SADiLaR

South African Centre for Digital Language Resources

NEWSLETTER

ENGLISH: NOVEMBER 2020

Prof Langa Khumalo appointed to lead digital language research infrastructure

Prof Khumalo, who is the former Director of Language Planning and Development at the University of KwaZulu-Natal, took the reins at SADiLaR on 1 August, leading the Centre – the first of its kind in Africa – to achieve the constitutional imperative to recognise all the South African languages as key resources in innovation and science.

READ ABOUT

In 2016, Prof Khumalo completed an intensive Advanced Leadership Programme for Higher Education, which made him acutely aware of the vast challenges higher education is confronted with in generating knowledge that is innovative and responsive to our ever-evolving human needs.
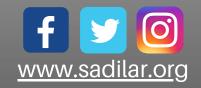
During the early stages of his career, Prof Khumalo was introduced to working with language data. **"Trained by many luminaries in corpus linguistics, including none other than the late Professor John Sinclair at the Tuscany Word Centre in Italy, I was one of the first editors to compile and publish the first corpus-aided monolingual dictionary of isiNdebele,"** he said.

Prof Khumalo has been involved in various Human Language Technology (HLT) projects, including the development of the isiZulu National Corpus, an isiZulu Oral Corpus, and the English-isiZulu Parallel Corpus. These corpora are carefully designed to enable HLT development.
His involvement in various exciting projects such as the development of disciplinary terminology for isiZulu, which culminated in the publication of glossaries such as the illustrated glossary of architecture terms and the glossary of law terms, has contributed to raising awareness of the profile and role of African languages in the academy.

"SADiLaR became an attractive and inevitable destination for me because of what it is doing, what it seeks to do in terms of developing language resources, and its national character and reach," Prof Khumalo stated.

Prof Khumalo has been involved with SADiLaR as a member of the Scientific Advisory Committee (SAC) since its inception in 2017.

Given the close association with SADiLaR, Prof Khumalo has insight into all of the projects that have been rolled out internally as well as projects driven by the various nodes.

"The Centre has a young and dynamic team and I hope that my experience and skillset will be a useful addition to an already amazing array of skills,"

Prof Khumalo noted on working with the SADiLaR team. At the core of it all, Prof Khumalo wants the Centre to deliver research, (technology) solutions, and (digital) resources that are impactful and have transformational effects in the communities where these languages are predominantly spoken and/or taught. He also aims to embrace the quality research, productivity culture, and the links that the Centre has established in its short period of existence.

"Language is at the centre of our human existence. It carries and conveys the identity of humanity. It records our human achievements and spurs on our innovation and development. It is notable that our Constitution recognises eleven official languages. It is a constitutional imperative to develop all of them in order to achieve 'parity of esteem' between them," said Prof Khumalo.

Since its inception, SADiLaR has made strong connections with similar research infrastructures internationally such as the European Common Language Resources and Technology Infrastructure (CLARIN), the European Language Resources Association (ELRA), and the Linguistic Data Consortium (LDC) of the USA. The Centre has been guided by experienced leaders in the field of Human LanguageTechnology (HLT), Linguistics, and Digital Humanities.

Lingua Africa: Health resources in the official languages of South Africa

Wouldn't it be great if healthcare providers could treat patients in their own languages? Could this lead to more effective ways of making a diagnosis? Wouldn't you be glad to have your health problems addressed in your mother tongue?

Do you have any suggestions for our website?
CLICK HERE

www.sadilar.org

One of the projects that SADiLaR is currently working on is "Health resources in the official languages of South Africa". The project runs under the Department of Health and Rehabilitation Sciences/Child Language Africa in the Faculty of Health Sciences at the University of Cape Town. This project sets out to document health resources available in the official languages of South Africa. Health resources are materials such as screening questionnaires and diagnostic assessments designed to be used in a healthcare setting with patients and families by a health professional (e.g. doctor, nurse, and physiotherapist). The purpose is to establish a database of these resources which healthcare workers can use to access available materials.

The project aims to support efforts to promote and develop local language resources and contribute to more effective healthcare services for all people in the country. In the long run the database will enable health practitioners to search for clinical materials available in local languages – and develop more resources where gaps exist.

The project leaders are Michelle Pascoe (Director and Associate Professor of Child Language Africa / Communication Sciences and Disorders), Olebeng Mahura (Lecturer in Communication Sciences and Disorders), and Jessica Dean (Lecturer in Communication Sciences and Disorders). During the course of the project they supervised 14 speech-language therapy final year/honours students.

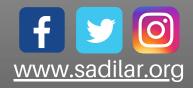To date the research team have conducted two literature reviews

answering theoretical questions about available resources and how they were developed. **"In future, we plan to develop scientific expertise and knowledge about cross-cultural adaptation of assessments",** says project leader Michelle Pascoe. The database currently contains over 400 resources with all official South African languages represented. **"We will continue to add resources to the database and are looking at ways of automating this process and inviting researchers to contribute directly to it",** stated Pascoe.

This part of the project focused on resources in the official languages of South Africa, but our larger project includes developing a database of health resources for Africa.

PUBLICATIONS:

Open access publication in *Health SA Gesondheid: Journal of Interdisciplinary Health Sciences;* and forthcoming book chapter in the Springer Handbook of Communication Disabilities and Language Development in Sub-Saharan Africa.
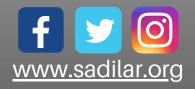
What are your expectations of SADiLaR as a research infrastructure?
CLICK HERE

3

www.sadilar.org

Bridging communication barriers in COVID-19 triage and screening: A mobile speech app

The Council for Scientific and Industrial Research (CSIR) and Aweza have collaborated to develop a unique mobile application to bridge communication barriers between healthcare providers and patients. Aweza is an internationally award-winning tech-based initiative that strives to inspire and empower South Africans to overcome language barriers across all sectors of society.

The mobile application, AwezaMed COVID-19, features localised speech technology such as speech recognition, machine translation, and text-to-speech developed by the CSIR and it works on any Android smartphone. It enables healthcare providers to access a phrase in English, translate it into any official South African language, and play the phrase in the selected language.

Originally developed with content pertaining to maternal healthcare and obstetrics, the application has been enhanced with COVID-19-related content and is available for download from the Play Store. There is no cost to users.

"By bridging the communication barrier, the trust relationship between the healthcare provider and patient can be improved. In addition, the patient's experience and the healthcare provider-patient confidentiality can be improved, and lives can possibly be saved," says Dr Karen Calteaux, CSIR Digital Audio-Visual Technologies Research Group Leader.

"The content of the application was developed in collaboration with health experts and is aimed at helping healthcare workers to communicate with patients at healthcare facilities, especially during COVID-19-related screening and triage."

Dr Calteaux explains that **"AwezaMed emanated from a project funded by the Department of Sport, Arts and Culture, that strives to bridge language barriers between healthcare practitioners and patients in clinics. A decision was taken to develop a version to address language barriers experienced by medical professionals working with COVID-19 patients."**

For more information on the CSIR, visit
www.csir.co.za

For more information on the Department of Sport, Arts and Culture, visit
www.dac.gov.za
or email Tshikani Mabasa at
Tshikanim@dsac.gov.za

www.sadilar.org

ABOUT THE CSIR:

The CSIR is one of the leading scientific and technology research, development, and implementation organisations in Africa. Constituted by an Act of Parliament in 1945 as a science council, the CSIR undertakes directed and multidisciplinary research, technological innovation, as well as industrial and scientific development to improve the quality of life of all South Africans.

CSIR
Touching lives through innovation

sport, arts & culture
Department:
Sport, Arts and Culture
REPUBLIC OF SOUTH AFRICA

ABOUT THE DEPARTMENT OF SPORT, ARTS AND CULTURE:

The mission of the Department of Sport, Arts and Culture (DSAC) is to transform the delivery of sport, arts and culture by ensuring equitable access, development and excellence at all levels of participation, and harnessing the socio-economic contributions that can create a better life for all South Africans. Through the Human Language Technologies directorate, DSAC has supported numerous projects that develop language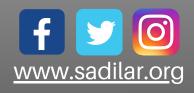 technologies to give South Africans access to information and government services in their home language, irrespective of their level of literacy. One such project is Mburisano: *A speech-to-speech system to bridge language barriers*. The mobile app AwezaMed COVID-19 is one of the deliverables of the Mburisano project.

What types of training workshops are you interested in?
CLICK HERE

www.sadilar.org

The Carpentries

The Carpentries project comprises the Software Carpentry, Data Carpentry, and Library Carpentry communities of instructors, trainers, maintainers, helpers, and supporters who share a mission to teach foundational computational and data science skills to researchers. The Carpentries have a vision to become the leading inclusive community that teaches data and coding skills, with a mission to build global capacity in essential data and computational skills for conducting efficient, open, and reproducible research.

Carpentry workshops either happen face-to-face or virtually, but the year 2020 has seen a series of virtual workshops. Virtual workshops have become the way of the future since face-to-face engagements have been put on hold because of the unprecedented events caused by the coronavirus pandemic. Even though these virtual workshops might be a new way of doing things for most, they do come with numerous benefits. These benefits include the convenience of participating in a professional workshop without leaving one's home. This experience is budget-friendly: all one needs is a working laptop and enough data bundles to escape travelling, accommodation, and catering costs. Accessibility is another benefit because one can participate in as many workshops as desired since most of them are free. This can lead to gaining even more knowledge than would have been possible under normal circumstances.

SADILAR'S INVOLVEMENT IN CARPENTRIES:

To date, eight carpentry workshops have taken place with 113 participants between the months of June and September 2020.

SADiLaR researchers have been involved in a few of them this year, either as supporting instructors or as helpers. This has been a great experience for the researchers because the concept of such a project is quite new to them, and as budding researchers in the field of Digital Humanities, they find that the experience plays a significant role in exposing them to further computational ways of doing their research.

During the week of 24 to 28 August, SADiLaR hosted its very first Library Carpentry workshop facilitated and run by some of the SADiLaR researchers, with the generous assistance of Dr Angelique van Rensburg, the Regional Consultant for Southern Africa at The Carpentries. During this workshop, those researchers

who had already received their certification to be Carpentry instructors led attendees through the entire workshop step by step. The programme was as follows:

Day 1:
Introduction to working with Data (Regular expressions) by Rooweither Mabuya, the isiZulu researcher, accompanied by Dimakatso Mathe, the Sesotho sa Leboa researcher.

Day 2:
Open Refine by Mmasibidi Setaka, the Sesotho researcher.

Day 3:
Continuation of Open Refine by Juan Steyn, Project Manager at SADiLaR.

Day 4:
The Unix Shell by Andiswa Bukula, the isiXhosa researcher.

Day 5:
Continuation of The Unix Shell by Juan Steyn.

This was indeed a great experience for the researchers, who are hoping that you will join the next Carpentry workshop.

THE
CARPENTRIES

- by Andiswa Bukula

For more about
The Carpentries,
please go to:
https://carpentries.org/

www.sadilar.org

Blogs and community engagement on SADiLaR's website

One way in which SADiLaR has actively tried to engage with the public, is through blogs mainly written by the language researchers. In these online blogs, a wide variety of topics is discussed. Researchers have the option to write the blogs in their own language, with a short summary in English for those who do not have command of the relevant language.

Topics that have been discussed in the blogs so far include impressions of workshops or events that were attended, interjections in Xitsonga, the history of the limited written works in isiNdebele, and historical literary works in isiZulu, to name but a few.

It is interesting to take a cursory look at some of the numbers relating to the blogs. Here is a tabled summary relating to the blogs as at 30 September 2020:

Number of blogs that have appeared:	35	
Blogs written in English:		21
Blogs written in Xitsonga:		3
Blogs written in Afrikaans:		3
Blogs written in Siswati:		2
Blogs written in isiNdebele:		2
Blogs written in isiXhosa:		1
Blogs written in isiZulu:		1
Blogs written in Sesotha sa Leboa:		1
Blogs written in Setswana:		1
Highest reads:	4664	
Average reads:	1008	

The blog with the highest number of reads is titled "Tinanatelo temaswati" and was written by SADiLaR's Siswati researcher, Muzi Matfunjwa. This blog is written in Siswati and is about clan praises and was published on the website on 9 June 2020.

The only other blog that has over 4000 reads so far, is an isiXhosa blog written by SADiLaR researcher Andiswa Bukula, titled "Ukubaluleka kolwimi nenkcubeko eluntwini". In this blog, aspects of culture, language, and identity as interrelated are discussed.

Which blogs would you like to read about?
CLICK HERE

www.sadilar.org

Besides blogs, SADiLaR also hosted a year of celebrations for all the official languages in 2019. Events that were organised involved school visits, theatre workshops, as well as talks by prominent figures in the respective language communities. Great care was also taken to host the events in different parts of South Africa to make them more accessible to different communities. A few pieces were also written about these events and celebrations and published on the SADiLaR website. Please visit the dedicated page on our website for an overview of the different events that took place. The use of blogs as an avenue for community engagement has definitely increased since the first blog was published on 26 November 2019. With an average of 1008 reads per blog, it is clear that language communities are curious about the different topics that members of the SADiLaR team choose to write about.

Currently there is still ample opportunity to discuss many topics, to collaborate with others and to write in one of our official languages. Here's to a fruitful future for blog-writing at SADiLaR!

- Benito Trollip

Release of the online Dictionary of South African English (DSAE) - MOBILE EDITION

This release presents a data-rich historical dictionary model on smartphones and tablets (iOS and Android supported, free access at http://dsae.co.za). It marks the completion of a major digitisation phase from early pilot online editions to a fully-fledged web application thoroughly adapted for desktop and mobile devices. The same advanced feature set is available on both platforms.

The dictionary was first published in print as A Dictionary of South African English on Historical Principles (Oxford University Press, 1996) – a 25-year project documenting the development of regional English usage over three centuries. It contains 4600 main entries with 45 000 It contains 4600 main entries with 45 000 bibliographically-referenced quotations covering 17 500 word forms and their histories, including first recorded use.

Following extensive digital enhancements, dictionary users can now manipulate this unrivalled record of the language variety

as a structured database. Words can be grouped in dictionary "slices", e.g. by date of entry into the language using an interactive visual slider, or filtered by language of origin, subject, or usage category (with over 180 subcategories in all). Individual word histories – often spanning centuries – can also be navigated visually by word sense or quotation evidence, revealing developmental patterns previously submerged in the written text.

DSAE Dictionary of South African English

KEY FEATURES:

- Fully responsive design, rescales to preserve advanced Desktop features on Android and iOS (most modern browsers supported)

- Graphical presentation and navigation of word histories and source quotation data (data aggregation devices reveal new insights and relations)

- Navigate complex entries structurally, not only by sequential reading of text

- Powerful filtering options presented in a simple way: no unwieldy 'advanced search' forms to complete; stack or unstack search filters easily without cognitive overload

- Refine search and filter results by date span with a visual slider; sort entries (or the entire unfiltered dictionary) by year of first recorded use.

To access the dictionary, go to
https://dsae.co.za

Web:
www.ru.ac.za/dsae/

Email:
dsae@ru.ac.za

The online DSAE is a project of the Dictionary Unit for South African English in Makhanda (Grahamstown), South Africa.

Click here to subscribe to our newsletter
OR
Click here to get in contact with us

www.sadilar.org