DEVELOPMENT OF AN INDIGENOUS LANGUAGE THROUGH CROWD-SOURCING ON SOCIAL MEDIA

Mpho Ngoepe, Kganathi Shaku & Napjadi Letsoalo

Introduction

Many commentators bemoan the dearth of African languages in the classroom, on television, radio, and even in social spaces. Many languages are vanishing from the face of the earth, according to some experts. Indeed, many languages are on the verge of extinction and must be preserved. According to the UNESCO *Atlas of the World's Languages*[1] shows that, there are approximately 3000 endangered languages worldwide, with indigenous languages facing the greatest threat of extinction. It is estimated that half of the world's 6000 languages will become extinct within the next century. In the United States, for example, it was reported in 1962 that approximately fifty-one American Indian languages had ten or fewer speakers, which might result in the languages becoming extinct.[2] This is also happening to minority languages with a relatively large number of speakers in the United States.

This is a problem that is also prevalent in South Africa. In South Africa, for example, Katrina Esau—also known as *Ouma* (aunt) Katrina—is the last speaker of the N/uu language, a San language. Despite the fact that she produced and released the only children's book ever written in the N/uu language on 24 May 2021, she is concerned that the language will be lost if no one speaks it after her. *!Qhoi n|a Tjhoi* (*Tortoise and Ostrich*) is a children's book that has been translated into English, Afrikaans, and isiXhosa[3]. Languages like N/uu are on the verge of extinction once the few remaining

---

[1]UNESCO. Atlas of the World's Languages. (2011). Paris: UNESCO.

[2]Daniel J Villa, "Integrating technology into minority language preservation and teaching efforts: an inside job" *Language Learning & Technology* 6, 2 (2002): 92-101.

[3]Katrina Esau, *!Qhoi n|a Tjhoi/Tortoise and Ostrich/Skilpad en Volstruis* (Cape Town: New Africa Books, 2021).

speakers die. Therefore, initiatives to promote and preserve such languages must be undertaken. Language preservation, promotion, and development, according to UNESCO, is a solution to language extinction.

As Kari Chew,[4] a professor of educational leadership and policy studies specializing in indigenous languages at the University of Oklahoma attests that indigenous communities, individuals, civil society, and governments are increasingly aware that languages are an important linguistic and cultural tradition. As a result, some individuals and civil society organisations have chosen to take measures to halt the deterioration of their heritage language. In South Africa, there are many initiatives of developing, promoting, and preserving indigenous languages, such as the national reading for enjoyment campaign, PUKU's Creative Incubator, and the national reading summit by organisations like *Nal'ibali*,[5] PUKU Children's Literature Foundation, and the National Library of South Africa (NLSA)'s Centre for the Book respectively. For example, the NLSA and the University of South Africa hosted the first national reading summit in 2021, where people working in the reading space were invited to share their experiences on the theme "strengthening the reading ecosystem." The summit aimed to achieve the following outcomes:

1. Assess national reading initiatives and campaigns in order to determine their effectiveness, impact and reach.

2. Develop strategies and interventions to strengthen and integrate reading initiatives in the country in order to reach significant levels of leisure reading.

3. Enable informed decision-making by officials involved in reading programmes.

PUKU, on the other hand, is a non-profit organisation that offers creative writing and book review workshops in indigenous languages to aspiring young writers, teachers, and librarians. The Republic of South Africa's Constitution recognises eleven official languages: Sepedi (also known as Northern Sotho), Setswana, siSwati, Tshivenda, Xitsonga, Afrikaans, English,

---

[4] Kari AB Chew, "#KeepOurLanguagesStrong: Indigenous language revitalization on social media during the Early COVID-19 Pandemic" *Language Documentation & Conservation* 15 (2021): 239 –266.

[5] Nal'ibali is a national reading-for-enjoyment campaign promoting reading, writing and sharing stories in all South African languages.

isiNdebele, isiXhosa, and isiZulu, with sign language and the recently gazetted Khelobedu[6] being the twelfth and thirteenth.[7] The Constitution states that all official languages must enjoy parity of esteem and must be treated equitably. This includes their application in both private and public settings. In addition, the National Language Policy calls for the promotion of all previously disadvantaged and marginalised indigenous African languages in South African society. As a result, the country has a multilingual population that is fluent in at least two languages, the most common of which are isiZulu and isiXhosa.

From 1652 to 1994, the official languages of South Africa were of European origin, namely Dutch, English, and Afrikaans (a creole language that evolved during the colonial era having roots in Dutch with influences from Malay, Portuguese and the languages of the Khoe and San people). African languages, which are spoken by at least 80% of the population, were largely ignored. Although a language with a large speaker population is in a unique position, Jozina van der Klok[8] cautions that this does not automatically make it immune to the risk of extinction, as there may be other factors that can lead to it not being widely spoken, such as increased mobility, a rising middle class, the influence of English, and globalisation. South Africa's new Constitution of 1996 granted official protection to all major languages. It should be noted, however, that South Africa has approximately thirty-four historically established languages, thirty of which are living languages and four Khoesan languages that are extinct. The nine indigenous languages are broadly classified as

- Nguni-Tsonga (isiNdebele, isiXhosa, isiZulu, siSwati & Xitsonga)
- Sotho-Makua-Venda (Sesotho, Sepedi, Setswana & Tshivenda).

---

[6] Khelobedu is a dialect spoken by over two million Balobedu people in Limpopo in the northern part of South Africa.

[7] President Cyril Ramaphosa's Cabinet has approved the Constitutional Eighteenth Amendment Bill for public comment, which will finally making South African Sign Language the country's 12th official language." Devon Thomas "South African Sign Language Approved as SA's 12th Official Language" *EWN Eyewitness News* (28 May 2022) [online resource] https://ewn.co.za/2022/05/28/south-african-sign-language-approved-as-sa-s-12th-official-language (accessed 28 May 2022).

[8] Jozina van der Klok, "The Javanese language at risk? perspectives from an East Java village" *Language Documentation & Conservation* 13 (2019): 300-45.

Within the first group, Xitsonga is the only Tswa-Ronga language, whereas isiZulu, isiXhosa, isiNdebele, and siSwati are Nguni languages.

South Africans are a code-switching people due to the fluidity of languages in the country as a result of migration, urbanisation, friendship, marriages, and other factors. The term "code switching" refers to the use of more than one language in a single conversation. Indeed, almost every adult South African does this at some point in their lives, whether consciously or unconsciously. A Northern Sotho/Sepedi speaker, for example, would say "*Ke tla phone*" instead of "*Ke tla go leletša*" which means "I will call you."

This study focused on Sepedi, also known as Northern Sotho. The language will be referred to as Sepedi for the purposes of this paper. Since the social media project is in this language, Sepedi was chosen as a case study. As Kari Chew argues, "social media have become a key space for both documenting and facilitating shifts, adaptations, and persistence of a language."[9] Indeed, the rise of social media has implications for language development and preservation. Although it is a double-edged sword, social media can help to develop, promote, and preserve a language if used properly. Facebook is the most popular social media platform with over 2.85 billion active monthly users worldwide. South Africa's Facebook user base was predicted to be around 28.83 million by December 2021.

The sheer speed with which social media has grown, on the other hand, can be a problem, particularly in areas where technology is out of reach for the majority of people. This is true in South Africa, where most rural areas lack network coverage. For example, according to a study by Mpho Ngoepe, Samuel Mojapelo, Noko Ngoepe and Thomas Van der Walt, in certain spots in South Africa, access to the network is problematic and smart phones do not work properly due to weak signals.[10]

The current study used online ethnography to explore the development of vocabulary in Sepedi through crowdsourcing in social media, especially the private Facebook page, *Sešego: Pukuntšutlhaloši ya Sepedi* (A Basket for Sepedi

---

[9] Chew, "#KeepOurLanguagesStrong," 239-266.

[10] Mpho Ngoepe, Samuel Mojapelo, Noko Ngoepe and Thomas Van der Walt, "Bridging the Information Gap Through Career Expos to Rural School Learners in the Makgabeng Village in the Limpopo Province in South Africa" *Innovation* 55, 2 (2017): 26-41.

Dictionary).[11] This is a private group with content that only current members can see, and membership must be requested and approved before a user can join.

Dr Napjadi Letsoalo of the University of South Africa founded the Facebook page on 16 October 2019, after realising that people's records of their traditions, culture, and way of life die with the oldest member of the community unless that record is memorised by subsequent generations. Even the introduction of a written record does not solve the problem. Writing down a story, for example, does not capture how skilled storytellers pass on cultural history or the language skills they use to do so. However, the development of new technologies such as social media has altered this limitation. The Facebook page was originally intended to share Sepedi vocabulary and definitions, but it has since expanded to include any Sepedi-related content. Members of the group debate words until they reach an agreement on their meaning. This is accomplished through crowdsourcing, which is the practice of obtaining information or input for a task or project by enlisting the help of a large number of people, typically via the internet. The participants in this project were not compensated. The group currently has 320 members and posts on topics range from new words—or words that members want to clarify—to the use of emoticons and emojis in the form of a quiz for Sepedi proverbs.[12]

According to Eli Dresner and Susan Herring, both emoticons and emojis[13] are any of various keyboard-generated or image symbols used in computer-mediated communication (as in text messages, email and social media)

[11]*Sešego: Pukuntšutlhaloši ya Sepedi* [online resource] *Facebook* https://www.facebook.com/groups/499183754266703 (accessed 29 May 2022).

[12]An emoticon is "a group of keyboard characters (such as :-)) that typically represents a facial expression or suggests an attitude or emotion and that is used especially in computerized communications (such as email)" *Merriam-Webster Dictionary* [online resource] https://www.merriam-webster.com/dictionary/emoticon. An emoji, on the other hand, is "any of various small images, symbols, or icons used in text fields in electronic communication (as in text messages, email, and social media) to express the emotional attitude of the writer, convey information succinctly, communicate a message playfully without using words, etc." *Merriam-Webster Dictionary* [online resource] https://www.merriam-webster.com/dictionary/emoji (accessed 31 May 2022).

[13]Eli Dresner and Susan Herring. "Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force" *Communication Theory* 20, 3 (2010): 249-268.

to express the emotional attitude of the writer, convey information succinctly, communicate a message playfully without using words, etc. The group is private, but anyone on Facebook can find it. Only members, however, can see who is in the group and what they post. As observed by the academic linguist Ayo Bamgbose, language development is an activity that is not only limited to the government authorities, as it can draw participation from different societal stakeholders such as individual authors, language commissions and societies, university departments, media houses, writers, and translators.[14] As advocates for the preservation, promotion, and development of indigenous languages, Sumathi Renganathan and Inge Kral contend that:

> Language is the medium linking people to knowledge and therefore, a community's connection with its history, its culture and tradition and its base of specific knowledge are lost when a particular language has ceased to exist anymore.[15]

The problem

Although the Constitution of the Republic of South Africa recognises eleven official languages, it appears that almost all of these languages, with the exception of English and Afrikaans to some extent, have been approved merely for ceremonial purposes because they are still marginalised in the government, in education, in the private sector and in the economic sector. Thembelihle Makhanya and Sibonsile Zibane argue that indigenous languages in South Africa remain on the periphery.[16] For example, these languages rarely feature in official government documents. It is even worse at universities, as they are reflected in university language policy documents or are taught as subjects—but are not a medium of instruction. This hinders

---

[14]Ayo Bamgbose, *Language and the Nation* (Edinburgh: Edinburgh University Press, 1991).

[15]Sumathi Renganathan and Inge Kral, "Digital Preservation of Language: Cultural Knowledge and Traditions of the Indigenous Semai" *SHS Web of Conferences* 53, 02001 (2018): [unnumbered pages 1-2].

[16]Thembelihle Makhanya and Sibonsile Zibane. "Students' voices on how indigenous languages are disfavoured in South African Higher Education" *Language Matters* 51, 1 (2020): 22-37 (especially page 22).

access to education by many, and it is contrary to the constitutional provision that "[e]veryone has the right to receive education in the official language or languages of their choice."[17]

Furthermore, few books are published in these languages. On 7 March 2022, the National Library of South Africa, in partnership with PUKU Children's Literature Foundation and the University of South Africa, launched the first catalogue of 100 children's books in isiXhosa. This marginalisation is exacerbated by the fact that indigenous languages, including Sepedi, are underdeveloped due to a lack of adequate vocabulary. This has an impact on the education sector as well, because identifying and obtaining materials for instructional purposes is a major issue in the teaching of indigenous languages in South Africa. Hence, academic specialist in multilingual education, Mbulungeni Madiba contends that:

> the use of indigenous South African languages in modern domains such as science, technology and business is hampered by a lack of modern terminology in these languages.[18]

This, therefore, raises the need for interventions such as crowdsourcing in social media for the development of these languages.

Methodology

This study used the qualitative research approach, which is associated with the constructivist paradigm. In qualitative research, which is inductive in nature, the researcher generally explores meaning and insight into a given situation or phenomenon. In this regard, the study adopted an online ethnography research design to explore the development of vocabulary in Sepedi through crowdsourcing on the Facebook page *Sešego: Pukuntšutlhaloši ya Sepedi*. In online ethnography, data are collected and analysed from virtual communities, online texts, and interactive chatrooms. As renowned social scientist, Alan Bryman contends, one of the key—and yet most difficult—

---

[17]South Africa. *Constitution*, Chapter Two, Bill of Rights: Education, Clause 29 (2), page 12 [online resource] https://www.justice.gov.za/legislation/constitution/pdf.html (accessed 30 May 2022).

[18]Mbulungeni R. Madiba "Strategies in modernization of Venda" Thesis, PhD (University of South Africa), 2000.

steps in ethnography is gaining access to a social setting that is relevant to the research problem in which one is interested.[19]

Although the internet is a placeless space, in this study we were part of the cyberspace community (the Facebook page). For the current study, the researchers participated in the activities of the page since it was created on 16 October 2019. They were part of the crowdsourcing members who contributed towards discussion on the Facebook page. In this regard, as argued by Bryman, the researchers immersed themselves in the group on the Facebook page for an extended period, observing contribution and engaging in online discussions, and asking questions. Since postings on the Facebook page, like letters in newspapers, are in the public domain, obtaining consent from participating members was unnecessary. The benefit of social media is that information is available even after the post has been deleted. In this regard, we were able to revisit the page and capture screenshots for the purpose of presenting results for data analysis. The names of the participants were removed from the screenshots so that they could not be identified.

Presentation and discussion of results

For analysis, data were categorised around emerging themes from the posts, which include vocabulary development, proverbs in emojis, language promotion, and nature, including names of trees, rain and stars. There is also an additional theme for "miscellaneous" regarding posts that are not language related. It is worth mentioning that, in some instances, new words were coined; for example, a word for "emoji" has been created in the group as *tshwantšhoema* through compounding the word *seswantšho* (image) and *emela* (represent) whereby the prefix of the former word and the suffix of the latter have been clipped off, while "webinar" is *bobinare* through compounding Bobi (web) and *nare*, which is a transliteration of seminar. Although these words are used, they are not yet registered with the language authorities. We also took a wide range of examples presented as a screenshot image of posts.

---

[19]Alan Bryman, *Social Research Methods*. 4th edition (Oxford: Oxford University Press, 2012).

Vocabulary development

Posts on this theme encouraged group members to post words they knew but did not know the meaning of, as well as words they knew in another language but not in Sepedi, and what they thought could be new words, such as *bobinare* for webinar and *tshwantšhoema* for emoji. Members will then discuss the meaning of the word until a consensus is reached. Fig. 2-1depicts some of the words discussed in this theme. Interesting words that are aligned to the latest developments include *dibaledi* (computers), *mararankodi* (internet), *marangrang* (network), *kiletšo ya mesepelo* (lockdown), *lengwalo la boeti* (visa), *tšenereitara* (generator), *sebeubeu* (solar plexus) and others. Some of the words such as *sebokuboku* reflected in Fig. 2-1 are very old. The word attracted a discussion, as it was not easy to find its meaning by members of the group.
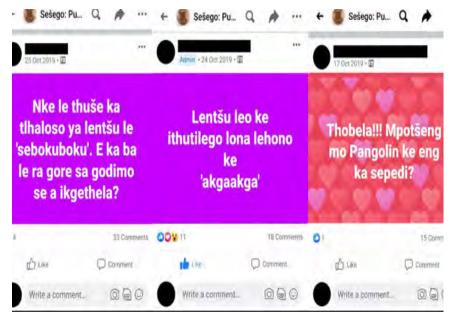

Fig. 4-1. Explanation of words in Sepedi

In one of the posts, as reflected in Fig. 2-2, a member wanted to know what "one million" was in Sepedi. A response was that there was only a borrowed word. This is so because Sepedi has words for counting to thousands, not

millions. So, only the borrowed word "*milione*" is used. Mona Baker describes borrowing as a translation strategy of taking a word or an expression straight from another language. Once a word or expression is borrowed into a language, the source language cannot predict or control its development or the additional meanings it might or might not take on.[20]

Fig. 4-2. Explanation of words in Sepedi

In addition, a pictorial entry of a dictionary with an explanation in Sepedi was posted. While some members felt that a dictionary with vulgar language is undesirable in Sepedi, others felt that such a dictionary is necessary because a dictionary should reflect the dynamic nature of a society, even if certain words may be regarded as crude or vulgar; such words can be indicated as vulgar. As a result, the language must embrace new developments as well as interaction with other languages and cultures.

---

[20]Mona Baker, *In Other Words: A Coursebook on Translation*. 3rd edition (London: Routledge, 2018).

Emoji proverbs in Sepedi

Sepedi is one of the African languages that is rich in proverbs. These proverbs are thought to be ancient words of wisdom passed down from the forefathers.

Fig. 4-3. Sepedi proverbs in emojis>>>

They are inspired by nature, animals, and living and non-living things and they include thoughts and advice. There are various forms of Sepedi prov-

erbs. One of the Facebook page's themes was Sepedi emoji proverbs. The administrator was the only one who posted the emojis in this case; the members would then indicate what the proverb was and provided their interpretation. Fig. 2-3 depicts some of the proverbs. Many people struggled to decipher the proverbs and provide the meaning.

The first proverb in Figure 2-3 depicting a person running away from rain to a heavy rain in Sepedi is *motšhabapula o tšhabela matlorotlorong*. It is equivalent to the English saying, "out of the frying pan into fire." The second emoji with a rhinoceros and a person climbing a tree is *gopola tšhukudu o namele mohlare*. The closest English equivalent is "speak of the devil, and he will appear." The third emoji, with a locust and a plane, represents flight, *tšie e fofa ka mošwang*; its English equivalent is "an empty sack cannot stand upright" meaning you can keep going if you have eaten. The last emoji is for the proverb *swana ya mošate wa e gapa o molato, wa e tlogela o molato*. The English version of this one can be "between the devil and the deep blue sea" or "damned if you do and damned if you don't" or even a "catch-22 situation." Members of the group felt that a book of proverb emojis should be written and possibly used in schools after the emojis were posted and discussed. We hope that the Facebook page's developers will consider writing the book in the future.

Nature

Another theme related to nature. In this theme, posts were about the names of plants, rain and months in Sepedi.

Fig. 2-4 reflects two types of trees that were posted in the group. It was argued that some proverbs, as well as surnames, were based on trees. For example, the first picture in Fig. 2-4 is for pearl millet. The proverb for this is *pudi ya ja leotša e fetetša tše dingwe*, meaning when an offence is committed all members of a group are punished irrespective of who committed that offence (common purpose). However, there were arguments that the plant depicted in the image was not pearl millet, so a correct plant, as shown in Fig. 2-5, was posted by one of the group's members.

Fig. 4-4. Names of trees in Sepedi

The other tree in Fig. 2-4 is known as *mogapa* (species of thorn-tree). Mogapa is also used as a surname among Pedi people, mostly found around Seshego township, in Limpopo Province. A popular idiom regarding the tree is *mogapa wo mogolo o wele* meaning "a giant has fallen."

Miscellaneous

There were also posts that had nothing to do with language development or the purpose of the Facebook page. For example, one group member consistently advertised his services as a mechanic while some members posted requests for help with their children's homework. It should be noted that such requests for assistance were in Sepedi and related to language. However, participation was of high quality, despite the fact that inputs were minimal. For example, it is not uncommon for two weeks to pass without anyone posting anything.

Fig. 4-5. The correct picture of pearl millet

Conclusion and recommendations

The purpose of this research was to look into the development of the Sepedi language through crowdsourcing on social media, especially Facebook page. The analysis of the social media page revealed four key themes in the posts:

☞ vocabulary development

☞ proverbs in emojis

☞ nature, including the names of trees and rain

☞ and miscellaneous for posts that were not related to the Facebook page.

The discussion shows that the group page is educational and has a lot of potential for developing the Sepedi language. There were numerous comments in which members expressed their joy at being a part of the group because they learn so much and contribute to the creation or discussion of

words. With the proliferation of social media, indigenous language speakers can take advantage of the opportunities provided by these platforms to develop their languages.

As the authors of a UNESCO report would contend, if a language is endangered or lost, that represents a profound loss of humanity because culture, social interaction, rituals and spirituality are possible through language, not just a language, but a mother tongue.[21]

In addition, "when a language of a community is lost, traditional knowledge of the community is also lost."[22] Hence, Patrick Ngulube, a scholar of indigenous knowledge systems and archival science, believes that publishers and writers are the major role players in knowledge production and, this reproduction chain may assist in promoting and preserving indigenous languages in general and in South Africa in particular.[23] Thus, when language is written and used in both print and digital spaces such as books (including dictionaries) and on social media platforms like Facebook, it becomes preserved and will not cease to exist. This also explains why the developers of the Facebook group *Sešego: Pukuntšutlhaloši ya Sepedi* created the page. Social media can play a role as a platform used to develop, preserve, and promote a language.

As it has been established that up to two weeks can pass without a single post in the group, administrators should consider making the group public so that it can grow and the language can be exposed to a larger community of speakers. Members can be encouraged to become active participants, if administrators establish connections with new members through individual welcome messages, keep the members hooked through constant polls, surveys, and quizzes and posting group-related photos. This will ensure the group's long-term viability and ability to achieve its objectives. Furthermore, the results of the activities on the Facebook page are not disseminated to a

---

[21]UNESCO, "Local and Indigenous Knowledge Systems" [online resource] https://en.unesco.org/links (accessed 3 June 2022).

[22]Sumathi Renganathan and Inge Kral, "Digital Preservation of Language, Cultural Knowledge and Traditions of the Indigenous Semai" *SHS Web of Conferences* 53 (2018) [online resource] https://doi.org/10.1051/shsconf/20185302001.

[23]Patrick Ngulube, "Revitalising and Preserving Endangered Indigenous Languages in South Africa tHrough Writing and Publishing" *South African Journal of Libraries and Information Science* 78, 1 (2012): 11.

wider audience in order to preserve and promote the use of the Sepedi language in official government, educational, and other economic institutions.

Still further, there was only one post on the page about the promotion of the language. As a result, the page's owners must register the project as a community-engagement project with a university. The products of such a page should be catalogued, and a word inventory or dictionary should be developed.

The words derived from the projects can then be registered with regulatory bodies such as the Pan South African Language Board, which has language development authority. Community radio stations—as well as regional and national radio stations—can be encouraged to use newly created words for dissemination.

Furthermore, words from Sepedi dialects should be considered for standardisation in order to enrich the language; that is, words from related dialects should be included in the Sepedi lexicon. Documents from the government could also be translated into the language. Failure to pursue initiatives such as this one will result in the extinction of marginalised languages in South Africa.

❧