Analyzing political formation through historical isiXhosa text analysis: Using frequency analysis to examine emerging African Nationalism in South Africa

Jonathan Schoots
LEAP
Stellenbosch University
jschoots@sun.ac.za

Abstract

This paper showcases new research avenues made possible by applying computational methods to historical isiXhosa text. I outline a method for isiXhosa computational text analysis which adapts word frequency analysis to be applied to isiXhosa texts focusing on root words. The paper showcases the value of the approach in a study of emerging political identities in early African nationalism, examining a novel dataset of isiXhosa newspapers from 1874 to 1890. The analysis shows how a shared identity of 'Blackness' (*Abantsundu* and *Abamnyama*) dynamically emerged, and follows the impact of leading intellectuals as well as African voter mobilization in shaping communal political discourse.

1 Introduction

The growing wave of digitization of African text sources and the creation of new digital African language archives creates exciting new possibilities for researchers. Materials which were once difficult to access, available only in hard copy in archives, are already more widely accessible than ever before, often in digital text format. As the pace and scope of digitization grows, such access will only increase. The availability of these materials comes alongside an academic and public demand for more knowledge about African intellectual histories and knowledge traditions. Questions and sources fit hand in hand in this exciting new opening for research. How can researchers who are interested in substantive and 'humanistic' (historical, cultural, literary, etc) questions, especially in African language contexts, seize the opportunities created by these sources?

This paper draws on a novel digital archive of historical isiXhosa text and showcases some possibilities that digital archives and computational text analysis can offer humanistic and social science scholars working in African languages. The study examines political identity formation in the period of emerging African nationalism in South Africa by looking at isiXhosa newspapers from 1874-1890. I use computational text analysis methods to explore the collective development of a shared language of political identity in the emerging 'proto-nationalist' movement in what is today the Eastern Cape of South Africa.

The paper has two goals. First, to outline the method used. I draw on frequency analysis over time, an approach which can overcome the current limitations in the Natural Language Processing (NLP) tools for isiXhosa, and remains accessible to qualitative and humanistic scholars because it operationalizes a research approach which is conceptually straightforward to non-experts in Natural Language Processing (NLP). IsiXhosa, like many other under-resourced languages, faces a lack of NLP tools which are available for English and other highly resourced languages. IsiXhosa is an agglutinative language, and the structure of isiXhosa morphology and grammar means that many NLP tools cannot be imported from other languages. In particular, tools for the accurate parsing of the subunits of compound words–including stemming or lemmatizing, parts of speech tagging, and stopword removal–cannot be imported. These 'preprocessing' steps often form the necessary basis for other more advanced NLP techniques (cf. word embeddings or topic models), which themselves are not built to deal with the range of variations created by the prefix, infix and suffix structure of agglutinative languages. Such tools are being developed by isiXhosa computational linguists (Mzamo et al., 2015; Puttkammer and Toit, 2021), but are not yet sufficiently advanced to be used for social science or humanities inquiry.

This paper engages this problem by drawing on the established method of frequency analysis (Baron et al., 2009) and highlighting its utility for analysis of isiXhosa and other agglutinative lan-

guages. Very little work has used computational text analysis on isiXhosa texts for social science or humanities research. This paper shows that by adapting a frequency analysis to the structure of isiXhosa grammar, it is possible to advance in new ways. I show that this approach can be tailored to resolve the problem of agglutinative grammar by focusing on root words. This can effectively bypass the need for NLP pre-processing tools, and thus can be applied to isiXhosa texts using currently available tools. A frequency analysis approach has added advantages for social science and humanistic scholars. It is analytically clear (simply counting the number of word occurrences) where more advanced approaches can remain an opaque 'black box' to non-experts. It is also relatively technically simple and thus has a lower technical barrier to entry. Nonetheless it creates new opportunities to analyze isiXhosa texts which have not been possible using the dominant close textual analysis approach. These features commend it to scholars interested in new social and humanistic answers.

The second goal is to showcase the analytical capacity of this approach through a specific example: an analysis of collective political concept formation in the early period of African Nationalism in South Africa. Here I examine the shifting ideas of collective identity in isiXhosa newspaper publications which shaped the debates of the nascent 'proto-nationalist' movement. I follow a 16-year period as this emerging movement developed from a series of social and political debates among African intellectuals into an increasingly interlocked set of political organizations mobilizing Africans to protect African political interests. To do so, I draw on a novel dataset of digitized isiXhosa newspapers, following *Isigidimi sama-Xosa* and *Imvo Zabantsundu*, which gave voice to the first ideas and organizations of early African nationalism.

This analysis reveals a diverse emerging ideological landscape with a surprising amount of flux in the conceptual language of the African political community. I show how different notions of Ethnicity, Nationhood, and Race ebbed and flowed as the focal conceptions of the political self. I then focus on isiXhosa speakers' own language of 'Blackness'—-*Ntsundu* and -*Mnyama*—demonstrating how frequency analysis reveals surprising new aspects of political development within this important proto-nationalist community.

2 Creating the digital text dataset

This paper draws on a novel dataset of digitized isiXhosa newspapers and books created by myself and a team of researchers at the University of Cape Town. To create this dataset, we scanned hardcopies of these newspapers held at The National Library of South Africa, and at Corry Library of Rhodes University. We then performed Optical Character Recognition to produce a digitized version of the newspaper. Every page of digitized text was checked by a team of first language isiXhosa speaking researchers to correct any errors. This produced a high-quality digital replica of the original newspapers. More information about this project, along with downloadable texts are available at https://ibali.uct.ac.za/s/isixit/.[1]

This analysis focuses on all available editions of *Isigidimi sama-Xosa* and *Imvo Zabantsundu* from 1874-1890. Details of these papers are reported in Table 1.[2] Some hard copies are missing from the archival record. Most notably, the whole year of *Isigidimi* 1886 is no longer available in hard copy. A small number of other individual editions are missing due to the inability to access hard copy scans. These missing editions seem sporadic and random and are not expected to introduce any systematic bias into the analysis.

This analysis focuses only on isiXhosa text. Since some articles written in *Imvo Zabantsundu* are in English, I identified and removed all English text to produce an isiXhosa only corpus.[3] Isigidimi and Imvo are comprised of 98.8% and 72.8% isiXhosa words respectively.

3 Methods of computational text analysis

3.1 The opportunities of 'distant' reading

This paper draws on computational text analysis techniques to examine the dynamics of emerging African political thought in ways that have not previously been possible, showing the shifting empha-

---

[1]The data is also permanently available for access and download from a repository, Isigidimi: https://doi.org/10.25375/uct.22332271 and Imvo: https://doi.org/10.25375/uct.22332268

[2]Included is the type-to-token ratio (TTR), a ratio of the number of unique words to the number of total words. This captures lexical diversity in the text. When considered for individual editions of both papers, this ratio remains very stable throughout the period ($Imvo : 0.608 \pm 0.036, Isig : 0.554 \pm 0.052$).

[3]This is done by identifying and excluding all paragraphs containing >50% English words using the English word list from the NLTK package nltk.corpus.words.

Newspaper Title	Date Range	Total Editions	Total Pages	Publication Frequency	Pgs per Edition	Total Words	Type to Token Ratio
Isigidimi sama-Xosa	Jan 1874 - Nov 1888	168	1,352	Monthly	8	1,403,237	0.156
Imvo Zabantsundu	Nov 1884 - Dec 1890	314	1,259	Weekly	4	2,081,047	0.125

Table 1: Details of digitized newspaper dataset used in this analysis

sis of a community over time as they created the foundations of a new political framework.

Recent scholarship has emphasized the importance of turning to African intellectuals and knowledge systems, an area which has historically been marginalized. Scholars of South Africa have increasingly highlighted the importance of African intellectuals who preceded the formation of the famous African National Congress (ANC) and who laid a foundation for new and creative African responses to colonialism and apartheid (Attwell, 2005; Mkhize, 2008, 2018; Mangcu, 2014; Masilela, 2013, 2014; Ndletyana, 2008; Nyamende, 2000; Odendaal, 2013; Schoots, 2020, 2021). Such scholarship has largely focused on individual intellectuals or intellectual lineages, primarily drawing on close textual analysis and biography to illuminate the larger currents of emerging African intellectual thought. This scholarship is immensely valuable in deeply examining the experiences, thoughts, and innovation of prominent African intellectuals as responded to colonialism, however this individualized approach makes it difficult to see the dynamics of a larger intellectual community.

This paper seeks to contribute a new perspective on this important tradition of research by drawing on computational text analysis. Using these tools I seek to explore two elements of intellectual transformation which are harder to see though the dominant methodological paradigm of close reading and individual biography.

First, this study emphasizes the *language use of a shared community of African writers*, and thus captures a much wider audience than close reading methods do. The African language newspapers still limit our view to African journalists and the contributions of literate African letter writers. Yet here we are able to see the shifting conceptual trends of a community much wider than the traditional focus on a few key intellectuals. This wider community was collectively instrumental in early proto-nationalist politics, and by looking at this collective

dialogue we avoid the possibility of generalizing any idiosyncratic emphases of individual leaders.

Second, this analysis *foregrounds dynamic change over time by emphasizing frequency and chronology*. Close reading methods must flatten time to some extent, as they synthesize the different writings of a person to form a fuller picture of their ideas. The analyses presented here loses this synthetic depth, yet clearly highlight larger trends in shifting attention and the influence of critical events and punctuated moments in reorienting attention and discourse. For these reasons, this paper offers a set of methods which go hand-in-hand with close reading approaches. The findings offered here only make sense when contextualized by the deeper dynamics revealed in close studies, yet this analysis is able to reveal dynamics which are surprising and unexpected by such close analyses.

3.2 Analysing collective concept formation with frequency analysis

To undertake this communal and temporal analysis of ideas, this paper adapts the the method of historical text frequency analysis (Jucker et al., 1999; Baron et al., 2009), highlighting how it is useful for isiXhosa. Frequency analysis consists simply of counting the number of words of a text. Using search pattern matching (here I use regular expressions) it is possible to define a 'root word' pattern to search and count. This approach is simple and powerful when applied to agglutinative languages like isiXhosa and can bypass the need for NLP tools which are not yet sufficiently developed for isiXhosa.

Due to the variations from prefixes, infixes and suffixes of isiXhosa agglutinative grammar, it is not possible to use words as tokens because the variation in word spelling masks the use of common terms or concepts. Stemming or lemmatizing and parts of speech (PoS) tagging are NLP approaches used to solve this challenge in well-resourced languages. However, these tools are still being developed for isiXhosa (although advances are be-

ing made (Mzamo et al., 2015; Puttkammer and Toit, 2021)) and other under-resourced languages. This paper applies one approach which can by-pass the need for these NLP tools. This is useful for isiXhosa and might also be applied to other under-resourced agglutinative languages. Despite the lack of more advanced NLP tools, I show that the approach can offer a significant new perspective, showing patterns which have been hidden in close textual analysis.

Studying frequency over time involves grouping texts into time periods, counting the frequency of selected root words in each period, and plotting these over time. In the analysis showcased below, frequency analysis offers a conceptually intuitive and technically simple way to quantify how much 'attention' is paid to different concepts over time.

Technically, I use the python coding language to group texts and count the number of occurrences of root words using regular expressions (detailed below). I use the pandas package to organize data and the plotly package to plot results. I normalize the frequency counts by reporting frequency as a percentage of all words used. This accounts for differences in the total volume of the text.[4] For this analysis I group each newspaper into months. For *Isigidimi*, which published monthly for most of its life, each month mostly represents one edition. For *Imvo*, which published weekly, each month groups mostly 4 editions.

3.3 Focusing on key root words

The key value of using frequency analysis for isiXhosa is the ability to define and examine root words over time. In practice a root word must be carefully and clearly defined and then counted for each grouped period. This comes with some limitations: first, the set of words analyzed is defined by the researcher, who must identify concepts which may be important or meaningful. It thus requires a sufficient qualitative, hermeneutic, and historical knowl-

edge of the texts and their contexts, as well as sufficient language competence. Thus, this method may be best used by humanities or social science scholars who likely already require such competencies. The second limitation is that the approach does not scale well, and is most fit to explore on the order of 10s to perhaps 100s of words. The approach must thus be guided by the researchers existing contextual knowledge and expectations and thus benefits from the support of qualitative and historical analysis which can guide the researcher's intuition.

Analysis of root words in this analysis took the following steps: 1. define the relevant focal concepts and terms, identifying the root word(s) which capture the usage of this term. This includes defining inclusions and exclusions in the search pattern to correctly select only relevant words. I used regular expression matching which offer powerful tools for pattern identification. 2. Check the words and frequencies identified by this pattern to ensure that the defined pattern collects only the desired word set. To achieve this, I output a list of the frequency counts of all unique words identified by the pattern in the whole corpus[5]. This makes it easy to check if words which do not match the concept are being included. An iterative process of defining and checking the search pattern ensures accuracy and confidence that the analysis reveals only the frequency of the desired words. A simple example: If *uhlanga* ('nation/race') is the desired concept/term, 'hlang' is the core root. It is desirable to capture many variations of this term (e.g. *yohlanga*, *bebehlanga*, etc) but other words such as *umhlanga* ('reed') or *-hlangana* ('gathering') should be excluded. The regular expression "(?<!m)hlanga\b" more clearly defines the desired word set than ".*hlang.*". Once defined these roots can be used in the approach outlined in the section above.

This approach has been developed for a specific analysis of isiXhosa texts. However, it may also be useful for other underresourced agglutinative languages which also use root words with modifications. A limitation of this paper is that this analysis has not tested the approach on any language other than isiXhosa. The grammatical similarities between isiXhosa and other Southern Africa and Bantu languages, and the structure of agglutinative

---

[4]I also check for correlation between this frequency percentage and total words as well as the type-to-token ratio at the level of newspaper edition. For the two focal words analysed below (-Ntsundu and -Mnyama) there is no statistically significant correlation between frequency percentage and total words. There is no statistically significant correlation between TTR and frequency percentage, except for a very low correlation magnitude between TTR and Mnyama in Imvo (r(311)=0.11, p=.0476) and for TTR and Ntsundu in Isigidimi (r(163)=.24, p=.002). These non-significant or low magnitude correlations suggest that changes in total text volume and lexical diversity (shown by the TTR) are not driving the changes seen in the analysis.

[5]I use a python dict, incrementing the count of each unique word matched by the pattern.

languages in general, theoretically suggest the possible extension. However, the author has not tested the extent and limitations of applications to other languages.

4 Case study: The shifting language of emerging political community

The paper turns to exemplify the utility of this approach by looking at emerging ideas of African nationalism in isiXhosa newspaper texts. The first recorded accounts of African nationalism and Pan-Africanism in South Africa began in isiXhosa intellectual communities. First appearing in print from the 1860s (Soga, 1983), these ideas developed in the 1870s in contexts like the *Isigidimi* newspaper as well as other social venues where communities of missionary educated African intellectuals developed an increasingly critical analysis of colonization and the government's legislation on Africans. In the 1880s these ideas were turned into political action: a range of the new African political organizations emerged and flourished, experimenting with various political forms and projects which laid the ground for African nationalism across Southern Africa (Odendaal, 2013, see for more).

The two papers studied here: *Isigidimi sama-Xosa* and *Imvo Zabantsundu* were central to this emerging movement. These were the first African language newspapers with an African head editor, and they both offered a mouth-piece to early African political and social organizations (Switzer and Switzer, 1979, 40–41, 45–46).

The frequency analysis approach outlined here offers a 'distant reading' perspective (see Moretti, 2013; Jänicke et al., 2015; Bode, 2017) of the early African language press, and makes visible how a community of African writers were involved in a collective process of shaping the foundations of emerging African nationalism. These texts show an unexpected fluidity of many concepts of identity that have since become fixed and taken for granted, and show how political identity was being dynamically and communally shaped in the period.

Figure 1 demonstrates this conceptual flux by looking at the *Isigidimi* newspaper, examining the changing frequency of three of the most central concepts of proto-nationalist political identity: Nationhood, Ethnicity, and Race.

Figure 1 plots the frequency of several different root words which mark different concepts of group identity (note that 1886 is missing data).
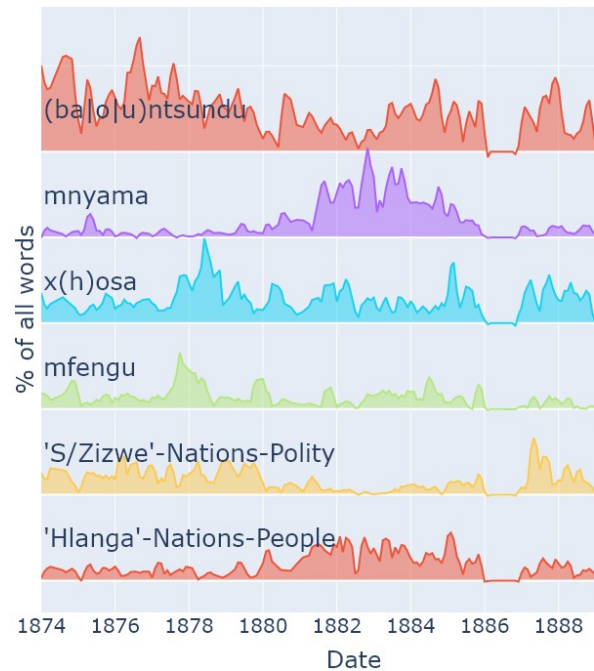

Figure 1: Frequency of identity words in Isigidimi

The first thing to note is that multiple isiXhosa terms cover what are considered in English to be unified concepts. Two root words for race are shown: *-Ntsundu*, meaning 'Dark Brown' (henceforth 'Brown') and *-Mnyama*, meaning Black. Both terms mark community based on skin colour. Yet their changing frequency reveals a great degree of conceptual transformation over the period. Ntsundu identity begins as the most frequent of all the identities plotted in Figure 2, yet falls away almost entirely by 1883, being replaced by a surge of usage of Mnyama. Yet from 1883 onward, Ntsundu returns, used alongside Mnyama for a time, and then spiking back to prominence in 1887.

This flux in 'racial' language is not unique. Ethnicity and Nationhood, two additional concepts central to emerging African nationalism, both show significant shifts in emphasis in this time period. *Xhosa* and *Mfengu*, the most politically central ethnic identities of this community, also shift in focus. Ethnic identification peaks in focus around 1877, and while both diminish, *Xhosa* identification remains more frequent than *Mfengu* identification. The figure also shows competing conceptions of Nationhood shifting through the period. The language of *Isizwe/Izizwe* (sg./pl.) ('Nation/Tribe/Kingdom') and *Uhlanga* ('Nation/People/Lineage/Race') compete for centre stage as the leading imagination of collective nationhood over the period. Both terms imply the community as a shared nation or people, but *Isizwe*

has the connotations of a polity—a territory or kingdom, where *Uhlanga* has the connotations of a people—a lineage. *Uhlanga* supersedes *Isizwe* as the focal concept in 1880, yet *Isizwe* returns strongly albeit briefly in 1887.

The shifting use of language offers a perspective on early political identification which might surprise even experts in this time period. The significant shift in frequency shows isiXhosa speaking communities collectively working with and transforming a range of conceptions of communal identity in their own languages as they worked to build a growing consciousness of their shared identity. These observations directly counteract some dominant theories of emerging nationalism. For example, here we see that the ideas Africanness, Blackness, or Nationhood were not simply an import from European nationalism, not merely a 'derivative discourse' (Chatterjee, 1986) or a 'colonization of consciousness' (Comaroff and Comaroff, 1991, 1997). Instead, these isiXhosa texts point to a real wrestle, even an internal conflict, over how African/Black/National identity should be defined within the proto-nationalist movement - a debate happening on isiXhosa, not European, terms.

5 Changing conceptions of Race in early African nationalism

To explore these shifting concepts of identity more deeply, I now focus in on the root terms -*Ntsundu*, and -*Mnyama* which offered two contrasting and competing conceptions of what today might be termed 'Black' or 'African'.

Figure 2 shows how the usage of words containing the root terms -Ntsundu and -Mnyama changed across 16 years of the emerging proto-nationalist movement.[6] Each dot here shows the percentage frequency of usage in one month of Isigidimi (orange) or Imvo (blue) with the line representing a moving weighted average. In the beginning of the period, Ntsundu identity was dominant, and its usage reached a local peak around 1877 when it began to decline. Mnyama usage remained infrequent until 1881 when it began to rapidly rise to prominence and peaked between 1883 and 1884, its peak beginning when Ntsundu reached its lowest usage late in 1882. Yet from 1884 Mnyama began to rapidly recede and Ntsundu usage dominated

[6]This analysis focuses on these terms as they are used in relation to people (in both singular and plural) by including the prefixes -ba, -o, or -u in the search.
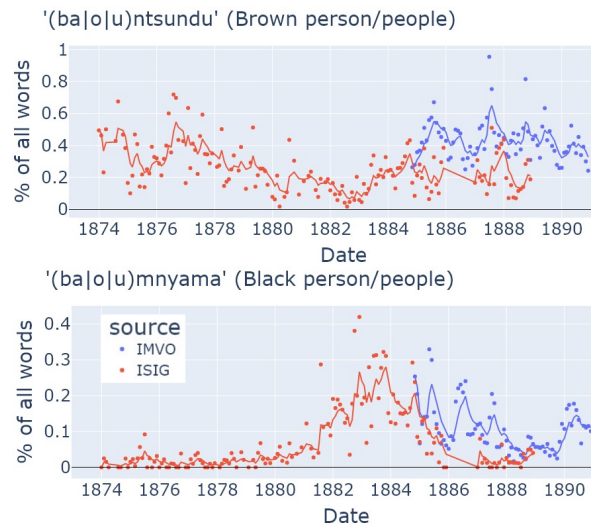

Figure 2: Frequency of Ntsundu and Mnyama

the final period, being used more frequently in the newly founded Imvo Zabantsundu paper than in Isigidimi.

These findings are surprising in light of both the historical analysis to date and present-day language usage. The expectation from the literature or from present day usage is that these terms for 'blackness' might be either equivalent and substitutable, or that Mnyama usage might be the dominant and more frequent expression of racial identity.

A brief explanation of these terms offers useful context. Some scholars, such as Ncedile Saule (2017), have argued that *Abantsundu* was the Xhosa people's own language and conception of skin colour or race ('dark brown') before colonial conceptions of Black and White (see Mzileni, 2019, 86 footnote 5). The term was also translated by missionaries and African writers to the colonial term 'native', showing that in addition to a pre-colonial resonance, the term was also connected to colonial conceptions of difference-yet ones not racially marked.[7] Mnyama more explicitly invokes a binary paring in a way Ntsundu does not: of both white-black (*mhlophe-mnyama*) and light-darkness (*ubukhanya-ubumnyama*). In this way, Mynama might more explicitly invoke the racial binaries of the colonial experience. The term might implicitly foreground a binary opposition with (*abamhlophe* ('white people') in a way that is not implicit in Abantsundu. Mnyama language may thus have more directly addressed the racial distinctions that colonists made between 'Black' and 'White'.

This discussion only begins to scratch at a se-

[7]Consider that 'native' was a global colonial term not simply connected to 'blackness' (cf. the US or New Zealand).

ries of new questions which the analysis presented in Figure 2 raises: what were the different meanings of these terms? Why were they used in such clear contrast where today they might be taken as synonyms? This paper will examine only the significant moments of linguistic shift. I leave for future analysis investigation into the rich historical and linguistic connotations of these two terms.

6 Editors, elites, and identity language

One key insight into the shaping of this discursive space becomes visible when we switch perspectives from newspapers to newspaper editors. Figure 3 now presents the four different editors who headed these papers over the period. In purple is James Stewart, the missionary editor who oversaw Isigidimi until 1876. In green is Elijah Makiwane, the first African newspaper editor who headed Isigidimi until 1881. In orange we see John Tengu Jabavu, in both his role as editor of Isigidimi from 1881 to 1884 and his shift to his own paper Imvo Zabantsundu, which he founded when he left Isigidimi in late 1884. Finally, William Wellington Gqoba, shown in blue, was the editor of Isigidimi from late 1884 until his death in 1888, and Isigidimi closed soon after his passing.

Looking at editorship, we see that Ntsundu was used frequently under the missionary leadership of Stewart. It was also resonant to an isiXhosa editor, spiking when Makiwane, the first African editor, took over. Yet the usage slowly declined under Makiwane's editorship. Figure 3 also shows that Mnyama identity was not used frequently under the editorship of either Stewart or Makiwane, and that Mnyama language did not rise to substitute Ntsundu language during its slow decline. Instead, we see the clearly visible impact that J.T. Jabavu had on promoting Mnyama identity. From the beginning of his tenure as the editor of Isigidimi, we see a sharp increase in the usage of Mnyama. This makes it clear that Jabavu played a large role in promoting a shared *Abamnyama* identity for the emerging proto-nationalist movement.

The historical context of Jabavu helps to contextualize this promotion of Mnyama identity. Jabavu was a key leader in the emergence of organized African political activity, but also had particularly deep exposure to colonial society. As editor of Isigidimi and later Imvo Zabantsundu, Jabavu mobilized a 'progressive' political strategy that looked to the African future instead of to the African
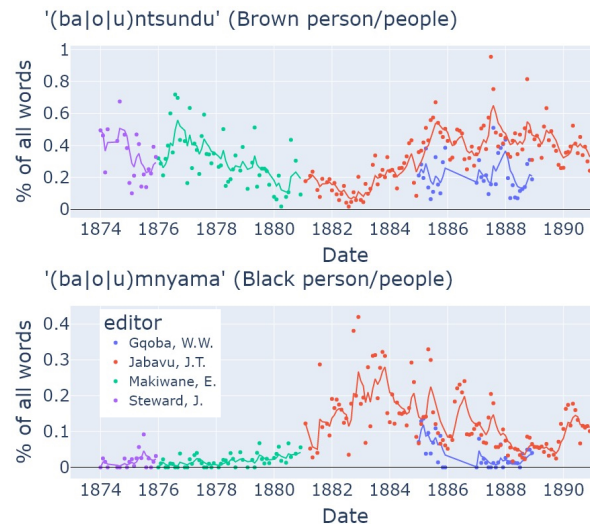

Figure 3: Editors and Ntsundu and Mnyama

past. Other leading Xhosa intellectuals of the period had promoted an African and Xhosa identity which deepened and renewed ties to Xhosa culture, history, and identity (Soga, 1983; Wauchope, 2008; Gqoba, 2015; Mqhayi, 2009). Jabavu instead sought African advancement through the mobilizing of Africans to participate more directly in colonial institutions, especially the Cape Parliament. He believed that to win gains for African people it was necessary to directly assert and defend African rights within the colonial political and legal system.

This history helps to contextualize the powerful shift of race language seen emerging under Jabavu's editorship. Mnyama language in this context appears to lean into colonial conceptions of African identity on the basis of Black vs White. Such an identification is useful when it is in direct dialogue with colonial institutions and authorities, highlighting and challenging colonial racism. Jabavu's early emphasis of Mnyama identity thus appears to be a "Black consciousness of Blackness" (Mbembe, 2017, 30), forged by a man who had been embedded in relationships with whites who viewed him as Black, and who sought to mobilize around racial identity to both challenge the racial inequalities which he confronted, and to build a new political movement pursuing Black modernity.

This editorial influence shows the powerful role that African intellectual and political elites, like Jabavu, had in shaping the discourse and dialogue of emerging proto-nationalist politics. Yet it does not tell the whole story. While it is clear that Jabavu set out to forge a collective sense of shared *Abamnyama* identity when he took over Isigidimi, *this focus on Mnyama language did not stick*. To un-

derstand why Mnyama language receded, we must again switch perspective to make visible another driving influence of the period: African voter mobilization and electoral participation in the Cape Parliament.

7 The impact of African franchise on shared identity

Figure 4 overlays key events in the history of African voter mobilization over the changes of word frequency. As I will show, mobilizing the African community to vote and opposing African disenfranchisement had a significant impact on the language of early Proto-nationalism.

The establishment of "Responsible Government" in the Cape in 1872 shifted power from a colonially appointed Governor to an elected Parliament. At this time the Cape Colony followed the British model of voter limitations based on class: there was no racial limitation, but voters had to qualify by holding property (land) or earning above a yearly income threshold. While few missionary educated Africans qualified to vote, proto-nationalist organizations saw an opportunity to mobilize more rural Xhosa people who were eligible due to their communal land tenure. Following voter-mobilization drives, by 1886 African voters made up 47 percent of all voters in 5 key Eastern Cape districts, the heartland of the proto-nationalist movement. In these districts, African voters could decisively swing the vote to send their chosen candidate to the Cape Parliament (Walshe, 1969; Odendaal, 2013, 96 for voter breakdown). In 1887 new laws (Parliamentary Voters Registration Act) were proposed which directly threatened African electoral participation. These laws, among other changes, would remove tribal land holdings as a basis to meet the property qualification to vote (see Odendaal 2013, 144, Walshe 1969, 587). The proto-nationalist community immediately responded and political activity reached a new peak across the eastern Cape with new connections forged between rural and urban African political communities (Odendaal, 2013). Although the bill passed in September 1887, political defeat only heightened the organizing efforts of the proto-nationalist movement which strongly mobilized and participated in the 1889 election.

This brief history offers the context to understand the significant shifts seen in figure 4. The vertical dotted blue lines show new elections to the Cape Parliament in 1879, 1884 and 1889. We see
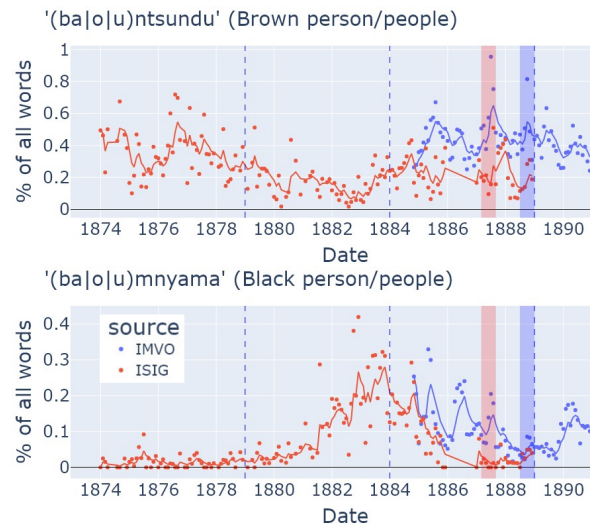

Figure 4: Elections and Ntsundu and Mnyama

no particular impact on Ntsundu and Mnyama language in 1879, when there was no effort by African leaders to mobilize Africans to vote. Yet, we see a significant impact on language use by the election of 1884. In the lead up to this election the Xhosa intellectual community, especially Jabavu, were deeply engaged in discussing the election, although registration numbers were not yet high enough for African's to swing the vote.

Two key shifts are visible: first, under Jabavu's editorship, discussion of Mnyama identity peaks in the run-up to the 1884 election. It is also at this moment that Ntsundu language reaches its inflection point: shifting from its lowest point at the end of 1882, and beginning to rise from 1883, right in the lead up period to the election. Ntsundu comes close to catching Mnyama language just before the election. Mnyama language peaks at approximately 0.275% of all words and Ntsundu language reaches approximately 0.225%. Second, we see the contrasting influence of elites and the wider community. Jabavu, who was very active in the 1884 election, appears to have pushed his focus on Mnyama identity to new heights to mobilize Africans during the election period. Yet the return of Ntsundu language may point to the wide communal resonance of the term which was activated in light of the election.

This wider social resonance of Ntsundu identity is made clear in 1887 and 1888. We see the highest spike in the frequency of Ntsundu language in 1887 during the Parliamentary discussion of the proposed disenfranchisement laws (red-shaded area). In Imvo the spike happens alongside the parliamentary debate period and in Isigidimi the

spike happens directly after the law was passed. We also see the impact of the 1889 election, with the 6-month lead up to the elections shaded blue. In this election campaigning period we see a spike in both papers in the use of Ntsundu identity.

This emphasis of Ntsundu identity fits a larger pattern. After the election of 1884, Ntsundu continued to rise and Mnyama usage swiftly declined. In the face of voter disenfranchisement, Ntsundu usage surges to nearly 1% of all words in one month, where Mnyama has a small rise which only reaches 0.2%. The lead up to the 1889 elections (blue shaded area) might even suggest a suppression of Mnyama language, reaching its lowest point in Imvo and remaining low throughout 1889.

These changes in language show a clear relationship to African voter mobilization and elections. This suggests that the language of shared 'Blackness' was not determined by intellectual elites alone. Instead Ntsundu identity appears to have had a significant resonance within the larger community that Mnyama did not have. Remember that it was largely African voters with rural or tribal land holdings who could register to vote, and it was these voters who were threatened with the loss of their voting rights. This suggests that Mnyama and Ntsundu language might have been meaningful for different reasons to different communities. Mnyama identity language seems to more explicitly foregrounded Black/White racism, and it may thus have resonated more with figures like Jabavu and other missionary educated or urban Xhosa who were daily in contact with white colonists and facing racism through the language of Black and White. However, it may be that Ntsundu language was indeed a conception which held deeper resonances for the broader proto-nationalist rural community, drawn from an older isiXhosa epistemological framework which existed before colonialism. Evidence from newspapers also shows that African organizations which had started in the 1880s were already using Ntsundu language far more than Mnyama language, and also amplified this identity in election periods.

This analysis also suggests that Jabavu was sensitive and responsive to this wider community resonance. Notably, although he had promoted Mnyama identity during his time at Isigidimi, the 1884 election may have shown Jabavu that this identity was not the one with the most resonance. When he chose the name for his new newspaper Jabavu himself turned to Ntsundu identity: Imvo

Zaba*ntsundu*. Overall, this analysis shows the impact that leaders like Jabavu could have on political discourse and identity. Yet it also shows that the larger community had a powerful and, for this period, defining impact on the language of political discussion. These findings show that isiXhosa speakers were engaged in debates about identity on isiXhosa epistemological terms, and reveals how individuals, events, and collective political engagement all shaped and reshaped core ideas of political self-hood in early African nationalism.

8 Conclusion

This paper outlines an application of text frequency analysis over time which focuses on root words and offers an accessible method to study isiXhosa texts. This method has particular advantages because it by-passes the need for other complex Natural Language Processing tools which are limited for isiXhosa. For this reason, the approach might also be useful for other under-resourced agglutinative languages. The approach is also useful for humanities and social science scholars as it is technically simple, yet requires sufficient contextual knowledge of the examined texts.

I showcase the utility of this approach through an analysis of shared political identity formation in early African Nationalism. Using the outlined methods, I visualize shifts in the usage of key identity terms. I have focused on competing isiXhosa conceptions of shared 'Blackness', -Ntsundu and -Mnyama. The paper has examined how individual leaders (like J.T. Jabavu), historical events (such as African voter mobilization and disenfranchisement), and the wider community (such as rural Xhosa voters) all played an influential role in shaping collective identity discourse. This case shows the opportunities created by applying computational text analysis methods to isiXhosa, as the study is able to reveal communal shifts in the identity language of early African nationalism which have remained invisible to close textual analysis or historical approaches to these materials. These approaches thus support and augment existing qualitative methods. This offers new opportunities to study isiXhosa and other African language materials, and new perspectives on important humanities and social science questions.

Limitations

The paper presents a method which is developed on isiXhosa texts, which the paper suggests might be useful for under resourced agglutinative languages, including African and other global languages. As highlighted in the main text, a key limitation is the application to only isiXhosa text analysis. It may be the case that grammatical features of other languages are not compatible with the approach. This limitation may be mitigated in some ways: isiXhosa shares many grammatical features of the larger group of Bantu languages including a common class system and thus the method outlined here is likely extendable. Agglutinative languages in general share a structure of root words modified by prefixes and suffixes which this method is designed to utilize. However, the author has not tested the extent and limitations of applications to other languages.

The analysis presented here is also limited: the case study is partial and cannot fully make visible the depths and limits of the historical and qualitative analysis which was used alongside the quantitative analysis due to space limitations. This limitation means the case study is best taken as an exemplar of possible research rather than a fully demonstrated historical claim. The author of this paper is not a first language isiXhosa speaker. Qualitative research which supports the argument has combine reading of sections of isiXhosa newspaper text which use the key words discussed, and has been supplemented with broader readings of primary literature in translation and secondary literature written in English.

Ethics Statement

This research has drawn on isiXhosa historical text which is in the public domain and has been digitized and made publicly available.

Human coders and language editors who worked on the digitized text analyzed here are all employed as members of the larger archival research project. All coders and editors are first language isiXhosa speakers who are students or researchers.

The method outlined in this paper may be of particular use to under resourced languages. The approach does not use cutting edge tools, but rather outlines how accessible methodological approaches can be used to by-pass more complicated NLP steps which may not be available for under resourced languages. For this reason, the author believes that

this research has limited potential to cause harm. Instead it may have potential to enhance computational research in under resourced languages, fostering more equal access to computational text research methods

Acknowledgements

References

David Attwell. 2005. *Rewriting Modernity: Studies in Black South African Literary History*. KwaZulu-Natal University Press, Pietermaritzburg.

Alistair Baron, Paul Rayson, and Dawn Archer. 2009. Word frequency and key word statistics in corpus linguistics. *Anglistik*, 20(1):41–67.

Katherine Bode. 2017. The Equivalence of "Close" and "Distant" Reading; or, Toward a New Object for Data-Rich Literary History. *Modern Language Quarterly*, 78(1):77–106.

Partha Chatterjee. 1986. *Nationalist Thought and the Colonial World: A Derivative Discourse?* Zed Books.

Jean Comaroff and John L. Comaroff. 1991. *Of revelation and revolution, volume 1: Christianity, colonialism, and consciousness in South Africa*, volume 1. University of Chicago Press.

John L. Comaroff and Jean Comaroff. 1997. *Of revelation and revolution, volume 2: The dialectics of modernity on a South African frontier*, volume 2. University of Chicago Press.

William Wellington Gqoba. 2015. *Isizwe esinembali: Xhosa histories and poetry (1873-1888)*. University of KwaZulu-Natal Press, KwaZulu-Natal.

Andreas H Jucker, Franz Lebsanft, and Gerd Fritz. 1999. Historical dialogue analysis. *Historical Dialogue Analysis*, pages 1–486.

S Jänicke, G Franzini, M F Cheema, and G Scheuermann. 2015. On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. *Eurographics Conference on Visualization (EuroVis)*, page 21.

Xolela Mangcu. 2014. *Biko : a life*. I.B. Tauris, London.

Ntongela Masilela. 2013. *An Outline of the New African Movement in South Africa*. Africa World Press, New Jersey.

Ntongela Masilela. 2014. *The Historical Figures of the New African Movement*. Africa World Press, New Jersey.

Achille Mbembe. 2017. *Critique of Black Reason*. Duke University Press.

Khwezi Mkhize. 2008. Carrying the Cross: Isaac William(s) Wauchope's Ingcamango Ebunzimeni. Master's thesis, University of the Witwatersrand.

Khwezi Mkhize. 2018. 'To See Us As We See Ourselves': John Tengo Jabavu and the Politics of the Black Periodical. *Journal of Southern African Studies*, 44(3):413–430.

Franco Moretti. 2013. *Distant Reading*. Verso Books.

Samuel E. K. Mqhayi. 2009. *Abantu Besizwe: Historical and biographical writings 1902-1944 SEK Mqhayi*. Wits University Press, Johannesburg.

Lulamile Mzamo, Albert Helberg, and Sonja Bosch. 2015. Introducing xgl-a lexicalised probabilistic graphical lemmatiser for isixhosa. In *2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, pages 142–147. IEEE.

Pedro Mzileni. 2019. (Re)imagining Mandela's Leadership esiXhoseni. *International Journal of Critical Diversity Studies*. Publisher: Pluto Journals.

M. Ndletyana, editor. 2008. *African intellectuals in 19th and early 20th century South Africa*. HSRC Press, Cape Town.

Abner Nyamende. 2000. *The Life and Works of Isaac William(s) Wauchope*. PhD Thesis, University of Cape Town, Cape Town.

André Odendaal. 2013. *The founders: The origins of the ANC and the struggle for democracy in South Africa*. The University Press of Kentucky, Kentucky.

Martin Puttkammer and Jakobus S. Du Toit. 2021. Canonical Segmentation and Syntactic Morpheme Tagging of Four Resource- scarce Nguni Languages. *Journal of the Digital Humanities Association of Southern Africa*, 3(03).

Ncedile Saule. 2017. King Hintsa memorial lecture. Public lecture.

Jonathan Schoots. 2020. S.E.K. Mqhayi and African social analysis: African sociological thought in colonial South Africa. *Social Dynamics*, 46(3):493–514.

Leo Jonathan Schoots. 2021. *Novelty, Networks, and the Rise of African Nationalism: African Intermediary Intelligentsia and the Making of Political Innovation in Colonial South Africa (1860-1890)*. Ph.D., The University of Chicago, United States – Illinois.

Tiyo Soga. 1983. *The journal and selected writings of the Reverend Tiyo Soga*. The Graham's Town series 7. A.A. Balkema, Cape Town.

Les Switzer and Donna Switzer. 1979. *The Black press in South Africa and Lesotho: a descriptive bibliographic guide to African, Coloured, and Indian newspapers, newsletters, and magazines, 1836-1976*. Boston: Hall.

A. P. Walshe. 1969. The Origins of African Political Consciousness in South Africa. *The Journal of Modern African Studies*, 7(4):583–610.

Isaac Williams Wauchope. 2008. *Isaac Williams Wauchope: selected writings 1874-1916*. Van Riebeeck Society, Cape Town.