



OPEN ACCESS

EDITED BY

Josiline Phiri Chigwada,
University of South Africa, South Africa

REVIEWED BY

Atif Ashraf,
University of Central Punjab, Pakistan
Masroor Ahmed,
Virtual University of Pakistan, Pakistan

*CORRESPONDENCE

Shahid Minhas
✉ shahidihsas@gmail.com

RECEIVED 03 June 2024

ACCEPTED 17 January 2025

PUBLISHED 29 January 2025

CITATION

Minhas S and Salawu A (2025) Wikipedia and indigenous language preservation: analysis of Setswana and Punjabi languages. *Front. Commun.* 10:1442935. doi: 10.3389/fcomm.2025.1442935

COPYRIGHT

© 2025 Minhas and Salawu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Wikipedia and indigenous language preservation: analysis of Setswana and Punjabi languages

Shahid Minhas* and Abiodun Salawu

Indigenous Language Media in Africa (ILMA), North-West University, Mahikeng, South Africa

The survival of a language is important for several reasons, some of which are to maintain cultural identity, tradition and wisdom. Therefore, people always try to protect their cultural identity, tradition, and wisdom, thus preserving and promoting their languages. Similarly, indigenous languages, Setswana and Punjabi face challenges in preservation and promotion in this digital world, but their survival is important for preserving cultural identity and heritage. This study examines Wikipedia's role in promoting and preserving Setswana and Punjabi. The research is framed by the Ethnolinguistic Vitality Theory (EVT), which suggests that language survival lies in reclamation, revitalization, and reinvigoration. A quanti-qualitative approach is used to investigate the issue, integrating quantitative metrics from Wikipedia's statistical pages with qualitative content analysis of the articles. Data were collected from May 2022 to May 2024, focusing on article counts, edits, active editors, new pages, top edited pages, and views. The study reveals significant disparities in content quality and community engagement between the Setswana and Punjabi Wikipedia editions. It highlights the challenges and opportunities of Wikipedia for enhancing the visibility and usage of indigenous languages. The findings show that Punjabi Wikipedia has a much larger content volume and user base, but comparatively lower recent activity and collaborative depth compared to Setswana Wikipedia. (Setswana) Tswana Wikipedia, while smaller in content volume, demonstrates a more engaged and active editing community, reflected by a higher depth score and a larger number of active users. This indicates a different stage of development and community involvement between the two language editions. Also, the study provides practical recommendations for Wikipedia contributors, language activists, and policymakers, suggesting that increased participation and strategic initiatives can significantly bolster the preservation efforts for these languages.

KEYWORDS

Wikipedia, indigenous language, language preservation, Setswana, Punjabi

1 Introduction

Wikipedia is one of the largest free online encyclopedias and the distinctive quality of Wikipedia is its collaborative spirit, which invites volunteers to join its large community (Piccardi et al., 2023). English has the most with over six million articles and sixty-two million articles across all languages. Wikipedia's goal is to democratize access to human knowledge, making it freely available to everyone. The content of Wikipedia differs from traditional printed references. It functions as a dynamic platform that is always changing and updating now. Wikipedia entries about recent articles appear within minutes, in contrast to hardcopy encyclopedias where the inclusion of new writings may take months or even years (Miquel-Ribé and Laniado, 2018). Comparing Wikipedia to other encyclopedias, its unequaled comprehensiveness is the result of its open nature, which allows anybody to add. Wikipedia continuously addresses misinformation, mistakes, and vandalism while also enhancing the

amount and quality of its pages with the help of contributors. This approach enables any reader to correct errors or add to the information, creating a cooperative atmosphere for the exchange and improvement of knowledge (Davis et al., 2023). Despite the absence of a central authority controlling content creation, it is still one of the most successful collaborative efforts on the internet. Rather, it functions according to predetermined content guidelines. Among these, the 'Neutral Point of View' bears essential importance, arguing for fair treatment of multiple opinions without editorial prejudice. Additionally, the idea of 'Notability' acts as a basic guideline, describing the criteria by which editors judge the relevance of topics worthy of article insertion (Kristiani, 2021). Hence, Wikipedia allows volunteers to contribute to the knowledge, it offers an opportunity of a free platform for indigenous languages to be alive and preserved.

Similarly, the United Nations noted that indigenous languages must be protected and promoted and for this purpose, states should recognize the linguistic rights of indigenous people and develop language policies. United Nations Permanent Forum on indigenous issues has expressed significant concern about the threat of extinction faced by indigenous languages. According to Minhas and Salawu (2024), almost 43% of the world's 7,000 languages are at risk of extinction by the end of the twenty-first century. Around 50% of these endangered languages are spoken by indigenous peoples. UNESCO ranks the threat of extinction on a scale of 0 to 5, with zero being already extinct languages with no speakers or recall. Languages classed as 1-critically endangered and 2-severely endangered are on the verge of extinction in the next 20 years because they are mostly spoken by the elderly. This makes up around 20%, or 1,400 languages, nearly doubling the already considerable losses to humanity's cultural history (Miquel-Ribé and Laniado, 2018). When a language disappears, a treasure of knowledge is undoubtedly lost, which is irreversible. Similarly, Guerretaz and Engman (2023) argue that losing a language is the end of that civilization.

Despite the significance of language in society and human civilization, many of the languages have become extinct and Setswana and Punjabi are such indigenous languages that are vulnerable and need to be preserved and archived using digital tools (Minhas and Salawu, 2024). Preserving and promoting the Setswana language is crucial for preserving cultural legacy, improving educational outcomes, encouraging social cohesion, and protecting the rights and dignity of Batswana and South African citizens. It ensures that the world's rich linguistic and cultural variety is preserved for future generations (Gwerevende and Mthombeni, 2023). Similarly, scholars noted that Punjabi is integral to the cultural identity of millions of Pakistanis. It embodies Punjabi culture's rich traditions, folklore, music, literature, and history. Punjabi makes a substantial contribution to Pakistani and global linguistic diversity. Each language represents a unique form of thought and expression, and maintaining Punjabi contributes to the global linguistic mosaic (Shah and Sahito, 2024). So, it is imperative to preserve and keep the indigenous languages alive. Therefore, this study aims to examine Wikipedia's role in preserving and promoting Setswana and Punjabi indigenous languages. By examining the presence, quality, and community engagement in these languages on Wikipedia, this study aims to:

- Evaluate the quantity and quality of Setswana and Punjabi content on Wikipedia.

- Analyze how Wikipedia contributes to preserving and revitalizing Setswana and Punjabi languages.
- Evaluate the effectiveness of Wikipedia as a tool for indigenous language promotion and education.
- Explore the opportunities and challenges faced by Setswana and Punjabi languages on Wikipedia.

By achieving these aims, the study seeks to contribute to the broader discourse on digital language preservation, offering insights and practical solutions to ensure the continued vitality and growth of indigenous languages in the digital age.

2 Research context

Indigenous languages are important because most of the indigenous people communicate through these languages (Mpofu and Salawu, 2018); the death of these languages is the death of indigenous culture, indigenous communication, and, most importantly, the death of language (Simpson, 2014). According to Minhas and Salawu (2024) 97% of the world's population communicates in 4% of the languages, while the remaining 3% speaks and communicates in the rest of the 96% of languages. Furthermore, this loss would signify not only the disappearance of linguistic diversity, but also the erosion of cultural heritage, knowledge systems, and identities tied to these languages.

Therefore, the preservation and promotion of indigenous languages have received significant attention in recent years due to their importance and the alarming rate at which these languages are disappearing. Keeping in mind that to promote and keep these indigenous languages alive, there is a dire need to collaborate with indigenous people, communities, government, and world organizations. Meanwhile, scholars noted that the digital revolution has significantly transformed the environment of language preservation (Udoiwang and Akpan, 2023). Digital platforms, particularly collaborative and user-generated ones like Wikipedia, offer new opportunities to document and revitalize endangered languages. Wikipedia, being a freely accessible and open-source encyclopedia, has proven to be an invaluable resource in these endeavors. It enables communities to generate, modify and manage information in their native languages, helping to preserve and promote them (Mlambo and Matfunjwa, 2024). This study is situated at the intersection of linguistics, digital humanities, and social sciences. It aims to fill a gap in the academic literature about the role of digital platforms in language preservation, with a specific focus on Wikipedia. By analyzing how Wikipedia contributes to the preservation of Setswana and Punjabi languages, the study provides insights into effective strategies for using digital tools for language revitalization. It also offers practical recommendations for policymakers, educators, and community activists engaged in language preservation efforts.

2.1 Setswana and Punjabi languages

Setswana is an indigenous African language that is part of the Bantu (Sintu) language family, namely the Sesotho group. It is the native language of Botswana, a group of Bantu tribes that make up a sizable majority of Botswana's population. Setswana is one of

South Africa's eleven official languages (Batibo, 2016); it is also spoken in Botswana, Namibia, and Zimbabwe. Setswana speakers are the fifth largest linguistic group in South Africa, with the majority found in the North-West region. More than 6 million people speak the language, which accounts for roughly 8% of South Africa's population (Gunnink, 2020). The first major book on Setswana was written by British missionary Robert Moffat, who lived among the Batlhaping people. He authored basic publications such as the Bechuana Spelling Book, the Bechuana Catechism, and various Bible volumes. Setswana has been researched and documented since the 18th century, and it retains the distinction of being the first South African language into which Bible texts and Shakespeare's works were translated (Gunnink, 2020). In post-colonial Botswana, Setswana enjoyed significant political support as the national language, but English remains the dominant language in education and the formal sector (Chebanne, 2022). This dual-language policy has socio-political implications, as English is often seen as the language of upward mobility, while Setswana holds cultural importance. Theledi and Masote (2024) noted that in South Africa, Setswana, despite its official status, competes with other indigenous languages, as well as English and Afrikaans, for visibility in media, education, and politics. Setswana is also deeply rooted in the cultural identity of the Batswana people, serving as a medium for traditional practices, oral literature, and cultural expressions. However, urbanization and the dominance of English in formal sectors have posed challenges to the language's long-term vitality. Efforts to revitalize and preserve Setswana, particularly through digital platforms like Wikipedia, are crucial for supporting its cultural significance in an increasingly globalized world.

Similarly, the Punjabi language is spoken not only in Pakistan and India but also in countries with significant Punjabi immigrant populations, such as Canada, the United Kingdom, and the United States. In Pakistan, Punjabi is the mother tongue of most of the population and the provincial language of Punjab (Manan and David, 2014; Abdul et al., 2020). The Punjabi-speaking community holds substantial influence in powerful institutions such as the army and bureaucracy. However, this dominance does not translate into the elevated status of their language. Historically, even before the establishment of Pakistan, the Punjabi language faced social, political, and economic challenges (Mir, 2010). Also, Murphy (2018) noted during the era of united India under Mughal rule, Persian was promoted as the language of power, limiting opportunities for Punjabi to flourish. Urdu also held a helpful position due to its linguistic proximity to Persian, mutual intelligibility with Hindi, and semantic affinity with Punjabi. In contemporary Pakistan, Punjabi is still a spoken language for millions but lacks institutional support. It is not taught as the primary medium of instruction in schools, nor is it widely used in government administration or media (Jabeen, 2023). Instead, Urdu and English dominate these spaces, positioning Punjabi as a less prestigious language in formal and professional domains. Many Punjabi speakers, particularly in urban areas, often raise their children speaking Urdu, perceiving it as more helpful for education and social mobility. According to Mirza et al. (2024) this linguistic marginalization is most evident in the educational system, where Urdu and English are prioritized. Punjabi is not taught as a compulsory subject in most schools, and there is little emphasis on developing written skills in the language. This lack of formal instruction in Punjabi has contributed to a decline in its written use, even though the language has a rich literary tradition that dates back centuries. Despite

its marginalization in formal settings, Punjabi retains a vibrant presence in the cultural landscape of Pakistan (Farooq, 2023). It is the language of choice for most of Pakistan's folk music, poetry, and oral traditions. The works of classical Punjabi poets like Bulleh Shah and Waris Shah remain widely celebrated, and modern Punjabi music continues to thrive, especially in the form of folk and popular music genres. In rural areas of Punjab, Punjabi remains the primary language of communication, with deep ties to local customs, traditions, and cultural expressions. Also, Khaira (2020) argued that Punjabi is also prominent in the Sufi tradition, where spiritual poetry and hymns in Punjabi are central to religious practices. The use of Punjabi in Sufi shrines and gatherings reinforces its role in Pakistan's religious and cultural identity, particularly in rural Punjab. In addition, the Punjabi diaspora, especially in countries like Canada, the United Kingdom, and the United States, has played an essential role in preserving and promoting the language through literature, music, and media. While the language faces challenges in Pakistan itself, the global diaspora continues to create new avenues for its expression and preservation (Mucina, 2024). To be precise, despite their relatively substantial number of speakers, both languages face challenges such as limited representation in digital media, educational resources, and global awareness.

3 Wikipedia as a language preservation platform

Wikipedia is unquestionably one of the most popular open-access resources available on the internet, constantly ranking among the top ten most visited websites. According to its data, it received almost twenty-five billion hits in April 2024 alone. Wikipedia serves as the flagship for a series of sister projects, including Wiktionary (a dictionary and thesaurus), Wikinews (free-content news), Wikiquote (an anthology of quotations), Wikibooks (free textbooks and manuals), Wikispecies (a directory of species), Wikisource (a free-content library), and Wikiversity (free learning materials and activities), along with several independent Wiki-based spin-offs (Baxter, 2009). Scholars noted that the defining feature of all these Wiki-based enterprises is their free content approach, which allows anybody to add and edit articles. Wikipedia follows a set of core rules as well as more restrictions governing article naming, category, and content. Users are subject to varied degrees of compliance with these standards, which include restrictions against copyright infringement, defamation, and failing to maintain a balanced point of view (Borges, 2022). Wikipedia is multilingual, with over sixty-two million articles in more than 329 languages, where English is at the top with more than six million articles (stats as of April 2024). Notably, several languages with over a thousand pages are not major world languages, proving Wikipedia's dedication to promoting linguistic variety. The rising availability of more affordable and quicker internet connections in many but not all parts of the world has boosted the popularity of online resources such as Wikipedia. This platform attracts millions of visitors and contributors worldwide. In areas where the technological infrastructure allows affordable and dependable internet access, Wikipedia provides a unique opportunity for indigenous languages, which are frequently regarded as less economically feasible (Vasconcelos et al., 2024). By overcoming the financial barriers associated with publishing, Wikipedia can aid serve the modern

indigenous language user community. Furthermore, where internet access is widespread and inexpensive, online open-access resources can serve as a valuable new tool for the preservation and advancement of linguistic diversity, preventing indigenous languages from being suffocated by larger ones due to cost-effectiveness and speaker count. Paradoxically, whereas other major languages may feel overwhelmed by English, indigenous languages might thrive on the Internet (Paolillo et al., 2005).

4 Justification of selected languages

Setswana and Punjabi face significant challenges about digital representation and resource availability. While Setswana is spoken by a substantial population in Southern Africa and Botswana, and Punjabi is widely spoken in Pakistan and India, both languages are underrepresented in digital contexts (Minhas and Salawu, 2024). This study highlights that Wikipedia can support the preservation and revitalization of these languages, thereby contributing to their sustained use and cultural preservation. Wikipedia presents a great opportunity for the preservation and promotion of indigenous languages. Wikipedia's open-access and collaborative architecture enables the development and distribution of information in a variety of languages, including ones which need to be preserved (Al-Khmisay et al., 2023). Research shows that there is a paucity of extensive research on how digital platforms such as Wikipedia can help with language preservation. This study empirical data and analysis on the effectiveness of Wikipedia as a tool for supporting and promoting Setswana and Punjabi, enriching academic discourse in linguistics, digital humanities, and social sciences. Preserving indigenous languages like Setswana and Punjabi is critical to preserving cultural diversity and social cohesiveness. By capturing and promoting these languages on Wikipedia, this study proves how digital platforms can help to pass down cultural heritage to future generations. This has broader social effects, increasing awareness of linguistic and cultural diversity. The study findings can help policymakers, educators, and community leaders identify effective measures to support indigenous languages. The insights collected can be used to influence the development of policies and projects that use digital tools for language revitalization, which have a tangible impact on language preservation efforts. By understanding how Wikipedia is used to preserve Setswana and Punjabi, this study provides practical recommendations for Wikipedia contributors, language activists, and developers of digital platforms. Identifying best practices and challenges will help improve the efficacy of online language preservation initiatives.

5 Literature review

Scholars have conducted extensive research on Wikipedia and language preservation, and other aspects of the platform, like content creation, community engagement, and Wikipedia for learning and education. Previous studies have provided a valuable understanding of various aspects of Wikipedia, (Miquel-Ribé and Laniado, 2019; Kristiani, 2021) studied indigenous knowledge production and the cultural gap of 300 languages on Wikipedia, (Shen et al., 2017; Das et al., 2024) studied quality assessment of Wikipedia articles and dynamics of Wikipedia articles on Brazilian indigenous languages,

(Baigutanova et al., 2023) researched referencing quality on Wikipedia while (Rama et al., 2022) studied users' interactions with images on Wikipedia. Additionally, scholars also focus on the oral citations on Wikipedia and noted that it is a viable addition to Wikipedia (Gallert et al., 2016), also research has focused on understanding readers' preferences (Lehmann et al., 2014; Singer et al., 2017).

Other studies have identified underrepresented languages (Graells-Garrido et al., 2015; Hoenen and Rahn, 2021; Mandiberg, 2023) and investigated the asymmetry in coverage across various language editions (Ford et al., 2015; Lemmerich et al., 2019; Roy et al., 2020; Ashrafimoghari, 2023) and noted that despite being the largest Wikipedia with the most editors, articles in English Wikipedia frequently lack critical details available in other language editions. Furthermore, half of the items evaluated in non-English Wikipedia lacked matching articles on English Wikipedia. When examining overlapping articles in different languages, significant content discrepancies were discovered. Therefore, this study finds the gap and aims to focus on indigenous language preservation specifically, Setswana and Punjabi languages.

6 Theoretical framework

The Ethnolinguistic Vitality Theory (EVT) (Giles et al., 1977) provides a vigorous framework for understanding the factors influencing the preservation and promotion of Setswana and Punjabi through Wikipedia. This theory describes how a language effort to evade the domination of other languages. By examining the interplay of status, demographic, and institutional support factors, this study aims to illuminate that digital platforms can effectively enhance the vitality of indigenous languages. According to Jamallullail and Nordin (2023), Giles et al. (1977) proposed the Ethnolinguistic Vitality Theory, which holds that an ethnolinguistic group's vitality dictates its ability to behave as a separate and active collective in intergroup settings. The theory outlines three key variables that influence a group's language vitality: status, demography, and institutional support.

Similarly, Grenoble et al. (2023) noted that a language group's economic resources can substantially impact its vitality. Setswana and Punjabi speakers must investigate the economic ramifications of language use in numerous sectors, including digital platforms. Also, a language's vitality is influenced by its status and acknowledgement in the larger community. This study investigates whether Wikipedia can improve the social standing of Setswana and Punjabi by increasing their visibility and usage. Bromham (2023) studied the demographic factor of language vitality and argued that the number of speakers and their geographic distribution are critical for language preservation. This study finds that Wikipedia can reach distant Setswana and Punjabi speakers, potentially boosting language cohesiveness. The age distribution of speakers and intergenerational transmission is critical to language vitality therefore Wikipedia's contribution to engaging younger generations and aiding language learning and use is evaluated. Community activities and grassroots movements are also critical to preserving language use. Wikipedia's ability to organize community activities and foster a collaborative environment for language documentation and promotion will be investigated (Ren et al., 2023). Precisely, Wikipedia, as a digital and collaborative platform, offers unique opportunities for enhancing the ethnolinguistic vitality of Setswana and Punjabi. By providing an accessible and

widely used platform for content creation and dissemination, Wikipedia influences the status, demographic dynamics, and institutional support of these languages.

7 Methodology

This study employs a quanti-qualitative research design (Neri et al., 2024) for robust data analysis, integrated metrics data with in-depth content analysis. This approach helps to comprehensively understand and analyse Wikipedia's contribution to the preservation and promotion of Setswana and Punjabi languages. According to Creswell and Creswell (2017) the quanti-qualitative research method is useful in improving the findings accuracy and guiding second sources of data gathering, leading to a broad and deep understanding of the phenomenon under research.

The quantitative data was collected from Wikipedia's statistical pages (May 2022 to May 2024), including the number of articles, edits, active users, active editors and views for Setswana and Punjabi Wikipedia editions. Metrics from other language editions are also collected for comparison to understand the relative position of Setswana and Punjabi in the broader Wikipedia ecosystem. Further, articles were assessed for quality using automated tools ORES (Objective Revision Evaluation Service) (TeBlunthuis, 2021) to provide a quantitative measure of article completeness, accuracy, and reliability. Meanwhile, qualitative data was collected by analyzing the content of the articles and different themes were developed using thematic analysis of the content of the articles in both languages. The data is available to the public on the Wikipedia website, but for ethical reasons, the personal information of the editors has been kept confidential.

8 Results and discussion

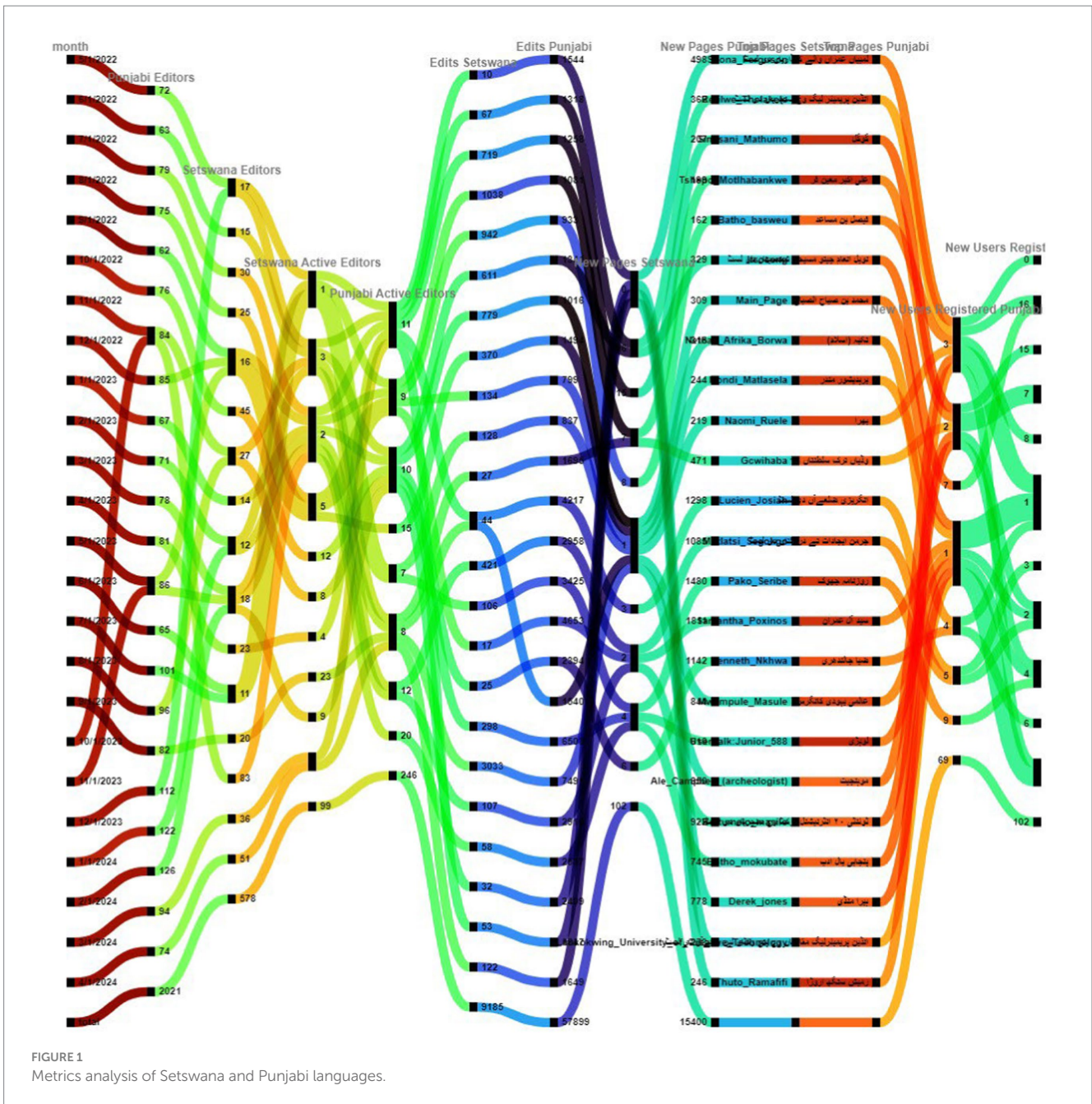
Given the quanti-qualitative nature of research, the findings are discussed below in two sections, first quantitative aspect of the data was presented and discussed, and then the qualitative data was presented through thematic aggregation.

8.1 Quantitative analysis

The metrics analysis of Setswana and Punjabi languages, shown in Figure 1 and Table 1, visualizes the number of editors, active editors, edits, new pages, new users registered and top edited pages for both languages from May 2022 to May 2024. The visualization highlights trends and fluctuations in editor activity over this period, providing insights into engagement levels for both languages. The number of Punjabi editors follows a stable pattern with some fluctuations. Peaks were observed in July 2023 (101 editors) and January 2024 (122 editors), with the highest number of editors (126) recorded in February 2024. There are noticeable dips in September 2022 (62 editors) and June 2023 (65 editors), indicating periods of lower engagement for the Punjabi language on Wikipedia. The number of Setswana editors fluctuates more compared to Punjabi editors. There was a significant increase in October 2023, with 83 editors. This spike in October 2023 indicates an event or intervention that led to the

boost in editor numbers. Generally, the Setswana editor numbers are lower with occasional spikes, suggesting less consistent engagement. The number of active Setswana editors is relatively low, but there is significant variation, especially with a spike in September 2023, with 23 active editors. This spike in September 2023 is notable and suggests a concentrated effort or event that led to increased activity. Otherwise, the numbers remain low, reflecting limited active engagement (Schroer and Hertel, 2009). The number of active Punjabi editors has remained relatively stable with minor fluctuations. The level of activity seems to be more consistent compared to the active editors of Setswana. There were peaks in December 2022 (15 active editors) and January 2024 (20 active editors), indicating periods of increased activity. The overall stability suggests a steady core of active editors who contribute regularly. The data indicates that Punjabi editors, and active editors are more consistent compared to their Setswana counterparts. This suggests that the Punjabi community is more established and engaged. The spikes in Setswana editor numbers, especially in October 2023, and active editor numbers in September 2023, due to specific events or campaigns that temporarily increase engagement. Investigating these periods could reveal strategies for maintaining higher engagement. The relatively low and fluctuating numbers for Setswana's active editors indicate difficulties in maintaining regular contributions. The data also shows that only 4 Punjabi editors contributed more than 1,200 edits, however, Setswana's top four editors only contributed 130 edits for the same period. It is beneficial to develop strategies to convert occasional editors into active contributors. The stable trends in Punjabi suggest a sustainable model for community engagement and language preservation (Jarusawat et al., 2018). Identifying the factors contributing to this stability could provide insights for other language communities.

Similarly, the number of edits in Punjabi Wikipedia exhibits significant variability, with major peaks in March, May, and October 2023, while the number of edits in Setswana Wikipedia is generally lower, with minor peaks in mid-2023 and significant activity in October 2023 and March 2024. These peaks overlap with specific editing drives, cultural events, or academic semesters when contributors are more active. Precisely, the high number of edits suggests a vibrant and active community, consistently adding and updating content. Accordingly, the lower number of edits corresponds to the smaller editor base. Also, the number of new pages created in Punjabi Wikipedia varies, with substantial increases in May, July, and October 2023. These increases indicate focused campaigns to expand content, such as thematic months or contests (Miquel-Ribé and Laniado, 2021). New page creation is a positive indicator of the growth and diversification of the content available in Punjabi. However, the number of new pages in Setswana Wikipedia is exceptionally low, with sporadic increases and many months with no new pages. This indicates challenges in content creation, possibly due to fewer contributors or less motivation and support for creating new articles. Efforts to increase the number of new pages could focus on community building and providing training for potential editors. Similarly, the number of new users registered in Punjabi Wikipedia is relatively stable, with occasional spikes in July, October 2023, and February 2024. These spikes may be due to outreach efforts, educational programs, or viral social media campaigns encouraging new registrations. Consistent registration of new users is crucial for sustaining the community and ensuring continuous content contribution (Chen et al., 2023). However, as expected the number of



new users registered in Setswana Wikipedia is generally low, with slight increases in June 2022, July, and August 2023. This reflects the challenges in attracting new contributors, likely due to limited awareness or incentives. Increasing registration rates could involve partnerships with educational institutions and local organizations to promote Wikipedia editing. The data indicates that Punjabi Wikipedia shows greater overall activity (Amina et al., 2021) compared to Setswana Wikipedia. This is evident in the higher numbers of editors, active editors, edits, new pages, and new users registered. Specific periods of increased activity in Punjabi Wikipedia suggest successful engagement strategies, while Setswana Wikipedia's lower and sporadic activity highlights the need for targeted initiatives to boost participation and content creation. The study's objective "analyze how Wikipedia contributes to preserving and revitalizing Setswana and

Punjabi languages" achieved as the data shows that Wikipedia can preserve and promote indigenous languages. However, the challenges like lack of digital literacy and language standardization are there for both languages.

Also, for Punjabi, the total page views during this period fluctuated while peaks were noticeable in March 2023 (6,123,342 views) and April 2023 (6,483,774 views) and the highest recorded view count is in June 2023 with 7,195,667 views. The stats show that the United States consistently shows the highest number of views, peaking at 222,000 in April 2022. Canada had a significant spike of 26,000 views in May 2022, but generally lower than the US. While India showed notable views in June 2022 with 16,000 views, reflecting the language's regional importance. The most critical point is that Pakistan is not one of the top visitors, which shows that Pakistani citizens

TABLE 1 Quanti-metrics of Setswana and Punjabi languages.

Metric	Setswana Wikipedia	Punjabi Wikipedia
Total editors (May 2022–May 2024)	Fluctuates, peaks at 83 in October 2023	More stable, peaks at 126 in February 2024
Active editors	Generally low, peaks at 23 in September 2023	Stable, peaks at 20 in January 2024
Edits (total)	Peaks in October 2023 and March 2024 (298, 44)	Peaks in March, May, October 2023 (1,695, 2,958, 6,503)
New PAGES CREATED	Sporadic increases, peak of 16 in October 2023	Major increases in May (1,085), July (1,811), and October 2023 (610)
New users registered	Generally low, peaks of 4 in October 2023	Stable with peaks in July (16), October 2023 (4), and February 2024 (4)
Top edited pages	<ul style="list-style-type: none"> - Shona Ferguson (10 edits) - Refilwe Tholakele (7 edits) - Simisani Mathumo (6 edits) 	<ul style="list-style-type: none"> - “تسول ید ایراڈھک ےل او رارم ع لایبمل” (358 edits) - “تسول ید ایرا چنس چو گیل ریئیم یرپ نیقنا” (11 edits) - “لگوگ” (8 edits)
Translation completion rate	10% overall, 83% in the “ISA” group	16% overall, 85% in “Lingua Libre SignIt”
Translation review rate	11% reviewed	1% reviewed
Top countries for views	Hong Kong, South Africa, Botswana	US, Canada, India, Germany, UK
Total page views	Continuous rise, peaks at 290,817 in May 2024	Fluctuates, peaks at 7,195,667 in June 2023

setting in Pakistan are less interested in Punjabi content. Precisely, European countries such as the Netherlands, Germany, France, and the UK also contribute to the viewership but at lower levels. Countries like Bulgaria (4,000 views in February 2023) and Poland (3,000 views in April 2023) have sporadic but noticeable contributions. It is noted that Countries like the US, Canada, and India show consistent engagement, indicating strong Punjabi communities or interest in these regions. Some smaller or unexpected countries like Bhutan, Brunei, and Panama show minimal but consistent views, which may reflect smaller diaspora communities or academic interests. The data highlights the global reach and varied interest in Punjabi Wikipedia, with significant engagement from the US and periodic spikes in other countries (Ahmed and Poulter, 2023). This suggests a broad and diverse audience base for Punjabi content. The peaks in views could be tied to specific events, cultural festivals, or academic activities that drive interest. Further qualitative analysis could identify the specific causes for these spikes and help target content better to sustain and grow viewership.

Meanwhile, the total page views for the Setswana language show a continuous rise, starting from April 2022 with 177,979 views, with record-high views in May 2024 with 290,817 views. The rise was noticed throughout the year. There is variability in total page views, with notable peaks in specific months such as October 2022 and January 2023. Overall, the page views show an increasing trend, albeit with more fluctuations compared to the Punjabi Wikipedia data. The data shows that Hong Kong and South Africa exhibit the highest view counts for the Setswana language on Wikipedia. United States, Botswana, Canada, and Germany also display high view counts but not like Hong Kong and South Africa. Also, a wide range of countries with smaller view counts (1,000 views) suggests global but sporadic engagement. The fluctuating page views suggest that global events or specific interests might drive sporadic spikes in engagement (Suh et al., 2009). The wide range of countries with small but consistent view counts indicates a broad international interest, though not as concentrated as in the Punjabi Wikipedia data. The data highlights the potential for increasing engagement through localized content and outreach, especially in countries with sporadic but notable view counts (Vinck and Pham, 2020). Interestingly, about the traffic on pages, the Wikipedia statistics for both languages show that most of

the traffic comes from mobile devices, indicating that users prefer accessing Wikipedia on the go. This is consistent with global trends, where mobile usage often surpasses desktop access. By these records, the study objectives “Evaluate the effectiveness of Wikipedia as a tool for indigenous language promotion and education,” and “Explore the opportunities and challenges faced by Setswana and Punjabi languages on Wikipedia” are achieved as Wikipedia seems very effective in the preservation and promotion of these languages. The engagement from different countries around the world and user’s interest in the articles shows that this platform provides an opportunity to preserve and promote, meanwhile, the below stats about the translation tool of Wikipedia is also effective in the promotion of indigenous languages.

The translation statistics for Punjabi and Tswana highlight distinct differences in their translation progress, completion rates, and review activities. The data shows that Punjabi has a higher overall completion rate of 16% compared to Tswana’s 10%. This indicates a more significant portion of the messages for Punjabi have been translated. Notably, Punjabi’s highest completion rates are seen in the “Lingua Libre SignIt” group at 85%, followed by “OpenHistoricalMap” at 70%, and “Wikipedia Preview” at 84%. These groups show focused translation efforts, contributing significantly to the overall completion rate. On the other hand, Setswana demonstrates a stronger review process, with 11% of messages reviewed compared to only 1% for Punjabi. This indicates that while Setswana has fewer translations completed, there is a higher level of quality control and validation for the translations that are done. Setswana’s highest completion rates are seen in the “ISA” group at 83%, followed by “Kiwix” at 57%, showing substantial progress in specific areas. Additionally, the “Glossary pages” group in Setswana has a completion rate of 10% and a high review rate of 89%, reflecting thorough translation and review processes. Both languages show minimal outdated translations, at 1%, indicating that most of their translations remain current and relevant. This aspect is crucial for maintaining the accuracy and usability of the translations over time (Das et al., 2024). Despite these high-performing groups, a huge portion of messages remains untranslated for both languages. Punjabi has 60,186 untranslated messages out of a total of 71,979, whereas Setswana has 16,910 untranslated messages out of 18,802. This highlights the substantial work still required to improve translation coverage for both languages. Punjabi shows a

higher overall translation completion rate, and Setswana excels in the review and validation process (Banerjee et al., 2023), ensuring higher quality for the completed translations. Both languages have areas of strength but also significant gaps that need to be addressed to enhance their translation efforts fully.

The analysis of the top-edited pages in Setswana and Punjabi shows that the page named Shona Ferguson leads with ten edits, followed by Refilwe Tholakele with seven edits. Other notable pages include Simisani Mathumo, Tshepo Motlhabankwe, and Batho basweu, each with six edits. The Main Page and various individual names like Itsoseng and Naomi Ruele are also among the top edited. In contrast, the Punjabi top edits list is dominated by page “راى بمل” (شسل ىد اى راڏهڪ ٺلاو نارم ع نيڏنا) with 358 edits. Other pages like “شسل ىد اى رچنس ڇو گيل رى ىم ىرپ” and “لگ گوگ” have significantly fewer edits, with 11 and 8, respectively. Pages on Ali Akbar Moenfar, Faisal bin Musaid, and Nobel Prize-winning Christians each have seven edits, showing a focus on diverse topics ranging from individuals to thematic lists. The data reveals a substantial difference in the focus and volume of edits between the two languages, with Punjabi edits being more concentrated on fewer pages, whereas Setswana edits are more evenly distributed across various pages. This indicates differing user interests and engagement levels in editing content for each language.

8.2 Qualitative analysis

The analysis of the quantitative data just discussed naturally leads to the qualitative aspect of our investigation, the top edited and new pages were analysed, and the below themes were developed.

8.2.1 Renowned personalities

The results of the top edited and new pages in both Setswana and Punjabi Wikipedia show a significant focus on prominent personalities and notable figures within their respective cultures. For Setswana, the top edited pages include entries on individuals like Ferguson, Refilwe Tholakele and Simisani Mathumo, these edits are reflective of the interest in documenting and updating information about influential figures in the Setswana-speaking community. In contrast, the top edited pages in Punjabi Wikipedia include entries such as “راى بمل” (شسل ىد اى راڏهڪ ٺلاو نارم ع ىل ع) (List of Oldest Living People), “رف ني عم ربڪا حابصلا” (حابص نب دمحم) (Ali Akbar Moenfar) and “مهد بن صباح السابح” (Muhd bin Sabah Alsbah), this indicates a broader focus that includes both prominent personalities both locally and internationally. When comparing the two, it is evident that both languages emphasize the importance of recording and maintaining information about influential personalities. However, Punjabi Wikipedia also shows a broader inclusion of international personalities, indicating a slightly wider scope in terms of content focus. This reflects different editorial priorities and perhaps varying user interests in each linguistic community (Khatri and Shaw, 2022). Overall, both Setswana and Punjabi Wikipedias are actively engaged in documenting the achievements and histories of prominent figures, although Punjabi Wikipedia includes a mix of notable lists alongside individual profiles.

8.2.2 Events

The top edited and new pages in both Setswana and Punjabi Wikipedias also cover various noteworthy events, reflecting the

diverse interests of their respective communities. In Setswana Wikipedia, notable events include updates on “Netball Afrika Borwa” (Netball South Africa) and “Gcwhihaba,” a famous cave system in Botswana, “Lebala la metshameko la University ya Botswana,” Pertains to events at the University of Botswana’s sports field, including sports matches and athletic events. “Motlakase Power Dynamos” this page is about events involving the Motlakase Power Dynamos football club, such as matches, tournaments, and team achievements. “Carrot cake” While not an event in itself, but this page is about recipes and includes discussions on culinary events or festivals where carrot cake is featured. Meanwhile, in Punjabi Wikipedia, the study found different events pages like “ىڙمول” which is related to the festival of Lohri, a popular winter Punjabi folk festival celebrated primarily by Sikhs and Hindus from the Punjab region. “شکرک لئش ىر نر نثنا ۲۰ ىڙنىوٺ” (List of Centuries in Twenty20 International Cricket), this page lists notable centuries scored in T20 international cricket matches, focusing on individual performances and records. “ىڙم اري” (Heera Mandi), this page provides details on the historical and cultural significance of Heera Mandi, a famous red-light district in Lahore, including its events and transformations over time. “شسل ىد اى لاء ىل او نىل راى لباقم گىل رى ىم ىرپ” (List of Five-Wicket Hauls in Indian Premier League Matches), this page is related to cricket matches in the IPL where players have taken five-wicket hauls, showcasing significant sporting events (Pentzold et al., 2017). Similarly, “اوانه پ ىباجنپ” (Punjabi Attire), the page describes traditional Punjabi attire, often discussed in the context of cultural festivals, weddings, and other ceremonial events. Punjabi Wikipedia shows a pronounced interest in sports events and historical records, highlighting achievements in cricket and human longevity. Setswana Wikipedia, on the other hand, emphasizes national sports, natural heritage sites, and local events, reflecting a blend of cultural, natural, and sports interests (Marwick and Smith, 2021). The content in Punjabi Wikipedia is more geared toward records and lists, such as sports statistics and notable personalities, which appeal to a broader audience interested in data and achievements. In contrast, Setswana Wikipedia content is more localized, focusing on specific places and national events that hold cultural significance. Precisely, comparing both languages, the Setswana and Punjabi pages show a rich diversity in topics, including noteworthy events, cultural festivals, and notable achievements in sports and academia. Setswana pages often focus on local sports and educational institutions, while Punjabi pages include international events and cultural traditions. Both sets of pages highlight the importance of events in shaping cultural identity and community activities (Marwick and Smith, 2021).

8.2.3 Sports

Both Setswana and Punjabi sports pages emphasize the importance of sports within their communities but differ in their focus areas. Setswana pages concentrate more on local sports personalities and facilities, highlighting the development of sports within Botswana’s educational and local contexts. For example, “Boitumelo Masilo” is about an athlete known for his achievements in track and field. “Motlakase Power Dynamos” page is about a prominent football club in Botswana, known for its significant contributions to local football. “Limkokwing University of Creative Technology,” this page also includes a focus on sports facilities and events held at the university. “Lebala la metshameko la University ya Botswana,” page discusses the sports stadium at the University of

Botswana, where various sports events are held. On the other hand, Punjabi sports pages are heavily focused on cricket, reflecting the sport's immense popularity in the region and its international significance, particularly through the Indian Premier League. For example, ریمی کرپ نیڈنا and ٹیسٹل ید ایراڈھک ےلاو رارم ع ایبمل, also میڈیٹس ٹکرک یلشیج نورا, ٹیسٹل ید ایراچنس چو گیل, and تسرف یک روٹکو چن اپ سیم ولباقم گیلری می کرپ نیڈنا. The Setswana sports pages provide insight into local heroes and community sports initiatives, fostering a sense of local pride and encouragement for future athletes. The Punjabi sports pages, with their emphasis on cricket statistics and achievements, cater to a broader audience and reflect the national and international stature of the sport in the region. Both sets of pages are crucial in documenting and promoting sports culture in their respective languages, serving as valuable resources for enthusiasts and historians alike (McDonough Dolmaya, 2017).

8.2.4 Cultural articles

Both languages on their respective Wikipedia also promoted cultural, literature and historical topics. The Setswana pages focus more on local literature, and cultural and historical aspects, providing a narrative centered on contemporary and historical figures who have made significant contributions to South Africa and Botswana's cultural and historical landscape. Some of them like, “Lebala La Motshameko Mo Botswana” This page focuses on the significance of sports stadiums in Botswana, not just as sports venues but also as cultural hubs where various events take place.

“Loungo Matlhaku,” page highlights the achievements of an influential personality in Botswana, emphasizing contributions to cultural or literary fields. Magdeline Moyengwa” this page provides details the impact of a notable individual on Botswana's cultural or historical landscape. “Mokgabo Thanda,” deflects on the cultural and possibly literary contributions of an important figure in Botswana. “Carrot Cake,” Discusses a culinary item, likely focusing on its cultural relevance and popularity within Botswana, and “Goitseone Seleka” which covers the achievements and influence of a significant personality in the cultural or literary domains. In contrast, the Punjabi pages offer a broader spectrum of cultural and historical topics, emphasizing traditional festivals, historical districts, and attire, reflecting the deep-seated cultural practices and historical significance of the region. For example, یڈمول, which also explores the cultural significance and celebrations of the traditional Punjabi festival. ارے میڈنم, chronicles the historical and cultural aspects of the famous district in Lahore. اوانہپ یداجنپ, describes traditional clothing, reflecting the cultural identity and customs of Punjab. The Punjabi content tends to have a more extensive historical scope, documenting events and practices that have shaped the cultural identity over centuries.

The Setswana pages focus on local personalities and cultural institutions, emphasizing contemporary and historical figures who have significantly shaped Botswana's cultural and literary landscape. These pages provide insights into the personal achievements of individuals and the cultural relevance of various elements within Botswana. In contrast, the Punjabi pages cover a broader range of topics, with a strong emphasis on traditional festivals, historical districts, and cultural attire, offering a rich historical context that spans centuries. The Punjabi content highlights deep-rooted cultural practices and their historical significance, showcasing a

comprehensive overview of Punjab's heritage. Both Setswana and Punjabi pages serve as valuable repositories of cultural knowledge, preserving and promoting their respective heritages. While Setswana pages highlight the impact of individual personalities and modern cultural institutions, Punjabi pages delve into historical and cultural traditions, offering a broader historical perspective. This comparison underscores the diverse ways in which different cultures document and celebrate their heritage (Santos and Cabral, 2009), with Setswana pages providing a modern, personality-driven narrative and Punjabi pages offering a traditional, historically rich account. The study's objective “Evaluate the quantity and quality of Setswana and Punjabi content on Wikipedia” is achieved as the data shows that most of the articles on these two Wikipedia editions have quality content. The results show a disparity in community engagement between Setswana and Punjabi on Wikipedia, therefore the critical analysis of the data revealed that the difference in community engagement is largely due to gaps in digital access, cultural attitudes toward technology, and institutional support. In many Setswana-speaking regions, particularly in rural Botswana, limited internet access and digital literacy hinder community participation in digital platforms. Additionally, cultural preferences for oral traditions and the lack of formal instruction in Setswana reduce its presence in educational and formal digital spaces. In contrast, Punjabi-speaking communities, especially in Pakistan and its diaspora, have better access to the internet and a stronger tradition of using technology for cultural preservation, which has fostered greater participation in Wikipedia. Moreover, historical, and political factors play a significant role. Punjabi benefits from the global diaspora, contributing to its use in online platforms, while Setswana faces marginalization in favor of English. The global Punjabi diaspora, particularly in countries like Canada and the UK, is also highly engaged in promoting the language online, whereas the Setswana diaspora is smaller and less active. These differences create a dynamic where Punjabi thrives on digital platforms through broader engagement, while Setswana struggles with lower participation due to systemic barriers and limited institutional support.

9 Broader implications

The study of Setswana and Punjabi on Wikipedia provides insights into the preservation of other indigenous languages. Many indigenous languages face challenges like limited digital access, lack of institutional support, and the dominance of global languages, particularly English. Key results from the study highlight the importance of ubiquitous internet access and digital literacy, especially in rural or economically disadvantaged areas. Policymakers must invest in digital infrastructure and implement tailored digital literacy programs to enhance engagement in these communities. Cultural attitudes toward digital platforms can create resistance in communities reliant on oral traditions. Emphasizing the integration of technology with traditional practices such as oral storytelling can encourage participation and content creation. Institutional support is crucial. While Setswana benefits from official backing in Botswana and South Africa, many indigenous languages at risk of extinction lack similar support. Policymakers should consider granting official recognition, funding educational programs, and integrating these

languages into national policies. The active engagement of the Punjabi diaspora highlights the potential for other indigenous languages to leverage their communities for online preservation. Efforts to engage smaller diasporas through social media and educational initiatives can strengthen digital contributions and support language revitalization.

10 Conclusion

This study aimed to examine the role of Wikipedia in preserving indigenous languages, with a focus on Setswana and Punjabi. The research concludes that Wikipedia is crucial for the preservation and promotion of these languages. By utilizing the Ethnolinguistic Vitality Theory (EVT) as the theoretical framework, the research examined the status, content factors, and community support contributing to the vitality of these languages on Wikipedia. Through a mixed-methods approach, combining quantitative metrics and qualitative content analysis, the study provided a comprehensive understanding of Wikipedia's impact on these languages. The objectives of this study were to assess the quantity and quality of content of Setswana and Punjabi Wikipedia editions, identify factors influencing their development, and offer recommendations for enhancing their use as tools for language preservation. Each objective was thoroughly addressed across the various chapters of this article. The quantitative analysis revealed significant differences between the Setswana and Punjabi Wikipedia editions. Punjabi Wikipedia showed higher activity levels, including more articles, edits, and user engagement, compared to Setswana Wikipedia. However, both editions face challenges related to content quality and community engagement. The application of EVT highlighted the critical role of institutional support in the vitality of these Wikipedia editions. For both languages, institutional support from educational and cultural organizations was identified as a key factor that could enhance content quality and increase user participation. Additionally, demographic factors such as the number of native speakers and their internet access levels significantly impacted Wikipedia activity. Despite the challenges of limited representation and resources, Wikipedia's open-access and collaborative nature allows for extensive community engagement and content creation. The findings reveal significant differences in the presence, engagement levels and content quality between the two languages, with Punjabi showing more consistent activity than Setswana. However, both languages benefit from the visibility and preservation opportunities provided by the platform. Qualitative content analysis demonstrated that both Wikipedia editions contain a mix of cultural, historical, and contemporary topics. For example, the Setswana articles included content on cultural events and local personalities, while the Punjabi Wikipedia featured articles on regional history, cultural practices, and notable figures. However, the depth and quality of these articles varied, highlighting the need for more comprehensive and accurate information. The study validated the applicability of EVT in the context of digital platforms such as Wikipedia. It demonstrated that the theory's components status, demographic, and institutional support are essential for understanding and improving the vitality of indigenous languages online. The findings suggest that improving these factors can significantly increase preservation efforts for any language on digital platforms. To address the identified challenges and leverage Wikipedia's potential for language preservation, the study offers

several recommendations, it recommends that collaboration with educational institutions to integrate Wikipedia editing into curricula is important also, and partnership with cultural organizations to enrich content quality and accuracy is needed. The study also recommends that for capacity building of editors and native speakers to contribute to language preservation on Wikipedia, workshops and training are needed, and there should be rewards for active contributors. Meanwhile, to improve the content quality, implementation of systematic content reviews is necessary to ensure accuracy and comprehensiveness and develop guidelines and resources to aid contributors in creating high-quality articles. Policymakers and language preservation advocates should recognize the potential of Wikipedia as a modern tool for linguistic preservation. Strategic initiatives, including funding and resource allocation for Wikipedia-related projects, can foster a supportive environment for these efforts. In addition, policies that promote digital literacy and access to the Internet among native speakers will further enhance participation and content development on Wikipedia. This study has provided a detailed examination of Wikipedia's role in preserving Setswana and Punjabi. It has highlighted both the opportunities and challenges associated with the use of digital platforms for language preservation. By addressing the key factors identified through EVT and implementing the recommended strategies, stakeholders can significantly enhance the effectiveness of Wikipedia as a tool for maintaining linguistic and cultural diversity. The study underscores the transformative potential of digital platforms in preserving indigenous languages and calls for concerted efforts to realize this potential.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SM: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdul, G., Umair, C. M., Shahid, M., Shehla, J., and Tasaddaq, H. (2020). Social Media as a Tool of Political Awareness and Mobilization: A Study of Punjab, Pakistan. *Creat. Innov. Manage.* 14, 1331–1343.
- Ahmed, W., and Poulter, M. L. (2023). Representation of non-western cultural knowledge on Wikipedia: the case of the visual arts. *Dig. Stud.* 13, 1–27. doi: 10.16995/dscn.8078
- Al-Khmisy, R., Hosman, L., and Nova, R. (2023). Curating an offline Wikipedia for schools in any language: a road map. *Int. J. Emerg. Technol. Learn. (ijET)* 18, 129–148. doi: 10.3991/ijet.v18i21.44313
- Amina, W., Warraich, N. F., and Malik, A. (2021). Usage of and learning from Wikipedia: a study of university students in Pakistan. *Global Know. Memory Commun.* 70, 282–292. doi: 10.1108/GKMC-04-2020-0042
- Ashrafimoghari, V. (2023). Detecting Cross-Lingual Information Gaps in Wikipedia. Companion Proceedings of the ACM Web Conference 2023.
- Baigutanova, A., Myung, J., Saez-Trumper, D., Chou, A.-J., Redi, M., Jung, C., et al., (2023). Longitudinal assessment of reference quality on wikipedia. Proceedings of the ACM Web Conference 2023.
- Banerjee, A., Kumar, V., Shankar, A., Jhaveri, R. H., and Banik, D. (2023). Automatic resource augmentation for machine translation in low resource language: EnIndic Corpus. *ACM Trans. Asian Low-Res. Lang. Inform. Process.*, 1–17. doi: 10.1145/3617371
- Batibo, H. M. (2016). The origin and evolution of Setswana culture: a linguistic account. *Botswana Notes Records* 48, 134–149.
- Baxter, R. N. (2009). New technologies and terminological pressure in lesser-used languages: the Breton Wikipedia, from terminology consumer to potential terminology provider. *Lang. Prob. Lang. Plan.* 33, 60–80. doi: 10.1075/lplp.33.1.04bax
- Borges, R. (2022). Sourcing data from Wikipedia for the study of language contact: the csbwiki. *Acad. J. Modern Philol.* 18, 7–22.
- Bromham, L. (2023). Language endangerment: using analytical methods from conservation biology to illuminate loss of linguistic diversity. *Cambridge Prisms Extinction* 1:e3. doi: 10.1017/ext.2022.3
- Chebanne, A. (2022). The prospect of languages in education in Botswana: a critical reflection. *Mosenodi J.* 25, 1–13.
- Chen, Y., Farzan, R., Kraut, R., YeckehZaare, I., and Zhang, A. F. (2023). Motivating experts to contribute to digital public goods: a personalized field experiment on Wikipedia. *Manag. Sci.* 1–61. doi: 10.1287/mnsc.2023.4852
- Creswell, J. W., and Creswell, J. D. (2017). Research design: qualitative, quantitative, and mixed methods approaches. Los Angeles, United States of America: Sage publications.
- Das, P., Johnson, I., Saez-Trumper, D., and Aragón, P. (2024). Language-Agnostic Modeling of Wikipedia Articles for Content Quality Assessment across Languages. New York, USA. doi: 10.48550/arXiv.2404.09764
- Davis, L. L., Sigalov, S. E., Maljković, F., and Peschanski, J. A. (2023). The Wikipedia education program as open educational practice: Global stories. Open Educational Resources in Higher Education: A Global Perspective. Singapore: Springer, 251–278.
- Farooq, M. (2023). The role of literary and cultural associations in the development of Saraiki ethnic identity in Pakistan. *J. Arts Ling. Stud.* 1, 837–860.
- Ford, H., Graham, M., and Meyer, E. (2015). Fact factories: Wikipedia and the power to represent. London, England: University of Oxford.
- Gallert, P., Winschiers-Theophilus, H., Kapuire, G. K., Stanley, C., Cabrero, D. G., and Shabangu, B. (2016). Indigenous knowledge for Wikipedia: A case study with an OvaHerero community in eastern Namibia. Proceedings of the First African Conference on Human Computer Interaction.
- Giles, H. B., Bourhis, R. Y., and Taylor, D.M. (1977). Towards a theory of language in ethnic group relations. In. ed. H. Giles. *Language, ethnicity and intergroup relations*. (London: Academic Press), 307–348.
- Graells-Garrido, E., Lalmas, M., and Menczer, F. (2015). First women, second sex: Gender bias in Wikipedia. Proceedings of the 26th ACM conference on hypertext & social media.
- Grenoble, L. A., Vinokurova, A. A., and Nesterova, E. V. (2023). Language vitality and sustainability: Minority indigenous languages in the Sakha Republic. *The Siberian World*: Routledge, 31–46.
- Guerrettaz, A. M., and Engman, M. M. (2023). Indigenous language revitalization, in *Oxford encyclopedia of race and education*. ed. P. G. Price. London: Oxford Research Encyclopedia of Education. doi: 10.1093/acrefore/9780190264093.013.559
- Gunnink, H. (2020). Language contact between Khoisan and bantu languages: the case of Setswana. *South. Afr. Ling. App. Lang. Stud.* 38, 27–45. doi: 10.2989/16073614.2020.1737158
- Gwewende, S., and Mthombeni, Z. M. (2023). Safeguarding intangible cultural heritage: exploring the synergies in the transmission of indigenous languages, dance and music practices in southern Africa. *Int. J. Herit. Stud.* 29, 398–412. doi: 10.1080/13527258.2023.2193902
- Hoenen, A., and Rahn, M. D. (2021). Migration of small and endangered languages into the Wikipedia. Proceedings of the Workshop on Computational Methods for Endangered Languages, 2
- Jabeen, S. (2023). Language planning and policy, and the medium of instruction in the multilingual Pakistan: a void to be filled. *Int. J. Multiling.* 20, 522–539. doi: 10.1080/14790718.2020.1860064
- Jamallullail, S. H., and Nordin, S. M. (2023). Ethnolinguistics vitality theory: the last stance for a language survival. *Sustain. Multilingual.* 22, 27–55. doi: 10.2478/sm-2023-0002
- Jarusawat, P., Cox, A., and Bates, J. (2018). Community participation in the management of palm leaf manuscripts as Lanna cultural material in Thailand. *J. Doc.* 74, 951–965. doi: 10.1108/JD-02-2018-0025
- Khaira, A. (2020). Exploring the impact of Sufi music for reconciliation between east and West Punjab. Alberta., Canada: University of Alberta.
- Khatri, S., and Shaw, A., S. Dasgupta and B. M. Hill (2022). The social embeddedness of peer production: a comparative qualitative analysis of three Indian language Wikipedia editions. Proceedings of the 2022 CHI conference on human factors in computing systems.
- Kristiani, I. (2021). Encouraging indigenous knowledge production for Wikipedia. *New Rev. Hypermedia Multimedia* 27, 245–259. doi: 10.1080/13614568.2021.1888320
- Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., and Kaltenbrunner, A. (2014). Reader preferences and behavior on Wikipedia. Proceedings of the 25th ACM conference on hypertext and social media. doi: 10.1145/2631775.2631805
- Lemmerich, F., Sáez-Trumper, D., West, R., and Zia, L. (2019). Why the world reads Wikipedia: Beyond English speakers. Proceedings of the twelfth ACM international conference on web search and data mining.
- Manan, S. A., and David, M. K. (2014). Mapping ecology of literacies in educational setting: the case of local mother tongues Vis-à-Vis Urdu and English languages in Pakistan. *Lang. Educ.* 28, 203–222. doi: 10.1080/09500782.2013.800550
- Mandiberg, M. (2023). Wikipedia's race and ethnicity gap and the Unverifiability of whiteness. *Social Text* 41, 21–46. doi: 10.1215/01642472-10174954
- Marwick, B., and Smith, P. (2021). World heritage sites on Wikipedia: cultural heritage activism in a context of constrained agency. *Big Data Soc.* 8:20539517211017304. doi: 10.1177/20539517211017304
- McDonough Dolmaya, J. (2017). Expanding the sum of all human knowledge: Wikipedia, translation and linguistic justice. *Translator* 23, 143–157. doi: 10.1080/13556509.2017.1321519
- Minhas, S., and Salawu, A. (2024). Preserving and promoting indigenous languages: social media analysis of Punjabi and Setswana languages. *J. Asia Afr. Stud.* doi: 10.1177/00219096241243061
- Miquel-Ribé, M., and Laniado, D. (2018). Wikipedia culture gap: quantifying content imbalances across 40 language editions. *Front. Phys.* 6:54. doi: 10.3389/fphy.2018.00054
- Miquel-Ribé, M., and Laniado, D. (2019). Wikipedia cultural diversity dataset: A complete cartography for 300 language editions. Proceedings of the International AAAI Conference on Web and Social Media, 13, 620–629.
- Miquel-Ribé, M., and Laniado, D. (2021). The Wikipedia diversity observatory: helping communities to bridge content gaps through interactive interfaces. *J. Internet Serv. Appl.* 12:10. doi: 10.1186/s13174-021-00141-y
- Mir, F. (2010). The social space of language: Vernacular culture in British colonial Punjab. London, England: University of California Press.
- Mirza, A., Hafeez, S., and Fatima, A. (2024). Analyzing the impact of English dominance on Urdu language pedagogy in Pakistani universities. *Pakistan Lang. Human. Rev.* 8, 293–306.
- Mlambo, R., and Matfunjwa, M. (2024). The use of technology to preserve indigenous languages of South Africa. *J. Lit. Crit. Comp. Ling. Lit. Stud.* 45:2007. doi: 10.4102/llt.v45i1.2007
- Mpofu, P., and Salawu, A. (2018). Re-examining the indigenous language press in Zimbabwe: towards developmental communication and language empowerment. *S. Afr. J. Afr. Lang.* 38, 293–302. doi: 10.1080/02572117.2018.1518036

- Mucina, M. K. (2024). A short history of Izzat among the Punjabi diaspora. *India Migration Report 2024*. India, Routledge, 199–211.
- Murphy, A. (2018). Writing Punjabi across borders. *South Asian History Culture* 9, 68–91. doi: 10.1080/19472498.2017.1411049
- Neri, M., Niccolini, F., and Martino, L. (2024). Organizational cybersecurity readiness in the ICT sector: a quanti-qualitative assessment. *Inform. Comp. Secur.* 32, 38–52. doi: 10.1108/ICS-05-2023-0084
- Paolillo, J., Pimienta, D., and Prado, D. (2005). Measuring linguistic diversity on the internet. Canada: UNESCO Institute for Statistics Montreal.
- Pentzold, C., Weltevrede, E., Mauri, M., Laniado, D., Kaltenbrunner, A., and Borra, E. (2017). Digging Wikipedia: the online encyclopedia as a digital cultural heritage gateway and site. *J. Comp. Cult. Heritage (JOCH)* 10, 1–19. doi: 10.1145/3012285
- Piccardi, T., Gerlach, M., Arora, A., and West, R. (2023). A large-scale characterization of how readers browse Wikipedia. *ACM Trans. Web* 17, 1–22. doi: 10.1145/3580318
- Rama, D., Piccardi, T., Redi, M., and Schifanella, R. (2022). A large scale study of reader interactions with images on Wikipedia. *EPJ Data Sci.* 11:1. doi: 10.1140/epjds/s13688-021-00312-8
- Ren, Y., Zhang, H., and Kraut, R. E. (2023). How did they build the free encyclopedia? A literature review of collaboration and coordination among Wikipedia editors. *ACM Trans. Comp. Human Interact.* 31, 1–48. doi: 10.1145/3617369
- Roy, D., Bhatia, S., and Jain, P. (2020). A topic-aligned multilingual corpus of Wikipedia articles for studying information asymmetry in low resource languages. Proceedings of the Twelfth Language Resources and Evaluation Conference.
- Santos, D., and Cabral, L. M. (2009). "GikiCLEF: Crosscultural issues in an international setting: asking non-English-centered questions to Wikipedia" in Cross language evaluation forum: working notes for CLEF 2009 (Corfu 30 Setembro-2 Outubro) springer. eds. F. Borri, A. Nardi and C. Peters (Springer).
- Schroer, J., and Hertel, G. (2009). Voluntary engagement in an open web-based encyclopedia: Wikipedians and why they do it. *Media Psychol.* 12, 96–120. doi: 10.1080/15213260802669466
- Shah, M. A., and Sahito, M. S. (2024). Cultural institutions in Pakistan: promoting cultural and National Identity. *Pakistan Lang. Human. Rev.* 8, 377–391. doi: 10.47205/phr.2024(8-II)33
- Shen, A., Qi, J., and Baldwin, T. (2017). A hybrid model for quality assessment of Wikipedia articles. Proceedings of the Australasian Language Technology Association Workshop 2017.
- Simpson, J. (2014). Teaching minority indigenous languages at universities. *Indigen. Lang.* 10, 54–58.
- Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., et al., (2017). Why we read Wikipedia. Proceedings of the 26th international conference on world wide web.
- Suh, B., Convertino, G., Chi, E. H., and Pirolli, P. (2009). The singularity is not near: slowing growth of Wikipedia. Proceedings of the 5th international symposium on wikis and open collaboration.
- TeBlunthuis, N. (2021). Measuring Wikipedia article quality in one dimension by extending ORES with ordinal regression. Proceedings of the 17th international symposium on open collaboration.
- Theledi, K., and Masote, S. (2024). Indigenous language policy in academic writing at south African higher education: the issue of publishing and accessing scientific materials in Setswana. *Af. J. Inter/Multidiscip. Stud.* 6, 1–10. doi: 10.51415/ajims.v6i1.1339
- Udoinwang, D., and Akpan, I. J. (2023). "digital transformation, social media revolution, and e-society advances in Africa: are indigenous cultural identities in danger of extinction?". *SSRN.* 19. doi: 10.2139/ssrn.4349795
- Vasconcelos, M., de Souza Mizukami, P., and Pinhanez, C. S. (2024). "Disappearing without a Trace: Coverage, Community, Quality, and Temporal Dynamics of Wikipedia Articles on Endangered Brazilian Indigenous Language." *Proceedings of the International AAAI Conference on Web and Social Media*, 18, 1531–1544. doi: 10.1609/icwsm.v18i1.31407
- Vinck, P., and Pham, P. N. (2020). Localizing outreach. *The International Criminal Court: Contemporary Challenges and Reform Proposals*. Boston: Brill Nijhoff, 65–70.