Digital footprint crucial for indigenous language preservation

By Birgit Ottermann , issued by North-West University (NWU)   23 Feb 2024

"The work that is being done by SWiP is highly commendable. The key message of today's meeting for me is the importance of preserving our languages. As a Zulu speaker, it is crucial that isiZulu does not die."

These are the words of Tholakele Nkwanyana, a lecturer in Education and Language Development at North-West University (NWU) and one of the panellists at the SWiP side event of Science Forum South Africa (SFSA) 2023, which took place at the Council for Scientific and the Industrial Research International Convention Centre in Pretoria on 6 December 2023.

SWiP, which is short for SADiLaR-Wikipedia-PanSALB, is a collaborative initiative by the South African Centre for Digital Language Resources (SADiLaR), Wikipedia and Pan South African Language Board (PanSALB), aimed at promoting all South African indigenous languages online. Launched in September 2023, the project endeavours to bring together communities of indigenous language users and give them the necessary skills to create and review content on Wikipedia. In doing so, participants will collectively increase their respective languages' digital footprint.

"There is a serious lack of online presence when it comes to South Africa's indigenous languages," Lihle Sosibo, SADiLaR's communication manager, said in her welcome and introduction to the SWiP meeting. "When we talk about language, we also talk about culture and identity. The SWiP project's objective is to empower anyone who is interested in contributing content to Wikipedia in their own language; and, ultimately, to promote and preserve our indigenous languages and protect them from disappearing over time."

Accessible knowledge for all

The one-day SWiP event comprised an exciting panel discussion by experts in the fields of preservation of languages,

culture, and digitisation of information, and an introductory mini workshop to Wikipedia focusing on editing, translating, and making content available online.

Besides Tholakele Nkwanyana, the six-member panel also included Dumisani Ndubane, monitoring and evaluation strategist at the Wikimedia Foundation; Julius Dantile, executive head of languages at PanSALB; Marissa Griesel, project manager of the African Wordnet and Multilingual Terminology projects at the University of South Africa (Unisa); Prof Menno van Zaanen, professor in Digital Humanities at SADiLaR; and Dr Laurette Marais, senior researcher at the Council for Scientific and Industrial Research (CSIR).

Prof Laurette Pretorius, emeritus professor of Computer Science at Unisa, facilitated the panel discussion which was centered on different aspects of the theme 'Preserving Languages & Scientific Information: Accessible Knowledge for All'.

"A language that doesn't cross the digital ocean is a dead language," the Wikimedia Foundation's Dumisani Ndubane said. "For us to achieve a greater online presence for our indigenous languages, we need to fuse Wikipedia with education. We know through research that if someone cannot speak their language, and speak it properly by the fourth grade, they are less likely to speak any other language better; and they are particularly less likely to have a good proficiency in English, which translates in fewer economic opportunities for them later," he explained.

According to Ndubane, the education system is currently experimenting with bilingualism, which is already showing promising results in the Eastern Cape where kids with an increased proficiency in their own languages are drastically increasing their ability to learn in the classroom.

"But for digital literacy, you need a good technology infrastructure in place, such as access to the internet and high-quality digital educational content in our indigenous languages," Ndubane continued. "This is where Wikipedia comes in. It has consistently been one of the top 10 most accessed websites over the past 20 years. It is free and easy for people to upload information in their own languages – a great platform to leverage in building the much-needed digital presence of our indigenous languages."

'If you want to preserve a language, use it!'

Julius Dantile, the executive head of languages at PanSALB, emphasised the need for language and culture preservation in South Africa. "Our country's constitution is celebrated in the world as an excellent one, but is it good for the preservation of our languages and our cultures? If you want to preserve a language, use it! You can't develop a language that you don't use. For us to promote multilingualism, we need to develop it while we use it. That is what PanSALB stands for: it was established to promote and create conditions for the development and use of our official languages," Dantile said.

Marissa Griesel, project manager of the African Wordnet and Multilingual Terminology at UNISA, discussed the use of African Wordnet to promote multilingualism in the classroom. "A wordnet is made up of basic building blocks called synsets. Just as in other electronic dictionaries, we have definitions and usage examples for each of the entries in a wordnet, but what makes this different is that the entries or concepts described are connected to each other via semantic relations.

"This is very useful if you are trying to do machine translation and other natural language processing tasks. When we talk about multilingualism, we often talk about English to our indigenous languages. But what makes African Wordnet special is that our indigenous languages are connected. A learner could learn another language from a language they already know. For example, they could learn isiZulu from Sesotho or Afrikaans, and thus acquire more knowledge in one of our other indigenous languages," she explained.

Tholakele Nkwanyana shone a spotlight on North-West University's strides towards multilingualism. "We have a very diverse community of students and staff members. We provide training for our lecturers and tutors, and pride ourselves in our interpreting services for educational purposes in our lecture rooms. Having interpreters comes with challenges, though, when it comes to terminology. So, we have embarked on a journey of developing (and standardising) terminology in the

different languages."

Machine translation systems and computational grammars

SADiLaR's professor in Digital Humanities, Menno van Zaanen shared his personal journey from computer science to discovering the field of computational linguistics (CL), explaining how CL is used to analyse and build tools that investigate patterns in languages. "Machine translation systems rely on a broad range of online information. The more information and descriptions are available in our indigenous languages online (on platforms such as Wikipedia), the better the quality of machine translation systems will become. Wikipedia is also an important source of terminology," he added.

The final panellist, Dr Laurette Marais from the CSIR, gave a brief overview of the role computational grammars play in language and knowledge preservation. "In very simple terms, computational grammars are structured descriptions of language that can run on computers. It's looking at the structures to define sets of rules in language that help to computationally preserve language and knowledge.

There are two kinds of computational grammars: morphosyntactic and semantic. The first explores what you can say in a language by looking at parts of speech (e.g. a noun, verb, and adjective) in a given sentence and how they relate to each other. The second looks at how specific concepts are expressed in a given language," she explained. "It is possible to build a grammar where the rules in a language are about concepts, and there are also ways of linking these two kinds of grammars to encode knowledge."

'We can achieve something huge'

After a short Q&A session, Bobby Shabangu, president of the Wikimedia ZA Chapter, presented a mini workshop in which he introduced participants to the use of Wikipedia and its role in preserving language and culture. This was followed by breakout sessions where participants discussed a total of 10 themes and explored the barriers and opportunities to improving Wikipedia.

"My biggest take of today's event was WordNet and Wikipedia," said Ben Nkhumane, a principal language practitioner at the North West Department of Arts, Culture and Recreation. "We have a lot of translation work to do, and this is another giant step for us to take translation to the next level, using a different format."

"The workshop was incredibly valuable," Marissa Griesel commented. "I think if communities can rally behind our languages and participate in this project, we can achieve something huge – not just to preserve our languages and the cultures behind them, but also to make technological advancements in artificial translation and search engines, and to represent all of our South African languages online."

*The SWiP event was live streamed on YouTube for participants who were unable to attend in person. Watch the YouTube Stream here.

- **GoAllOut student chapter breaks Guinness World Record** 17 May 2024
- **Seven Eagles to soar at the 2024 Olympic Games in France** 15 May 2024
- **Child speech database research project attracts international attention** 7 May 2024
- **Siya celebrates Rassie** 7 May 2024
- **"Yster" Rassie Erasmus receives honorary doctorate** 7 May 2024