

# Exploring *African* Digital Humanities Using the Journal of the Digital Humanities Association of Southern Africa

**Johannes Sibeko**

Nelson Mandela University  
Digital Humanities Hub, Faculty of Humanities  
Gqeberha, South Africa  
johanness@mandela.ac.za

## Abstract

Digital Humanities scholarship is often framed through paradigms developed in the Global North, leaving African-specific practices and epistemologies underexplored. In this article, I use topic modelling and lexical analysis to investigate what constitutes African DH by analysing 41 Southern African DH articles. The findings indicate that the majority of publications in JDHASA engage deeply with language-related topics. The field combines advanced computational methods with a strong grounding in local languages, cultural heritage, and socio-historical realities. It also reflects responsiveness to evolving digital social realities, addressing themes such as online harm, misinformation, and affective communities. This article contributes to the theorisation of African DH by identifying thematic tendencies and methodological patterns specific to the Southern African context. It highlights the dual focus on computational innovation and cultural rootedness, offering an empirically grounded foundation for further critical engagement with what African DH is and what it can become.

## 1 Introduction

Summer schools, hackathons, conferences, workshops, ignitions, the recruitment of professors, and South African Chairs Initiatives—these are just some of the many efforts driving the development of Digital Humanities (DH) in Southern Africa. Since 2016, such initiatives have been further formalised through the founding of the Digital Humanities Association of Southern Africa and the launch of its journal, *The Journal of the Digital Humanities Association of Southern Africa* (JDHASA). While the United States, Germany, and England remain dominant contributors to global DH scholarship (Su, 2020), recent momentum in Southern Africa signals the emergence of a growing and vibrant field. Even so, questions persist—particularly from

funders, institutions, and scholars—about the relevance and local specificity of these developments. Given DH's origins in Western academic traditions, an important question emerges: how are African scholars engaging with, adapting, and contributing to the field?

Despite this momentum, there is a lack of empirical research that systematically traces thematic and collaborative trends within the Southern African DH context. Much of the current literature reflects global trajectories based on large datasets from long-established Western journals (Sula and Hill, 2019), institutional surveys (Sula et al., 2017), or subfields such as distant reading (Jänicke et al., 2015) and library roles in DH (Sula, 2013). While valuable, these studies do not address the region-specific concerns and emerging scholarly identities that are taking shape in Southern Africa.

This article analyses DH research articles published in JDHASA between 2021 and 2024 to investigate what DH means in Southern Africa. JDHASA's development is closely aligned with national initiatives, including the Department of Science, Innovation and Technology's South African Centre for Digital Language Resources (SADi-LaR), which organises DH Ignite workshops in all nine provinces. The biennial DHASA Summer Schools and DigiMethods hackathons—hosted by institutions like the University of Eswatini and Rhodes University—further support skills development and methodological innovation. Taken together, these efforts represent a concerted attempt to build DH capacity and foster a cohesive scholarly community. As such, JDHASA is not just a site of publication but also a reflection of how DH is being institutionalised, practiced, and redefined in the region.

The article is structured as follows: Section 2 introduces the central research question. Section 3 reviews relevant literature on mapping Digital Humanities scholarship. Section 4 describes the data

This work is licensed under CC BY SA 4.0. To view a copy of this license, visit

The copyright remains with the authors.



sources and analytical methods employed. Section 5 presents the key findings, organised by thematic focus and patterns of institutional participation. Section 6 concludes with reflections on the implications of the study for the theorisation of Digital Humanities in Africa.

## 2 Research Question

The central question guiding this study is: *What is African<sup>1</sup> Digital Humanities?* Differently put, one can also ask what African scholars are doing in the DH. While DH has been widely theorised within Euro-American contexts, the contours of DH as it is practiced, conceptualised, and experienced in Africa remain underexplored. This study seeks to empirically ground an understanding of African DH by analysing textual discourse specifically emerging from South African DH scholarship and practice.

The article aims to move beyond imported definitions of DH and instead uncover how African DH articulates itself in relation to local languages, cultural heritage, infrastructural realities, and social-political commitments. Ultimately, this research contributes toward a theory of African Digital Humanities that is empirically rooted, context-sensitive, and attentive to the unique epistemologies and challenges within the continent.

## 3 Related Literature

### 3.1 Methodologically-related Literature

A number of international studies have sought to map the intellectual structure, thematic evolution, and institutional networks of DH. For example, [Sula and Hill \(2019\)](#) traced the early history of the field by examining publications in two foundational journals—*Computers and the Humanities* and *Literary and Linguistic Computing*. Through an analysis of media types, author disciplines and affiliations, and teaching-related content, the study provided a data-driven account of how DH developed during its formative years. The present article adopts a similar focus on journal-based output, but narrows its scope to a single regional publication, JDHASA, in order to surface Southern African trends and research identities.

---

<sup>1</sup>I use this term with caution, recognising that Southern African DH is not a peripheral or derivative offshoot of the field. Rather than employing “African DH” in a way that implies marginality, I treat it as a situated and theoretically generative practice that contributes meaningfully to the global landscape of Digital Humanities.

Using large-scale bibliometric data, [Wang \(2018\)](#) provided an overview of DH research by examining outputs indexed in the Web of Science Core Collection. Their analysis covered language distribution, key institutions, prominent journals, and topic areas, using visualisation tools like VOSviewer and CiteSpace. Similarly, [Gao et al. \(2017\)](#) analysed citation patterns across three major DH journals to construct networks of co-cited authors, thereby revealing the intellectual structure of the field and how it has evolved over time.

In a more methodologically varied approach, [Joo et al. \(2022\)](#) employed text mining and topic modelling techniques—including keyword co-occurrence analysis, bigram analysis, and biterm topic modelling—to examine the thematic development of DH scholarship over the past decade. Their work highlights shifts in dominant research topics and introduces scalable computational methods for mapping knowledge domains, which also inform this study’s use of topic modelling via Voyant Tools.

A longitudinal perspective is offered by [Tang et al. \(2017\)](#), who investigated five dimensions of intellectual cohesion in DH over time. Their study employed co-authorship analysis to trace collaboration patterns, co-citation and bibliographic coupling to assess thematic proximity, and Social Network Analysis (SNA) to evaluate the field’s structural coherence. They also used modularity maximisation to identify key research clusters within citation networks—an approach that parallels this article’s attention to patterns of authorship and theme within JDHASA.

Lastly, [Su \(2020\)](#) examined global collaboration networks in DH by calculating descriptive and network statistics, such as author centrality and network density, to identify regional hubs and cross-national research communities.

Together, these international studies provide a solid foundation for understanding how DH can be empirically analysed. However, they are often shaped by Western infrastructural and scholarly contexts. The current article extends these methods to a Southern African context, using them to explore how DH is being practiced, conceptualised, and institutionalised within a regional publication space.

### 3.2 African Perspectives

This section reviews selected African scholarship that foregrounds regional approaches to DH. In do-

Table 1: Frequently Used Content Words in the Corpus

Word	Frequency	Word	Frequency
data	817	digital	576
language	502	african	412
south	404	languages	402
africa	362	text	357
words	283	research	272
work	269	model	268
humanities	263	texts	239
word	237	learning	236
knowledge	226	english	218
media	199	social	191
analysis	189	study	181
resources	178	number	169

ing so, it positions the specific contribution of this article: a focused analysis of how DH is emerging in Southern Africa, as reflected in authorship patterns and thematic priorities within JDHASA.

In a doctoral study, [Chadha \(2024\)](#) investigated how scholars in African Digital Humanities navigate ethical challenges when using digital visualisation tools to represent sensitive topics such as slavery. Their qualitative analysis of semi-structured interviews identified limited technical skills and concerns about ethical representation as key challenges in DH research involving visualisation ([Chadha and Bartlett, 2024](#)).

Although focused on West Africa, [Anigala \(2025\)](#) provides a valuable comparative perspective for this study’s interest in how Digital Humanities (DH) is conceptualised in Southern Africa. Anigala examines how Nigerian literature reflects hybrid postcolonial identities shaped by indigenous traditions and colonial legacies. By employing digital tools such as text mining, sentiment analysis, network mapping, and stylometry, the study illustrates how DH methods can be tailored to address locally specific literary questions. This localisation of DH, through the integration of digital and traditional interpretive approaches, aligns with the current article’s exploration of how DH is taking shape in Southern Africa.

According to [Smit and Chetty \(2014, 2018\)](#), the journal *Alternation*, hosted by the University of KwaZulu-Natal, has been engaging with DH themes since at least 2002, marking a transition from humanities computing to a broader DH focus within Southern African scholarship.

However, while these studies highlight important regional developments and methodological adaptations, they do not specifically address the evolving patterns of authorship, thematic concerns, and in-

tellectual networks within Southern African DH publications themselves. This article fills that gap by analysing the *Journal of the Digital Humanities Association of Southern Africa (JDHASA)* to establish what DH means in this context, thereby offering new insights into the regional shaping of the field.

## 4 Methodology

Digital Humanities (DH) is not yet formally recognised as a field in major citation databases, making it challenging to comprehensively identify the literature that constitutes its knowledge base ([Tang et al., 2017](#)). Against this backdrop, JDHASA serves as a valuable focal point for this study, as it is the only dedicated DH journal currently operating in Southern Africa.

As of writing, the journal has published a total of 72 papers<sup>2</sup>, including short and full-length research articles. Its issues comprise a mixture of general submissions, proceedings from DHASA conferences, and contributions from the Resources for African Indigenous Languages (RAIL) workshops.

### 4.1 Corpus Description

For this article, I excluded issues that included RAIL contributions to avoid overrepresenting language resource research, which, although valuable, is not always explicitly grounded in DH. I also excluded abstract-based issues, such as volumes 1 and 2 of the journal. Furthermore, only articles presenting original research were included in the corpus; editorials, conference reports, and other non-research contributions were excluded from the analysis.

In the end, I compiled a collection of 41 articles published in JDHASA between 2021 and 2024. The final corpus comprises 152,494 tokens and 15,454 unique word forms. It spans a diverse range of topics.

### 4.2 Lexical Distribution in the Corpus

An analysis of word frequency provides an initial indication of the thematic focus and lexical patterns across the corpus. Table 1 presents the most frequently used content words, offering a quantitative overview of dominant concepts in the dataset.

<sup>2</sup>See <https://upjournals.up.ac.za/index.php/dhasa> for the JDHASA’s volumes and issues

Table 2: Institutional Contributions to DH Research (2021–2024 Corpus)

South African Institutions		International Institutions	
Institution	Papers	Institution	Papers
University of Cape Town	6	University of Verona (Italy)	2
Nelson Mandela University	6	Koforidua Technical University (Ghana)	2
CSIR	5	University of Turin (Italy)	1
North-West University (NWU)	4	Fundação de Apoio à Escola Técnica (Brazil)	1
Stellenbosch University	3	Universidade Federal Rural do Rio (Brazil)	1
SADiLaR	3	University of Cambridge (UK)	1
University of KwaZulu-Natal	3	University of Leicester (UK)	1
University of Johannesburg	2	York University (Canada)	1
University of Limpopo	2	The University of Western Ontario (Canada)	1
University of South Africa	1	Indian Institute of Technology Delhi (India)	1
University of Pretoria	1	SRM University Andhra Pradesh (India)	1
University of Fort Hare	1	Jilin University (China)	1
Tshwane University of Technology	1	Polish Academy of Sciences (Poland)	1
SA Centre for Artificial Intelligence Research	1	Göttingen Centre for Digital Humanities (Germany)	1

These frequencies reflect an emphasis on digital scholarship, language technologies, and Southern African contexts. The prominence of terms such as data, language, digital, and text indicates a sustained engagement with corpus-based research, possible computational methods, and linguistic resources. Similarly, the frequent appearance of African, South, and languages suggests a regional anchoring of DH activity. Furthermore, terms such as model, learning, and analysis indicate methodological interests in machine learning and data-driven approaches, while concepts like knowledge, research, and resources align with broader academic and infrastructural concerns.

### 4.3 Topic Modeling

To identify latent thematic structures within the articles, I applied Latent Dirichlet Allocation (LDA) topic modelling using the Python scikit-learn library. Each file was read and converted to lowercase. To reduce noise, all punctuation was removed using regular expressions, ensuring that only words and whitespace remained.

The preprocessed documents were transformed into a document-term matrix using the CountVec-torizer class. English stop words were excluded to focus on meaningful content words. Terms appearing in fewer than two documents ( $\text{min\_df}=2$ ) or more than 95% of documents ( $\text{max\_df}=0.95$ ) were omitted to balance specificity and generality in the feature set.

I applied LDA with 15 topics. The LDA model was initialised with a fixed random seed ( $\text{random\_state}=42$ ) for reproducibility and fit to the document-term matrix. For each topic, the top 15

words with the highest contribution weights were extracted to aid in my semantic interpretation. The LDA model identified the dominant topic per document (article) as the one with the highest posterior probability and recorded this along with the associated confidence score.

## 5 Results

### 5.1 Authorship

I make two notable observations regarding authorship patterns in JDHASA. First, the publications surveyed in this article were written by a total of 66 authors. Some of these scholars emerge as major contributors to the JDHASA. Notably, Johannes Sibeko (Nelson Mandela University), Laurette Marais (Council for Scientific and Industrial Research (CSIR)), and Menno van Zaanen (SADiLaR) have each authored or co-authored at least three articles in the journal. Their research interests align with dominant themes in the corpus, as reflected in Table 1, including language, literacy, digital tools, and indigenous knowledge systems.

Second, I examine the institutional affiliations of these contributors. Table 2 presents the range of home institutions represented in the journal, including both South African and international universities and research organisations.

As shown, Nelson Mandela University and the University of Cape Town are the most prominent contributors, each associated with at least six papers. They are followed closely by the CSIR with five contributions, and North-West University with four. Stellenbosch University, the South African Centre for Digital Language Resources (SADi-



Table 3: Document-Topic Assignments with Confidence Scores

Article (Author)	Topic	Confidence
Asiedu et al. (2024)	1	0.885
Sibeko (2024)	1	0.867
Mahlaza and Khumalo (2024)	3	0.94
Cappellotto and Cioffi (2024)	3	0.758
Heyns and van Zaanen (2022)	3	0.999
Pannach et al. (2022)	3	0.683
Morris (2024)	3	0.603
Bam-Hutchison et al. (2023)	4	0.789
Miller and Pretorius (2023)	4	1.0
Wadhwa (2024)	4	0.881
Deliwe (2024)	4	0.999
Ogrodniczuk (2024)	4	0.999
Wilken and Marais (2024)	4	0.883
Sibeko (2022)	5	0.763
Xulu (2024)	5	0.556
Du Toit (2024)	6	0.999
Roberts (2024)	7	0.916
Charumbira and Turkel (2024)	8	0.86
Vezzani and Rebora (2024)	8	0.96
Kaganof (2023)	9	0.962
Laubscher (2022)	9	0.953
Ngcungca and Sibeko (2024)	9	0.914
Raborife et al. (2024)	9	0.932
Badenhorst and de Wet (2023)	10	1.0
Mkhwanazi and Marais (2024)	10	0.999
Nkuna et al. (2024)	10	0.801
Mak et al. (2024)	11	0.999
Ghosh (2024)	12	0.848
Hartley (2024)	12	0.999
Li and Chen (2024)	12	0.999
Bessa and Azevedo (2023)	13	0.629
Marais and Wilken (2022)	13	0.918
Mabokela Koena et al. (2024)	13	1.0
Junck (2023)	13	0.596
Setaka and Trollip (2023)	13	0.705
Van Huyssteen (2022)	13	0.999
Sibeko (2022)	14	0.999
Mabaso-Nkuna et al. (2024)	14	0.999
Makhenyane (2024)	14	0.999
Cele et al. (2024)	15	0.99
Eiselen and Van Huyssteen (2022)	15	0.999

LaR), and the University of KwaZulu-Natal each contributed to three papers. The University of Johannesburg and the University of South Africa follow with two contributions each. Notably, two international institutions, Koforidua Technical University from Ghana and the University of Verona from Italy, also contributed to at least two publications each, indicating growing global scholarly engagement with DH in Southern Africa.

## 5.2 Trending Topics in JDHASA

The thematic and confidence analyses (Tables 4 and 3) offer valuable insights into the character and trajectory of Digital Humanities research in Southern Africa as represented in JDHASA.

The prominence of topics related to language

resources, corpus linguistics, and computational modelling (for instance, Topics 3, 4, 10, and 11) signals a strong investment in developing digital linguistic infrastructure that supports the documentation, analysis, and revitalisation of African languages. This focus aligns with scholarship on regional priorities around multilingualism, language preservation, and digital inclusion, underscoring how DH here is deeply tied to linguistic and cultural sustainability (see for instance [Khumalo and Mchombo \(2018\)](#); [Ngubane and Khohliso \(2024\)](#), and [Nhongo \(2024\)](#)).

Equally significant is the sustained engagement with indigenous knowledge systems, archival digitisation, and postcolonial critique (Topics 1, 6, 8, and 12). These themes reflect Southern African DH's commitment to decolonial methodologies and the recovery of historically marginalised voices and cultural heritage. The clear focus on digitising archives such as the Bleek and Lloyd collections, and the integration of indigenous storytelling, illustrates how DH in this region serves as a tool for cultural reclamation and critical intervention, not just technological advancement.

The evolving research on social media, sentiment analysis, and digital cultural expression (Topics 9 and 13) highlights the responsiveness of Southern African DH to contemporary digital social realities. It points to a growing interest in how digital platforms shape affective communities, social discourse, and cultural identity (see, for instance [Ngcungca et al. \(2024\)](#) and [Dlamini \(2024\)](#)). This focus is increasingly important as careers and economies are being shaped by social media interactions, the datafication of everyday life, and the movement of data across platforms and publics.

Notably, as shown in Table 3, none of the publications were strongly associated with Topic 2. The near absence of this topic suggests that such creative expressions in Southern African Digital Humanities are typically embedded within broader scholarly agendas rather than treated as discrete areas of inquiry. I assume that this pattern points to a distinctively integrative approach in the region, where cultural and artistic elements are mobilised to support interdisciplinary explorations of language, history, and identity.

These findings reveal two things about DH in this region. First, the field is technologically innovative, with the participation of scholars who have computational expertise and interests. Second, the field is grounded in local epistemologies, linguistics,

tic realities, and sociohistorical contexts. In this way, I can propose that DH in this region is not simply the application of global digital tools, but a site of critical engagement. That is, it actively shapes, preserves, and challenges cultural narratives while expanding the horizons of academic inquiry through digital means.

### 5.3 N-gram Analysis

To gain further insight into the thematic emphases of the articles, I conducted an analysis of frequent multi-word expressions. Table 5 presents the results for expressions comprising three or more words. Nonetheless, key insights from the bigram analysis are also woven into the discussion that follows.

The most common bigrams—such as *South Africa* (163), *South African* (160), and *African languages* (108)—highlight a strong regional grounding. This trend aligns with the journal's stated mission to foreground Digital Humanities (DH) research in the Southern African context.

Beyond geographic markers, several n-grams reflect the corpus's methodological orientation. Terms such as *sentiment analysis* (50), *LDA topics* (42), *predictor variables* (41), and *data sets* (38) suggest the increasing adoption of computational techniques, particularly within natural language processing and machine learning. This methodological inclination is consistent with the topic modelling and lexical frequency findings presented earlier in the article.

Pedagogical concerns also emerge prominently. The appearance of *school governing bodies* (11) in Table 5, alongside bigrams such as *teaching learning* (57) and *text readability* (34), suggests sustained engagement with educational contexts, especially in the teaching and promotion of African languages.

Finally, the presence of expressions such as *hate speech detection* (17) and *social media platforms* (15) underscores the applied dimension of Southern African DH, where scholars mobilise digital tools to address pressing social issues such as online harm, misinformation, and digital publics. At the same time, expressions like *phonological conditioning rules* (16), *nine African languages* (10), and *indigenous South African* (8) signal a deep commitment to linguistic diversity, language technology, and the preservation of local knowledge systems.

## 6 Conclusion

Much of the theoretical framing of Digital Humanities has emerged from the Global North, often shaping the dominant narratives and definitions of the field. As such, there remains a need to theorise DH from within African contexts. This study responds to that need by offering a contextualised and empirically grounded account of DH as practised in Southern Africa, focusing on the discourses represented in JDHASA. By analysing thematic patterns across published scholarship, the article identifies defining features that contribute to a more situated and expansive understanding of the field.

A striking finding is the overwhelming prominence of language-focused research, spanning computational linguistics, corpus development, multilingual education, and sentiment analysis. This emphasis underscores that Southern African DH is deeply intertwined with language as both a cultural cornerstone and a practical research domain. Such a focus distinguishes the region's DH from more archival, computational, or Eurocentric models and highlights its commitment to developing digital tools responsive to local linguistic realities.

Moreover, the integration of critical perspectives—particularly decolonial methodologies, archival digitisation, and social media studies—demonstrates DH's role as a site for contesting and reshaping cultural narratives. This dual orientation towards technological innovation and critical contextualisation affirms that DH in this region is not a mere transplantation of global digital methods but an active, situated practice that both preserves heritage and addresses contemporary social challenges.

These insights contribute to theorising African DH as a dynamic interplay between language, culture, technology, and social justice, offering a vital framework for shaping future scholarship and practice across the continent. Importantly, this study's data-driven approach provides an empirical foundation for such theorisation, moving beyond imported definitions to reflect the unique epistemologies and priorities of Southern African and African DH.

Although this article is limited by corpus size and linguistic scope, it opens pathways for expanded research and greater collaboration within African DH communities. Its findings call for supportive policies that prioritise linguistic plurality, open access, and community engagement—ensuring that DH remains a tool for epis-

Table 4: Thematic Topics and Related Terms

Topic - Title	Keywords (Themes)
Topic 1 - Indigenous Storytelling, Machine Translation, and Decolonisation	storytelling, knowledge, indigenous, translations, readability, texts, data, machine, stories, decolonization, article, translation, measures, decolonizing, african
Topic 2 - Artistic Expression and Cultural Narratives	briefly, ascertain, complexity, engaged, spans, artists, songs, sure, dispersed, fluent, brand, depicts
Topic 3 -Computational Linguistics: LDA, Dictionaries, and Morphological Modelling	lda, topics, text, word, rules, transition, number, dictionaries, sequence, dictionary, phonological, words, morpheme, results, algorithm
Topic 4 -Language Resources, Archives, and Digital Knowledge Platforms	data, language, archive, knowledge, using, resources, languages, learning, information, lms, south, available, project, process, resource
Topic 5 -Text Complexity, Language Assessment, and Technology in African Contexts	grade, scores, texts, african, indigenous, correlations, significant, indicate, language, text, technology, combined, complexity
Topic 6 -Bleek and Lloyd Collection: Visual and Textual Materials in Southern Archives	lloyd, collection, bleek, materials, cape, works, content, material, visual, collections, knowledge, south, tree, african, national
Topic 7 -Hate Speech Detection and Annotated Social Media Corpora	data, hate, speech, dataset, using, analysis, section, classification, hateful, annotation, posts, random, number, detection, collection
Topic 8 -Postcolonial Literature and Reader Reception in Southern Africa	readers, southern, reviewers, literary, books, scores, language, effect, african, languages, like, literature, historical, postcolonial, figure
Topic 9 -Music, Sentiment, and Social Media in South Africa	music, south, sentiments, sentiment, performance, members, africa, social, media, tweets, comments, live, ensemble, issue
Topic 10 -Text Segmentation and Language Modelling in DH Corpora	data, text, words, models, languages, language, model, table, nchlt, news, adaptation, test, segmentation, drama, word
Topic 11 -Broadcast Speech Corpora and Evaluation of South African Languages	data, model, news, nchlt, models, table, baseline, broadcast, languages, isizulu, training, set, afrikaans, speech, test
Topic 12 -Digitisation of Historical and Oral Archives in South African Scholarship	south, african, history, relations, languages, global, historical, social, diplomatic, scholars, digitization, media, oral, records, archives
Topic 13 -Sentiment Analysis and Social Media Corpora in isiZulu and English	data, language, social, sentiment, media, analysis, corpus, spoken, articles, bias, repositories, written, south, tweets, isizulu
Topic 14 -Language Learning, Readability, and Education in African Schools	learners, languages, african, south, language, learning, teaching, english, readability, africa, education, metrics, texts, school, schools
Topic 15 -Quantitative Analysis in Heritage and Cultural Film Studies	variables, predictor, model, variable, outcome, documentary, heritage, data, film, african, cultural, word, assumption, test, statistically

temic justice and inclusive knowledge production in African contexts.

This study offers a data-driven foundation for understanding African Digital Humanities. Even so, future research is needed to deepen and broaden the theoretical articulation of the field. One promising avenue would be a comparative analysis between African Digital Humanities and its Global North counterparts (e.g., in the US, UK, and Europe). Such work could illuminate points of convergence and divergence in thematic priorities, infrastructural conditions, and epistemic commitments.

That said, I acknowledge the tension in suggesting such a comparison while having deliberately avoided it in this study. My hesitation stems from a belief that DH in Southern Africa is not isolated from global DH, nor should it always be framed through comparison. Instead, I contend that it constitutes a meaningful and generative part of the global field in its own right.

## Limitations

While [Sula and Hill \(2019\)](#) analysed a large corpus of 1,334 research articles spanning several decades, my study is based on a more recent and focused corpus of 41 papers published between 2021 and 2024 in a single journal dedicated to Digital Humanities in Southern Africa. The limited size and temporal scope of the corpus may restrict the generalisability of findings and reduce the ability to observe long-term trends. However, this narrower focus offers a valuable snapshot of current developments, priorities, and regional dynamics within an emerging DH community.

A limitation of this study is the inclusion of all articles regardless of the confidence scores produced by the LDA model. While the majority of topic assignments demonstrated high confidence ( $\geq 0.80$ ), a small number of articles were assigned to topics with lower scores. The confidence score distribution is as follows:  $\geq 0.90$  — 26 articles (high con-

Table 5: Most Frequent Multi-word Expressions in the Corpus

Multi-word Expression	Count
south african languages	21
hate speech detection	17
phonological conditioning rules	16
south african context	15
social media platforms	15
test data sets	11
nchlt text corpora	11
statistically significant effect	11
school governing body	11
lwazi tts text corpus	10
nine african languages	10
data harvesting procedure	9
use african languages	9
social media data	9
phonological conditioning rule	9
natural language processing	8
news adaptation data	8
san tsf khoen digital archive	8
indigenous south african	8

fidence), 0.70–0.89 — 6 articles (moderate-high), 0.60–0.69 — 4 articles (borderline), and < 0.60 — 2 articles (Topic 5: 0.556; Topic 13: 0.596). Lower confidence values may reflect weaker thematic cohesion, which could reduce the precision of topic classification. Nevertheless, all articles were retained to preserve the representativeness and thematic diversity of the corpus. Topic interpretations for lower-confidence articles are presented with appropriate caution.

These limitations notwithstanding, the study lays the groundwork for a theory of African Digital Humanities grounded in actual practice and discourse within the South African context.

## Acknowledgements

This research was supported by funding from the National Research Foundation (NRF) of South Africa under the Thuthuka Post PhD Track (Grant Number TTK240411213481). The views and conclusions expressed in this article are those of the author and do not necessarily reflect those of the NRF or Nelson Mandela University.

## References

Ogochuku Oritsematosan Anigala. 2025. Digital humanities: New directions in nigerian literary scholarship. *Existential: Conrescene Journal of Arts and Humanities*, 3(1).

Richard Asiedu, Michelle Stewart, and Sfundu Cele. 2024. Investigating the role of digital arts in decolonizing knowledge and promoting indigenous standpoints. *Journal of the Digital Humanities Association of Southern Africa*, 5(1).

Jaco Badenhorst and Febe de Wet. 2023. Gauging the accuracy of automatic speech data harvesting in five under-resourced languages. *Journal of the Digital Humanities Association of Southern Africa*, 4(02).

June Bam-Hutchison, Shamila Abrahams, and Eshcha Adams. 2023. Restoring african feminist indigenous knowledge in the southern african human languages technologies project: An action research case study of the san tsf khoen digital archive 2020–2022. *Journal of the Digital Humanities Association of Southern Africa*, 4(02).

Alessandra Castro Fiorini Bessa and Layla Azevedo. 2023. Reflection on fair principles and digital humanities through scientific pre-initiation projects in vocational education. *Journal of the Digital Humanities Association of Southern Africa*, 4(02).

Anna Cappellotto and Raffaele Cioffi. 2024. Towards an inclusive and accessible digital scholarly editing: A critical assessment. *Journal of the Digital Humanities Association of Southern Africa*, 5(1).

Sfundu Cele, Michelle Stewart, Richard Asiedu, and Kurt Marais. 2024. Digital archival preservation and cultural heritage: A practice-led study and preservation project of the work of late sculptor bonginkosi michael gasa. *Journal of the Digital Humanities Association of Southern Africa*, 5(1).

Kartikay Chadha. 2024. Visualization tools for african digital humanities: Scholars’ perspectives on ethics and morality. In *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*, pages 1–3.

Kartikay Chadha and Joan Bartlett. 2024. Ethical and moral evaluation of african digital humanities: Methodology and preliminary results. In *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*, pages 1–3.

Ruramisai Charumbira and William J. Turkel. 2024. A minimal computing approach to southern african language resources. *Journal of the Digital Humanities Association of Southern Africa*, 5(1).

Ayanda Deliwe. 2024. The importance of a learner management system in implementing data-driven instruction in higher education institutions. *Journal of the Digital Humanities Association of Southern Africa*, 5(1).

Gabby Sipho Dlamini. 2024. 4 public displays of affection. *Digital Technologies and Gendered Realities*, page 62.

Magdaleen Du Toit. 2024. Convening context: Reading the bleek and lloyd !xun collection digitally. *Journal of the Digital Humanities Association of Southern Africa*, 5(1).



- Roald Eiselen and Gerhard B Van Huyssteen. 2022. [Using ordinal logistic regression to analyse self-reported usage of, and attitudes towards swearwords](#). *Journal of the Digital Humanities Association of Southern Africa*, 3(01).
- Jin Gao, Oliver Duke-Williams, Simon Mahony, M Ramdarshan Bold, and Julianne Nyhan. 2017. The intellectual structure of digital humanities: An author co-citation analysis. In *Digital Humanities 2017*. Alliance of Digital Humanities Organizations (ADHO).
- Arjun Ghosh. 2024. [Recovering knowledge commons for the global south](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Thomas Hartley. 2024. [Re-discovering narratives of south african defence force servicemen through the informal digital archive](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Nuette Heyns and Menno van Zaanen. 2022. [Finding topic boundaries in literary text](#). *Journal of the Digital Humanities Association of Southern Africa*, 3(01).
- Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, Gerik Scheuermann, and 1 others. 2015. On close and distant reading in digital humanities: A survey and future challenges. *EuroVis (STARs)*, 2015:83–103.
- Soohyung Joo, Jennifer Hootman, and Marie Katsurai. 2022. Exploring the digital humanities research agenda: a text mining approach. *Journal of Documentation*, 78(4):853–870.
- Leah Junck. 2023. [‘But what for?’: On the Potentials of Acknowledging Scholastic Positionalities](#). *Journal of the Digital Humanities Association of Southern Africa*, 4(02).
- Aryan Kaganof. 2023. [The herriverse: Introducing a new kind of research method, one that is structural or even meta- insofar as it exists in the reader’s navigation of the curated space and the possible contingent connections as much as in the objects being curated; an epistemic construction therefore, that is obliquely but absolutely determined by ontologically unpredictable exchanges](#). *Journal of the Digital Humanities Association of Southern Africa*, 4(02).
- Langa Khumalo and Sam Mchombo. 2018. The role of language in human existence, education, innovation and research, and the intellectualisation of african languages. *Alternation Special Edition 38b*, pages 533–541.
- Melandri Laubscher. 2022. [Uplorc: A networked, live coding laptop orchestra based in south africa](#). *Journal of the Digital Humanities Association of Southern Africa*, 3(01).
- Chensi Li and Xi Chen. 2024. [Advocating for the digitization of the history of china-africa diplomatic relations](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Raesebe Florah Mabaso-Nkuna, Nyikelani Davis Mabasa, and Brian Valoyi. 2024. [The possibility of using african languages as media of teaching and learning in south africa](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Ronny Mabokela Koena, Mpho Primus, and Linda Mahlobo. 2024. [Uncovering media bias in eviction reporting: A comprehensive analysis utilising sentiment analysis framework and social media data](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Zola Mahlaza and Langa Khumalo. 2024. [Algorithm for assisting grammarians when extracting phonological conditioning rules for nguni languages](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Franco Mak, Avashna Govender, and Jaco Badenhorst. 2024. [Exploring asr fine-tuning on limited domain-specific data for low-resource languages](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Lukhanyo Elvis Makhenyane. 2024. [The use of augmented reality in the teaching and learning of isixhosa poetry](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Laurette Marais and Ilana Wilken. 2022. [Morphology-based investigation of differences between spoken and written isizulu](#). *Journal of the Digital Humanities Association of Southern Africa*, 3(01).
- Grant Miller and Laurette Pretorius. 2023. [Institutional repositories in the linked open data cloud](#). *Journal of the Digital Humanities Association of Southern Africa*, 4(02).
- Sthembiso Mkhwanazi and Laurette Marais. 2024. [Generation of segmented isizulu text](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Lorna Morris. 2024. [Peds for south african schools](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Nkazimlo Ngcungca and Johannes Sibeko. 2024. [Unmasking deception: An exploratory study of viewers’ attitudes towards romantic betrayal](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Nkazimlo N Ngcungca, Johannes Sibeko, and Sharon Rudman. 2024. A qualitative inquiry into the south african language identifier’s performance on youtube comments. In *Proceedings of the Fifth Workshop on Resources for African Indigenous Languages@ LREC-COLING 2024*, pages 45–54.
- Nomalungelo Ngubane and Xolani Khohliso. 2024. Intellectualisation of african indigenous language at the university of the free state and academic implications for the speakers. *Journal for Language Teaching=*

- Ijenali Yekufundzisa Lulwimi*= *Tydskrif vir Taalonderrig*, 58(1):1–16.
- Raphael Nhongo. 2024. The intellectualisation of african languages at institutions of higher learning in zimbabwe. In *Language Matters in Contemporary Zimbabwe*, pages 135–153. Routledge.
- Blessing Nkuna, Thipe I Modipa, and Simon P. Ramalepe. 2024. [Developing a code-mixed sentiment analysis dataset of xitsonga-english music reviews](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Maciej Ogrodniczuk. 2024. [Towards including south african hansard papers in the parlamint schema](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Franziska Pannach, Francois Meyer, Edgar Jembere, and Sibonelo Zamokuhle Dlamini. 2022. [Nla-post2021 1st shared task on part-of-speech tagging for nguni languages](#). *Journal of the Digital Humanities Association of Southern Africa*, 3(01).
- Mpho Raborife, Blessing Ogbuokiri, and Kehinde Aruleba. 2024. [The role of social media in xenophobic attack in south africa](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Ethan Roberts. 2024. [Automated hate speech detection in a low-resource environment](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Mmasibidi Setaka and Benito Trollip. 2023. [Resource repositories and linking resources: An exploratory study](#). *Journal of the Digital Humanities Association of Southern Africa*, 4(02).
- Johannes Sibeko. 2022. [Exploring linguistic complexity of English First Additional Language grade 12 and grade 11 exam texts](#). *Journal of the Digital Humanities Association of Southern Africa*, 4(2):01–12.
- Johannes Sibeko. 2024. [Harnessing google translations to develop a readability corpus for sesotho: An exploratory study](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Johannes A Smit and Denzil Chetty. 2014. Reimagining the humanities in the twenty-first century: Towards an interdisciplinary and collaborative ‘digital humanities’ in africa. *Alternation*, 21(2):176–206.
- Johannes A Smit and Denzil Chetty. 2018. 23/25 years of alternation, and the african digital humanities: Capacity, communication, and knowledge-power1. *Alternation Journal*, 25(1):8–30.
- Fangli Su. 2020. Cross-national digital humanities research collaborations: structure, patterns and themes. *Journal of Documentation*, 76(6):1295–1312.
- Chris Alen Sula. 2013. Digital humanities and libraries: A conceptual model. *Journal of library administration*, 53(1):10–26.
- Chris Alen Sula, Sarah E Hackney, and Phillip Cunningham. 2017. A survey of digital humanities programs. In *DH*.
- Chris Alen Sula and Heather V Hill. 2019. The early history of digital humanities: An analysis of computers and the humanities (1966–2004) and literary and linguistic computing (1986–2004). *Digital Scholarship in the Humanities*, 34(Supplement\_1):i190–i206.
- Muh-Chyun Tang, Yun Jen Cheng, and Kuang Hua Chen. 2017. A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics*, 113(2):985–1008.
- Gerhard B Van Huyssteen. 2022. [Swearing in south africa: Multidisciplinary research on language taboos](#). *Journal of the Digital Humanities Association of Southern Africa*, 3(01).
- Gabriele Vezzani and Simone Rebori. 2024. [Empathic engagement and aesthetic appreciation between readers’ ethnicity and narratives’ literary prestige](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Soni Wadhwa. 2024. [Personal archives for community building: Lessons learned from pg sindhi library](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Qing Wang. 2018. Distribution features and intellectual structures of digital humanities: A bibliometric analysis. *Journal of Documentation*, 74(1):223–246.
- Ilana Wilken and Laurette Marais. 2024. [Bringing children’s dictionaries to digital life](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).
- Zolile Celiwe Xulu. 2024. [Integrating technology in teaching african languages in south african universities: A call for digitalisation](#). *Journal of the Digital Humanities Association of Southern Africa*, 5(1).