

News in focus



EYE UBQUITOUS/ALAMY

Student teachers in Kampala. Researchers want to expand scientific terms in African languages including Luganda, spoken in East Africa.

AFRICAN LANGUAGES TO GET MORE BESPOKE SCIENTIFIC TERMS

Many words common to science have never been written in African languages. Now, researchers from across Africa are changing that.

By Sarah Wild

There's no original isiZulu word for dinosaur. Germs are called *amagciwane*, but there are no separate words for viruses or bacteria. A quark is *ikhwakhi* (pronounced kwa-ki); there is no term for redshift. And researchers and science communicators using the language, which is spoken by more than 14 million people in southern Africa, struggle to agree on words for evolution.

IsiZulu is one of approximately 2,000 languages spoken in Africa. Modern science has ignored the overwhelming majority of these

languages, but now a team of researchers from Africa wants to change that.

A project called Decolonise Science plans to translate 180 scientific papers from the AfricArXiv preprint server into 6 African languages: isiZulu and Northern Sotho from southern Africa; Hausa and Yoruba from West Africa; and Luganda and Amharic from East Africa.

These languages are collectively spoken by around 98 million people. Earlier this month, AfricArXiv called for submissions from authors interested in having their papers considered for translation. The deadline is 31 August.

The translated papers will span many

disciplines of science, technology, engineering and mathematics. The project is being supported by the Lacuna Fund, a data-science funder for researchers in low- and middle-income countries. It was launched a year ago by philanthropic and government funders from Europe and North America, and Google.

Languages left behind

The lack of scientific terms in African languages has real-world consequences, particularly in education. In South Africa, for example, less than 10% of citizens speak English as their home language, but it is the main teaching

News in focus

language in schools – something that scholars say is an obstacle to learning science and mathematics.

African languages are being left behind in the online revolution, says Kathleen Siminyu, a specialist in machine learning and natural language processing for African languages, based in Kenya. “African languages are seen as something you speak at home, not in the classroom, not showing up in the business setting. It is the same thing for science,” she says.

Siminyu is part of Masakhane, a grass-roots organization of researchers interested in natural language processing in African languages. Masakhane, which means ‘we build together’ in isiZulu, has more than 400 members from about 30 countries on the continent. They have been working together for three years.

The Decolonise Science project is one of many initiatives that the group is undertaking; others include detecting hate speech in Nigeria and teaching machine-learning algorithms to recognize African names and places.

Eventually, Decolonise Science aims to create freely available online glossaries of scientific terms in the six languages, and use them to train machine-learning algorithms for translation. The researchers hope to complete this project by the beginning of 2022. But there’s a wider ambition: to reduce the risk of these languages becoming obsolete by giving them a stronger foothold online.

Terminology creation

Decolonise Science will employ translators to work on papers from AfricaRxiv for which the first author is African, says principal investigator Jade Abbott, a machine-learning specialist based in Johannesburg, South Africa. Words that do not have an equivalent in the target

language will be flagged so that terminology specialists and science communicators can develop new terms. “It is not like translating a book, where the words might exist,” Abbott says. “This is a terminology-creating exercise.”

But “we don’t want to come up with a new word completely”, adds Sibusiso Biyela, a writer at ScienceLink, a science-communication company based in Johannesburg that is a partner in the project. “We want the person who reads that article or term to understand what it means the first time they see it.”

“The big idea is cultural ownership of science.”

Biyela, who writes about science in isiZulu, often derives new terms by looking at the Greek or Latin roots of existing scientific words in English. Planet, for example, comes from the ancient Greek *planētēs*, meaning ‘wanderer’, because planets were perceived to move through the night sky. In isiZulu, this becomes *umhambi*, which also means wanderer. Another word for planet, used in school dictionaries, is *umhlaba*, which means ‘Earth’ or ‘world’. Other terms are descriptive: for ‘fossil’, for example, Biyela coined the phrase *amathambo amadala atholakala emhlabathini*, or ‘old bones found in the ground’.

In some scientific fields, such as biodiversity research, researchers trying to find the right terms will need to tap into spoken sources. Lolie Makhubu-Badenhorst, acting director of the Language Planning and Development Office at the University of KwaZulu-Natal in

Durban, says that the absence of a scientific word from written data sets does not mean that it does not exist. “You’re written-centred, I’m oral-centred. The knowledge is there, but it is not well documented,” says Makhubu-Badenhorst, who is not part of the Decolonise Science project.

Decolonise Science’s terminology specialists will come up with a framework for developing isiZulu scientific terms, says Biyela. Once that’s complete, they will apply it to the other languages.

The team will offer its glossaries as free tools for journalists and science communicators, as well as national language boards, universities and technology companies, which are increasingly providing automated translation. “If you create a term and it isn’t being used by others, it isn’t going to permeate into the language,” says Biyela.

Big tech: ‘we need your help’

Masakhane’s researchers say that global technology companies have historically ignored African languages, but in recent years, they have begun funding research in the field.

“We’re aware that the many thousands of African languages are currently under-represented in translation software,” a Google spokesperson told *Nature*. The tech giant wants to expand Google Translate to include more African languages, including Twi, Ewe, Baoulé, Bambara, Fula, Kanuri, Krio, Isoko, Luganda, Sango, Tiv and Urhobo, they added. However, it needs “speakers of those languages to help us improve the quality of our translations” so they can be integrated into the service.

“The big idea is cultural ownership of science,” Biyela explains. Both he and Abbott say it is crucial to decolonize science by allowing people to do research and speak about science in their own languages. At the moment, it is possible to use African languages to talk about politics and sport, but not science, says Biyela.

Similarly, English is the dominant language of environmental stewardship and conservation – but unless people understand the meaning of specific terms and concepts and can talk about them in their home languages, they can feel disconnected from government efforts to preserve ecosystems and species, says Bheka Nxele, a programme manager for restoration ecology, environmental planning and climate protection in the eThekwin Municipality of South Africa.

The researchers are concerned that if African languages are not included in online algorithms, they could eventually become obsolete and forgotten. “These are languages [people] speak. These are languages they use every day, and they live with and see the reality that in x number of years, their language might be dead because there is no digital footprint,” says Siminyu.



Google is calling for help to improve the quality of its African-language translations.

CRISTINA ALDEHUELA/AFP VIA GETTY