Linguistics & Applied Language Studies, Rhodes University, Makhanda, South Africa
Dion NKOMO (d.nkomo@ru.ac.za)
African Languages section, School of Languages, Rhodes University, Makhanda, South Africa
Bulelwa NOSISLELA (b.nosilela@ru.ac.za)
African Languages section, School of Languages, Rhodes University, Makhanda, South Africa

The *Greater Dictionary of isiXhosa* (GDX) is among the most expansive and comprehensive

dictionaries produced to date for any African language (Nkomo et al. 2022). Across its three volumes, the dictionary preserves a vast amount of lexical, phonological, semantic, and etymological data that remains underutilised as a source of linguistic information. This paper presents a proposed digital humanities project aimed at transforming the GDX into a searchable lexical database capable of supporting large-scale investigation of phonological and morphological patterning in isiXhosa. In particular, the project explores whether the isiXhosa lexicon exhibits evidence of lexical stratification, whereby different classes of lexical items display distinct phonological and morphological behaviours reflecting different historical layers of borrowing and language contact.

The concept of lexical strata emerged within generative phonology in the 1970s and 1980s and has been widely used to explain why words within a single language may pattern differently regarding affixation, lexical derivation, and phonological rules. The idea is rooted in Chomsky and Halle's (1968) The Sound Pattern of English (SPE), which sought to derive the surface sound patterns of English. In English, morphophonological alternations such as *profane ~ profanity* have historically been analysed as reflecting distinctions between inherited and borrowed lexical layers, where synchronic phonological structure preserves traces of historical development.

Similar phenomena have been documented cross-linguistically. Japanese, for example, contains recent loanwords from English alongside older Sino-Japanese forms derived from Chinese and words from the native Yamato lexicon. These strata exhibit distinct phonological properties: for instance, the sound [p] may occur word-initially in Sino-Japanese vocabulary but not in Yamato words. Lexical stratification therefore provides a useful framework for examining how histories of contact become encoded within the synchronic structure of a language.

There are strong reasons to expect isiXhosa to exhibit comparable stratification effects. Historically, the language has undergone sustained contact with multiple linguistic communities. Earlier contact with Khoekhoe and San languages was likely extensive, while colonial expansion later brought isiXhosa into contact with Dutch, Afrikaans, and English. These different contact histories appear to have left distinct traces within the lexicon. Claughton (1992:145) observes that the tonal patterns of English and Afrikaans loanwords are generally predictable, whereas loanwords from Zulu or Khoekhoe display less uniformity. He further notes that word-final consonants in loanwords are typically adapted through vowel insertion, as in English *bed* > isiXhosa *ibhedi*, although older Khoekhoe borrowings may resist this pattern, e.g. Khoekhoe *hagub* 'pig' > isiXhosa *ihagu*. Borrowed verbs from English and Afrikaans are sometimes suffixed with -isha, as in *park* > *ukupakisha*, while such morphology is less common in indigenous verbs except those derived from ideophones. Words containing click consonants are likely to reflect Khoekhoe influence, whereas consonant clusters are more characteristic of English or Afrikaans borrowings. Such patterns suggest that the isiXhosa lexicon may contain multiple synchronically active strata whose properties can be investigated systematically through large scale lexical analysis.

This presentation proposes the creation of a lexical database derived from the GDX. The initial stage involves extracting lemmas from scanned volumes using Optical Character Recognition (OCR) software. These lemmas can then be cross-referenced (Nkomo et al. 2022) to encode relationships such as derivation and free variation, while also incorporating phonetic annotation and updated etymological information (Christie 2023, 2025). The resulting database will enable searches by phonological structure, segmental sequence, morphological category, borrowing source, and related linguistic features. Beyond its value

as a digital resource, the project demonstrates how historical contact phenomena may be quantitatively investigated through lexicographic data and contributes to broader efforts in African language documentation and digital humanities.

References

Chomsky, N. & Morris, H. 1968. *The Sound Pattern of English*. New York: Harper and Row.
Christie, C.R. 2023. Click consonants in contact: a comparative sociohistorical analysis with special reference to Nama-Afrikaans contact. Unpublished PhD dissertation. Makhanda: Rhodes University.
Christie, C.R. 2025. Click words which isiXhosa, isiZulu and Afrikaans have in common. *Stellenbosch Papers in Linguistics Plus* 69: 77-106.
Claughton, J.S. 1992. The tonology of Xhosa. Unpublished doctoral dissertation. Grahamstown: Rhodes University.
Nkomo, D., Nosilela, B. & Gambushe, W. 2022. How Great is thy Dictionary? Cross-referencing as a Lexicographic Device in *The Greater Dictionary of (isi)Xhosa*. *Lexikos* 32(3): 166–190.
Tshabe, S.L., et al. (eds.). 2006. *The Greater Dictionary of IsiXhosa*. Volumes 1-3, previously published in 1989, 2003. Alice: IsiXhosa National Lexicography Unit, University of Fort Hare.

*****

When Lexicographical Communication Malfunctions: The Case of Fraction Names in the *Afrikaanse Woordelys en Spelreëls*[i]

Herman L. BEYER (hbeyer@unam.na)
School of Humanities, Society and Development, University of Namibia, Windhoek, Namibia

This paper examines instances of dysfunctional lexicographical communication through a case study of fraction names in the *Afrikaanse woordelys en spelreëls* (AWS), the authoritative reference work on Standard Afrikaans spelling. The analysis is grounded in a lexicographical communication model, which distinguishes between three types of results: effects in the usage situation, outcomes in the user situation, and impact in the broader socio-linguistic context. Within this framework, successful lexicographical communication co-depends on the ability of target users to locate, interpret, and apply the lexical information encoded and implied in lexicographic texts. However, the model also allows for dysfunctional and non-functional results, which may arise when inconsistencies or inadequacies in lexicographic texts disrupt this process. Against this theoretical background, the AWS is investigated as a complex lexicographic resource whose internal coherence is essential for facilitating functional communication.

The study adopts a qualitative, comparative textual analysis of four consecutive editions of the AWS, spanning the eighth edition (1991) to the eleventh edition (2017). The analysis focuses specifically on the spelling rules governing the orthographic representation of fraction names, as well as their interaction with related articles in the word list and the terminology list in the back matter. By tracing diachronic changes across these editions, the study seeks to identify both the motivations for rule changes and their linguistic and lexicographical implications.

Across the editions examined, the prescribed representation of fraction names undergoes a notable shift. According to AWS8 (1991), fraction names may be written either as two separate words (e.g. *twee derdes* 'two thirds') or as hyphenated forms (e.g. *twee-derdes*). In later

[i] Accepted but could not present due to medical issue.

editions, this variation is progressively reduced, culminating in AWS11 (2017), where fraction names are prescribed predominantly as single-word forms (e.g. *tweederdes*), with exceptions only when the coordinating conjunction en is involved (e.g. *'n vier-en-twintigste*; "a four-and-twentieth" = a twenty-fourth).

The analysis yields two main findings. Firstly, none of the spelling rules governing fraction names across the editions appear to be adequately supported by the underlying syntactic and morphological structure of Afrikaans fraction expressions. From a linguistic perspective, fraction names exhibit properties that resist straightforward classification as single orthographic words, particularly given their structure as phrasal constructions. The prescriptive shift towards single-word forms therefore lacks a consistent linguistic motivation and may obscure rather than clarify the structure of these expressions for users.

Secondly, the study identifies instances of dysfunctional implicit intertextuality within the AWS. Specifically, inconsistencies arise between the spelling rules section, the word list, and the terminology list in the back matter. These inconsistencies lead to the encoding of conflicting or linguistically inaccurate lexical data in individual word list articles. As a result, users consulting different parts of the AWS may encounter divergent representations of the same linguistic phenomenon, undermining the reliability of the resource as a whole.

Interpreted within a lexicographical communication model, these findings suggest a complex pattern of results in lexicographical communication. On the one hand, users may still be able to retrieve and apply the information provided in the AWS, resulting in functional effects in the usage situation. On the other hand, the application of inconsistent or linguistically unsupported information is likely to produce dysfunctional outcomes in the user situation, particularly in tasks involving the correct written representation of fraction names. The persistence and dissemination of such usage patterns may, in turn, give rise to a dysfunctional impact in the broader context, affecting conventions of written Afrikaans over time.

The paper therefore argues that the case of fraction names in the AWS illustrates how shortcomings in lexicographic design and internal coherence can lead to systemic disruptions in lexicographical communication. It highlights the importance of aligning prescriptive rules with linguistic structure and of maintaining consistent intertextual relationships within complex reference works. More broadly, the study demonstrates the value of applying a communication-oriented framework to the evaluation of lexicographic resources, enabling a more nuanced understanding of how users interact with and are influenced by such texts.

References

AWS8 = Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 1991. *Afrikaanse woordelys en spelreëls*. Eighth edition. Cape Town: Tafelberg.

AWS11 = Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 2017. *Afrikaanse woordelys en spelreëls*. Eleventh edition. Cape Town: Pharos.

*****

Lexicographic Treatment and Discursive Legacy: The French Lexeme *nègre*

Ruth DE OLIVEIRA (ruth.deoliveira@uct.ac.za)
School of Languages and Literatures, University of Cape Town, Cape Town, South Africa

In the first editions of the *Dictionnaire de l'Académie française* (hereafter DAF; Académie française 2020), *blanc* ("white") and *noir* ("black") are primarily recorded as lexemes of colour. *Blanc* develops as a rich and productive colour term, carrying positive associations before its extension to human reference. *Noir*, for its part, develops a range of figurative extensions,

mostly connoting negativity, while not yet functioning primarily as a human designation. By contrast, this human reference appears from the outset with *nègre*, which, from its first entry in the 4th edition, is attached to a marked and subordinated human referent through the phrase *traiter quelqu'un comme un nègre* ("to treat someone like a nigger"), glossed as *traiter quelqu'un comme un esclave* ("to treat someone like a slave"). This asymmetry remains visible throughout the lexicographic history of nègre and helps explain its later servile, racialized, and discursively persistent trajectory within the dictionary.

The French lexeme *nègre* therefore offers a particularly revealing case for diachronic lexicographic study. Across the successive editions of the DAF, the term undergoes definitional, grammatical, and contextual reformulations, yet remains tied to a semantic history shaped by demeaning racialized representations. This paper builds on earlier work on *nègre* in French lexicography and translation (De Oliveira 2022, 2024) and argues that the DAF does not merely register its lexical evolution; it also records the persistence, displacement, and partial reformulation of the ideological frameworks in which the term has circulated. From its indirect appearance in the 4th edition to its explicit association with colonial slavery in the 5th, its broader racial generalization in the 6th and 7th, its human reformulation in the 8th, and its present-day historicization and restriction to marked expressions in the current edition, the treatment of *nègre* reveals a long process of semantic reclassification and ideological continuity.

Methodologically, the paper combines diachronic analysis of the successive DAF entries with selective contextual reading of materials that illuminate the lexical and ideological afterlives of the term. Particular attention is given to one of these discursive legacies, namely *parler petit nègre*. This expression makes visible the extension from the naming of racialized subjects to the racialization of speech itself. In this perspective, lexicographic evidence is read alongside a colonial pedagogical document, the 1916 booklet *Le français tel que le parlent nos tirailleurs sénégalais* ("French as Spoken by Our Senegalese Riflemen"). In its preliminaries, the text presents *petit nègre* as a language allegedly created by Black *tirailleurs* in contact with their European instructors and governed by "*règles fixes* / fixed rules." Its stated aim is practical and disciplinary: to help European superiors make themselves understood rapidly, give their instructions in a form said to be intelligible to all, and intensify military training. To that end, the booklet prescribes a reduced model of French: articles are suppressed, gender and number are neutralized, verbs are reduced to the infinitive, and syntax is compressed into what the text calls the "*forme très simple*" ("very simple form") of the "*dialectes primitifs de notre Afrique Occidentale*" ("primitive dialect of our French West Africa"). The document is therefore relevant as an explicit colonial codification of a hierarchical and demeaning representation of African speech.

The paper concludes by considering later attempts to curb or replace some inherited uses of the term, such as official recommendations concerning *nègre littéraire* ("ghostwriter"; DGLFLF 2017) and public controversies around names such as *tête de nègre* ("nigger's head" as a historical cake name). These cases show that the inherited meanings attached to *nègre* remain active in present-day French. Their relevance is sharpened by a contemporary social context in which racialized inequality remains measurable: recent data on metropolitan France show that Black people continue to report very high levels of racial discrimination (Richardot 2023). The preliminary conclusion is that later metalinguistic interventions have not severed *nègre* from the regimes that shaped its history in French: slavery, colonial pedagogy, and hierarchical racial classification. As such, the term remains marked by racial segregation.

References
Académie française. 2020. *Dictionnaire de l'Académie française*. Available: https://www.dictionnaire-academie.fr/. Accessed on 17/04/2026.
DGLFLF = Délégation générale à la langue française et aux langues de France. 2017. NÈGRE (LITTÉRAIRE). Available: https://www.culture.fr/FranceTerme/Recommandations-d-usage/NEGRE-LITTERAIRE. Accessed on 17/04/2026.
De Oliveira, R. 2022. Sur l'in/visibilité du traducteur dans les notes explicatives de l'édition française de Capitães da areia / Capitaines des sables de Jorge Amado: Analyse des emplois de la lexie "Nègre". *Revista (Con)Textos Linguísticos* 16(35): 132–149.
De Oliveira, R. 2024. Traduction et usage du N... word en français: De Dix petits nègres à Ils étaient dix. *Cadernos de Tradução* 44(1): 1–16.
Le français tel que le parlent nos tirailleurs sénégalais ("French as Spoken by Our Senegalese Riflemen"). 1916. Paris: Imprimerie-Librairie Militaire Universelle.
Richardot, R. 2023. Racisme: 91 % des personnes noires en métropole se disent victimes de discrimination. *Le Monde*, 15 February. Available: https://www.lemonde.fr/societe/article/2023/02/15/racisme-91-des-personnes-noires-en-metropole-se-disent-victimes-de-discrimination_6161879_3224.html. Accessed on 17/04/2026.

*****

Restructuring Semantic Information for Cognitive Accessibility: The Case of *e-Le Petit Robert*

Emmanuel Sylvain FOMAT (sylvainfomat4@gmail.com)
Department of English and German Philology, University of Santiago de Compostela, Santiago de Compostela, Spain

"Even in highly respected dictionaries, there remain basic problems of access" (Lew: 2013:1). Despite the transition from print to digital formats, many online dictionaries continue to reproduce linear and densely structured microstructures inherited from print lexicography. This paper investigates how semantic information in digital dictionaries can be systematically restructured to improve cognitive accessibility, particularly for non-experts and language learners, using *e-Le Petit Robert* as a case study. The central research question guiding this study is: How can the microstructure of digital dictionary entries be redesigned to align with users' cognitive processes, thereby reducing cognitive load and improving semantic accessibility?

To address this question, the study integrates cognitive linguistics and user-centered lexicography. Cognitive linguistics, particularly prototype theory (Rosch 1978; Lakoff 1987) and Langacker's (1987) cognitive grammar, provides a framework for understanding how meaning is cognitively organized around prototypical and peripheral senses. Within this perspective, lexical meaning is not static but structured in graded categories that influence how users interpret dictionary entries. The study also conceptualizes the dictionary as an interconnected system in which definitions, examples, and related information function together to support meaning construction. On the lexicographic side, it builds on user-centered approaches initiated by Wiegand (1977) and operationalized into the function theory by Bergenholtz and Tarp (2003) and Tarp (2008), which emphasizes that dictionary structures must be designed according to specific user needs and consultation situations.

The methodology is qualitative and comparative. It involves a microstructural analysis of selected entries from *e-Le Petit Robert*, followed by a redesign-based comparative analysis between original and restructured versions. The sample comprises entries such as *escargot* (snail), *hémisphère* (hemisphere), *banque* (bank), *voiture* (car), *ver* (earthworm), *valise* (suitcase), *sifflet* (whistle), and *feu* (fire). These entries were selected according to three criteria: frequency

Accepted but not Presented.

in everyday language, degree of polysemy, and complexity of their semantic networks. The analysis evaluates entries using three parameters: visual hierarchy, semantic transparency, and navigational efficiency.

A closer qualitative analysis of the entry *feu* further illustrates structural misalignment between semantic units and usage examples in *e-Le Petit Robert*. In the current microstructure, examples are not consistently linked to specific senses but are instead grouped in a separate "*EXEMPLES*" section, which weakens the cognitive connection between definition and contextualization. As a result, users must infer independently which example corresponds to which sense, increasing interpretive effort.

To address these limitations, a restructuring model is proposed based on three principles: (1) segmentation of meanings into cognitively manageable units supported by collapsible digital structures; (2) prioritization of prototypical senses; and (3) multimodal enhancement through typographic differentiation and semantic tagging. A central feature of the model is the systematic alignment of each sense with a corresponding example, ensuring that definition and usage are directly connected within the same microstructural unit. Figurative idioms are relocated into clearly labeled sub-sections to preserve lexical richness while improving navigability and reducing ambiguity.

The comparative analysis between original and restructured entries shows consistent improvements across all parameters. In the redesigned *feu* entry, the prototypical literal sense is immediately paired with a contextually accessible example, while figurative idioms are clearly separated and labeled according to meaning type. Across all sampled entries, the restructured versions reduce information density through clearer segmentation, enhance semantic transparency through direct definition–example alignment, and improve navigational efficiency through structured visual hierarchy and guided pathways.

The study concludes that semantic restructuring guided by cognitive principles can substantially improve the accessibility of digital dictionaries. The comparative analysis shows that direct definition–example alignment, clearer visual hierarchy, and the prioritization of prototypical senses enhance semantic transparency and navigational efficiency across the sampled entries. These findings suggest that accessibility issues in e-lexicography are not solely technological but also structural, arising from traditional microstructural conventions that do not fully accommodate digital consultation practices. The proposed model contributes to ongoing discussions in e-lexicography by offering a practical framework for aligning semantic organization with users' cognitive processing patterns.

Accepted but not presented.

References:

Bergenholtz, H. & Tarp, S. 2003. Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes – Journal of Language and Communication in Business* 31: 171–196.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Langacker, R. 1987. *Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.

Lew, R. 2013. Online dictionaries of English. Fuertes-Olivera, P.A. & Bergenholtz, H. (eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 230–250. London: Bloomsbury.

Rosch, E. 1978. Principles of categorization. Rosch, E. & Lloyd, B.B. (eds.). *Cognition and Categorization*: 27–48. Hillsdale, NJ: Lawrence Erlbaum.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge*. Berlin: De Gruyter.

Wiegand, H.E. 1977. Nachdenken über Wörterbücher: Aktuelle Probleme der Wörterbuchforschung. Drosdowski, G., Helmut, H. & Wiegand, H.E. (eds.). *Nachdenken über Wörterbücher*: 51–102. Mannheim/Wien/Zürich: Bibliographisches

Institut/Dudenverlag.

*****

Teaching Lexicography at Undergraduate Level: Student Perceptions and Dictionary Culture in an isiXhosa Classroom

Wanga GAMBUSHE (wanga.gambushe@uct.ac.za)
School of Languages and Literature: African Languages, University of Cape Town, Cape Town, South Africa

Hartmann & James (1998:41) define dictionary culture as "the critical awareness of the value and limitations of dictionaries and other reference works in a particular community". Central to this understanding is the notion that dictionaries become meaningful linguistic resources not merely through their existence, but through sustained user awareness, exposure, and training in dictionary use. In this regard, Gouws (2018) argues that language and linguistics departments should incorporate lexicography into their curricula as part of developing a comprehensive dictionary culture. Such a culture equips users with knowledge of different dictionary types, their functions, and the skills required to retrieve and interpret information effectively. Gouws further suggests that dictionary didactics should be directed at dictionary users, future lexicographers, and future trainers of lexicography.

Within the South African context, Nkomo (2020:12) argues that the inclusion of dictionary skills in the national curriculum potentially creates favourable conditions for the development of dictionary culture. However, various contextual factors continue to shape the extent to which dictionary use and lexicographic awareness develop within African language contexts. These include questions of dictionary access, availability, exposure, and dictionary skills development.

Against this background, this paper reflects on the teaching of a second-year lexicography and terminology development course offered in isiXhosa at the University of Cape Town. The paper proceeds from the observation that lexicography remains a relatively small component (if available at all) within many African Languages programmes, where greater emphasis is often placed on formal language studies, literature, and applied language studies such as translation. Although lexicography teaching and dictionary culture have received scholarly attention in South Africa, fewer studies have explored undergraduate students' perceptions of lexicography and their engagement with dictionaries within African Languages programmes.

Using exploratory before-and-after questionnaires administered to second-year isiXhosa students enrolled in the course, the study examines students' understandings of lexicography, their previous engagement with dictionaries, and their expectations of the course. The before-course responses revealed uneven and largely intuitive understandings of lexicography, with most students associating the discipline primarily with dictionaries, word meanings, and vocabulary study, while demonstrating limited familiarity with lexicography as a formal academic field. The responses also suggested relatively limited prior engagement with dictionaries among participants, raising broader questions around dictionary culture and dictionary literacy within African language education contexts.

At the same time, students' expectations of the course extended beyond technical dictionary use and reflected broader interests in isiXhosa language history, fluency, terminology development, meaning-making, and language change. Some responses further revealed emerging critical awareness regarding questions of linguistic authority and

dictionary construction.

The after-course responses suggested shifts in students' understandings of lexicography. While many students continued to associate the discipline with dictionary compilation, their responses demonstrated expanded awareness of lexicography as involving processes of language standardisation, dictionary planning, editorial decision-making, and the historical development of dictionaries. Several students also reflected on having previously viewed dictionaries and dictionary-making as uninteresting or narrowly technical prior to the course. The paper argues that the teaching of lexicography in African Languages departments may serve not only as technical linguistic training, but also as an important pedagogical space for developing critical language awareness and fostering engagement with African language intellectualisation.

References

Hartmann, R.R.K. & James, G. 1998. *Dictionary of lexicography*. London: Routledge.
Gouws, R.H. 2018. Woordeboekdidaktiek binne 'n omvattende woordeboekkultuur. *Literator* 29(1): 49–65.
Nkomo, D. 2020. Dictionary culture in African language communities: research, development, challenges and prospects. *Lexicographica* 36: 11–37.

*****

The Role of Dictionaries in the Intellectualisation of isiNdebele

Pearl Linda KABINI (pearl.kabini@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa
Nontokozo Gladys MDHLULI (nontokozo.mdhluli@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa

The intellectualisation of African languages has been recognised as a significant concern within South African language planning and higher education scholarship (Alexander 2007; Webb 2002; Maseko 2011; Khumalo 2017). Intellectualisation refers to the process of developing and expanding a language to enable its effective use in specialised domains such as higher education, science, technology, administration, and research (Alexander 2007; Maseko 2011). Though isiNdebele is designated as one of South Africa's official languages, it continues to be used in the educational and professional sectors at lower levels than other "official" languages like English. With ongoing university emphasis placed upon multilingualism and the incorporation of African languages into instruction and learning, questions about the provision of available linguistic resources for developing isiNdebele as a language of scholarship arise. This article presents evidence that dictionaries are instrumental in the intellectualisation of isiNdebele through facilitating term formation, language standardisation, access to knowledge, and preservation of linguistic identity. While considerable scholarship has examined the intellectualisation of African languages in South Africa, limited attention has been given to the specific role that isiNdebele dictionaries play in supporting this process. This study addresses that gap by examining how dictionaries contribute to the intellectualisation of isiNdebele and identifying areas where current lexicographic resources remain inadequate.

Alexander (2007), Webb (2002), and Maseko (2011) argue that African languages can only function effectively in higher education when they are supported by developed terminologies, standardised linguistic resources, and sufficient academic materials for teaching, learning, and research. Maseko (2011) adds that the intellectualisation of indigenous languages is contingent upon purposeful action toward expanding their utility through the development of terminology, corpora, and lexicographic resources. Dictionaries serve as foundational

instruments for this process, providing standardised vocabularies, definitions, spellings, and usages that facilitate communication across various spheres of activity. Historically, lexicographical efforts for African languages have primarily addressed common language use rather than specialised or current uses demanded by modern education and knowledge production.

This research employs a qualitative research design that combines descriptive lexicography and theory surrounding the intellectualisation of languages. Data was collected from a documentary analysis of several selected isiNdebele dictionaries and articles written by scholars engaged in the development of African languages. The research addresses two primary areas: how dictionaries are utilised to develop and standardise isiNdebele vocabulary within academic settings, and whether dictionaries currently available contain adequate amounts of contemporary terminology specific to emergent domains of knowledge, such as education, research, technology, and social sciences. Data collection and subsequent analysis reveal that dictionaries have provided substantial support to preserve and document isiNdebele, and to enhance the codification of isiNdebele vocabulary. While current dictionaries assist students in acquiring language skills, establishing spelling consistency, and clarifying meanings for both native speaker populations, students, and educators, there were additional areas where current dictionaries demonstrated considerable value in promoting isiNdebele as a legitimate language capable of communicating beyond colloquial speech. Current lexicographic contributions through new terminology and formalised vocabulary contribute to increasing the breadth of isiNdebele language capabilities while improving the academic viability of the language.

Although current dictionaries are beneficial to learners and users, they were found to have constraints limiting their potential contribution to the intellectualisation of isiNdebele. Several examples exist wherein current dictionaries fail to include an adequate amount of specialised terminology for higher education or professional purposes. For example, many modern concepts, e.g., technology, research methods, science, and digital communications, were found lacking in representation or completely absent in many cases. These shortcomings necessitate that speakers, educators, and translators utilise English terminology or create unstandardised translations when faced with these limitations. Additionally, the results indicated that lexicographic advancements made in isiNdebele have not kept pace with changing needs for the language in today's sociolinguistic environment.

In addition to demonstrating the importance of lexicographic development in the context of language policies supporting the intellectualisation of isiNdebele, this article emphasises that lexicographic development cannot be accomplished solely through policies designed to promote the utilisation of African languages. Rather than serving as static collections of words, dictionaries must evolve as responsive linguistic tools addressing changes in the social, educational, and technological context in which they are utilised. Therefore, greater cooperation among lexicographers, educators, linguists, translators, and higher education institutions is needed to produce continuously updated and relevant isiNdebele lexicographic resources. This article contributes to larger discussions concerning African lexicography by illustrating the connection between dictionary development and the intellectualisation of languages. Ultimately, this study demonstrates the indispensable nature of dictionaries as mechanisms for establishing isiNdebele as a viable academic/educational/knowledge-producing language.

References

Alexander, N., 2007. The role of African universities in the intellectualisation of African

languages. *Journal of higher education in Africa* 5(1): 29–44.
Khumalo, L., 2017. Intellectualization through terminology development. *Lexikos* 27: 252–264.
Maseko, P., 2011. Intellectualisation of African languages with particular reference to isiXhosa. Unpublished PhD dissertation. Grahamstown: Rhodes University.
Webb, V. 2002. *Language in South Africa: The Role of Language in National Transformation, Reconstruction and Developmen*t. Amsterdam: John Benjamins.

*****

Generative AI and Lemma Treatment in Sepedi: Testing M365 Copilot (GPT-5), ChatGPT (GPT-5.5) and DeepSeek-v3

Napjadi Eugene LETSOALO (letsone@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa

Generative Artificial Intelligence (GenAI) has significantly transformed linguistic practice, particularly in lexicography, by enabling the automated production of textual, semantic, and grammatical content. Powered by large language models trained on extensive datasets, these systems can generate dictionary-like outputs such as definitions, examples, and lexical relations with notable speed and fluency (Oluwagbenro 2024; Lew 2024). However, within the context of African languages, especially Sepedi, the application of GenAI raises critical concerns regarding linguistic accuracy, cultural fidelity, and alignment with established lexicographic principles. This study examines the performance of three GenAI systems: M365 Copilot (GPT-5), ChatGPT (GPT-5.5) and DeepSeek-V3 in generating lemma treatments in Sepedi, focusing on whether they can approximate the standards expected of trained lexicographers.

The study is guided by the question: what happens when GenAI systems are instructed to "think like Sepedi lexicographers"? To answer this, a systematic literature review is combined with a mixed-method comparative analysis. Data generated from the three systems are evaluated alongside insights from scholarship in lexicography, corpus linguistics and artificial intelligence. The analytical framework is grounded in core lexicographic principles, including noun-class integrity, access structure coherence, reversibility, cultural fidelity and corpus-based naturalness (De Schryver & Prinsloo 2001).

Findings indicate that while GenAI systems can rapidly generate lemma treatments, they do not independently produce lexicographically sound and grammatically reliable Sepedi entries. At a surface level, the systems perform well in producing fluent definitions, paraphrases, and example sentences, confirming their strength in microstructural tasks such as definitional drafting and example construction (Lew 2023; De Schryver 2023). However, significant limitations arise in deeper lexicographic functions. None of the models consistently satisfies essential requirements, particularly in handling noun-class morphology, which is central to Sepedi grammar and meaning.

A major challenge concerns access structures, where GenAI inconsistently balances stem-based and word-based lemmatisation, often hindering user accessibility and weakening reversibility in bilingual resources (Prinsloo & De Schryver 1999, 2002). Cultural and semantic fidelity is also problematic, as Sepedi meanings are deeply embedded, yet GenAI systems frequently impose Anglocentric interpretations, leading to oversimplification and misrepresentation (Kosch 2016; Pinhanez et al. 2024). Furthermore, the lack of corpus grounding limits reliability, as models generate plausible but sometimes unattested forms, resulting in overgeneralisation and inauthentic examples, contrary to corpus-based

Accepted but not presented.

lexicographic principles (Faaß 2022).

Despite these limitations, the study identifies notable potential for GenAI in supporting Sepedi lexicography. Each system demonstrates distinct strengths: Copilot generates rich microstructures with multiple examples; DeepSeek shows strong morphological awareness and noun-class marking; and ChatGPT offers clearer access-oriented guidance. These complementary capabilities suggest that GenAI should be viewed as an assistive tool within a human-governed editorial framework.

The study concludes that GenAI should not replace lexicographers but can enhance lexicographic processes through efficiency and drafting support. When integrated into a corpus-first, user-centred and culturally grounded framework, with strong human validation, GenAI can contribute to the expansion and sustainability of lexicographic resources. A hybrid approach that combines the strengths of multiple systems offers the most effective path forward.

References

De Schryver, G.M. & Prinsloo, D.J. 2001. Corpus-based activities versus intuition-based compilations by lexicographers: The Sepedi lemma-sign as a case in point. *Nordic Journal of African Studies*, 10(3): 25–51.

De Schryver, G.M. 2023. Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography* 36(4): 355–387.

Faaß, G. 2022. Applied corpus linguistics for lexicography: Sepedi negation as a case in point. *Lexikos* 32(2): 82–104.

Kosch, I.M. 2016. Lemmatisation of fixed expressions: The case of proverbs in Northern Sotho. *Lexikos* 26: 145–161.

Lew, R. 2023. ChatGPT as a COBUILD lexicographer. *Humanities and Social Sciences Communications*, 10(1): 1–10.

Lew, R. 2024. Dictionaries and lexicography in the AI era. *Humanities and Social Sciences Communications* 11(1): 426.

Oluwagbenro, M.B. 2024. Generative AI: Definition, concepts, applications, and future prospects. *TechRxiv*, 4 June. https://doi.org/10.36227/techrxiv.171746875.59016695/v1

Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., Vasconcelos, M., Domingues, P., Mizukami, P.D.S., Grell, N. & Gongora, M. 2024. Harnessing the power of artificial intelligence to vitalize endangered Indigenous languages: Technologies and experiences. *arXiv preprint arXiv:2407.12620*.

Prinsloo, D.J. & De Schryver, G.M. 2002. Reversing an African-language lexicon: The Northern Sotho Terminology and Orthography No. 4 as a case in point. *South African Journal of African Languages* 22(2): 161–185.

*****

Data Management for an e-Lexicography Project

Jimmy MABINA (mabinj@unisa.ac.za)

Department of Anthropology and Archaeology, University of South Africa, Pretoria, South Africa

Makwarela MUDAU (mudaum@unisa.ac.za)

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

This study examines the central role of effective data management in the development, usability, and long-term sustainability of online dictionaries within the field of e-lexicography. As lexicographic practices continue to shift from print to digital environments, the complexity of structuring, organising, and preserving lexical data has increased significantly. Without robust data management frameworks, digital dictionaries risk becoming inconsistent, difficult to maintain, and vulnerable to technological obsolescence (Atkins & Rundell 2008:67-69). The study therefore focuses on how well-designed systems and standards can address these challenges while enhancing both editorial workflows and user experience.

Accepted but not presented.

A key concern identified is the need for coherent and scalable data structures that support interoperability across platforms and systems. Online dictionaries increasingly interact with other linguistic resources, natural language processing (NLP) tools, and digital infrastructures, making standardisation essential. The adoption of recognised metadata frameworks such as the Dublin Core Metadata Initiative (DCMI) and the Lexical Markup Framework (LMF) enables consistent data description and exchange (Baker 2012; ISO 2008). These standards support interoperability, facilitate data reuse, and ensure that lexical resources remain adaptable to evolving technological environments.

The study also explores the role of tools used in lexicographic workflows, including database systems, spreadsheets, and specialised lexicographic software. Database management systems provide efficient storage, retrieval, and manipulation of large-scale structured data, while spreadsheets offer accessible solutions for initial data organisation (Bergenholtz & Gouws 2010:110). More advanced lexicographic software integrates these functions into unified platforms that support linguistic annotation, cross-referencing, and digital publication (Granger & Paquot 2012:15-18). The combined use of these tools contributes to improved efficiency, consistency, and overall quality in e-lexicographic production.

Despite these advancements, several challenges persist. Inconsistent data entry remains a major issue, often resulting from a lack of standardised editorial guidelines or insufficient quality control (Atkins & Rundell 2008:50-60). Scalability is another concern, particularly for projects that expand rapidly without appropriate infrastructure. Furthermore, the rapid pace of technological change presents ongoing risks, as file formats, software systems, and platforms may become obsolete (Conrad 2016:20-22). To mitigate these challenges, the study emphasises the importance of flexible, standards-based data models, continuous system evaluation, and strong editorial oversight.

Accepted but not presented.

In addition, the study highlights the growing importance of multimedia integration and user-centred design in modern online dictionaries. Digital dictionaries increasingly incorporate audio, images, and video to enhance user comprehension and engagement (Lew 2011:2-3). Effective data management must therefore accommodate diverse data types and ensure their seamless integration within dictionary entries. Accessibility is also a critical consideration, requiring compatibility with assistive technologies and responsiveness across multiple devices. By aligning technical infrastructure with user needs, lexicographers can improve both usability and inclusivity.

Digital preservation is identified as a crucial factor in ensuring the long-term sustainability of lexical resources. Without appropriate preservation strategies, valuable linguistic data may be lost due to hardware failure, software obsolescence, or data corruption. Best practices include the use of standardised formats, regular backups, and persistent identifiers to ensure ongoing access and integrity (Conrad 2016:25; ISO 2012). Preservation efforts must be supported by institutional policies and long-term planning to remain effective.

The findings of this study indicate that structured metadata, scalable tools, and robust preservation strategies significantly enhance the integrity, accuracy, and usability of online dictionaries. These practices lead to improved workflow efficiency, better data consistency, and increased user engagement. However, challenges remain, particularly in maintaining editorial quality, adapting to technological change, and addressing the needs of diverse linguistic communities.

The study concludes that effective data management is fundamental to modern e-lexicography. It argues that the successful development and sustainability of digital

dictionaries depend on the integration of technological innovation with sustained human editorial oversight. By adopting standardised frameworks, leveraging appropriate tools, and implementing comprehensive preservation strategies, lexicographers can ensure that digital dictionaries remain accurate, relevant, and accessible in the long term.

References

Atkins, B.T.S., & Rundell, M. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.

Baker, T. 2012. Dublin Core Metadata Initiative: Overview and principles. Available: http://dublincore.org. Accessed on 30/04/2026.

Bergenholtz, H., & Gouws, R.H. 2010. A new perspective on the access process. *Hermes – Journal of Language and Communication Studies* 44: 103–127.

Conrad, S. 2016. Data management and preservation in the digital age. *International Journal of Digital Curation* 11(1): 15–28.

Granger, S., & Paquot, M. 2012. *Electronic lexicography*. Oxford: Oxford University Press.

ISO. 2008. ISO 24613:2008 Language resource management — Lexical Markup Framework (LMF). Geneva: International Organization for Standardization.

ISO. 2012. ISO 14721:2012 Open archival information system (OAIS) reference model. Geneva: International Organization for Standardization.

Lew, R. 2011. Studies in dictionary use: Recent developments. *International Journal of Lexicography* 24(1): 1–4.

*****

Comprehensive Research on Terminology Development, Verification and Authentication in Sesotho sa Leboa[i]

Mapula Joyce MADIBA (joyce@pansalb.org)

Pan South African Language Board, Polokwane, South Africa

This study explores the processes of terminology development, verification and authentication within the context of Sesotho sa Leboa, one of South Africa's twelve official languages. As an important component of the country's multilingual framework, Sesotho sa Leboa contributes to linguistic diversity, cultural identity, and inclusive communication across various domains. The study examines how terminology is verified, authenticated, and integrated into the language to meet the evolving demands of sectors such as education, science, technology, governance, and law (Khumalo 2017). Moreover, particular attention is given to the role of institutional structures, especially the Pan South African Language Board (PanSALB), in facilitating these processes. The study also considers how National Language Bodies (NLB) and other stakeholders ensure that newly developed terms are linguistically accurate, culturally appropriate, and socially acceptable to language users.

The lack of verified, authenticated, and standardised subject specific terminology in Sesotho sa Leboa prevents the language from being used in the wide-ranging domains in which English and Afrikaans are used.

The general problem this research addresses is the verification and authentication of semantic terms from different institutions. Terminology creation efforts have been undertaken by various stakeholders and verified, authenticated, and standardised by the Sesotho sa Leboa NLB, which is a structure of PanSALB. The coining of new terms in Sesotho sa Leboa is still a challenge. The equivalent for a headword tends to be explanatory rather than a term.

Terms are frequently coined in isolation, without proper verification and authentication by relevant stakeholders, leading to multiple equivalents for the same concept and confusion

[i] Accepted but could not present due to medical issue.

among language users (Mojapelo, 2018). Moreover, when coming to other domains such as health, there is a difference between the indigenous knowledge system and Western scientific knowledge. The term borrowing and direct coining do not always reflect what the indigenous language and culture mean (Tembane 2019).

To address these challenges, this study undertook a comprehensive investigation of terminology development strategies and authentication processes in Sesotho sa Leboa. The study employed a qualitative methodological approach, combining corpus analysis of existing terminology resources with a comparative examination of term-creation strategies across multiple specialised domains and the authentication processes used by the Sesotho sa Leboa National Language Body.

The study observed how terms are coined by the Department of Arts and Culture through various processes such as harvesting, recording, compounding, semantic extension, borrowing, translation, and consultation by which these terms are verified, authenticated and standardised and accepted as legitimate components of the language's lexicon (Mojapelo, 2018; Molatou, 2024). Special attention was given to the management and internal standardisation of terminology, including the terminological processing of technical source texts prior to translation and the development of models for terminology management in the absence of readily available standardised terms (Taljard & Nchabeleng 2012).

This study will also examine the following terminology lists: Covid 19 Multilingual, Multilingual Statistical, Multilingual Soccer and Multilingual Pharmaceutical terminologies. The data was collected through textual analysis. The outcome of this study reveals that successful terminology development in Sesotho sa Leboa requires an organised stakeholder approach, involving collaboration among terminologists, lexicographers, subject experts, language practitioners, and language users throughout the terminology development process.

The verification and authentication emerge as a critical but often neglected phase; terms that undergo formal verification and authentication by the Sesotho sa Leboa National Language Body achieve greater validity and wider usage than those circulated without an authentication process (Mojapelo 2018; Shaku 2022). The comparative analysis of indigenous and Western health terminology further demonstrates that effective term creation must account for both the semantics of Sesotho sa Leboa and the conceptual frameworks of the relevant discipline (Tembane 2019).

The findings of this study are substantial for language planning, education, and language policy in South Africa. Establishing standardised processes for terminology development, verification, and authentication will strengthen Sesotho sa Leboa's functional capacity in education and public administration. The findings advocate for the institutionalisation of terminology development, verification, and authentication processes that involve private sector participation, universities, and dedicated terminology institutes (Shaku 2022).

References

Khumalo, L. 2017. Intellectualization through terminology development. *Lexikos* 27: 252–264.

Mojapelo, W.M. 2018. An Investigation of term creation in Northern Sotho. Unpublished MA Thesis. Polokwane: University of Limpopo.

Molatou, I.T. 2024. A critical study of the use of loan words in Sesotho translations: a case study of Mangaung region. Unpublished MA thesis. Cape Town: University of the Western Cape.

Shaku, K. 2022**.** Language Development during Covid-19. The Case of Sesotho sa Leboa Language in South Africa. *Africa Insight* 52(2): 48–61.

Taljard, E. & Nchabeleng, M. J. 2011. Management and internal standardization of chemistry terminology: A Northern Sotho case study. *Lexikos* 21: 194–216.

Tembane, S.M. 2019. A Comparative Study of Medical and Health Terms with Special Reference to Sesotho sa Leboa and Western Terminology. Unpublished PhD dissertation.

Pretoria: University of South Africa.

*****

Digital Dictionaries as Inclusive Knowledge Infrastructures: Reimagining Lexicographic Accessibility in the Global South

Muimeleli MAEBA (muimelelimaeba1@gmail.com)
Department of Arts and Social Sciences, University of Venda, Thohoyandou, South Africa

Language remains a fundamental mechanism through which individuals access knowledge, participate in education, and engage in broader social and economic life. In multilingual societies across the Global South, however, language often functions as a barrier rather than a bridge to inclusion. Many students learn through languages that are not their mother tongue, while educational resources frequently privilege dominant languages at the expense of indigenous and minority languages. Within this context, dictionaries continue to play an important role in facilitating language learning and knowledge acquisition. Yet traditional lexicographic resources have often been conceived as static repositories of lexical information, offering limited responsiveness to the diverse linguistic, cultural, and educational realities of multilingual users.

This paper argues that digital dictionaries should be reimagined as inclusive knowledge infrastructures rather than merely digital versions of conventional dictionaries. Such a reconceptualisation expands the role of lexicographic resources beyond word definition and translation, positioning them as dynamic tools capable of supporting equitable access to knowledge in multilingual educational environments. The paper is situated within ongoing debates concerning language, technology, educational inclusion, and epistemic justice, with particular attention to the challenges and opportunities facing users in the Global South.

The study draws on an interdisciplinary theoretical framework that combines epistemic justice theory, digital inequality perspectives, and sociolinguistic theories of language and power. Epistemic justice scholarship highlights the ways in which individuals and communities may be excluded from knowledge production, interpretation, and participation because of structural inequalities and unequal recognition of their linguistic resources (Fricker 2013). This perspective is particularly relevant in multilingual educational settings where students may experience disadvantages when academic knowledge is mediated through dominant languages. Digital inequality perspectives further suggest that meaningful inclusion cannot be reduced to access to digital technologies alone. Rather, inclusion depends on whether technologies are usable, relevant, culturally appropriate, and responsive to the needs of diverse users (Mariën & Prodnik 2014). In addition, sociolinguistic understandings of language and power draw attention to how educational systems often privilege particular linguistic forms while marginalizing others, thereby reproducing unequal patterns of knowledge access and participation.

Methodologically, this paper adopts a conceptual qualitative approach based on a critical review of literature in lexicography, digital learning, educational technology, and inclusion studies. The analysis examines selected examples of contemporary digital language tools and considers their implications for lexicographic accessibility. Particular attention is given to issues of usability, adaptability, localization, multilingual support, and cultural relevance. The paper also considers the extent to which current digital dictionaries accommodate indigenous languages and local knowledge systems, as well as the implications of these design choices for educational participation.

Accepted but not presented.

The discussion explores the transformation of lexicography in the digital era, including the movement from print-based reference works to interactive, mobile-first, and user-centered platforms. These developments create new possibilities for supporting multilingual learners through enhanced search functions, contextualized definitions, translation features, multimodal content, and personalized learning experiences. At the same time, the paper recognizes that technological innovation does not automatically produce inclusion. Digital lexicographic resources may continue to reproduce existing inequalities when they fail to address linguistic diversity, cultural specificity, and varying levels of digital literacy.

The paper concludes that understanding digital dictionaries as inclusive knowledge infrastructures offers a productive framework for addressing linguistic inequalities in multilingual educational contexts. By foregrounding accessibility, cultural relevance, and epistemic inclusion, digital lexicography can contribute to more equitable forms of knowledge access and participation. Furthermore, the integration of indigenous languages and locally grounded knowledge into digital lexicographic systems has the potential to support broader processes of educational transformation and epistemic justice in the Global South. The paper contributes to ongoing debates on language, technology, and social inclusion by arguing that the future of digital lexicography lies not only in technological advancement but also in its capacity to foster more inclusive and socially responsive knowledge systems.

References

Fricker, M. 2013. Epistemic justice as a condition of political freedom? *Synthese* 190(7): 1317–1332.

Mariën, I. & Prodnik, J.A. 2014. Digital inclusion and user (dis)empowerment: A critical perspective. *Info* 16(6): 35–47.

*****

Maximising the Development of Tshivenda Terminology as a Tool for Sustainable Development Goal 4 (Quality Education) in Rural Limpopo

Mashudu Peter MAKHADO (peter.makhado@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa

According to the Agenda for Sustainable Development (United Nations 2015), the Sustainable Development Goals (SDGs) are 17 global objectives adopted by the United Nations to end poverty, protect the planet, and ensure peace and prosperity by 2030. Lexicography can respond to them by documenting, standardising, and disseminating the evolving vocabulary of sustainability, thereby supporting education, policy-making, and cross-cultural communication, including Sustainable Development Goal 4 (Quality Education), which promotes inclusive education that respects linguistic and cultural diversity. Indigenous terminology plays a crucial role in supporting cultural development through language preservation. Lexicography also plays a key role in advancing the SDGs by standardizing terminology, promoting multilingual access, thus ensuring translation and harmonizing SDG-related vocabulary across languages, tracking language evolution and supporting education and awareness by developing dictionaries and glossaries tailored to SDGs can be used in schools, universities, and community programs to spread awareness and foster global citizenship.

In Limpopo, Tshivenda terminology represents not only cultural heritage but also a practical tool for advancing rural development. Tshivenda, as one of the dominant indigenous languages in Limpopo—alongside Northern Sotho and Xitsonga—has a critical role in enhancing sustainable development in the province, particularly in the Vhembe District.

Accepted but not presented.

This is a qualitative study. Data was collected from documents, including policy papers, academic studies, and educational resources, and analysed through the content analysis method. This approach allows for systematic examination of how Tshivenda terminology is documented, standardised, and applied within rural education and development contexts. Data analysis was completed through terminology extraction focusing on terms related to education, sustainability, education, sustainability and SDG 4, and semantic analysis where gaps will be identified to determine the need for new terminology.

This paper argues that the systematic documentation, standardisation, and application of Tshivenda terminology can be used as a tool to enhance quality education, thereby aligning with South Africa's broader development agenda. This paper aims to show the crucial role of lexicography as a terminology development domain in enhancing sustainable development in rural areas in the Limpopo Province.

Embedding Tshivenda terminology in rural schools enhances comprehension and reduces dropout rates by allowing learners to engage with content while being grounded in early literacy and comprehension in mother-tongue education (Nethengwe 2022). As mother-tongue instruction improves cognitive development and supports multilingual education policies (Alexander 2023), terminology development becomes a valuable enabler. Standardised terminology can bridge gaps between indigenous knowledge and modern curricula, ensuring that rural learners are not disadvantaged in national education systems.

This study also emanates from challenges observed in the status of Tshivenda terminology. These include limited documentation in technical fields, inadequate digital resources, and inconsistent policy implementation. Few dictionaries or mobile applications support Tshivenda, limiting accessibility for younger generations, while a lack of textbooks with standardised terminology further constrains progress (Netshivhumbe & Mudau 2024). Consequently, without standardised terminologies, rural communities risk linguistic marginalisation. The burden therefore rests on lexicographers, terminologists, and translators to strengthen initiatives aimed at developing proper terminology to enhance quality mother-tongue education in rural areas.

Accepted but not presented.

This paper provides proposals to strengthen the quality and accessibility of Tshivenda terminology as a tool for sustainable development. It advocates for terminology development in local languages to ensure accessibility and sustainable growth, as suggested by Makgopa (2023). Developing mobile applications and e-learning platforms that incorporate Tshivenda terminology will assist in embedding it into rural development programmes, ensuring inclusivity in skills development and local economic projects. Bilingual dictionaries are vital for development in rural areas. They help translate complex global concepts into local languages, ensuring communities understand and apply them in education, agriculture, health, and governance. This supports SDG 4 (Quality Education), SDG 8 (Decent Work), and SDG 11 (Sustainable Communities).

Tshivenda terminology is not merely a linguistic resource but a strategic instrument for sustainable rural development. By integrating it into education, agriculture, governance, and digital platforms, rural communities in Limpopo can achieve inclusive growth while preserving cultural heritage. Sustainable development requires cultural responsiveness, and Tshivenda terminology provides the foundation for a development model that is both locally grounded and nationally aligned. Finally, this paper makes a scholarly contribution by demonstrating how indigenous language lexicography can serve as a critical framework for aligning linguistic innovation with global development agendas, thereby enriching both language planning and SDG-focused research.

References
Alexander, N. 2023. *Language policy and multilingual education in South Africa*. Pretoria: HSRC Press.
Makgopa, M.S.F. 2023. Information and Communication Technology Model to Enhance Small-Medium Enterprises Development Using Indigenous Languages in Limpopo Province. Unpublished doctoral dissertation. Durban: University of KwaZulu-Natal
Mudau, P. P. 2025. Respect and politeness as ingredients of indigenous address terms: A phenomenological study of Tshivenda speakers from Mavhulani, Limpopo. *African Identities* (January): 1–18
Nethengwe, T. 2022. Linguistic Features of Tshivenda. Unpublished prestatio. Reading Benchmarks for African Languages*:* Department of Basic Education.
Netshivhumbe, N. P., & Mudau, A. V. 2024. Perceptions, challenges, and opportunities in the development of the Tshivenda scientific register for physical sciences. *South African Journal of Education* 45(2): Art. #2477.
Phalanndwa, A.C. 2024. Teachers' Perceptions on Development of Tshivenda Reading Skills at Selected Primary Schools in Vhembe East District of Limpopo Province. Unpublished M.Ed. thesis. Polokwane: University of Limpopo.
United Nations. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development*. Available: https://sdgs.un.org/2030agenda.

*****

Evaluating the Status of Translation Equivalents in Selected Northern Sotho-English Bilingual Dictionaries

Mmagonkahloleng Brudence MAKUA (mmakua@ufh.ac.za)
Department of Languages and Communication Studies, University of Fort Hare, East London, South Africa

This paper seeks to evaluate problems of translation equivalence in bilingual dictionaries with Northern Sotho and English as language pair. Although translation equivalence has been studied comprehensively, especially in the older literature, Northern Sotho confronts lexicographers with numerous problems that need to be solved. Translation dictionaries deal with two or more languages and provide target language translation equivalents of source language lexical items, as opposed to statements about the meaning of source language items, as users and even some lexicographers often believe. Bilingual dictionaries in South African languages have existed for many years. These dictionaries may serve different purposes depending on the communicative needs of the dictionary user and the amount of data supplied by the compiler.

Nefale (2009:15) defines "translation equivalence" as the coordination between a word or a phrase in one language and its translation in another language. Translation equivalence represents a relation that holds between a source language item and one or more items in the target language that can be used to translate a specific occurrence of the source language item.

For this research, two of the leading bilingual dictionaries with English and Northern Sotho as the language pair were selected. They were evaluated, and a model was suggested for an improved treatment of translation equivalence in future bilingual dictionaries. Qualitative research methods make use of documents and texts, among others, as sources of data (Myers 2009). Thus, the evaluation of translation equivalence in the two selected bilingual dictionaries formed the point of departure. In this connection, the analysis, discussion and interpretation, presentation and treatment of translation equivalents and the lack of translation equivalents was based on data found in the *Popular Northern Sotho Dictionary* (Kriel, Prinsloo & Sathekge 2011) and *Pukuntšu ya Sekolo School Dictionary* (De Schryver & Mogodi 2015) bilingual dictionaries. The study investigated whether the lexicographers followed such an approach and how the articles with insufficient treatment of translation equivalence and the

presentation of items to ensure equivalent discrimination to support the genuine purpose and the functions of these dictionaries can be improved. The discussion focused on scientific terms, cultural terms, and words of general use in the language as presented in the selected dictionaries.

Key findings indicate that there currently is a significant gap in solutions provided by the prevailing theory and its applications in the lexicographic practice, and the challenges found when planning and compiling bilingual dictionaries with Northern Sotho as one of the treated languages. Implications of these findings are profound. Consequently, existing Northern Sotho dictionaries often do not satisfactorily treat equivalence.

References

De Schryver, G.-M. & Mogodi, M. (eds.). 2015. *Pukuntšu ya Sekolo School Dictionary.* Cape Town: Oxford University Press South Africa.

Kriel, T.J., Prinsloo, D.J. & Sathekge, B.P. (eds.). 2007. *Popular Northern Sotho Dictionary.* Cape Town: Pharos Dictionaries.

Myers, M.D. 2009. *Qualitative Research in Business & Management.* London: SAGE

Nefale, S.J. 2009. The role of equivalence in the creation of terminology in Tshivenḓa: A case of National Curriculum Statement in Grades R-9 (2002). Unpublished MA thesis. Turfloop: University of Limpop.

*****

Examining How Technological, Scientific and Cultural terms are Treated in Terms of Translation Equivalents in Bilingual Dictionaries

Matlala Rivonia MAKWALA (makwalamr@tut.ac.za)

Department of Applied Languages, Tshwane University of Technology, Polokwane, South Africa

Mmagonkahloleng Brudence MAKUA (mmakua@ufh.ac.za)

Department of Languages and Communication Studies, University of Fort Hare, East London, South Africa

This paper will discuss the presentation of translation equivalents in the bilingual dictionaries *Pharos Popular Northern Sotho Dictionary* (PPNSD) and *Oxford Popular School Sekolo Dictionary* (OPSSD). The paper will examine different types of equivalent relations and how the two assessed bilingual dictionaries have presented their translation equivalents with regard to congruence, divergence, surrogate equivalence and the treatment of general words and culture-bound terms. This research will use a qualitative research method which makes use of documents and texts, among others, as sources of data (Myers 2009). Thus, in this paper, the examination of how technological, scientific and cultural terms are treated in bilingual dictionaries is based, as its point of departure, on the analysis of the PPNSD and OPSSD bilingual dictionaries. In this connection, the identification, analysis, discussion and interpretation of the presentation will be based on the two bilingual dictionaries.

Mehlape (2007:2) points out that most lexicographers of Northern Sotho English/English-Northern Sotho bilingual dictionaries often have a problem providing an equivalent for technical, scientific, and general language terms and cultural terms in their dictionaries. Al-Kasimi (1983:61) suggests that in the treatment of translation equivalence bilingual dictionaries should not only provide translation equivalents but also items giving meaning discrimination to enable a user to select the appropriate equivalent. He further maintains "… unless the problem of meaning discrimination is solved systematically, the bilingual dictionary cannot be a dependable guide to the proper equivalents".

Different types of equivalent relations can be identified in translation dictionaries. These

differences are the result of varying degrees of equivalence that hold between source and target language items. A distinction can be made between full equivalence, partial equivalence and zero equivalence. This distinction motivates the existence of the following types of equivalent relations: congruence (full equivalence), lexical and semantic divergence (partial equivalence) and surrogate equivalence (zero equivalence). According to Gouws (1996), both source and target language forms have the same meaning and the same stylistic and pragmatic value. This can be seen in examples like the following from *Oxford Pukuntšu ya Sekolo School Dictionary* (OPSSD):

> **aeroplane** noun (pl. **aeroplanes**) sefofane Our **aeroplane** will leave at one o'clock. *Sefofane sa rena se tla tloga ka iri ya pele*.
> **coffee** noun (pl. **coffees**) *kofi*. Would you like to drink **coffee** or tea? *Na o tla thabela gon nwa kofi goba teye?*
> **kankere** noun 9/10 (pl. **dikankere**) cancer. Ke gopola go dira diteko tša **kankere** ya letswele. *I'm thinking of undergoing tests for breast cancer*.

Congruence implies that the translation equivalent can substitute the lemma in all its uses. For the lexicographer, this type of equivalent relation only holds a few problems. An equivalent relation of congruence could lead to a deviation from the idea of a homogeneous article structure with a similar treatment allocated to all lemmas. According to Gouws (2002:196), full equivalence implies that the source language and the target language lexical items are equivalent lexically, pragmatically and semantically, which means that the target language lexical item can replace the source language lexical item in all contexts and uses. From the lexicographic perspective, full equivalence is considered relatively unproblematic, but as Zgusta (1971:312) points out, absolute equivalents are quite rare.

Divergence is characterized by a one to more than one relation between source and target language forms. Gouws (1998:12) states that for a given lemma the translation equivalent paradigm will contain more than one translation equivalent. Different types of divergence occur. According to Gouws (2002:195-209), where lexical divergence prevails, the equivalents are target language synonyms. The lexicographer must make sure whether the translation equivalents are full synonyms or not. If they are full synonyms, which is very seldom the case, the lexicographic treatment can be the same as in the case of congruence. But often the equivalents are only partial synonyms. In such cases, the user should not only be informed about the fact that these partial synonyms can substitute the source language word but should also be cautioned that the target language equivalents cannot substitute one another in all contexts. Here, meaning discrimination comes into play. As a result, the lexicographer should make use of some sort of contextual and co-textual items to indicate the typical pragmatic or syntactic environment of the translation equivalents. In conclusion, the study highlights the importance of a systematic approach to the treatment of translation equivalents, particularly in the case of technological, scientific and cultural terms.

References

Al-Kasimi, A.-M. 1983. The interlingual/translation dictionary. Hartmann, R.R.K. (ed.). *Lexicography: Principles and practice*: 153–162. London: Academic Press.
De Schryver, Ğ.-M. (ed.). 2007. *Oxford Bilingual School Dictionary: Sesotho sa Leboa–Seisimane, English–Northern Sotho*. First Edition. Cape Town: Oxford University Press Southern Africa.
Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
Gouws, R.H. 1998. Equivalent Relations in Translation Dictionaries. MPhil Study Guide.

University of Stellenbosch, Stellenbosch.
Gouws, R.H. 2002. Equivalent Relations, Context and Co-text in Bilingual Dictionaries. *Hermes – Journal of Language and Communication Studies* 28: 195-209.
Kriel, T.J. Prinsloo, D.J. & Sathekge, B.P. 2007. *Popular Northern Sotho Dictionary*. Cape Town: Pharos.
Mehlape, P.P. 2007. The role of lexicographic definition in Northern Sotho-English/English-Northern Sotho. Unpublished MA thesis. Polokwane: University of Limpopo.
Myers, M.D. 2009. *Qualitative Research in Business & Management.* London: SAGE
Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Thirty Years of African Lexicography: The Contribution of AFRILEX to Indigenous Language Development in South Africa

Masila Joshua MASIPA (masilajoshua.masipa@univen.ac.za)
Department of Arts and Social Sciences, University of Venda, Thohoyandou, South Africa
Nontokozo Gladys MDHLULI (nontokozo.mdhluli@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa

The development of indigenous African languages in South Africa remains closely linked to lexicography, terminology development, and language intellectualisation. Since its establishment in 1995, the African Association for Lexicography (AFRILEX) has played a significant role in advancing African-language lexicography through scholarly collaboration, dictionary compilation, terminology standardisation, and the promotion of multilingualism. This paper examines AFRILEX's contribution to the development of indigenous South African languages over the past three decades and evaluates its impact within the broader context of language planning and decolonisation.

The study employs a qualitative historical-analytical approach based on documentary analysis of AFRILEX conference proceedings, *Lexikos* publications, policy documents, and selected dictionaries produced during the democratic era. The analysis is informed by theories of language planning, intellectualisation, and lexicography as an interdisciplinary field.

The findings indicate that AFRILEX has made three notable contributions to indigenous language development. First, it has strengthened the scholarly legitimacy of African-language lexicography by creating sustainable academic networks and research platforms. Second, it has contributed to corpus development and language standardisation through the promotion and dissemination of dictionaries and terminological resources in indigenous languages. Third, it has supported the training and mentorship of emerging lexicographers, thereby ensuring continuity in the field. The study further finds that AFRILEX's activities have contributed to the preservation of indigenous linguistic heritage and have aligned closely with contemporary decolonial efforts aimed at repositioning African languages within knowledge-production systems. However, the impact of these achievements has been constrained by persistent challenges, including the dominance of English in higher education and digital communication, limited funding for indigenous-language projects, inadequate institutional support, and the slow pace of digital lexicography development. Despite these obstacles, AFRILEX has remained a key driver of indigenous language advancement and intellectualisation in South Africa.

The paper concludes that greater investment in digital lexicography, artificial intelligence applications, multilingual education policies, and cross-border linguistic collaboration is necessary to consolidate and extend the gains achieved over the past thirty years.

*****

Accepted but not presented.

Towards Inclusive Lexicography: The Need for User-Friendly Tshivenḓa Dictionaries in Education and Society

Mashudu MAṬHABI (mashudu.nthambeleni@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa

Since the beginning of democracy in South Africa in 1994, there has been a growing focus on developing and promoting indigenous African languages such as Tshivenḓa. This shift has been influenced by national language policies that support multilingualism and aim to redress the historical marginalisation of African languages during the apartheid era. As a result of this increased attention, several Tshivenḓa dictionaries have been produced over the years to support language development, teaching, and learning. Despite the growing number of published dictionaries, there remains uncertainty about whether these lexicographic works truly meet the needs of their intended users. In particular, there are concerns about whether learners, teachers, and ordinary speakers of Tshivenḓa are able to use these dictionaries effectively in real-life situations. Many dictionaries appear to focus more on academic and linguistic accuracy than on user-friendliness and accessibility. This raises an important question about the actual usefulness of these dictionaries in education and everyday communication. This paper argues that the value of Tshivenḓa dictionaries should not only be measured by the number of publications available, but also by the extent to which they are easy to use, accessible, and suitable for different categories of users. In other words, a dictionary should not only contain correct linguistic information but should also present that information in a way that is clear and practical for its users. The paper is guided by user-centred lexicographic theory, such as the Modern Theory of Lexicographic Functions (Bergenholtz & Tarp 2003), which emphasises that dictionaries should be designed according to the needs, skills, and consultation situations of their intended users. It also draws on pedagogical lexicography, which views dictionaries as important tools for teaching and learning, especially in formal educational contexts (Tarp 2011). These theoretical perspectives support the idea that dictionaries must be simple, well-structured, and responsive to the needs of learners and other non-specialist users. The paper adopts a qualitative research methodology, using a descriptive and evaluative research design. This approach allows for a detailed and interpretive examination of selected Tshivenḓa dictionaries published after 1994. The analysis focuses on key aspects that influence user-friendliness, such as the clarity and simplicity of definitions or equivalents, the use and quality of example sentences, the representation of pronunciation guidance, and the ease with which users can locate information within the dictionary. These aspects are important in determining whether a dictionary can effectively support its users in understanding and using the language. Purposive sampling was used to select the dictionaries included in the paper. The dictionaries were carefully chosen based on their availability, relevance, and widespread use in academic and educational environments. The sample includes printed dictionaries published after 1994. This selection was made to ensure that the paper covers commonly used lexicographic resources that are accessible to learners, teachers, and the general Tshivenḓa-speaking public.

The findings of the paper indicate that many Tshivenḓa dictionaries remain relatively complex in structure and content. They are often designed with academic or specialist users in mind, rather than ordinary speakers or learners of the language. As a result, their accessibility is limited, particularly for learners in schools who require clear and simple explanations. This reduces the effectiveness of these dictionaries as tools for language

Accepted but not presented.

learning, reading development, and everyday communication. The problem is especially evident in rural and under-resourced schools, where learners often depend on basic and easy-to-use learning materials. Furthermore, the limited availability of digital Tshivenḓa dictionaries presents an additional challenge. In the modern digital age, many young users rely on mobile phones, tablets, and online platforms to access information. The lack of user-friendly digital dictionary resources therefore limits access and reduces opportunities for independent learning. The paper also relates these findings to South Africa's language policy, which promotes equality and the development of all official languages. It argues that if dictionaries remain inaccessible and difficult to use, Tshivenḓa may continue to be marginalised in practice, despite its official status. This highlights the need for inclusive lexicography that prioritises usability, accessibility, and real-life relevance for all categories of users.

In conclusion, the paper recommends that future Tshivenḓa dictionaries should be developed with a stronger focus on user needs. This includes producing more learner-friendly dictionaries, improving collaboration between lexicographers and educators, and expanding the development of digital dictionary platforms. Such improvements would make Tshivenḓa dictionaries more effective educational tools and contribute to the continued growth and empowerment of the language in a democratic South Africa.

Accepted but not presented.

References

Bergenholtz, H. & Tarp, S. 2003. Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic function. *Hermes – Journal of Language and Communication Studies* 31: 171–196

Tarp, S. 2011. Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. *Lexikos* 21: 217–232.

*****

An Approach to Digital African Lexicography That Can Help Bridge the Accessibility Gap

Nontokozo Gladys MDHLULI (nontokozo.mdhluli@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa
Pearl Linda KABINI (pearl.kabini@univen.ac.za)
Department of African Languages, University of Venda, Thohoyandou, South Africa

This study looks at how AFRILEX has contributed to the development of digital African lexicography during the last thirty years, with a focus on closing the accessibility gap in African language resources. AFRILEX is a group of academics and lexicographers that was founded in 1995 and whose combined efforts have aided in the production, digitisation, and distribution of African language lexicographic resources. Therefore, rather than focusing only on individual members, the study evaluates AFRILEX's institutional and scientific contributions to African lexicography and assesses the degree to which advancements in digital lexicography have increased access to indigenous language resources.

The study is guided by the Functional Theory of Lexicography (Bergenholtz & Tarp 2003), which emphasises the importance of dictionaries meeting the cognitive and communicative demands of their users. This theoretical framework is supplemented by Digital Humanities approaches, which encourage the incorporation of technological innovation into the preservation and transmission of linguistic and cultural knowledge (cf. Wang et al. 2025). These frameworks are utilised to interpret how digital technologies and lexicographic practices contribute to accessibility, multilingualism, and language preservation, which are

directly relevant to the study's objectives.

A qualitative desktop research design was used. Data were gathered from scholarly articles, conference proceedings, policy documents, online dictionaries, and AFRILEX-related publications. The data were analysed using thematic content analysis, with an emphasis on patterns linked to digitalisation, accessibility, indigenous language preservation, and lexical innovation.

The findings show that AFRILEX has played an important role in advancing African lexicography through conferences, research outputs, terminology development projects, and collaborative initiatives involving African linguistic institutes. Online dictionaries, smartphone applications, and electronic terminology databases have all made it easier to access African language information. However, barriers such as limited finance, insufficient technological infrastructure, digital illiteracy, and unequal representation of minority languages continue to impede accessibility. These data demonstrate that accessibility difficulties go beyond availability and include usability, price, and language inclusion.

According to the study's findings, digital African lexicography significantly advances multilingual education, language preservation, and the decolonisation of knowledge systems. To improve accessibility and utility of lexicographic resources, however, lasting improvement necessitates increased funding, more robust institutional cooperation, and the incorporation of cutting-edge technology like artificial intelligence.

References

Bergenholtz, H. & Tarp, S. 2003. Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes – Journal of Language and Communication in Business* 31: 171–196.

Wang, G., Khalid, Y.I.A.M., Abdullah, N., Xu, Y.-S. & Yuan, Y. 2025. Tradition meets innovation: Research trends in libraries, archives, and museums convergence from a digital humanities perspective. *Knowledge Organization* 52(4): 395–414.

*****

The *Greater Dictionary of IsiXhosa* and Standardisation in IsiXhosa Writing

Buyiswa Mavis MINI (buyi.mini@gmail.com)
School of Languages and Literatures: African Language Studies, Rhodes University, Makhanda, South Africa.

In 2023, the isiXhosa language reached its 200-year milestone as a language of writing and printing (Mona 2023:1). This milestone in 2023 presupposes a stable level of standardisation in its orthography and other writing conventions. The history of the (then unconscious) language planning processes the language underwent, since 1823 (Mona 2023:2), shows a picture of the occurrence, to an extent, of the following four types of language planning: "*status planning* (... social standing ...), *corpus planning* (the structure of a language), *language-in-education planning* (learning), and *prestige planning* (image)" (Nordquist 2025, online update). It became the standard church and school language, medium of some school and church materials, and various literary genres and dictionaries. This paper focuses on corpus planning. While the other forms of language planning have not seen the desired level of progress, a few scholars have employed the language for research and writing dissertations. Also, a steady amount of literature continues to be produced over the years. Many writers across all age groups and genders have also emerged in the current post-1994 era. The scope of the paper does not allow much specific details in this regard.

The increase, especially in recent literary production in the language, has exposed the existence of knowledge gaps, inconsistencies and instances of reversal in orthographical and

other writing conventions which had long become well standardised. These are now threatening the quality of writing. Using document study to compare older and more recent literary works, this paper presents a description of these areas of concern with view to attracting constructive debate about them, leading to lasting improvements. The existing lexicography works in the language, especially the more comprehensive *Greater Dictionary of IsiXhosa* (GDX), should be part of the resources to be appealed to when engaging in such debates. The paper therefore calls for engagements and debates geared towards solutions by posing the question: Do isiXhosa dictionaries in general, and the GDX, in particular, provide adequate guidance for authors, learners, students, educators, language planners and policy makers and other language practitioners, in these aspects wherein accuracy and consistency are lacking and standardisation gaps are evident, affecting quality work in, and through, the language?

The glaring inconsistencies, and inaccuracies are observable in a number of literary works, such as *Umfundisintsapho Ulibulele Ikamva Lam* (Ngqokelela 2023.), Ubuhlobo BukaZazi noLili (Pakama 2013), to name a few. They include the following: word division, punctuations, lexical misrepresentation (e.g. **iziqulatho* vs *isiqulatho* for "contents"), women clan names, factual and language inaccuracies, inconsistency in writing of titles and headings, capitalisation, hyphenation, presentation of dramatic characters, meanings of figures of speech, idioms, and other figurative expressions. Referring to Haugen's (1983) corpus planning framework, Baldauf Jr., (2008) summaries these corpus planning aspects into two categories, namely, normative and those that relate to linguistic functions.

For a simple conceptual framework, the paper adapts Mahanta's (2019) idea that a dictionary is an important source of valuable information for, among others, meanings, spellings, correct pronunciation, punctuations, and other language conventions. It provides "... the readers the best and most comprehensive information about words" (Mahanta 2019:269). This is one of the guiding reasons for this paper's belief that the GDX will fulfil this ideal for the improvement of writing quality in, and the digitalisation of, isiXhosa.

One of the components of the General Theory of Lexicography that Mahanta (2019:270) mentions, is the "genuine purpose of dictionaries". As part of his conclusion about the latter, Mahanta (2019:271) makes another important observation that "Wiegand considers lexicography as a linguistic reconstruction of information." It is the assertion of this paper that the GDX adequately fulfils this ideal of information reconstruction, and that the evaluation, simplification and dissemination to stakeholders, of the guidelines it contains, will be a valuable contribution to the improvement of the quality of literary and other forms of writing in isiXhosa.

The paper concludes by pointing out that the GDX, in its macrostructure, microstructure and extra matter, is a source of considerable information concerning the above-mentioned aspects of isiXhosa. The design of the GDX is already part of the solution, in the non-lemma front and back matter sections of each of the volumes, available for language users.

References

Baldauf Jr., R.B. 2008. Language planning: Corpus planning. *Annual Review of Applied Linguistics* 10: 3–12.

Mahanta, M. 2019. A study of two theories of lexicography: Their contribution to lexical meaning. *Journal of Emerging Technologies and Innovative Research* 6(6): 269–274.

Mona, G.V. 2023. Celebrating 200 years of IsiXhosa in writing and printing. *Journal for African Writing: Nourishing the mind* 1(6): 1–10.

Nordquist, R. 2025. What is Language Planning? https://www.thoughtco.com/what-is-language-planning-1691098. Accessed: 12 June 2026.

*****

User Studies in the South African Lexicographic Laboratory: Methodologies and Case Studies

Lorna Hiles MORRIS (lorna@lemma.co.za)
Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa

User studies are an important part of lexicography. They guide lexicographers on the design of the dictionary in the planning stage, a dictionary's effectiveness and how it is received. They also help lexicographers know more about the users of their dictionaries. There are many different methods of studying dictionary use and users, and this paper presents different types of user studies and their applications, particularly within a South African environment. The nature of user studies has changed with the advent of digital lexical tools, as well as digital options of user studies, such as mass mailing of questionnaires, and survey software being available tools.

This paper is placed in a South African context because of the country's unique challenges such as historic inequalities, low literacy rates, and multilingual teaching environments. We also have a lack of dictionary culture and low levels of dictionary use and skills (Nkomo 2020). People do not have experience using dictionaries and may not have good consultation skills. However, "[due] to its multilingual and multicultural nature, South Africa can be regarded as a lexicographic laboratory that lends itself to different experiments … regarding dictionary use" (Gouws 2019:35).

This paper explains how user studies are useful at every stage of dictionary-making: planning, production, and post-production.

During the planning stage a needs analysis is vital. Lexicographers need to know who their potential users are, and what their needs are.

During production, different features or design decisions can be tested with pilot studies, that can lead to the establishment of the dictionary design and style guide. Sample entries can be tested with users.

Post-production studies include field testing the completed dictionary, with short term studies or longitudinal studies showing the effectiveness of the dictionary on vocabulary acquisition and retention, or improved results in school or university courses.

This paper also discusses what can be established from user studies

- what triggers a dictionary consultation,
- user behaviour,
- users' opinions and preferences on dictionaries.

This paper shows what decisions can be made based on user studies, whether it's the design of the dictionary, entry components, the inclusion and treatment thereof, as well as lemma selection.

The following types of user studies are presented: questionnaires, interviews, observation, experiments, tests, case studies, and focus group discussions. This study also discusses the limitations of user research, along with ways of overcoming these. Some of the limitations of user research include social desirability in subjects, observer's paradox, fallibility of instruments and devices, and potential language barriers between the interviewer and interview subjects (Lew 2015).

Social desirability is where the subject tries to give the interviewer the answers they are looking for, rather than honest answers, or they fill in a questionnaire with the answers they think the lexicographer wants. Observer's paradox is where the subjects perform differently,

or less naturally, when they are being observed. Instruments can be unreliable in a test situation, and there need to be backups in case of failures. Recording devices may also fail.

After this overview, three case studies of user studies from my own research are presented. These are: classroom observations, focus group discussions, and interviews. These are from a study undertaken in 12 South African primary schools, with learners being the subjects of the classroom observations and focus group discussions, and teachers being the subjects of the interviews.

The classroom observations took place in Grade 5 and 6 classes. Learners were given a stand-in dictionary device with some sample dictionary entries loaded. The learners were observed as they looked through the entries and activated audio and hyperlinks. The observations were done by the researcher, a research assistant, and in most, but not all cases, the class teacher. Following the observations, a group of learners was selected by the class teacher to take part in a focus group discussion with the researcher. Their opinions of the dictionary entries as well as those of a dictionary device were discussed. Finally, teachers were interviewed telephonically. Reasons for these methods of user studies are explained in the paper. The case studies will cover sample selection, research tools, preparation, the actual studies, outcomes, and lessons learned.

This paper concludes with recommendations for the practical application of user studies in South African lexicography.

References

Gouws, R.H. 2019. Metalexicographic models for multilingual online dictionaries in emerging e-societies. *Studies on Multilingual Lexicography* 157: 29–45.

Lew, R. 2015. Opportunities and limitations of user studies. *OPAL-Online publizierte Arbeiten zur Linguistik* 2015(2): 6–16.

Nkomo, D. 2020. Dictionary culture in African language communities: research, development, challenges and prospects. *Lexicographica* 36: 11–37.

*****

Digital Indigenisation of Terminology: How Community-Based Media Platforms in Limpopo Generate and Disseminate Economic Neologisms in Tshivenda and Xitsonga

Mutshidzi Portia MUNYAI (Munyaimutshidzi54@gmail.com)
Department of English, Media Studies and Linguistics, University of Venda, Thohoyandou, South Africa

Accepted but not presented.

This paper investigates the digital indigenisation of terminology by examining how community-based media platforms in Limpopo contribute to the creation, adaptation, and circulation of economic neologisms in Tshivenda and Xitsonga. Located at the intersection of lexicography, sociolinguistics, and media studies, the study argues that digital platforms such as community radio stations, Facebook pages, and WhatsApp groups function as participatory spaces for bottom-up terminology development. In many multilingual African contexts, economic discourse remains largely mediated through English, limiting access for speakers of indigenous languages. This study contends that local communities actively address this gap by generating contextually grounded terminology that reflects lived economic realities, thereby expanding linguistic accessibility and participation (Crystal 2011; Banda 2018).

A qualitative research design is employed, combining discourse analysis of selected broadcasts and online interactions with semi-structured interviews conducted among radio presenters, social media content creators, and audience participants in Limpopo. The analysis

focuses on widely used community platforms, including local radio talk shows and informal digital discussion forums where economic issues are debated in indigenous languages. Through this approach, the study traces how new economic terms emerge, are negotiated, and gain social acceptance within everyday communication. The findings indicate that speakers draw on multiple linguistic strategies, including borrowing, semantic extension, calquing, and code-switching, to construct meaning in Tshivenda and Xitsonga. For example, terms related to "inflation" are often rendered through descriptive phrases that metaphorically convey rising prices, while "entrepreneurship" is localised through expressions that emphasise self-reliance and informal business practices. Similarly, concepts such as mobile money and digital transactions are adapted through hybrid lexical forms that combine indigenous morphology with English-derived stems, making them more intelligible within local communicative contexts (Nkomo & Wababa 2013).

The study further demonstrates that audience participation plays a crucial role in shaping and legitimising emerging terminology. Interactive features such as call-in segments, listener feedback, comment threads, and message sharing enable collaborative processes of meaning-making, through which terms are proposed, debated, and gradually normalised. In this sense, community-based media platforms operate as informal yet influential mechanisms of lexical innovation and diffusion. Unlike traditional lexicographic processes, which are often centralised and slow to respond to linguistic change, these digital environments allow for rapid experimentation and adaptation. However, the study also identifies significant challenges associated with this decentralised model of terminology development. These include inconsistencies in usage across platforms, absence of standardised orthographic conventions, and limited documentation of emerging terms within formal lexicographic resources. As a result, while these platforms foster creativity and inclusivity, they also produce instability that may hinder wider institutional adoption.

The paper argues that these developments have important implications for lexicography and language planning in Africa. It suggests that lexicographers should engage more actively with community-based digital platforms as valuable sources of linguistic data and innovation. Incorporating such data into formal terminology development processes could enhance the relevance, responsiveness, and inclusivity of lexicographic outputs. Furthermore, the study contributes to broader debates on linguistic decolonisation by demonstrating how speakers of Tshivenda and Xitsonga assert epistemic agency through localised knowledge production. By enabling economic concepts to be expressed in indigenous languages, these practices challenge the continued dominance of colonial languages in specialised domains and support more equitable knowledge dissemination (Ngũgĩ wa Thiong'o 1986; Finnegan 2015).

In conclusion, this paper positions community-based digital media platforms as key agents in the ongoing evolution of African languages, particularly in specialised domains such as economics. It emphasises that terminology development is not solely an institutional activity, but a socially embedded process shaped by everyday language users. The study therefore calls for stronger collaboration between linguists, media practitioners, and local communities to harness the potential of digital environments for sustainable and inclusive terminology development.

Accepted but not presented.

References

Banda, F., 2018. *African multilingualism and the decolonisation of language*. London: Routledge.
Crystal, D., 2011. *Internet linguistics: A student guide*. London: Routledge.
Finnegan, R., 2015. *Communicating: The multiple modes of human communication*. 2nd ed. London: Routledge.
Ngũgĩ wa Thiong'o, 1986. *Decolonising the mind: The politics of language in African literature*.

London: James Currey.
Nkomo, D. & Wababa, Z. 2013. The impact of translation activities on the development of African languages. *Lexikos* 23: 251–272.

*****

Exploring Female Colloquial Reference in South Africa: A case of *Mzansi Taal* Online Dictionary

Neliswa MZIMBA (mzimbn1@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa
Napjadi Eugene LETSOALO (letsone@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa

Contrary to the apparent role that women have in society as bearers of children, home makers, sole providers, leaders, visionaries and community builders, women are often portrayed negatively and as inferior. Linguistically, the notion that women are unimportant is conveyed through toxic derogatory terms and these terms are unceasingly coined and habitually used to characterise women's behaviour and construct their identity in society.

Lexicography provides significant understandings into the ways that language influences and reflects societal norms, beliefs, and perceptions. In the latest South African online dictionary, Mzansi Taal, which has documented colloquial terms used in black communities in informal everyday conversations, a similar trend is observed. The entries identified frequently reflect and reinforce gendered stereotypes and negative representations of women. Gendered colloquial language use in South Africa is an important but relatively understudied area of research. This paper will focus on the use of gendered colloquial terms and their social implications.

The study draws on Social Constructionism, Discourse Analysis and Critical Discourse Analysis (CDA) as its theoretical frameworks. Social Constructionism (Andrews 2012) emphasizes that knowledge, reality, and social phenomena are not naturally given or fixed but are produced and maintained through social interactions, cultural practices, historical contexts, and shared understandings. From this perspective, what people accept as truth or reality is shaped collectively through social processes and is therefore subject to change across different societies and historical periods. Social Constructivism on the other hand offers a lens through which to understand the role of language in shaping identities and societal perceptions. According to Jung (2019) the world is shaped by shared ideas, beliefs, and norms and interests are built through collective meanings, identities and social interaction. Social Constructivism highlights the critical role of language, communication, and interpretation in the development of knowledge and understanding. While both viewpoints recognize that knowledge is socially shaped, social constructionism is mainly concerned with how societies communally create and sustain realities, while social constructivism focuses on how individuals develop understanding and meaning within those social contexts. These frameworks provide a useful lens for examining how language, discourse, and social interaction influence knowledge, culture, and identities.

Discourse Analysis is utilised to explore how gender is represented in language use, with discourse understood as language in context, where linguistic expressions carry socially embedded meanings that inform perceptions of gender roles (Bucholtz 2003). Discourse analysis provides a methodological approach for examining the broader social contexts in

Accepted but not presented.

which language operates and on the other hand CDA provides a robust framework for examining the implications of language use in relation to gender. CDA focuses on the ways in which language constructs and perpetuates ideologies, particularly concerning gender. CDA places gender ideologies at the forefront of analysis, examining texts that normalise and naturalise inequalities (Bucholtz 2003). Focusing on black township contexts, the study demonstrates how specific linguistic choices contribute to the construction and reinforcement of gendered perceptions and social realities.

The study uses a qualitative research approach with thematic analysis to explore the representation of women in the Mzansi Taal Online Dictionary. Thematic analysis, as a qualitative method, systematically examines and interprets textual content to identify patterns, themes and meanings (Liamputtong 2009). The study employs a lexicographic and linguistic analysis to examine the interplay between language, gender, and culture. It interrogates colloquial gendered entries in Mzansi Taal, their features of toxicity including the language that is employed to describe these terms which support the distorted representation of women. Twenty-five (25) female-gendered terms are selected and categorised according to themes that emerged from the collected terms in the dictionary. The analysis identifies terms categorically and analyses their literal and connotative meanings, as well as the social context in which they are used.

Accepted but not presented.

Lastly the study conducts an assessment of the level of toxicity in the language used to describe women, shedding light on the intersection of language, culture, and gender biases in the broader South African society. The study lays bare negative perceptions towards women that still exist today in various social groups based on language use, particularly among young black men and women in South African townships.

References

Andrews, T. 2012. What is Social Constructionism? *The Grounded Theory Review* 11: 39–46.

Bucholtz, M. 2003. Theories of discourse as theories of gender: Discourse analysis in language and gender studies. *The Handbook of Language and Gender*: 43–68.

Jung, H. 2019. The evolution of social constructivism in political science: *Past to present* 9(1): 1–10.

Liamputtong, P. 2009. Qualitative data analysis: Conceptual and practical considerations. *Health Promotion Journal of Australia* 20(2): 133–139.

*****

Implementation Progress of the Language Plan of Action for Africa in Namibia: A Lexicographer's Perspective

Eventhough NDLOVU (evennthough@yahoo.co.uk)
Department of Humanities and Arts, University of Namibia, Windhoek, Namibia;
Department of African Languages, University of the Free State, Bloemfontein, South Africa

The Language Plan of Action for Africa (henceforth, the Plan) was adopted to facilitate and coordinate the formulation and implementation of language policies in African Union member states. Among other things, the Plan enjoins all African Union member states to adopt clearly defined language policies which liberate Africans from undue reliance on ex-colonial languages. It also mandates member states to ensure that their national universities, research, and related institutions contribute to the practical promotion of African languages through the compilation of technical and general dictionaries in African languages. Against this background, this paper examines the implementation progress of the provisions relating to the compilation of technical and general dictionaries in Namibia's indigenous languages. Observations, document analysis and interviews with publishers and experts in Namibia's

indigenous languages show that not much has been done in Namibia insofar as the compilation of technical and general dictionaries in indigenous languages is concerned. Some indigenous languages continue to exist in pre-lexicographical environments. Oshiwambo has one dictionary, Mbenzi (2019). In Khoekhoegowab one bilingual dictionary exists, Haacke and Eiseb (2002) and a glossary, by Haacke and Eiseb (1999). Oshikwanyama has three bilingual dictionaries, Tobias and Turvey (1954), Turvey, Zimmermann and Taapopi (1977) and Oshikwanyama Curriculum Committee (2019). In Otjiherero only two dictionaries, Viljoen and Kamupingene (2010) and Nguaiko (2010) exist. Rukwangali has two dictionaries, Kloppers, Nakare, Isala and Bredell (1994) and Ausiku (2023). In Oshindonga two dictionaries exist, Viljoen, Amakali and Nauandi (1984) and Mbenzi and Iitheto (1998). Thimbukushu has one glossary, Legére and Mungande (2004). None of the indigenous languages have general purpose monolingual dictionaries. Only Oshiwambo has a technical dictionary, a dictionary of medical terms. Most languages have bilingual dictionaries. Only Oshiwambo has a children's dictionary. Khoekhoegowab has an online dictionary. The lack of an enabling language policy and excessive reliance on English accounts for the lack of technical and general dictionaries in Namibia's indigenous languages. A critical discourse analysis of past and present language policy documents reveals lack of an enabling language policy environment in Namibia to stimulate and sustain the growth and development of lexicography in indigenous languages. The policy of official monolingualism marginalised and marginalises indigenous languages. Article 3 of the Constitution employs a discriminatory language policy which favours English. It creates a hierarchical order of languages which reflects a hierarchy of power, politics, ideologies, and status differentials. This creates, perpetuates, and sustains systems of linguistic inequality, which do not create the impetus to compile dictionaries in indigenous languages since they are not used in high function domains. It was observed that a language policy which upholds monolingualism curtails the growth and development of lexicography in the excluded languages. The results of this study further confirm that lexicography and language policy have a symbiotic relationship (Bergenholtz & Gouws, 2006; Ndlovu & Moyo 2021). Namibia must, thus ensure that through appropriate legal provision and practical promotion, indigenous languages are used in high function domains alongside English to support the successful implementation of the provisions under study.

References

Ausiku, D.K. 2023. *Basic English - Rukwangali Dictionary*. Altona: FriesenPress

Bergenholtz, H. & Gouws, R.H. 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13–45.

Government of Namibia. 2015. *Constitution of the Republic of Namibia*. Windhoek: The Directorate of Print Media Affairs.

Haacke, W.H.G. & Eiseb, E. 1999. *Khoekhoegowab-English English-Khoekhoegowab Glossary/Mîdi Saogub*. Windhoek: Macmillan.

Haacke, W.H.G. & Eiseb, E. 2002. *A Khoekhoegwab Dictionary with an English-Khoekhoegowab Index*. Windhoek: Macmillan.

Kloppers, J.K., Nakare, D., Isala, L. M. & Bredell, A.W. 1994. *Bukenkango Rukwangali-English English Rukwangali dictionary*. Windhoek: Gamsberg Macmillan.

Legère, K. & Mungande, R. 2004. T*himbukushu-Thihingirisha English-Thimbukushu Maňandorandatha ghoThikuhonga Subject Glossaries*. Windhoek: Gamsberg Macmillan

Mbenzi, P.A. & Iitheto, O.N. 1998. *English-Oshindonga Dictionary for Primary Schools Embwiitya IyOshiingilisa noOshindonga IyOopilimaskola*. Windhoek: Out of Africa.

Mbenzi, P.A. 2019. *Oshiwambo - English Medical Dictionary & Phrasebook: Embwiitya lyOshiwambo nOshiingilisa lylitya nOmatumbulo gUunamati*. Windhoek: Namibia Book Market.

Ndlovu, E. & Moyo, W. 2021. The Nexus between Language Policy and Lexicography in Zimbabwe. *South African Journal of African Languages* 41 (3): 294–303.

Nguaiko, N.E. 2010. *The New Otjiherero Dictionary: English-Herero Otjiherero-Otjiingirisa*. Bloomington: AuthorHouse.
Oshikwanyama Curriculum Committee. 2019. O*shikwanyama - English English - Oshikwanyama Dictionary*. Windhoek: Namibia Publishing House.
Tobias, G.W.R. & Turvey, B.H.C. 1954. *English-Kwanyama Dictionary*. Johannesburg: Witwatersrand University Press.
Turvey, B.H.C., Zimmermann, W. & Taapopi, G. (eds.). 1977. *Kwanyama-English Dictionary*. Johannesburg: Witwatersrand University Press.
Viljoen, H.P., Amakali, P. & Namuandi, M.J. 1984. *Oshindonga/English English/Oshindonga Embwiitya Dictionary*. Windhoek: Gamsberg Macmillan.
Viljoen, J.J. & Kamupingene, T.K. 2010. *Otjiherero Woordeboek Dictionary Embo Romambo*. Windhoek: Gamsberg Macmillan.

*****

The Van Warmelo Dictionary Card Collection: The Use of Artificial Intelligence for Lexicographic Data Enrichment

Danie Prinsloo (danie.prinsloo@up.ac.za)
Department of African Languages, University of Pretoria, Pretoria, South Africa
Elsabé Taljard (elsabe.taljard@up.ac.za)
Department of African Languages, University of Pretoria, Pretoria, South Africa

This paper presents follow-up research on the Van Warmelo card collection, a set of approximately 70,000 Sepedi lexicographic index cards compiled by N. J. van Warmelo (1904–1989). Building on earlier work (Taljard, Prinsloo & Goosen 2024), the study investigates the use of artificial intelligence (AI) to enrich incomplete lexicographic data and to support the development of a structured electronic dictionary database. It is argued that AI can be systematically applied to enhance both the completeness and the lexicographic quality of legacy data resources. The study adopts a mixed methodology combining manual metadata analysis, database design, and AI-assisted data enrichment using large language models (LLMs).

In the initial phase of the project, each card was manually analysed to produce a detailed metadata description. The second phase extends beyond basic metadata description toward the development of a comprehensive database that preserves the original card data while enabling the extraction of lexicographic information, for example, for use in an online dictionary. Where specific data categories for a given lemma are incomplete, these gaps may be addressed by supplementing the data through the increased use of AI.

A key preliminary step involves designing an access strategy to the data. For instance, a user could input a Sepedi word and be directed to a PDF version of the corresponding lexicographic card, thereby enabling consultation of the original source material. The primary objective of this study, however, is to compile a comprehensive database comprising approximately eleven lexicographic fields derived from the cards and enriched with AI-generated information, from which a standard electronic dictionary entry could be extracted. At present, the approximately 70,000 cards are not readily accessible for information retrieval, as they are neither digitised nor alphabetically organised within the archival storage system.

In addition to filling gaps in the data, this paper demonstrates that AI can make a substantial contribution not only to the completion of empty data fields, but also to the enhancement of the overall lexicographic quality of the content. Furthermore, AI shows considerable potential in supporting the compilation and formulation of model dictionary entries for both print and electronic dictionaries. It will also be shown that a significant number of lemmas and their treatments on these cards have not been documented in existing

Sepedi dictionaries and are absent from Sepedi corpora. This highlights the value of the collection as a repository of previously undocumented lexical material and underscores the importance of its systematic digitisation and enrichment.

The paper first illustrates the capabilities of AI in card interpretation, even for a low-resource language such as Sepedi. This includes the ability to interpret and process lexicographic data, generate translation equivalents and definitions, and perform optical character recognition (OCR), even on handwritten cards. Particular attention is given to AI performance across a range of card qualities, from clearly typed entries to handwritten and, in some cases, barely legible cards produced on poor-quality paper. For example, the ability of AI systems such as ChatGPT to read and interpret handwritten cards in English, Sepedi and German is demonstrated.

The central focus of the paper, however, is on AI-driven enrichment of database fields. Consider, for instance, the data fields associated with the Van Warmelo card for the lemma *sehlare*. This card contains minimal lexicographic information, providing only the lemma and a single-word English translation equivalent, i.e. 'bush'. We illustrate how minimal card data can be expanded by AI into a full lexicographic entry by providing information for empty data categories. ChatGPT was prompted to analyse the card and generate a more comprehensive entry, providing information on part of speech, noun class, plural form, synonyms, cross-references to related Sepedi words, English translations, definitions, usage examples, and corresponding English translations of these examples.

ChatGPT also offered a formal lexicographic (dictionary-style) entry for *sehlare*, modelled on standard Sepedi dictionary conventions.

In conclusion, it is argued that the utilisation of AI holds significant potential for enhancing the data contained in the Van Warmelo card collection and for supporting the development of a comprehensive lexicographic database, ultimately contributing to the compilation of an electronic dictionary.

References

Taljard, E., Prinsloo, D.J. & Goosen, M. 2024. The Van Warmelo dictionary card collection – a lexicographic appraisal. Unpublished paper presented at the *28th International Conference of the African Association for Lexicography*, 1 – 4 July 2024, University of Pretoria, Pretoria, South Africa.

*****

Proverbs and Idioms as Knowledge Systems: How Should They be Treated Lexicographically?

Tshifhiwa RAMURUNZI (tshifhiwa.ramurunzi@ul.ac.za)
Department of Language Education, University of Limpopo, Sovenga, South Africa
Shumani Leonard TSHIKOTA (stshikota@univen.ac.za)
Department of Professional and Curriculum Studies, University of Venda, Thohoyandou, South Africa

Proverbs and idioms form essential components of African indigenous knowledge systems, embodying cultural values, moral philosophies, and communal worldviews in Tshivenḓa and other African languages (Kaschula 2017:45–47). Beyond functioning merely as stylistic expressions, they preserve historical memory, social norms, and epistemological frameworks that shape community identity and social conduct (Ngwenya 2020). However, lexicographic treatments frequently fragment these expressions through literal translations and reductive glosses that fail to communicate their cultural, pragmatic, and symbolic

meanings (Prinsloo 2017:21–22). This study investigates how proverbs, idioms, and oral genres may be represented lexicographically while preserving their status as indigenous knowledge systems. Drawing on ethnolinguistics, which explores the relationship between language and culture (Finlayson 2008), cultural hermeneutics, which emphasises socio-cultural interpretation (Mtuze 1992), and metalexicography, which focuses on dictionary structures and functions (Gouws & Prinsloo 2005), the study interrogates language use, meaning construction, cultural encoding, and intergenerational knowledge transmission.

The study adopts a qualitative research methodology. Data were collected through the analysis of 50 selected Tshivenḓa proverbs and idioms, including examples such as *Tsimba i ṱamba nga thuri* ('*tsimba* cleanses itself with *thuri*'), which metaphorically refers to a person who conceals wrongdoing by falsely accusing someone younger or less powerful. Data sources include text analysis, published literature, existing dictionaries such as the *Tshivenḓa Dictionary* (Tshivenḓa National Lexicography Unit 2022), and semi-structured interviews conducted with 20 native Tshivenḓa speakers, elders, and language experts from Limpopo communities. The analysis focuses on semantic depth, metaphorical meaning, contextual usage, and cultural functions of the expressions (Ngubane 2022:112–115). Current lexicographic practices are critiqued for prioritising reductive equivalence over cultural complexity. For example, entries such as *kholomo ya nḓila a i fhedzi hatsi* ('a grazing cow along the path cannot finish the grass'), which idiomatically conveys the idea that visitors should be offered food because they cannot exhaust available resources, are often reduced to simplified glosses without explanatory cultural notes (Tshivenḓa National Lexicography Unit 2023:78).

The findings reveal that conventional lexicographic approaches undermine the epistemic value of proverbs and idioms by omitting contextual information, usage notes, performance aspects, and cultural annotations. Such omissions marginalise indigenous knowledge systems and weaken the preservation of oral traditions (De Schryver 2023). Dictionaries rarely provide thematic cross-referencing between related proverbs, such as those expressing wisdom, caution, morality, or conflict resolution. Furthermore, most lexicographic resources exclude performative and oral dimensions such as pronunciation guides or audio recordings of elders reciting the expressions, despite the central role of oral transmission within African societies (Otlogetswe & Ramaeba 2024). Interviews conducted during the study further demonstrated that elders continue to use proverbs and idioms in mediation, dispute resolution, moral instruction, and community leadership contexts, confirming their juridical and educational functions within Tshivenḓa culture (Kaschula 2017:50).

Based on these findings, the paper proposes a culturally sensitive and decolonial model for the lexicographic treatment of proverbs and idioms. The proposed model recommends enriched dictionary entries that include:

1. Contextual examples presented in complete sentences;
2. Cultural notes explaining hermeneutic and symbolic meanings;
3. Narrative explanations linking expressions to worldview and lived experience;
4. Multimedia integration, including audio recordings of elder recitations;
5. Thematic cross-referencing connecting related proverbs and idioms; and
6. Participatory involvement of community custodians through collaborative workshops.

This participatory framework aligns with decolonial approaches to language documentation and knowledge preservation (Ngwenya 2020:200–202). It seeks not only to improve dictionary usability, but also to support the revitalisation and safeguarding of indigenous African knowledge systems in the context of globalisation, urbanisation, and

language shift.

References

De Schryver, G.-M. 2023. *Lexicography and Indigenous Knowledge Systems*. Oxford: Oxford University Press.
Finlayson, R. 2008. Language and Culture in African Contexts. *Southern African Linguistics and Applied Language Studies* 26(2): 145–156.
Gouws, R.H. & Prinsloo, D.J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN Press.
Kaschula, R.H. 2017. African Oral Traditions and Knowledge Systems. *African Studies Review* 60(3): 45–52.
Mtuze, P.T. 1992. *Cultural Hermeneutics in African Languages*. Cape Town: Maskew Miller Longman.
Ngubane, H. 2022. Semantic and Cultural Layers in African Proverbs. *Journal of African Languages* 14(1): 112–115.
Ngwenya, T. 2020. *Decolonising African Lexicography*. Johannesburg: Wits University Press.
Otlogetswe, T. & Ramaeba, M. 2024. Digital Preservation of African Oral Literature. *Lexikos* 34: 88–103.
Prinsloo, D.J. 2017. Challenges in Translating African Idioms. *Lexikos* 27: 21–34.
Tshivenḓa National Lexicography Unit. 2022. *Tshivenḓa Dictionary*. Thohoyandou: TNLU.
Tshivenḓa National Lexicography Unit. 2023. *Advanced Tshivenḓa Lexicographic Studies*. Thoh̓oyandou: TNLU.

*****

A Corpus-Based Approach to the Lemmatization of Anthroponyms in a Xitsonga Monolingual Dictionary

Ndzalama VUMA (vuman@unisa.ac.za)

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

Madala Crous HLUNGWANI (Crous.Hlungwane@univen.ac.za)

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

This study investigates the treatment of anthroponyms in the Xitsonga monolingual dictionary *Tihlungu ta Rixaka Dikixinari ya Ririmi ra Xitsonga* (Marhanele & Bila 2016). The primary objective is to examine how different types of personal names are lemmatised, classified, and defined, and to evaluate the adequacy of their linguistic representation. Specifically, the study identifies the types of anthroponyms included in the dictionary, analyses how these names are treated in terms of phonetics, morphology, and semantics, and proposes a corpus-based approach to improve their lexicographic treatment.

The study adopts a qualitative research design within an interpretivist paradigm, employing corpus-based methods to systematically analyse selected anthroponyms. Purposive sampling is used to extract relevant entries that align with the study's objectives, ensuring a focused and contextually grounded analysis (Memon et al. 2025:2). The theoretical framework is informed by the Function Theory of Lexicography, which emphasises the user-oriented distribution of lexicographic information, and linguistic theory, particularly structuralist and generative perspectives, which account for the internal structure and formation of lexical items (Tarp 2008:40-41; Chomsky 1965:31; Alexiadou & Lohndal 2016:1). These frameworks guide the analysis of how anthroponyms are represented in terms of phonetic transcription, grammatical classification, semantic interpretation, and morphological processes such as derivation and inflection.

The findings reveal that the dictionary incorporates a range of anthroponyms, which can be broadly categorised into male, female, and gender-neutral names. Furthermore, these names are divided into two major types: native (Xitsonga-origin) anthroponyms and

Accepted but not presented.

borrowed anthroponyms, particularly from neighbouring languages such as isiZulu. Within these categories, additional distinctions are observed based on morphological structure, including derived forms (e.g., names originating from verbs or adjectives) and inflected variants. However, the treatment of these anthroponyms is found to be inconsistent and, in some cases, linguistically inadequate.

In terms of treatment, the study demonstrates that many anthroponyms are presented with limited or incomplete linguistic information. For instance, phonetic representations are largely absent, which limits accurate pronunciation and standardisation. Grammatical information such as noun class and part of speech is inconsistently applied, while semantic definitions are often oversimplified or inaccurate. In several cases, definitions merely indicate the gender of the name bearer without providing the underlying meaning or etymology of the name. For example, a name such as *Hlamalani*, which is morphologically derived from a verb meaning *hlamala* 'to be amazed,' is inadequately defined without capturing its semantic motivation. Similarly, borrowed anthroponyms such as *Soshangana* are treated descriptively rather than lexically, with emphasis placed on historical or referential information rather than their linguistic structure, meaning, or phonological adaptation into Xitsonga.

The study further finds that the dictionary does not adequately account for the morphological processes underlying anthroponyms, particularly derivation and inflection. This omission obscures the relationship between base forms and derived names, thereby limiting the user's understanding of how personal names are formed and used within the language. Additionally, the lack of systematic treatment across entries results in inconsistencies that reduce the overall usability and reliability of the dictionary as a linguistic resource.

Accepted but not presented.

In response to these challenges, the study proposes a corpus-based approach to the lemmatisation of anthroponyms in Xitsonga monolingual dictionaries. This approach advocates for the use of authentic language data to inform the selection, classification, and description of anthroponyms. It emphasises the inclusion of comprehensive linguistic information, including phonetic transcription, morphological analysis, and contextually grounded semantic definitions. Furthermore, the approach calls for the standardisation of lexicographic conventions, ensuring consistency in the treatment of both native and borrowed names. By integrating corpus evidence with established lexicographic principles, the study aims to enhance the accuracy, depth, and usability of anthroponym entries.

Therefore, this study contributes to the advancement of Xitsonga lexicography by highlighting the need for a more systematic and linguistically informed treatment of anthroponyms. It provides practical recommendations for the development of improved monolingual dictionaries and lays the foundation for future research on the intersection of onomastics, corpus linguistics, and African language lexicography.

References

Alexiadou, A., & Lohndal, T. 2016. Editorial: The grammar of Multilingualism. *Journal of Frontiers in Psychology* 7: Article 1397.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.

Marhanele, M.M. & Bila, V. 2016. *Tihlungu ta Rixaka Dikixinari ya Ririmi ra Xitsonga*. Polokwane: Timbila Poetry Project Publishers.

Tarp, S. 2008. Lexicography in the borderland between knowledge and in general lexicographical theory with particular focus on learner's lexicography. *Lexicographica Series Maior* 134. Tubingen: Niemeyer.

Memon, M.A., Thurasamy, R., Ting, H. & Cheah, J.H. 2025. Purposive sampling: a review and guidelines for quantitative research. *Journal of Applied Structural Equation Modeling* 9(1): 1–23.