# SADiLaR

## South African Centre for Digital Language Resources

# NEWSLETTER

# ENGLISH: November 2021

## READ ABOUT :

# SOUTH AFRICAN UNIVERSITIES – KEY ROLE PLAYERS IN SADILAR'S STRATEGIC PLAN

### - Prof Langa Khumalo

The vision of SADiLaR is to stimulate and enable digital research and development in South Africa's 11 official languages, to cultivate and grow Digital Humanities scholarships in South Africa and the continent, while creating strategic synergies with local and international partners with similar or complementary infrastructure.

The universities of South Africa are therefore strategic centres that have the capacity to enhance the vision and mission of SADiLaR. They are in this sense characterised as key strategic partners. The executive director of SADiLaR, Prof Langa Khumalo, has initiated university visits in order to roll out a key part of the strategic engagement focus area of the Centre, and as also articulated in the SADiLaR strategic plan.

As a result of the Covid-19 pandemic, all the visits have hitherto taken place virtually through online platforms such as Zoom or Microsoft teams. At these meetings, Prof Khumalo articulated the mission and vision of SADiLaR. The aim of the visits has been, inter alia, to initiate collaboration between these universities and SADiLaR based on the respective institutions' identified areas of need or niche areas that these universities want to grow. The outcomes have been varied and interesting with a potential for success. The engagements with Stellenbosch University, University of the Western Cape, the Cape Peninsula University of Technology, and University of the Witwatersrand, University of

Pretoria and University of Venda have led to further meetings in order to concretise ties or projects of mutual interest and benefit. Visits with the University of KwaZulu-Natal have culminated in the introduction of a digital humanities module in the School of Arts. This module will be conducted entirely online, which will be tested in the second semester of the 2021 academic year.

SADiLaR is engaged with these institutions in order to finalise a memorandum of understandings that would frame ongoing and future engagements. With the advent of the New Language Policy Framework for Higher Education, Prof Khumalo, working with Universities South Africa (USAf), is leading a three-part colloquium series with the Vice-Chancellors and Deputy Vice-Chancellors of all 26 public institutions and other key stakeholders in order to position SADiLaR as a key player in providing technical and strategic resource support in the successful implementation of the New Language Policy Framework for Higher Education. The university visits have thus far been a huge success.

**Click here to get in contact with us:**

# COVID-19 MULTILINGUAL TERMINOLOGY PROJECT

## - Nomsa Skosana and Respect Mlambo

Earlier this year, the City of Tshwane approached the South African Centre for Digital Language Resources (SADiLaR) to assist in developing a Covid-19 multilingual terminology list for all the South African languages.

SADiLaR was approached because of the Centre's strong enabling function with a clear focus on the development of all South African official languages, and its support of research and development in the domains of language technologies and language-related studies in the humanities and social sciences.

Terminology lists were collected by both the City of Tshwane and SADiLaR. Definitions and explanations of the terms were provided solely in English. Thereafter, the City of Tshwane and SADiLaR provided the term equivalents in all of South Africa's official languages. The City of Tshwane provided these equivalents in six languages, namely Sepedi, Setswana, Afrikaans, isiZulu, Tshivenda and Xitsonga while SADiLaR provided the equivalents in the remaining four languages of isiNdebele, Siswati, isiXhosa and Sesotho. All the terms in all the languages were then edited by both the City of Tshwane and SADiLaR. The terminology verification was done by terminology verification committees under the National Language Bodies (NLB) and the process was financed by SADiLaR while the City of Tshwane coordinated the process. The verification process involved not only NLB members, but also language practitioners and researchers for all the languages, as well as field experts.

The project team consisted of the following members for each language:

- **Afrikaans**: 4 NLB members, 5 language practitioners and researchers and one expert from the Department of Health.
- **IsiXhosa**: 13 NLB members, 1 language practitioner and researcher.
- **Siswati**: 9 NLB members, 2 language practitioners and researchers and one field expert.
- **IsiNdebele**: 11 NLB members, two language practitioners and researchers and one health expert from the Department of Education.
- **Setswana**: 13 NLB members and 3 language practitioners.
- **Sesotho**: 8 NLB members, 1 language practitioner and/or researcher and one field expert from the Department of Health.
- **IsiZulu**: 12 NLB members, 3 language practitioners and researchers and one field expert.
- **Xitsonga**: 12 members, 3 language practitioners and researchers and one field expert from the Department of Health.
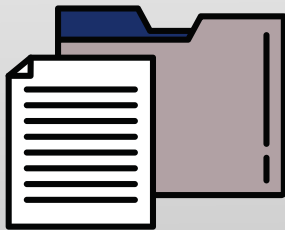
**SADiLaR's repository**

- **Sepedi**: 11 NLB members and 3 language practitioners.
- **Tshivenda**: 10 NLB members, 1 language practitioner and/or researcher and one field expert.
- **English**: 1 researcher and one language practitioner.

All 136 NLB members, language practitioners, researchers, other members of SADiLaR, and those from the City of Tshwane contributed immensely to the project's success.

The authentication and the standardisation of the multilingual terminology list was then undertaken by the NLB under PanSALB as a funder of this process.

The City of Tshwane completed the last stages of the project, which included concatenating the terms, printing the booklets, as well as distribution of the booklets in soft and hard copy.

The City of Tshwane officially published the COVID-19 multilingual terminology list as an open source in July 2021 through their website www.tshwane.gov.za, and printed copies were distributed to various departments. The terminology list is also available on SADiLaR's repository at https://hdl.handle.net/20.500.12185/544 .

# DIGITAL DICTIONARY RESOURCES FOR N|UU

### - **Kerry Jones, Bonny Sands, Menno van Zaanen**

N|uu is a severely endangered !Ui-Taa language which was historically spoken by the ǂKhomani people of South Africa. The ǂKhomani were traditionally hunter-gatherers from the southern Kalahari region. Today, they lead a more sedentary lifestyle.

Unfortunately, as far as linguists are aware, there is only one fluent mother tongue speaker of N|uu alive today. Her name is ǂXuu ǀEeki Katrina Esau. In recent years, "Ouma", as she is affectionately known, has lost close family and friends with whom she shared the N|uu language.

In an effort to consolidate historical and linguistic resources produced in N|uu, African Tongue together with the N|uu Language Authority and SADiLaR are collaborating on a project called

"Digital Dictionary Resources for N|uu". The primary focus of this project is to combine historical and contemporary research on the N|uu language into a dictionary that will be published in three formats, namely a limited edition physical dictionary, a digital dictionary that can be accessed via a dictionary portal and a dictionary app. These resources will be freely available upon completion in early 2023.

**Project collaborators:**

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

NORTHERN
ARIZONA
UNIVERSITY

sport, arts & culture

Department:
Sport, Arts and Culture
REPUBLIC OF SOUTH AFRICA

**Contact us:**

## BRIEF BACKGROUND TO THE EARLY DOCUMENTATION OF N|UU:

In the 1970s, the N|uu language was thought to be extinct in southern Africa. However, in the 1990s during a cultural heritage audit process implemented through UNESCO, 26 speakers of N|uu identified themselves. Linguists such as Nigel Crawhall, Tony Traill and Levi Namaseb verified that these speakers were in fact speaking N|uu and not Nama as previously thought by local residents. Crawhall, Traill and Namaseb then fervently began efforts to document the language as a matter of national importance. A detailed account of this process in collaboration with local community members can be found at the University of Cape Town, Special Collections under the name "‡Khomani San | Hugh Brody Collection (BVF41)".

Since then the N|uu language has been of local and international interest by many linguists curious about endangered southern African languages. The most recent data collection and verification was done in May 2021 in Upington by Kerry Jones, Ouma Katrina Esau (the last N|uu speaker), Claudia Snyman (grand-daughter of Ouma Katrina Esau), David van Wyk (Secretary of the N|uu Language Authority), Sussie Bock (Kalahari member of the N|uu Language Authority and descendant of N|uu speakers) and Bonny Sands (who assisted online from the USA).

## COLLATED DATA FOR RESEARCH AND DEVELOPMENT:

Linguistic data collected by Bonny Sands, Chris Collins, Amanda Miller, Johanna Brugman, Levi Namaseb and Alena Witzlack-Makarevich has generously been shared for the purposes of this project. To date, the data set is comprehensive and consists of over 1560 lexical items. These were recorded with mother tongue speakers who have since passed away: Anna Kassie, Hanna Koper, Griet Seekoei, |Una Rooi, Kheis Brou, Hannie Koerant, Andries Olyn, Vytjie |Abaka Koper and Simon Sauls. Ouma Katrina Esau was also included in these historical recordings and she continues to assist to this day with further verification of translations and clear audio recordings.

All these documented lexical entries have been translated into Afrikaans, English and Khoekhoegowab (the Namibian variety as recorded by Prof Levi Namaseb). Additionally, there are IPA transcriptions (describing the pronunciation of the words) of the N|uu entries recorded. The data set also offers 4860 recordings of lexical items in context as well as 20,000 additional recordings (e.g., diphthongs, targeted lists, and primers). Despite scant documentation of N|uu in its later years, these data sets offer a comprehensive description and documentation of the language. All data is now in the process of being cleaned and digitally archived in

**Project collaborators:**

RHODES UNIVERSITY
*Where leaders learn*

AFRICANTONGUE
Professional linguistic consultancy

**Contact us:**

order to be made available on the SADiLaR repository according to FAIR principles to ensure that it is findable, accessible, interoperable and reusable.

## PROJECT OUTPUTS:

By 2023 the following outputs are planned for delivery to the South African public and the world at large:

### 1.PHYSICAL DICTIONARY
Approximately 5000 limited edition hard copy dictionaries containing entries in all four languages described above will be made available to the ǂKhomani descendants, selected libraries, academics and language activists. The dictionary will include the biographies of the N|uu speakers who participated in the project as well as the linguists who worked on these data sets.

### 2. MOBILE DICTIONARY APP
The mobile dictionary app version of the dictionary will be available for both iOS and Android. The app will provide access to the entries in the different languages and provides an easy to use search interface. Current developments indicate that the app will feature two modes: an offline and online version with slightly different functionality. A selection of the audio recordings will also be made available through the app.

### 3. WEB DICTIONARY
The web version of the dictionary will have more features than the app version as it will be less restricted by space and mobile constraints. It will be accessible via web browsers and will be made available as a dictionary portal on SADiLaR's website. It will have similar functionality to the mobile app with additional audio recording also made available. For both the mobile dictionary app and the web dictionary, specific functionality needs to be developed, for instance, to deal with the input of symbols representing clicks that occur frequently in the N|uu language.

## THE NEXT STEPS IN THE PROJECT:

Currently, the data set is in the process of being cleaned and the dictionary entries need to be verified with the remaining N|uu speaker. Then the orthography needs to be consolidated from the work of previous researchers/linguists to ensure consistency throughout the entire data set. Additional audio recordings will be made where needed and the meta data of the data set also needs to be checked and corrected where needed. As soon as the data set has been cleaned and checked for consistency and completeness, it will be made available on the SADiLaR repository. This allows interested researchers to freely access the data. This data set will then be converted into the different formats that are needed to create different modalities (physical dictionary, dictionary app and dictionary web portal).

## Project collaborators:

## Contact us:

The project faces a range of challenges. For instance, only one speaker of N|uu can provide reliable information on the entries in the data set. The current pandemic makes traveling difficult as the health of participants needs to be taken into account. On the technical side, the diacritics/symbols used to indicate the clicks and sounds during pronunciation provide interesting challenges. For instance, the input of these characters into the digital versions of the dictionary is difficult as these symbols are not found on a keyboard. Providing suggestions (e.g., in case of spelling mistakes) is also challenging. All of these challenges are currently being addressed using a range of creative solutions. These challenges and solutions indicate that the development of these types of resources is non-trivial and ad hoc solutions need to be identified by the research team.

**For further information about his project please go to the following link for a free presentation:** https://figshare.com/articles/presentation/Development_of_a_digital_dictionary_for_N_uu/15104322



Caption: Early linguist fieldwork in the 1990s documenting the N|uu language. From left to right, Dawid Kruiper (previous traditional leader of the ‡Khomani), Andries Olyn (speaker of N|uu) and Nigel Crawhall (linguist and human rights activist).

**For further information about his project please go to the following link for a free presentation:**

Caption: Dawid Kruiper working with Namibia linguist, Levi Namaseb, on traditional place names in the Kalahari region.



**Contact us:**

Caption: ǂKhomani elders outside what is known today as the Kgalagadi Transfrontier Park. They were dispossessed of this land to make way for a national park. They have since successfully won their land back in and around the Kgalagadi Transfrontier Park in the Northern Cape and also have access to the park.

Caption: The Swartkop sisters are speakers of the N|uu language, who came forward in the 1990s and later became activists for their language, culture and heritage.



Caption: Linguist Bonny Sands working with N|uu speaker, Simon Sauls, in a makeshift Kalahari office, 2006.

**Contact us:**
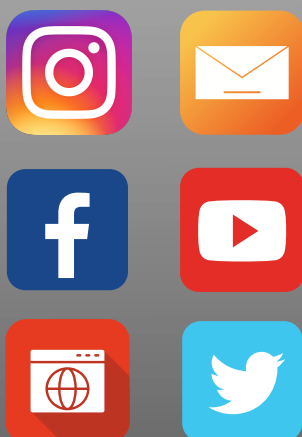
**9 August 1956**

WWW

**Contact us:**



Caption: Data verification in Upington with Kerry Jones, Ouma Katrina Esau, Sussie Bock, David van Wyk and Claudia Snyman, May 2021.

# #WOMENSMONTH2021: GROWING A COMMUNITY OF TECHNOLOGY EMPOWERED WOMEN IN HUMANITIES AND SOCIAL SCIENCES

## - Anelda van der Walt

"South Africa commemorates Women's Month in August as a tribute to the more than 20 000 women who marched to the Union Buildings on 9 August 1956 in protest against the extension of Pass Laws to women. The Government of South Africa declared August Women's Month and 9 August is celebrated annually as Women's Day." (https://www.gov.za/womens-month)

This year, ESCALATOR celebrated Women's Month by launching the EMPOWER track of the Digital Champions Initiative on 5 August. To kick off, EMPOWER offered a series of events under the title "Finding your place in an increasingly digital world as a woman in Humanities or Social Sciences". The events were open to women affiliated with South African research and academic institutes at all career stages and with any level of digital or technical experience.

**Digital Champions
Initiative Youtube
channel**

**Contact us:**

The series aimed to show women in Humanities and Social Sciences why they need to become part of the tech conversation, the value of digital and computational research, how they can learn and grow, and where they can get support. We specifically showcased communities that are committed to growing inclusivity and diversity in technology, such as WanaData, RLadies, The Carpentries, and Youthmappers. Our speakers also included technology practitioners within diverse research fields such as education, anthropology, international relations studies, and psychology.

According to a report from 2018 published by the Organisation for Economic Co-operation and Development (OECD) titled "BRIDGING THE DIGITAL GENDER DIVIDE INCLUDE, UPSKILL, INNOVATE" (2018), women are still far less likely to use technology. In the foreword, the authors say the following: "Today the digital transformation provides new avenues for the economic empowerment of women and can contribute to greater gender equality. The Internet, digital platforms, mobile phones and digital financial services offer 'leapfrog' opportunities for all and can help bridge the divide by giving women the possibility to earn additional income, increase their employment opportunities, and access knowledge and general information. We need to seize this opportunity to foster greater gender equality in the labour market, boost economic growth and build a more inclusive, digital world."

And building a more inclusive digital world is precisely one of the big aims of the ESCALATOR programme. The Digital Champions Initiative, a flagship activity of ESCALATOR, is a six-track mentorship programme to support Humanities and Social Sciences researchers in their digital and computational journeys. The six tracks are all aimed at different audiences with the EMPOWER track specifically focusing on supporting women.

The EMPOWER track launch series included four sessions with two speakers per session. During the sessions, we paired tech community leaders from communities that actively seek to support women and other minorities, with researchers from a Social Science or Humanities background. Recordings of talks and speaker slide decks are available via our YouTube channel.

The series was very well attended and participants contributed valuable comments and questions during and after the sessions. Organisations that were represented include the HSRC, UP, UCT, CPUT, SPU, UFH, UKZN, NWU, DUT, UWC, UNIZULU, UFS, TUT, WITS, and more! We look forward to continuing the engagement with this growing community of women in Humanities and Social Sciences who are interested in the intersection of research and teaching and technology.

**International Gender and Language Association's Conference**
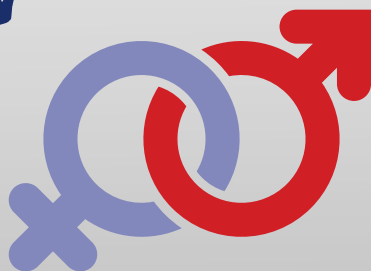
# RESEARCH ON AFRIKAANS GENDER TERMINOLOGY AT SADILAR

### - Benito Trollip

At the end of 2020 two of SADiLaR's researchers, Benito Trollip and Deon du Plessis, embarked on a journey to do research on Afrikaans gender terminology. Their interest was sparked by an Afrikaans blog Benito wrote with Risha Lötter earlier in 2020, discussing the challenge of expressing oneself or one's gender identity when the vocabulary for that purpose does not exist.

Focusing on gender terminology in Afrikaans, as well as the possibilities afforded by digital means to research them, represented an opportunity that these researchers did not want to let slip through their fingers.

The first order of business was to write an abstract to submit to an international conference specifically aimed at gender and language. The abstract they wrote and submitted to the 11th International Gender and Language Association's Conference (available online) was the first step in bringing this important topic to the foreground. The video presentation has also recently been added to SADiLaR's YouTube channel. Their discussion includes different interesting aspects of the normative history of Afrikaans that possibly played a role in the lack of terminology pertaining to the expression of gender. These researchers also highlight efforts by members of the language and other communities to develop terminology, like those of the Tshisimani Centre for Activist Education and the Commission of Gender Equality in collaboration with PanSALB's Afrikaans national language board. Several avenues for further investigation are mentioned in their presentation.

Since the conference in July 2021, the collaboration has continued, more specifically on the topics specified for further research. The project is currently being expanded in several ways that include a more detailed discussion on the variety of social, linguistic and possibly political aspects that influenced the development of Afrikaans, the expansion of words that were featured in the abovementioned conference presentation, and a discussion of possible pejorative or demeaning terminology or contexts in which terms relating to LGBTQIA+ persons are used. The project will investigate whether these terms are being reclaimed by the community or still only used to insult or demean.
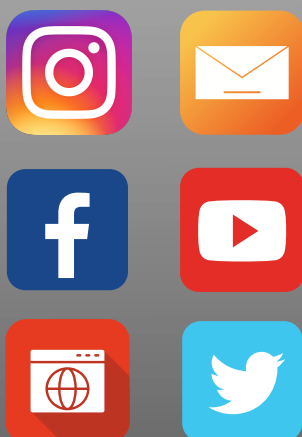
View the video presentation **HERE**.

### Contact us:

**Click here to subscribe to our newsletter**

**Contact us:**

# UPCOMING EVENTS

## Digital Humanities Association of Southern Africa (DHASA) Conference

### 29 November–3 December 2021

The DHASA conference is an interdisciplinary platform for researchers working on all areas of Digital Humanities (including, but not limited to language, literature, visual art, performance and theatre studies, media studies, music, history, sociology, psychology, language technologies, library studies, philosophy, methodologies, software and computation). It aims to create the conditions for the emergence of a scientific Digital Humanities community of practice.

Online registration open until 28 November 2021: https://dh2021.digitalhumanities.org.za/registration/

## Resources for African Indigenous Languages Workshop

### 29 November–3 December (during the DHASA conference)

SADiLaR is organising the second RAIL workshop in the field of African indigenous language resources. This workshop aims to bring together researchers who are interested in showcasing their research and thereby boosting the field of African indigenous languages. This provides an overview of the current state-of-the-art and emphasises availability of African indigenous language resources, including both data and tools. Additionally, it will allow for information-sharing among researchers interested in African indigenous languages and also start discussions on improving the quality and availability of the resources.

The RAIL workshop will be co-located with the DHASA conference, and therefore registration will run through the DHASA website: https://dh2021.digitalhumanities.org.za/registration/
Participants will need to register for the conference and choose to attend the RAIL workshop during the registration process.