

A computational study of media bias in South African online political news reporting over the period 2021 - 2023

Nonhlanhla Nomusa Ngwenya (0109877X)

Supervisor(s): Professor Rod Alence



A Research Report submitted in partial fulfilment of the requirements
for the degree of Master of Arts in the field of e-Science

in the

School of Social Sciences

University of the Witwatersrand, Johannesburg

7 July 2024

DECLARATION

I, Nonhlanhla Nomusa Ngwenya, declare that this report is my own, unaided work. It is being submitted for the degree of Master of Arts in the field of e-Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.



Nonhlanhla Nomusa Ngwenya

7 July 2024

ABSTRACT

The study examined the presence of tonality bias in South African political news reporting over the period 2021 until mid-2023. The study employed the methods of the Lexicoder Sentiment Dictionary, a lexical-based method, and Latent Semantic Scaling, a semi-supervised machine learning method. Sentiment was utilised as a proxy for tonality. Online commercial media publishers were contrasted against the state-owned news publisher to ascertain how online news reporting contributed to shaping the national agenda, and the framing of political actors and their respective political parties. The Lexicoder Sentiment Dictionary and the Latent Semantic Scaling evidenced that commercial media publishers exhibited positive tonality bias for the Democratic Alliance during the 2021 Municipal Elections. South African media publishers were found to exhibit consistent negative tonality bias when reporting on protest action. The state-owned media publisher was found to drive a pro ruling party sentiment whereas commercial media publishers' sentiment was anti-populist and agenda-setting. The congruency in political news reporting gave grounds to the call for diversity in publishing.

ACKNOWLEDGMENTS

I would like to thank my supervisor Professor Rod Alence for his guidance and feedback, throughout the research process.

“The conqueror writes history, they came, they conquered, and they wrote. Now you don’t expect people who came to invade us to write the truth about us. They will always write negative things about us, and they have to do that because they have to justify their invasion in all the countries”.

— Miriam Makeba

**Ngwenya
Mntimande,
Bambolunye,
Zingaba mbili,
Zifuze konina,
Ekhabonina,
Mabuya,
Bengasabuyi
Baye babuya
Emangwaneni,
Nungunde,
Wakhothe,
Bayosala beziloyanisa!**

CONTENTS

DECLARATION	I
ABSTRACT	II
ACKNOWLEDGMENTS	III
CONTENTS.....	IV
LIST OF FIGURES	VI
LIST OF TABLES	VII
LIST OF ABBREVIATIONS.....	VIII
1. CHAPTER ONE- INTRODUCTION	1
1. INTRODUCTION.....	1
1.1. BACKGROUND	1
1.2. LITERATURE REVIEW	2
1.3. SENTIMENT ANALYSIS	6
1.4. PURPOSE OF THE STUDY.....	7
1.5. RESEARCH QUESTION.....	7
1.6. LIMITATIONS	8
1.7. OVERVIEW.....	8
2. CHAPTER TWO – METHODS AND MATERIALS.....	9
2. INTRODUCTION.....	9
2.1. RESEARCH METHODOLOGY	9
2.2. DATA	9
2.3. SENTIMENT ANALYSIS	12
2.4. ANALYSIS.....	14
2.5. LIMITATIONS	14
2.6. ETHICAL CONSIDERATIONS.....	15
3. CHAPTER 3: DESCRIPTIVE ANALYSIS	16
3. INTRODUCTION.....	16
3.1. DESCRIPTIVE ANALYSIS.....	16
4. CHAPTER 4: LEXICODER SENTIMENT DICTIONARY ANALYSIS	21
4. INTRODUCTION.....	21
4.1. IS BIAS SYSTEMATIC?	21
4.2. LEXICAL-BASED SENTIMENT ANALYSIS BY MEDIA PUBLISHER	26
4.3. DISCUSSION	32
5. CHAPTER 5: LATENT SEMANTIC SCALING ANALYSIS.....	34
5. INTRODUCTION.....	34
5.1. TESTING FOR TONALITY BIAS ACROSS SOUTH AFRICAN MEDIA.....	34
5.2. TONALITY BIAS BY PUBLICATION: APPLYING REGRESSION TO LATENT SEMANTIC SCALING OUTCOMES.....	40
5.3. DISCUSSION	47
6. CHAPTER 6: PROTEST ACTION IN THE NEWS	49
6. INTRODUCTION.....	49

6.1.	LEXICODER SENTIMENT DICTIONARY: ARE THE MARGINALISED ALLOWED TO PROTEST?	49
6.2.	LATENT SEMANTIC SCORING: SOUTH AFRICAN MEDIA AND PROTEST ACTION	51
6.3.	DISCUSSION	55
7.	CHAPTER 7: CONCLUSION	56
7.	FUTURE STUDY CONSIDERATIONS	57
	REFERENCES	58

LIST OF FIGURES

Figure 2.1: Breakdown of news articles by online media publisher	10
Figure 2.2: Text transformation	11
Figure 3.1: Frequency chart aggregated across all publishers	17
Figure 3.2: Frequency chart of 13 South African online media publishers	18
Figure 3.3: Text similarity correlation plot	19
Figure 4.1: LSD Sentiment Chart: President Cyril Ramaphosa and the ANC	22
Figure 4.2: LSD Sentiment Chart: John Steenhuisen and the DA	23
Figure 4.3: LSD Sentiment Chart: Julius Malema and the EFF	24
Figure 4.4: LSD Sentiment Chart: Comparative analysis across party lines	25
Figure 4.5: LSD Sentiment Chart by publisher: ANC	27
Figure 4.6: LSD Sentiment Chart by publisher: DA	28
Figure 4.7: LSD Sentiment Chart by publisher: EFF	29
Figure 4.8: LSD Sentiment Chart by publisher: Ramaphosa	30
Figure 4.9: LSD Sentiment Chart by publisher: Steenhuisen	31
Figure 4.10: LSD Sentiment Chart by publisher: Malema	32
Figure 5.1: LSS Text plot terms: ANC	35
Figure 5.2: LSS-LSD Text plot terms: ANC	35
Figure 5.3: LSS Text plot terms: DA	36
Figure 5.4: LSS-LSD Text plot terms: DA	37
Figure 5.5: LSS Text plot terms: EFF	37
Figure 5.6: LSS Text plot terms: Ramaphosa	38
Figure 5.7: LSS-LSD Text plot terms: Ramaphosa	39
Figure 5.8: LSS Text plot terms: Steenhuisen	39
Figure 5.9: LSS Text plot terms: Malema	40
Figure 5.10: Regression interaction effect plots: 2021 Municipal Election party news	44
Figure 5.11: Regression interaction effect plots: 2021 Municipal Elections party leader news	47
Figure 6.1: LSD Sentiment Chart: Protests	50
Figure 6.2: LSD Sentiment Chart by publisher: Protests	51
Figure 6.3: LSS Text plot terms: Protests	52
Figure 6.4: LSS with LSD Text plot terms: Protests	52
Figure 6.5: Regression interaction effect plots: 2023 National Shutdown party leader news	55

LIST OF TABLES

Table 2.1: List of online media publishers and ownership	11
Table 2.2: Data search criteria	11
Table 3.1: Distribution of articles	16
Table 3.2: Frequency of identified political terms	19
Table 4.1: Political leader top term frequency	26
Table 5.1: 2021 Municipal Elections regression models: political party news	43
Table 5.2: 2021 Municipal Elections regression models: party leader news	46
Table 6.1: Term frequency: Protests	50
Table 6.2: 2023 National Shutdown regression models: party leader news	54

LIST OF ABBREVIATIONS

ANC:	African National Congress
DA:	Democratic Alliance
DFM:	Document Feature Matrix
ECB:	European Central Bank
EFF:	Economic Freedom Fighters
EWN:	Eyewitness News
IOL:	Independent Online
LSD:	Lexicoder Sentiment Dictionary
LSS:	Latent Semantic Scaling
MMA:	Media Monitoring Africa
NDR:	National Democratic Revolution
OLS:	Ordinary Least Squares
SABC:	South African Broadcasting Corporation
SANEF:	South African National Editors' Forum
quanteda:	Quantitative Analysis of Text Data

1. CHAPTER ONE- INTRODUCTION

1. Introduction

Lippmann (1922) wrote of a pseudo-environment that exists between man and his understanding of the real world. This pseudo-environment plays a role in influencing his behaviour. These are the “pictures in our heads” and “by the same mechanism through which heroes are incarnated, devils are made” (Lippmann, 1922, p. 10). In his seminal work, Lippman (1922) identified the growing intermediary role of the media in opinion setting. This would later be theorised as agenda-setting and framing. With the proliferation of the media in political news reporting, Strömbäck (2008) stressed that the chief responsibility of the media should be limited to providing objective information on political matters and their representatives without imparting influence. This was in response to the growing risk posed by mediatisation resulting in the loss of political independence (Strömbäck, 2008). With growing evidence linking political news reporting to electoral outcomes, examining media bias in the transmission of political news is of growing importance, particularly on the eve of a South African election in a charged political climate.

South Africa, a democracy under strain stands on the eve of what will be a heavily contested National Elections, provides an opportunity to examine the role of online news media publishers within political communication. With a media maintaining its objectivity amid accusations of bias from political parties and leaders, debate continues to focus on the role of the media. This has stimulated research on the role of news media in political discourse, particularly during an election cycle and the representation of political actors. The study is thus an undertaking to examine media bias through the lens of tonality and the framing of leaders and parties. This is of significance as it has the ability to shape electoral outcomes. The study is facilitated through the use of new computational methods and the growing availability of unstructured data.

The dawn of internet 2.0 and the advancements in technology have vastly transformed how society interacts and receives information. The accessibility of technology has continued to transform society; with developments advancing new platforms of communication that have charted new pathways of engagement and influence. News media has had little choice but to evolve due to the decline of print media and the rise of the internet, impacting margins and viability.

Not only has the advancement in technology provided new channels of communication, but it has also increased the variety, sources, and types of new data available to researchers. Of particular interest to political scientists is the meaning hidden in words and how these are conveyed, perceived, or manipulated (Rodriguez et al., 2023). Advancements in technology have also meant increasing computational ability to meaningfully analyse unstructured and voluminous data. What before relied on content analysis or numerical analysis, now includes the analysis of text data to determine sentiment.

This study is thus an undertaking in the computational analysis of text data utilising sentiment techniques, to determine the presence of tonality bias in the political news reporting of South African media publishers.

1.1. Background

Post-apartheid South Africa remains politically fractured with rising tensions spurred by demands for faster paced socioeconomic transformation and rising concerns over democratic legitimacy. This has resulted in progressively frequent protest action as a reflection of public discontent in the ruling party’s ability to deliver and the perceived levels of corruption eroding government. Nonetheless, South Africa is not new to polarising politics with a history that stems from colonisation and an oppressive apartheid regime. The preceding apartheid roots of present-day democracy has resulted in embedding persisting structural racial socioeconomic inequality leaving a disenchanted majority.

This has not abated due to the largely neoliberal macroeconomic policies adopted since the dawn of democracy (Chiumbu et al., 2018).

The ongoing challenges have become currency for political discourse on the role of different actors in influencing electoral outcomes. Accusations of bias are of concern as this has been posited to influence voting decisions and play a role in polarising citizens. As the political landscape has evolved, so has the role of the South African media, which has moved from repression to democratic watchdog. Conversely, the journey for the media has been turbulent with accusations of bias in the transmission of political news. News media maintains its critical role in advancing democratic principles and the need for independence and press freedom.

In the subsequent section, the study provides detailed analysis of the literature on bias and political communication. The relationship between bias and electoral outcomes is explored in depth and the many definitions of bias are contextualised within theory. Evidence of bias is provided to give legitimacy to theory. This is followed by an examination of South African political news reporting in the context of a regime change and the turbulent relationship with governance. This is amid accusations of bias versus a declaration of objectivity by South African media pundits. Lastly, the literature review, details the choice of computational methods and more specifically, the use of sentiment analysis for the study.

1.2. Literature Review

1.2.1. Political communication: objectivity versus bias

In democracies, political scientists have long advanced the view that the media plays a stirring intermediary role in decimating political news to the voting public. The claim is centred on the distinction between media communication and political communication. Bias is a concern to political scientists, as the public gain their knowledge on political parties, candidates, and positions through the media. Thus, partisanship has a detrimental impact and belies the assumed role of balance and objectivity. The ideological role of political news media is objectivity, whilst its antithesis, bias is a common criticism. Bias in political news is the slant or the distortion of news favouring one side or presenting an imbalanced and one-sided news story (Hackett, 1984). Trust in media is debated among political spectrums who put forward bias as its opposition.

In his review of political communication versus media communication, Strömbäck (2008) writes of three ages of political communication that began in the 1950s and vastly differs from the current mediated form. Mediated politics abstract the public from the originator with the media becoming the intermediary source for political news whether it be by newspaper, television, or the internet. This abstraction from source and the delivery of political news can distort or create a new reality through its depiction from a media lens. Discontent lies in the increased intensity in the mediation of political news in recent times (Strömbäck, 2008). By contrast, mediatisation carries a more serious consequence as the media is now central to political communication and how the public views and understands politics. Strömbäck (2008) defines four phases of mediatisation, and the most damning being the final phase, in which political actors adapt their strategies to media logic in their campaigning, policy making and strategising. It is at this stage that the media is an internalised function of governance.

1.2.2. Theory underpinning bias

Mullainathan and Shleifer (2002) define bias along two paths; one as an ideology which seeks to influence reader beliefs, and another as spin which reflects a desire for great storytelling. In this context, competition is insufficient to bring about a reduction in bias, as media outlets work to maximise the welfare of owners, as a function of their ideology and profitability. Thus, increased competition drives greater bias by catering to reader's *a priori* beliefs to secure patronage (Mullainathan and Shleifer, 2005). Instead, it is reader heterogeneity that is more likely to decrease bias through competing interest and views requiring more balanced reporting (Mullainathan and Shleifer, 2005).

Gentzkow and Shapiro (2006) build a theoretical model which proves that a Bayesian consumer is more likely to ascribe higher competency to a source that confirms his *a priori* beliefs if he is uncertain about the validity of information. This results in bias with the media reporting in favour of the consumer which leaves an ecosystem that is sub-optimal for all participants. This distortion occurs in order to build reputational trust with consumers. Media firms will use methods of omission, word selection and the level of credibility assigned to primary sources in ascribing bias (Gentzkow and Shapiro, 2006). A secondary finding by Gentzkow and Shapiro (2006) is the need for increased competition to reduce bias unlike that of Mullainathan and Shleifer (2005). Thus, greater competition results in a more diverse reader base which positively impacts the reduction of bias in political news reporting (Gentzkow and Shapiro, 2006).

In dispute with Lippman (1922), Herman and Chomsky (2008) proposed a propaganda model that theorised that the interests of both the state and commercial media can result in delivering propaganda. Propaganda is conventionally understood in the context of state censorship and media control whilst increased media competitiveness is associated with greater objectiveness and critique. However, this is subject to access and resources which is oft dominated by the elite (Herman and Chomsky, 2008). Thus, a media regime that is centralised within the elite and marginalises dissenting arguments will result in propaganda even when it maintains objectivity as its central doctrine. Well established filters will serve to determine what is deemed newsworthy, build, and reinforce a system that constricts what is considered news. Herman and Chomsky (2008) conclude that media coverage is subject to political dichotomisation which affects which stories are told and how they are told i.e. tone, coverage, and placement.

Eberl et al. (2017) defines the three forms of media bias as visibility bias; tonality bias; and agenda bias. Each form of bias impacts individual preferences when forming opinions on political actors and making political choices.

- Visibility bias is defined as the increased coverage of one candidate disproportionately to other candidates. This is a relative measure as it measures the disproportionate coverage of a candidate in relation to others.
- Tonality bias: The framing of a political actor in a favourable or unfavourable light in comparison to other candidates which in turn impacts their electoral odds.
- Agenda bias: The amplification of political party agendas in the media. It is not only important for a party or a candidate to appear in the media but also to have the opportunity to have their agenda positively amplified, and to inform the public of policy positions.

The theories on bias inform the study on how normative beliefs internal to the media publisher influence the reporting of political news thus introducing bias into the system. This is achieved through the choice of words, omissions, which stories are told and the representation of political actors.

Bias is thus distinctly a function of word selection and is aligned to the use of sentiment analysis as per the study methods. This is best evidenced in the definition by Eberl et al. (2017) of tonality bias which will serve as the basis for computational analysis and the choice of sentiment methods.

1.2.3. Evidence and implications of bias in political news

Political news reporting is a contested topic with the requisite standards of journalism being objectivity, fairness, and impartialness, particularly in developed economies (Endersby, 2011). However, research often bemoans the presence of bias along ideological lines. This section reviews empirical studies that have examined the topic of bias and media reporting in political communication and the associated impact at a country level.

A study by Chiang and Knight (2011) examined the role of newspaper endorsements in electoral outcomes in the North American elections of 2000 and 2004. The study was undertaken to determine media importance in electioneering. The authors utilised an econometric model and determined that voters had an awareness of possible bias in left- or right-wing newspapers. Voters filtered

endorsements and thus were more likely to assume more credibility in a neutral or oppositional paper to mitigate bias.

Hopmann et al. (2010) evidenced the role of the media in influencing election outcomes of parties in Denmark as directly linked to their visibility and tonality. Significant to the findings was the importance of visibility and a positive tone in improving electoral performance. The impact of biased media reporting was conflicting, as it resulted in voters feeling demoralised, reinforced partisan bias, or swayed undecided voters, who were the most vulnerable to persuasion. Hopmann et al. (2010) concluded that this was an increasing phenomenon in western countries with a growing population of undecided voters as party partisanship declined.

Isolating for leader effects, Takens et al. (2015) concluded that political leaders fared more favourably in elections with greater personalised prominence in the media, impacting public voting decisions. Thus, more positive appraisals of political leaders served to increase their weighted consideration by the electorate. This could serve to outweigh that of political party dynamics or policy issue accordance with popularity divergence (Takens et al., 2015). Thus, the visibility and prominence given to political leaders by the media has a direct relationship in impacting leader effects and voting. The media plays a significant role in the personalisation of political leaders and how they are viewed by the voting public in their reporting. Conversely, this can have a negative effect if the leader is negatively represented, and if this is amplified by greater visibility.

Haselmayer et al. (2017) found that bias played a gatekeeping role in determining whether political parties received greater visibility in reporting. The finding was based on an analysis of reporting on political press releases during the 2013 Austrian elections. Important to electioneering, is the ability of political party content or press releases to reach the electorate. This occurred when ideologies aligned with the partisanship of readers, rather than the merits of the content. Utilising content analysis, the authors found news reporting to be driven by partisan bias, and thus political parties that greater mimicked the ideologies of readers were more likely to be magnified. Of concern was the rise in gatekeeping which reflected the ability of bias to further polarise voters politically and limit access (Haselmayer et al., 2017).

Much of the research discussed above has focused on advanced economies, however a study was completed in India during the 2014 elections (Barclay et al., 2015). A content analysis of four major English newspapers was conducted over the month leading up to the elections of 2014, and each paper was scored individually for bias by Barclay et al. (2015). Three of the four papers were found to be positively biased towards the victorious party and thus potentially serving as an early prediction indicator.

McCarthy and Dolfsma (2014) found that approximately 90 percent of journalists perceived themselves as objective and neutral, following a review of journalist surveys over a 40-year period. However, a literature review confirmed that journalists are far from neutral and are more than likely incentivised in their reporting. The findings indicated that journalists are responsible for influencing negative perceptions by heightening perceived risk due to sensationalism and overreporting, influencing electoral outcomes, individual and firm reputations, and overall market sentiment (McCarthy and Dolfsma, 2014).

A significant volume of the body of work has evidenced the importance of the media in influencing electoral outcomes. It would be remiss not to further the analysis in the South African context. South African political news reporting has a contentious past much like the history of the country and its journey to democracy. The role of the media is closely intertwined with its political past and lays the basis for continued distrust with the public. This is insufficient to blanketly dismiss the critical role the media has played in the new dispensation. The privately owned commercial media maintains its objectiveness and need for press freedom. The next section explores South Africa's political news reporting and its role in politics.

1.2.4. A history of discontent: South African political news reporting

The South African media landscape is largely a legacy market. South African media has historically developed from an authoritarian apartheid regime with regulations on reporting and the promulgation of propaganda through state owned news. Post democracy, news media has become an independent state watchdog but not withstanding questions of bias (Wasserman, 2020). It is characterised by three main groupings which are public, privately owned commercial media and community media. Private commercial media is said to be largely skewed by an agenda that serves the interest of the wealthy white minority elite who are attractive to advertisers (Wasserman, 2020). This is not reflective of the vastly disparate and unequal socioeconomic landscape of the country. This has seen the development of community news to bridge the gap alongside the role of the public broadcaster (Wasserman, 2020). Further muddying the waters is the oligopolistic nature of commercial media consolidated into few hands; namely Naspers (Media24), Independent Media, Caxton publishers and Arena Holdings.

The relationship between the media and the state has grown increasingly tumultuous with pressure for greater regulation beyond self-regulation. Whilst the South African media has withstood the attacks, the accusations of imbalances remain. The contention being a need for political reporting to be more balanced and impartial (Soontjens et al, 2021). South African news media asserts the ideals of objectivity and defender of democracy and any contestation of bias as an attack on press freedom. The South African national editor's forum (Sanef), established in 1996, asserts its role as important to ensuring continued democracy; a champion for press freedom; promoting ethical journalism and readdressing racial and gender disparities in journalism (Sanef, 2020).

South African news media experts assert that it is its ability to keep the state accountable that has driven discontent with South African news media, and the rise in accusations of bias. The current news media landscape continues to reflect a tenacious media that has defied censorship and understands its vital role in a viable democracy. Daniels (2013) ascribes the growing tension between news media and the ruling party as a requirement to keep the public free from secrecy and ignorance, which is a hindrance to democracy. Alternately, South African media's adversarial role has resulted in a call for greater regulation, with the ruling party's proposal for a government media appeals tribunal in 2007 (Daniels, 2013).

On the contrary, South African has seen a decline in media freedom to "partly free" (Plaut, 2018). A call of bias is instead laid at the feet of government with interventions that include the establishment of a pro government paper, the New Age, which was in the vein of Apartheid style propaganda (Plaut, 2018). This resulted in the plundering of state funds in an attempt to halt waning electoral support. It is the state-owned media publisher that has suffered from the Apartheid styled interventions and without the resilience of South African journalism, would not have led to the expose of censorship by the SABC8 (Radebe, 2019). Radebe (2019) likens this process as a move to an authoritarian democracy, with limits on media freedom. Within this construct, South African news media publishers are under significant pressure to uphold their role in a democracy, whilst resisting increasing commercial pressure and threats from political actors.

Concern remains on whether democratic aspirations have been realised in the newsroom and how this impacts political reporting. Chiumbu (2016) challenges that even with the changeover to democracy and the racial diversification of ownership in commercial media, it remains driven by capitalism and has not addressed the inherent racial structural schism. Friedman (2011) finds validity in the critique of media bias in South Africa with reporting guilty of reflecting the views of the White minority elite. Friedman (2011) further lays claim that the White minority elite may inherently distrust a ruling Black government and ignore the struggles of the African majority. This is evidenced in the coverage of service delivery protests with the poor often being characterised as "rioting mobs" (Friedman, 2011, p. 111). He evidences this in the failure of the press to recognise or report on ongoing service delivery protests thus leaving the poor without a platform to voice their concerns. Key to this diagnostic is the assertion that South African media is guilty of middle-class bias and

viewing the country through a suburban filtered lens (Friedman, 2011). One defining event in biased reporting was at the onset of Marikana massacre (Duncan, 2013).

The reporting at the onset of the Marikana massacre on the 16th of August 2012 was a telling instance of media bias. Early coverage focused on the views of business and trade unions without deliberating with the protesting miners (Duncan, 2013). The miners were characterised as violent rioters threatening business operations, yet the strike would end with the death of 34 miners and over 70 injured (Duncan, 2013). Without adequate coverage, increasing police brutality went unchecked. The agenda-setting in the early coverage of the Marikana massacre provides evidence of editorial failure, which led to the silencing of the poor, maintenance of the status quo, and preserving the demands of the elite. This would also highlight the importance of academia in highlighting the inaccuracies and disparities in reporting. This reinforced elitist bias in commercial media that is pro capital and largely ignores the majority who are seen as inherently violent protestors (Chiumbu, 2016).

Inherent to the critique is the lack of contextualisation of the deep-seated challenges inherited from the apartheid regime that have left a significantly unequal society (Chiumbu et al., 2018). Without this framing and awareness, the stories that make the headlines lack the perspective of the marginalised and poor. This points to how the media views the poor and their challenges as evidenced in the reporting on protests (Chiumbu, 2016). Greater attention is paid not to alienate the white minority elite readership and business which often results in protests being portrayed negatively and as an inconvenience. This pattern of bias is also evidenced in the reporting on service delivery protests, which are characterised as harmful and violent. Protest action is thus depicted as delegitimate overall (Wasserman et al., 2018).

Having examined the above, the current body of research in South Africa relies on content analysis and is often case-specific. Notwithstanding, Cornelissen et al. (2019) conducted a computational study on media bias in South Africa. The authors analysed the behaviour of five media publishers by collecting text data through web scraping covering the leading political parties. The text data was analysed for visibility and tonality bias through the application of sentiment analysis and topic modelling. The authors found evidence of under and over engagement by topic, differences in sentiment polarity but no consistent bias with an overall neutral sentiment across media publishers (Cornelissen et al., 2019). The study revealed that whilst News24 over reported on Nkandla, the state media entity, the South African Broadcasting Corporation (SABC) under reported on the same matter (Cornelissen et al., 2019). Only one broadcaster was found to have a consistent negative bias whereas all media publishers were deemed to be unbiased by tone and visibility overall. Of concern with the Cornelissen et al. (2019) study was the scraping process potentially resulting in biased rather than a representative news corpus. The authors further acknowledge the limitations of only utilising lexicon-based sentiment analysis and the need to provide statistical inference.

The utilisation of sentiment analysis by Cornelissen et al. (2019) provided further strength to the validity of the methods chosen to examine bias for the study. This study aims to extend upon these computational studies in examining bias in South African news reporting by expanding the computational methodologies applied.

1.3. Sentiment analysis

Sentiment analysis has a long-standing history in the political sciences and the broader social sciences, as a tool to measure polarity or tonality of text (Haselmayer and Jenny, 2017). It has been used significantly in the study of media negativity, political legislation, and electioneering (Haselmayer and Jenny, 2017). It has increasingly been dominated by automated processes versus human coded content analysis and has proven useful for the analysis of large sets of text.

This has been led by the growth of text data through advancements in technology and digitisation which has seen a significant rise in the application of sentiment analysis (Ribeiro al., 2016). Sentiment analysis utilising dictionary defined lexicon is long established and popular due to its

quantitative simplicity in application. This has improved with the availability of off-the-shelf programmes, as they provide ease of application, particularly with small corpora and in domain specific contexts (Watanabe, 2021).

In this study, a dictionary-based method and a semi-supervised machine learning method are utilised. To overcome the criticism of domain specificity, the Lexicoder Sentiment Dictionary (LSD) by Young and Soroka (2012) will be utilised. Young and Soroka (2012) developed LSD for political text hence its applicability and choice for sentiment dictionary analysis. This is contrasted against the Latent Semantic Scaling (LSS), proposed by Watanabe (2021), for expanding sentiment analysis given the growth in demand across languages, regions, and domains. Bias in the context of South African online news reporting is defined as tonality bias for the purposes of the study.

1.4. Purpose of the study

The purpose of the study is to determine if there is evidence of tonality bias in South African online political news reporting amongst leading commercial media publishers. This is contrasted against the state-owned media publisher.

Whilst South African researchers have studied the issue of bias, many of the studies have relied on content analysis and case studies. This study aims to expand the available research by marrying computational methods together with the abundance of unstructured text data in the political sciences.

Tonality bias in this study is assumed to be sentiment and is measured by polarity. This study seeks to add a computational approach to the current body of work given the expansion in data and the growth in new methods. The research relies on evidencing tonality bias in the form of sentiment analysis by measuring tonality (polarity). A secondary objective is to examine whether lexical or machine learning methods result in alternate findings.

1.5. Research Question

The study is undertaken to determine the presence of tonality bias in South African online political reporting by utilising sentiment analysis, namely a lexicon-based approach and a semi-supervised learning approach.

The research question is as follows:

Is there computational evidence of tonality bias in South African online political news reporting over the period 2021 until 2023?

1.5.1. Research Aims

The study aims to address the gap between content and computational analysis in South African research and broadening the number of issues included in the scope of analysis. Through this, the study will ascertain how online news reporting contributes to shaping the national agenda and the framing of political actors and their respective parties.

1.5.2. Research Objectives

The objectives of the study are:

- Determine the presence of tonality bias in online political news reporting utilising a lexicon-based approach.
- Determine the presence of tonality bias in online political news reporting utilising a semi-supervised learning approach.
- Determine the presence of tonality bias in online political news reporting on protest action.

1.6. Limitations

The study is limited to online news media and events related to South African politics over the period January 2021 until June 2023. This includes seminal issues such as the 2021 Municipal Elections and the 2023 National Shutdown. The study is further limited to English online news media due to the current availability of sentiment tools and the utilisation of online news platforms. It is important to note that the chosen online media publishers have the highest levels of online traffic, with non-English media sites having significantly lower levels of online traffic.

The data is limited to the sample size which includes a total of 50 000 articles over the two-year period. The sample includes the top online news sites by online traffic which are News24, IOL, Times Live, the Daily Maverick and the Citizen (Similarweb, n.d.). These media sites will thus be the focus of the comparative analysis.

1.7. Overview

The next sections will detail the steps to be followed in the research process and the analysis to be undertaken. Chapter Two will detail the methodology of sentiment analysis using a lexicon-based approach, as detailed by Young and Soroka (2012), and the semi-supervised learning approach by Watanabe (2021). This will include a discussion of the limitations of each method and data considerations. Following the methods chapter, significant emphasis is placed on the results of the computational methods. The discourse will focus on the analysis of the outcomes of each method at a topic, and media publisher level to effectively address the question of tonality bias in political news reporting. Lastly, the conclusion reiterates the context and importance of the study in bringing about discourse on the topical issues of political reporting given the timing of the upcoming National Elections and media reporting on electioneering.

2. CHAPTER TWO– METHODS AND MATERIALS

2. Introduction

The focus of this chapter is to detail the methods followed, the process of data collection and processing as a prequel to answering the research question. The computational methods detailed in Chapter two are to be applied to ascertain the presence of tonality bias in South African online political news reporting, while also contrasting different sentiment methods. The chapter commences by detailing the research design in relation to the research question. This is followed by a theoretical discussion of the methodologies undertaken in the study and their relevance. Subsequently, an examination of the data requirements and data processing is undertaken in preparation for the analysis to follow.

The study utilises sentiment as the proxy for approximating bias as per the definition of tonality bias by Eberl et al. (2017):

Tonality bias: The framing of a political actor in a favourable or unfavourable light in comparison to other candidates which in turn impacts their electoral odds.

2.1. Research Methodology

2.1.1. Research design

The growth in the size, availability and complexity of unstructured data has expanded due to advancements in technology. This has led to the adoption of new methods in analysing atypical and unstructured data for meaningful insight. The study of text has grown vastly as a field to broaden the understanding of information and patterns embedded in text (Hossain et al., 2021). This very analysis of text data has grown significantly, particularly in the social sciences, through text mining and text analytics.

Text mining is defined as the process of applying computational techniques to analyse text data (Silge and Robinson, 2017). This includes the treatment of text as data as to transform, determine patterns and summarise (Silge and Robinson, 2017). Hossain et al. (2021) expands the concept further by defining text mining and analytics as the process of information retrieval through lexical or machine learning methods that allow visualisation and prediction. One such area of growth includes sentiment analysis, which utilises lexicon-based methods, and semi-supervised or unsupervised machine learning techniques.

Sentiment analysis provides the opportunity to determine the tone of political news content and hence bias and influence. The growth of political information in the age of digitisation provides researchers with the opportunity for increased discourse on the role and impact played by different political actors. Political news plays a significant role beyond the transmission of information and is said to be influential regarding elections and political party choices.

The study will thus be undertaken by applying text mining techniques to online political news articles to determine the validity of the bias assertions. The application of sentiment analysis will allow the determination of tonality bias.

2.2. Data

The study utilised the original unstructured text data from online news articles covering the period 01 January 2021 until 30 June 2023. A total of 50 000 articles were made available as per the breakdown detailed in figure 2.1. The articles were made available through the support of Media Monitoring Africa (MMA), an organisation aimed at promoting free, fair, and ethical media in South Africa (Media Monitoring, n.d.). The utilisation of a news repository and dedicated tool optimises

data collection and reliability by ensuring comprehensive coverage across all the online news platforms. The news articles are segmented by individual online media publisher.

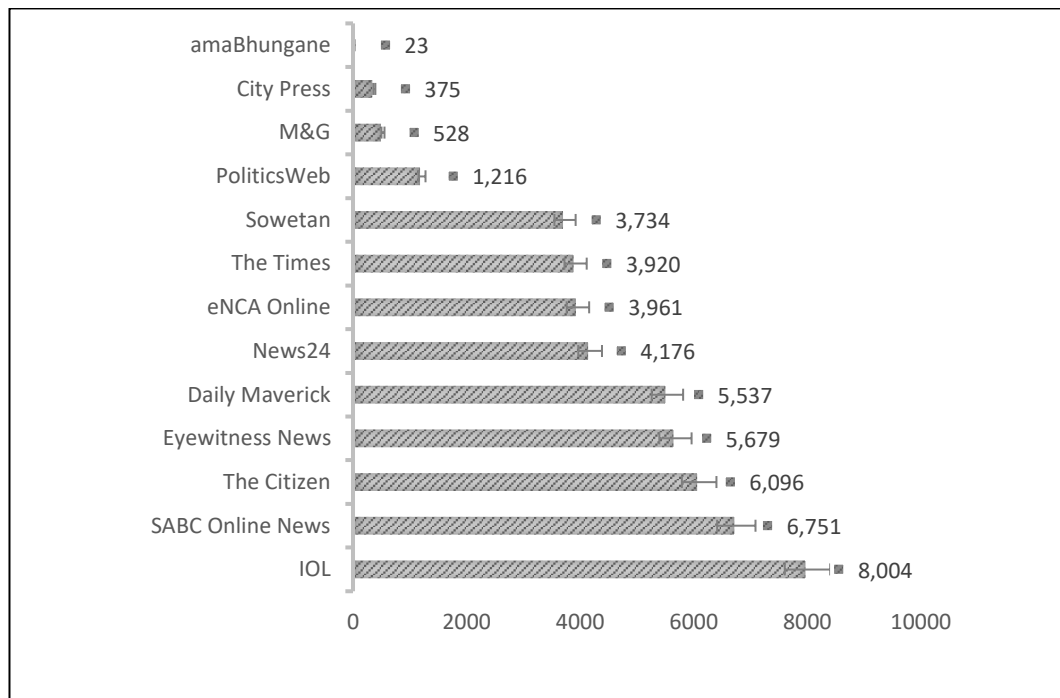


Figure 2.1: Breakdown of news articles by online media publisher

(Media Monitoring Africa, n.d.)

2.2.1. Data processing

The collection of news articles underwent pre-processing prior to transformation into text data for analysis. The process included the following:

- The removal of URLs, digits, and blank spaces.
- The removal of punctuation and common English words (stop words).
- The text was reduced to lowercase.
- The replacement of political party names with their acronyms.

Performing sentiment analysis on unstructured text data, required a process of tokenisation into tidydata i.e. the process of turning text into units (Silge and Robinson, 2017). The token takes the form of a word or a sentence for meaningful analysis (Silge and Robinson, 2017). Prior to applying any method, the text data was transformed sequentially into a corpus, tokens, and a document feature matrix as per figure 2.2. This was achieved using the package, quantitative analysis of text data (quanteda) (Benoit et al., 2018) in the program language R (R Core Team, 2022). The transformation of news articles into a corpus allowed for the capturing of text in a data frame. The subsequent tokenisation improved efficiency by transforming the corpus into vectors for pre-processing and positional analysis. The text data was further transformed into a document feature matrix (DFM) to analyse frequencies (Benoit et al., 2018).

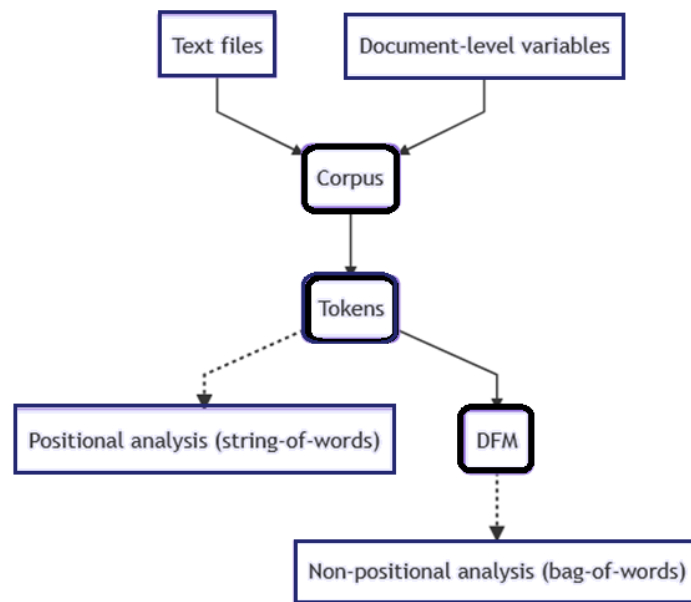


Figure 2.2: Text transformation

Benoit et al., 2018

2.2.2. Data processing: subset

The available dataset included articles across 13 online news media publishers, totalling 50 000 articles as provided. Table 2.1 represents the subset of leading online media publishers by online traffic and details ownership. The top five media publishers have been selected out of the 13 to be the focus of analysis of the commercially owned publishers. The SABC is included as it is the state-owned online news media publisher and will be contrasted for bias against commercial media publishers.

To locate the analysis in political news, the study focused on mining and analysing text data related to South African political leaders and parties over the period 1 January 2021 to 30 June 2023. The leading parties were identified as per their electoral presence. To contain the scope of analysis, the search items only included the heads of each party who were identified as President Cyril Ramaphosa (leader of the ruling party), John Steenhuisen of the Democratic Alliance (DA), and Julius Malema of the Economic Freedom Fighters (EFF) as per table 2.2.

Table 2.1: List of online media publishers and ownership

Online Media Publisher	Ownership	Commercial/Government
News24	Media24/Naspers	Commercial
IOL	Independent Media/Sekunjalo Group	Commercial
The Daily Maverick	No majority shareholder	Commercial
Times Live	Arena Holdings	Commercial
The Citizen	Caxton and CTP Publishers and Printers Ltd	Commercial
SABC	State-owned	Government entity

(Sanef, 2022; Wasserman, 2020)

Table 2.2: Data search criteria

Political head	Political affiliation	Search terms
Cyril Ramaphosa	ANC	"Ramaphosa", "Cyril Ramaphosa", "anc"
John Steenhuisen	DA	"Steenhuisen", "John Steenhuisen", "da"
Julius Malema	EFF	"Malema", "Julius Malema", "eff"
Protest	-	"protest", "shutdown", "loot*", "riot"

2.3. Sentiment analysis

2.3.1. Lexicon based approach: Lexicoder Sentiment Dictionary (LSD)

The analysis of tonality bias is undertaken utilising two sentiment methods. The first of these methods is a lexicon-based method. Sentiment analysis utilising a lexicon-based approach works by determining frequencies and occurrences of specific words in context. The processed text data is classified by polarity to be either negative, positive, or neutral through the application of a predefined dictionary. This process is performed using the packages *quanteda* (Benoit et al., 2018) in the programming language R (R Core Team, 2022). The focus of the lexical-based analysis is the application of the Lexicoder Sentiment Dictionary (LSD), designed for assessing political text (Young and Soroka, 2012).

The LSD was developed as a response to the growing volume of political information available through digitisation, and the need to develop an automated measurement for sentiment geared at political text (Young and Soroka, 2012). This was in line with the need to investigate the impact of political communication as a growing research field. At the time, many dictionaries were developed for specific or generic uses outside of political communication. Young and Soroka (2012) tested the dictionary against nine other lexicon dictionaries, and human coded news content; and found it performed better in analysing political text.

The version of the Lexicoder sentiment dictionary (LSD) applied is as per *quanteda* (Benoit et al., 2018) and consists of 2,858 “negative” words and 1,709 “positive” words. It further includes net negations words totalling 1721. These are excluded as they’ve been found to bring marginal benefit in improving sentiment scoring (Benoit et al., 2018).

Assessing sentiment as per the Lexicoder Sentiment Dictionary, required the pre-processing of the corpus, which was then tokenized and transformed into a DFM (figure 2.2). The LSD is subsequently applied to determine polarity from which net sentiment is calculated as a proxy of tonality bias.

The LSD has been proven effective in research as evidenced in the study of partisan bias in American economic news reporting (Merkley, 2019). The study examined bias in economic reporting between Democratic and Republican presidential terms. The outcomes evidenced positive bias towards Democratic presidencies for inflation and unemployment reporting (Merkley, 2019). Haselmayer (2021) extended the use of the LSD by applying its French equivalent. The adapted LSD was applied to a longitudinal dataset of political manifestos of French presidential electoral candidates over the period 1965-2017. This was to determine whether sentiment had grown increasingly negative over time. Haselmayer (2021), found that there was no evidence of increasing negativity in French elections and that the differences could be attributed to candidate attributes. The LSD has thus been shown to be effective in analysing both political content and bias.

2.3.2. Semi-supervised machine learning: Latent Semantic Scaling

Latent Semantic Scaling (LSS) is a semi-supervised machine learning model which comprises of both dictionary analysis and machine learning (Watanabe, 2021). LSS assigns both polarities i.e. sentiment as per the lexicon dictionary methodology and improves upon this by estimating proximity. Proximity is estimated by collocating the words using an independent scale (proximity to seed words¹). This allows for user defined dimensions of analysis instead of relying on a set dictionary and is a useful tool for longitudinal text data (Watanabe, 2021). LSS utilises a large corpus to estimate proximity of words through word-embedding and unsupervised algorithms to produce

¹ Small set of seed words

Positive	good	nice	excellent	positive	fortunate	wrong	superior
Negative	bad	nasty	poor	negative	unfortunate	wrong	inferior

word vectors. Thus, the LSS methodology allows for the analysis of a large corpus without the need for manual coding, due to only requiring a small set of seed words.

The methodology of LSS requires the following:

- a sizeable corpus: “The sizes of corpora required for LSS depends on the type of texts and measurements, but it is usually 5,000 to 10,000 documents for news articles which is roughly 200,000 to 400,000 sentences” (Watanabe, 2021, p. 86). The study includes 50 000 news articles in the sample.
- a small set of seed words: “Ideal seed words are subject-independent generic words in order to achieve their portability across projects” (Watanabe, 2021, p. 86). Seed words utilised in sentiment analysis are required to have negative and positive polarity to provide the weighting for the words in the corpus. The sentiment score of the subsequent text is calculated as a function of relative proximity. The study utilises a small set of generic seed words as per quanteda ((Benoit et al., 2018).
- a choice of model terms that are relevant to the target dimension.

The LSS methodology works similarly to that of the unsupervised machine learning techniques as seen in word embeddings by determining semantic proximity between words. A singular value decomposition is applied to a document feature matrix to factorise it into three components to obtain word vectors, $v_1, \dots, v_f$ (Watanabe, 2021). Once word vectors are obtained, the LSS computes polarity of words, $g_1, \dots, g_f$ based on a small set of seed words ($s \in S$) given their weighted proximity (ρ_s) (Watanabe, 2021, p. 87). The cosine is thus a measure of the similarity between word vectors.

$$g_f = \frac{1}{|S|} \sum_{s \in S} \cos(v_s, v_f) \rho_s$$

Watanabe, 2021, p. 87

Whilst the LSS is a new development, it is increasingly being utilised in political communication studies. Brändle and Eisele (2023) examined European border communication over the period 2009-2020, with governments increasing border control measures during crisis. An examination of Germany and Austria revealed that both governments evidenced a negative sentiment trend over time to border openness. The Austrian government focused on border protection and the German government focused on the transborder crisis in their communication. An alternate application by Rauh (2022) examined the tonality of speeches given by supranational actors in the European Union. Utilising the LSS, Rauh (2022) found that supranational actors utilised alarmist tonality in their executive speeches, particularly executive members of the European Central Bank (ECB). The study found evidence of supranational actors pro-actively engaging in emergency politics (Rauh, 2022). The utilisation of the LSS has a growing body of literature due to its applicability in political communication, only requiring a small set of seed words to compute a machine learning model.

2.3.3. Latent Semantic Scaling and Ordinary Least Squares regression

Watanabe (2017) further proposes the utilisation of ordinary least squares (OLS) regression. The method is analogous to a difference in differences design on continuous sentiment scores, to measure changes in news bias. This requires the identification of pivotal events ($e_1 \dots e_8$) as change points in the dataset. The continuous LSS scores are thus regressed on the chosen identified indicators, following the pivotal event to determine the presence of political bias (Watanabe, 2017). A dummy variable is created for the issue of interest and its interaction with the change point is captured as $e_1 g \dots e_8 g$ and the coefficients ($\gamma_1 \dots \gamma_8$) are thus estimated changes in bias as they capture tonality bias.

$$Y = \beta_0 + \beta_1 e_1 \dots \beta_8 e_8 + \delta_g + \gamma_1 e_1 g \dots \gamma_8 e_8 g + u + \varepsilon$$

2.4. Analysis

The subsequent chapters will detail the outcomes of the above methods as applied to the transformed text data. The analysis will be split by methodology. This will entail detailing the outcomes of the lexical sentiment approach, followed by the semi-supervised machine learning approach. As a precursor to the analysis, a descriptive analysis will be performed on the text data.

2.4.1. Descriptive analysis

As a fundamental first step, a descriptive analysis is undertaken to provide an overview of the data, by examining the distribution of articles between individual media publishers over the period 2021 until 2023. This is performed to ensure that the sample is representative of the leading commercial publishers as determined by online traffic. The descriptive analysis will then focus on examining word frequency within the dataset. This will be utilised to examine congruency or sameness across media publishers. Important to the descriptive analysis is establishing similarity across publishers as a point of departure in the discourse on tonality bias. A significant level of sameness will amplify tonality bias across the political reporting spectrum.

Frequency and visibility are an important factor in the discourse and term frequency will be examined as an aggregate and by individual media publisher. Once the descriptive analysis is established, it will be followed by the statistical analysis with a focus on tonality as a measure for bias present in political reporting.

2.4.2. Statistical analysis

The statistical analysis is performed in two parts, by utilising both a lexical-based approach and machine learning approach. The lexicon-based approach utilises a dictionary developed for political text, the Lexicoder Sentiment Dictionary, to determine the tonality. The lexical sentiment analysis will be split by political party and political leader. The study will thus examine sentiment by online media publisher towards political parties and leaders over the period 2021 to 2023 to establish the presence of tonality bias. A similar approach will be applied in examining the results of Latent Semantic Scaling, a semi-supervised method. Utilising two methodologies allows for a comparative analysis and an examination of consistency.

A third component of the analysis will be examining tonality bias directed at protest action within media reporting. This is a key aspect of the analysis as protest action has been a tool for the marginalised to voice their discontent and bring attention to their needs.

Lastly, a regression analysis will be applied to test for significance in tonality bias in political news bias reporting in South Africa. This will test for both party and political leader tonality bias by utilising significant event change points, identified through the analysis of the Lexicoder Sentiment Dictionary, combined with the findings of the Latent Semantic Scaling.

The analysis should thus give a view of overall visibility and tonality, to determine the extent of bias in South African online political news, and a comparative understanding across selected online media publishers. The LSD and LSS results will thus be combined to make a determination of the extent and presence of tonality bias in South African media publishers, and the impact across political parties and their leaders.

2.5. Limitations

Sentiment analysis has gained popularity given the increase in the availability of unstructured text data and the need to decode the meaning of communication. This is amidst increasingly polarising global politics and multiplying media sources. However, there remains concern of the cross application of dictionary lexicons without consideration of the applicability within a domain. A

further constraint is the availability of dictionaries across different languages with extensive work required to adapt dictionaries.

Ribeiro et al. (2016) applied a quantitative approach in reviewing the many sentiment analysis methods by testing performance between three (negative, neutral, and positive), and two (negative and positive) classes of polarity methods. Different lexicons performed poorly with certain datasets whilst performing strongly in others across diverse fields. Ribeiro et al. (2016) concluded that there is a trade-off in predictive ability and researchers need to find the best fit for their research dataset and field.

Of further contestation is the utilisation of lexicon methods versus machine learning based methods in analysing sentiment. Drus and Khalid (2019) contrasted differing methodologies to determine their strengths and weaknesses by way of a literature review. The lexicon-based methods were found to be computationally faster, more adaptable to different datasets and simpler. The machine learning based methods required training data sets and were computationally time consuming without necessarily improving efficacy (Drus and Khalid, 2019).

An important consideration in the application of dictionary-based sentiment analysis is the scope of the words available and their importance in determining polarity. Although English language dictionaries were found to be accurate, they were limited by scope in comparison to their machine learning counterparts (Haselmayer and Jenny, 2016). However, machine learning methods exhibited higher error rates. Young and Soroka (2012), determined a need for domain specific dictionaries, this may require extensive effort particularly transferring across languages.

In this study, a dictionary-based methodology and a semi-supervised machine learning methodology are utilised. To overcome the issue of domain specificity, the Lexicoder Sentiment Dictionary (LSD) by Young and Soroka (2012), developed for political text was chosen for sentiment dictionary analysis. This is contrasted against the Latent Semantic Scaling (LSS), proposed by Watanabe (2021) for expanding sentiment analysis given the growth in demand across languages, regions, and domains. The advancement in technology and software allows for the processing of large corpus without the loss of dimensionality encountered in utilising dictionary analysis. The LSS has proven to perform as well as the LSD in the English language, and in a language where the LSD was unavailable (Watanabe, 2021). Thus, these methods are chosen to overcome the limitations on sentiment analysis by applying a domain specific lexicon which is contrasted against a semi-supervised machine learning method.

Whilst the methods are used comparatively, there is a lack of a measurement to determine performance across the two methods to make a determination on the best fit on the outcomes for the study.

2.6. Ethical Considerations

The data was sourced from a secondary supplier, the MMA organisation. No risk is identified in the study and no ethics clearance or waiver is required as human participants are not involved or impacted. Individuals who are not publicly or politically known will not be identified in the research output.

3. CHAPTER THREE- DESCRIPTIVE ANALYSIS

3. Introduction

Chapter three details the descriptive analysis of text data collected from 50 000 articles over the period 2021 to mid-2023. The chapter begins by presenting the aggregated results from examining the topics of word frequency and article distribution. The analysis is followed by a test for text similarity across online media publishers and lastly, a discussion of the findings. The analysis was undertaken in the programming language R (R Core Team, 2022), utilising the package quanteda (Benoit et al., 2018).

3.1. Descriptive Analysis

The dataset contains 50 000 articles from 13 online news publishers over the period 1 January 2021 to 30 June 2023. The articles are cross cutting and are not limited to political news thus a search function was utilised to refine the articles for political news.

3.1.1. Distribution

The distribution of 50 000 articles is skewed right, with the majority of the sample from the period 2021 and 2022, totalling 80 percent, as per table 3.1. The IOL had the highest number of articles included in the samples totalling 8004, followed by the SABC and the Citizen with 6751 and 6096 articles respectively. The Daily Maverick and News24 came in at position five and six with 5537 and 4176 articles respectively. The online news media publisher, Times Live, is underrepresented in the sample, coming in at the eighth position with only 3920 articles, as a top five competitor.

Table 3.1: Distribution of articles

Publisher/Year	2021		2022		2023		Total	
	%	Number	%	Number	%	Number	%	Number
amaBhungane	61%	(14)	39%	(9)	0%	(0)	100%	(23)
City Press	28%	(104)	44%	(166)	28%	(105)	100%	(375)
Daily Maverick	39%	(2,187)	41%	(2,260)	20%	(1,090)	100%	(5537)
eNCA Online	42%	(1,668)	41%	(1,630)	17%	(663)	100%	(3,961)
Eyewitness News	38%	(2,184)	41%	(2,340)	20%	(1,155)	100%	(5,679)
IOL	45%	(3,600)	39%	(3,091)	16%	(1,313)	100%	(8,004)
M&G	31%	(164)	40%	(212)	29%	(152)	100%	(528)
News24	33%	(1,363)	34%	(1,420)	33%	(1,393)	100%	(4,176)
PoliticsWeb	41%	(501)	35%	(430)	23%	(285)	100%	(1,216)
SABC Online News	45%	(3,005)	39%	(2,600)	17%	(1,146)	100%	(6,751)
Sowetan	47%	(1,737)	36%	(1,359)	17%	(638)	100%	(3,734)
The Citizen	32%	(1,973)	46%	(2,804)	22%	(1,319)	100%	(6,096)
The Times	42%	(1,635)	38%	(1,479)	21%	(806)	100%	(3,920)
Total	40%	(20,135)	40%	(19,800)	20%	(10,065)	100%	(50,000)

3.1.2. Frequency analysis

A frequency analysis was performed across all articles to determine the leading themes in South African news reporting over the period of analysis. This yielded insight into the prevalence of South African political news in daily reporting across South African online media publishers. Figure 3.1 details a list of the top 20 most frequent terms utilised across the 50 000 articles, following text preprocessing.

As per figure 3.1, political news was as a dominant feature of discourse in news reporting, particularly in relation to governance and the ruling party. This was evidenced in the prominence of terms such as “government”, “state”, “president”, “court” and “police” in the top ten most featured words. The ruling party was the only political party to feature in the most frequent terms.

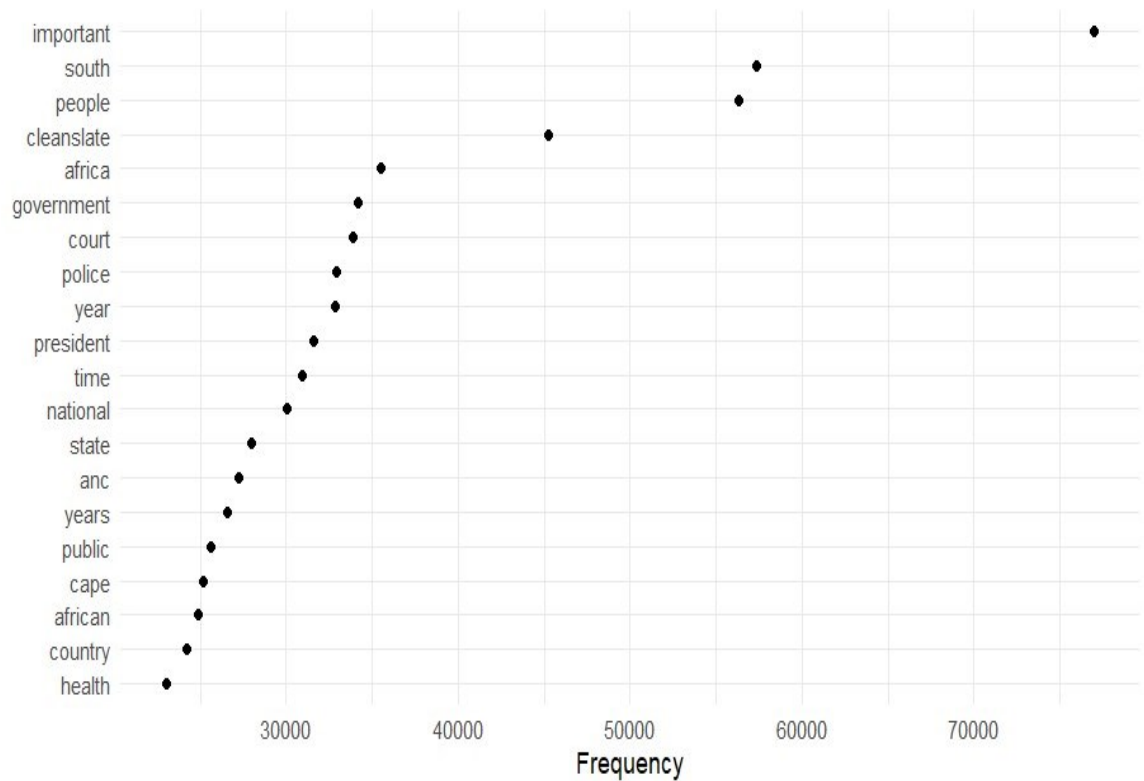


Figure 3.1: Frequency chart aggregated across all publishers

3.1.3. Frequency Analysis: media publisher

A frequency analysis is estimated by individual media publisher to determine the prevalence of political news at a disaggregated level (figure 3.2). The IOL, which accounts for the highest number of articles in the sample, evidences a diverse frequency plot in comparison to other online news publishers, with governance related terms not featuring in the top five most frequent terms. This was also the case for the Citizen. However, the Citizen is among the few online media publishers to feature the president by name in their frequency plot. The state media publisher’s frequency plot revealed a greater focus and visibility of government related terms such as “president” and “government” as the second and fourth most frequent words respectively. The frequency plots by individual online media publishers largely align with the aggregate frequency plot as the terms are chiefly governance related. A strong focus of South African media appeared to focus on the actions of government, the president and matters related to South African politics. These terms garner the greatest visibility as per the frequency plots.

the content is dictated by editorial and commercial interests rather than the readership base. Overall, South African online media publishers bare a high level of text similarity with high correlation statistics. This is of concern with consumers having limited access to varied content to establish independent decision making. This could be a sign of bias skewed by editorial or consumer beliefs of the white minority elite.

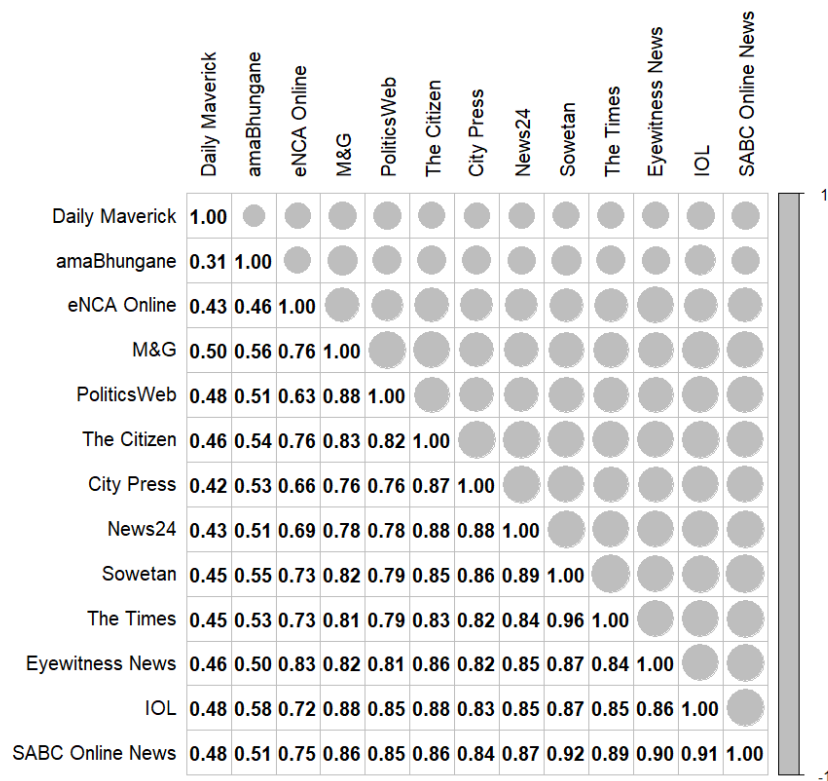


Figure 3.3: Text similarity correlation plot

3.1.5. Political visibility

The visibility of political actors is posited as important, particularly during in an election cycle, to garner support and convey political agendas to the electorate. In this vein, a frequency analysis of was undertaken of both political parties and leaders. It revealed that the ruling party and the president had significantly higher mentions in online media publications, both at the party and individual level, as per table 3.2. Although it is expected that the ruling party and the president to have higher visibility, this is not always good news. This may result in increased bias, if increased visibility is accompanied by increased negative tonality. The leader of the EFF, Malema, had marginally higher visibility than that of fellow opposition party leader, Steenhuisen. However, the DA had a higher frequency count as a party over the EFF.

Table 3.2: Frequency of identified political terms

Mentions	ANC	EFF	DA	Total
Party	53.42	18.58	28.00	11397
Leader	80.14	11.81	8.05	7220

Following the above descriptive analysis evidencing issues of congruence and similarity, a lexical-based sentiment methodology was utilised. The discussion on tonality bias across online media publishers is expanded in the subsequent chapter.

4. CHAPTER FOUR- LEXICODER SENTIMENT DICTIONARY ANALYSIS

4. Introduction

Chapter four details the data analysis stemming from a lexical-based methodology, the Lexicoder Sentiment Dictionary (LSD), as detailed in chapter 2. The LSD was selected due to its domain specificity as it was developed for political text (Young and Soroka, 2012). The chapter begins by presenting the results from the application of the LSD, followed by a discussion of the findings. The text data is analysed to determine the presence of tonality bias in South African online political news reporting utilising sentiment analysis.

The descriptive analysis in chapter three evidenced congruence amongst South African online media publishers with high text correlation. To focus methodically on political news, party and political search terms are utilised as per table 2.1 in chapter 2 for the lexical-based analysis. These were identified as per the top three political parties and their respective leaders in South African politics.

Thus, the main themes of chapter four will examine the following:

- The LSD is utilised to examine systemic bias across all newspapers within the entire sample for the selected political parties and their respective leader. This is to establish whether South African online news publishers express the same patterns of bias as a whole, and to determine whether there are winners and losers.
- The analysis is subsequently devolved into the five top online news publishers to determine the presence of bias by individual publisher for individual political parties and their respective leaders.

Lastly, the discussion will contrast the above findings in determining the presence of bias at a whole versus an individual publisher role within online news media in South Africa.

4.1. Is Bias systematic?

4.1.1. The African National Congress and party leader, President Cyril Ramaphosa

The LSD was estimated utilising separate search terms to test for tonality bias and leader effects for each party leader and their associated political party. The outcomes are depicted in the LSD charts which track net sentiment over time as a proxy for tonality. Net sentiment is calculated as total positive sentiment less total negative sentiment. Figure 4.1 depicts the net sentiment chart for the president of the republic, President Ramaphosa juxtaposed against that of his political party, the ANC. President Ramaphosa's sentiment graph over the period 2021 until mid-2022 is largely positive and trends above zero. This is aligned with the period in which South Africa began to slowly recover from the world epidemic, Covid-19, and the lifting of government restrictions. The dip in sentiment in 2021 is aligned with the social unrest and riots which resulted in the loss of life and brought the country to a halt. Overall, net sentiment is at its lowest in the period approximating the halfway mark of 2022 and again in the period ending 2022. The sharp decline is likely attributable to the failed no confidence motion tabled in parliament on March 30th, 2022 (Ferreira, 2022). This was at the behest of the leader of the opposition, Steenhuisen, tabling a motion of no confidence during the state of nation in February 2022, accusing Ramaphosa of failing to lead the country, by handing over his duties to the private sector (Masson, 2022). Towards the period ending 2022, the president avoided impeachment, however his net sentiment going forward does not recover.

In comparison, the ruling party's net sentiment graph is less volatile than that of its leader. It is also less negative than that of the President with higher peaks and lower troughs. In the period preceding the 2021 Municipal Elections, the net sentiment of the party declines, only to recover post the

elections. This is indicative of negative political news reporting, particularly close to an election and possible negative tonality bias. However, this does not appear to have been the case for the president. The president and the party exhibit substantially different sentiment charts and leader effects appear to be present as evidenced with sentiment disparities.

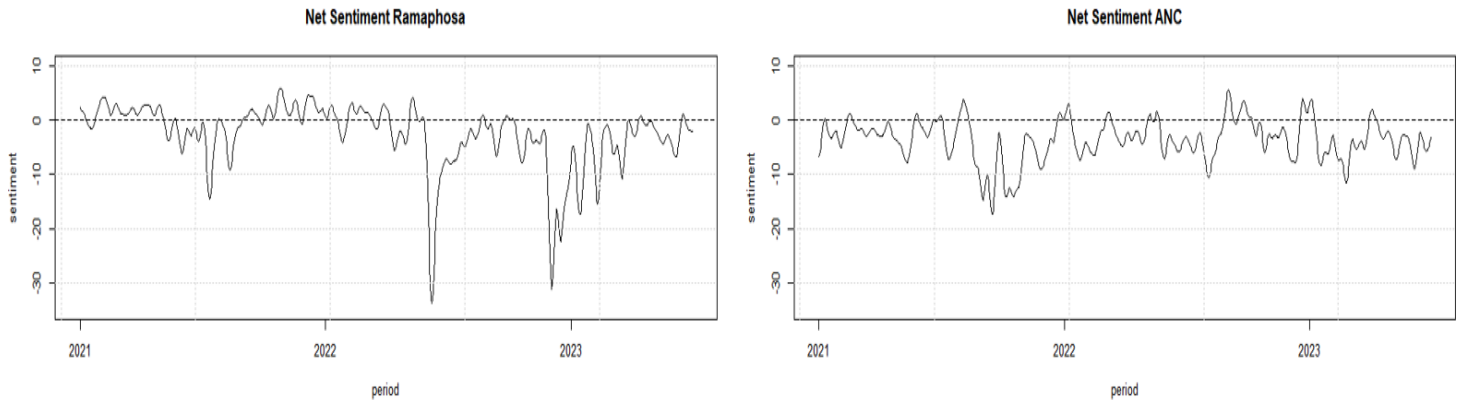


Figure 4.1: LSD Sentiment Chart: President Cyril Ramaphosa and the ANC

4.1.2. The Democratic Alliance and party leader, John Steenhuisen

The leader of the DA, Steenhuisen has lower visibility in the media in comparison to his political home, the DA. His net sentiment chart is muted in comparison to his party as per figure 4.2. The party leader's net sentiment chart rises close to the 2021 Municipal Elections and is sustained until quarter one of 2022 when it dips into negative territory.

The DA sentiment chart mimics that of the party leader, with greater volatility. The party peaks and troughs are larger than those of Steenhuisen but remain aligned, an indication of the absence of leader effects. Party sentiment peaks in the third and fourth quarter of 2021 and in the first to second quarter of 2023. The fourth quarter peak in 2021 can be related to the 2021 Municipal Elections with the party likely receiving positive political coverage during the election cycle. The skew in sentiment during an election has been proven to impact voting patterns, particularly partisan votes, and this appears to be an endorsement for the party and its political leader.

The rise in sentiment in the second quarter of 2023 can once again be related to the internal political processes of the DA receiving positive political coverage. The timing of the rise in sentiment is estimated to be aligned with the federal congress of the party. Ebert et al. (2017) defines bias in three parts, which includes the ability of political parties to convey their political agenda's positively in the press as evidenced with the DA. The party appears to benefit from both positive tonality and agenda-setting with online media publishers.

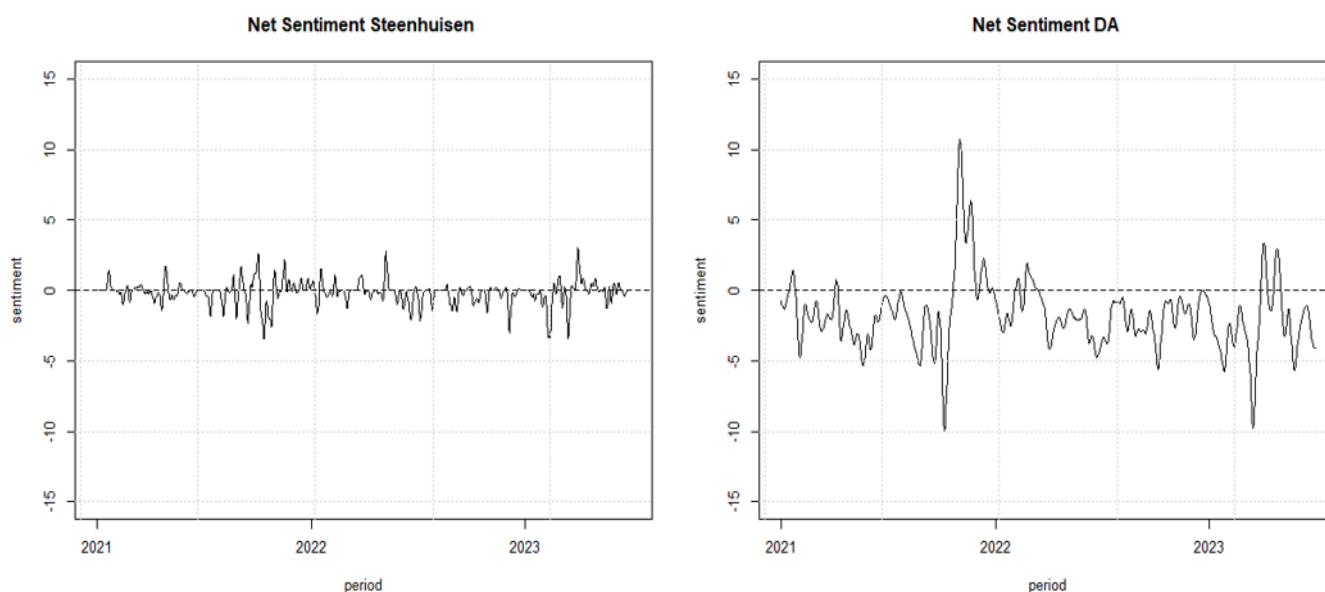


Figure 4.2: LSD Sentiment Chart: John Steenhuisen and the DA

4.1.3. The Economic Freedom Fighters and party leader, Julius Malema

Malema and the EFF sentiment charts are largely negative over the period of analysis as depicted in figure 4.3. As the leader of the EFF, he makes marginal positive gains prior to the elections in the latter half of 2021 but this sharply declines in early 2022. This represents the highest peak in sentiment for the leader prior to a sustained decline. A secondary dip in net negative sentiment is realised in the first quarter of 2023 and this is aligned with the EFF led National Shutdown of 20 March 2023. On the 29th of January 2023, Julius Malema had announced a nationwide shutdown set for the 20th of March 2023, as a call to the end of the protracted loadshedding experienced by South Africans due to the embattled parastatal power utility, Eskom, and as a call for the resignation of the president of South Africa, Cyril Ramaphosa, led government (EFF South Africa, 2023). As evidenced above, the EFF fails to achieve positive agenda-setting in the media, as evidenced with negative tonality bias.

The net sentiment of the EFF is congruent to that of party leader, Malema. Whereas the party and its leader net sentiment charts mimic each other, he has improved net sentiment outcomes. Without the party leader, the EFF fares more negatively with lower troughs in sentiment. Malema appears to have marginal leader effects as he evidences improved sentiment over his political home, the EFF.

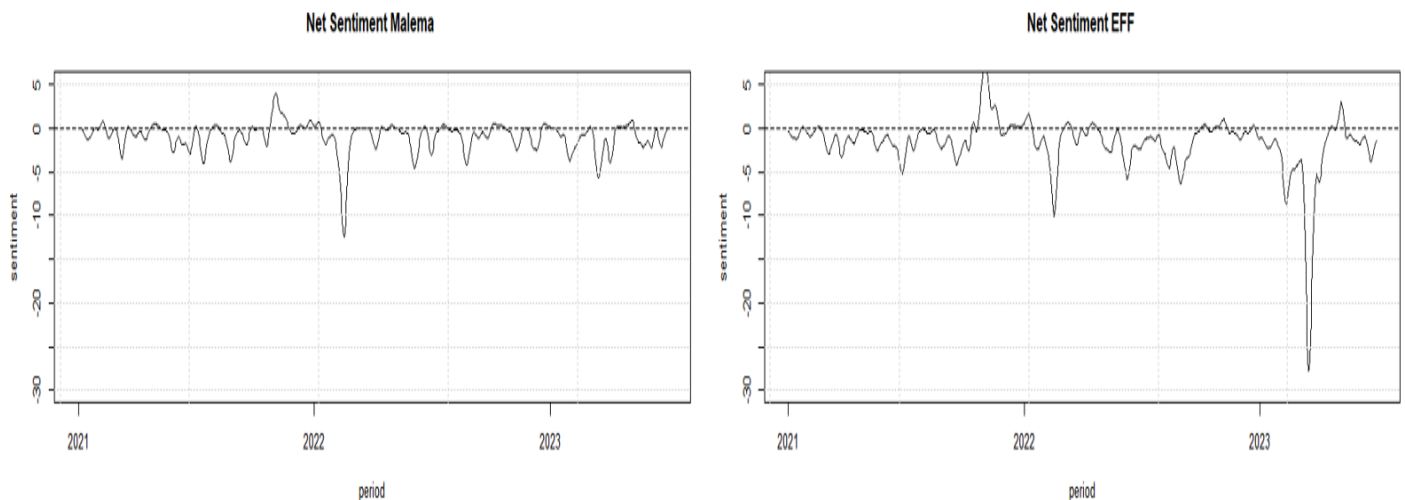


Figure 4.3: LSD Sentiment Chart: Julius Malema and the EFF

4.1.4. Bias across political lines

A comparative analysis is computed across political lines to ascertain trends in net sentiment and to determine the presence of tonality bias. Figure 4.4 depicts the LSD charts of all political parties and political leaders. The net sentiment charts are depicted as per a standardised scale to allow for comparative ease.

The first row of charts in figure 4.4 depict the net sentiment of political parties over the period of analysis. The ANC, has the most volatile net sentiment over the period, followed by the EFF. The DA outperforms all the parties with the highest peak in sentiment and least net negative sentiment. The DA is the most favourably depicted political party by online media publishers, on aggregate. The EFF fares poorly, with sentiment largely below the zero margin over the period and experiences a sharp decline in sentiment in line with the National Shutdown of early 2023. Figure 4.4 evidenced that the two most populist parties attained the highest level of negative tonality with bias skewed towards the more market friendly political party.

In turn, a comparative analysis across political leaders depicts a more disparate outcome with the president of the republic, exhibiting a significantly volatile and net negative sentiment chart, as depicted in figure 4.4. Malema, leader of the second largest opposition party, the EFF fares marginally better than the president but not as well as the leader of the DA. Steenhuisen of the DA, has a markedly different net sentiment outcome as compared to that of Ramaphosa with, sentiment over the two- and half-year period largely trending positively with a smaller bandwidth around the zero margin. However, he has lower visibility in the media in comparison to the president, overall. Whilst the leader of the DA has cyclical sentiment, it is a significant improvement to his biggest rivals, averaging between two and negative two. Malema of the EFF exhibits a comparative advantage in net sentiment over his political party. Conversely, President Ramaphosa, fares poorly in comparison to his party with online media publishers.

The comparative analysis indicates that on aggregate the tonality is positively biased towards the DA, and against President Ramaphosa, the ANC, and the EFF. The president of the republic is the only leader to evidence distinctive leader effects with his sentiment chart in contrast with that of his political home.

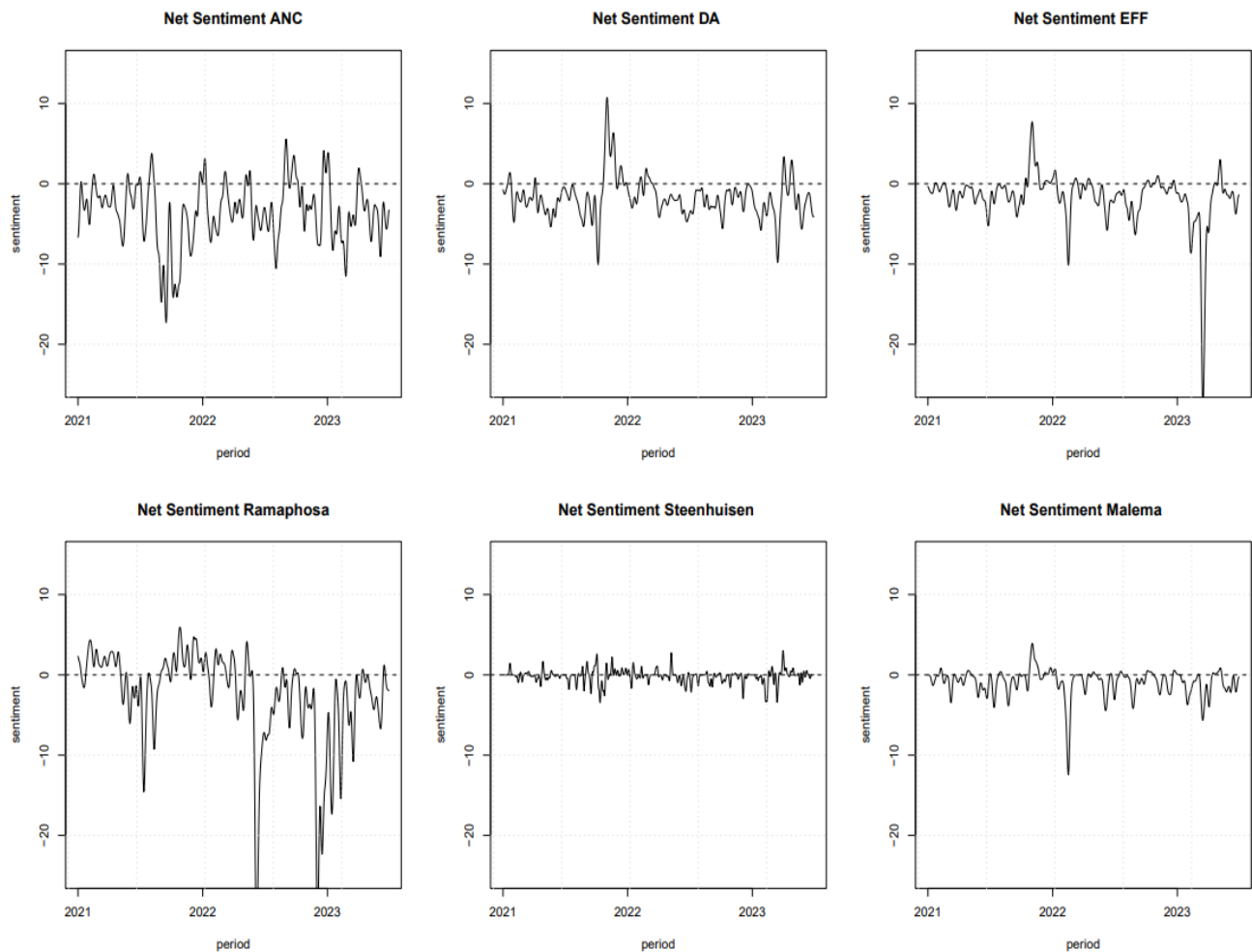


Figure 4.4: LSD Sentiment Chart: Comparative analysis across party lines

4.1.5. Frequency and tonality

The lexical-based sentiment analysis utilising the Lexicoder Sentiment Dictionary painted disparate representations of how the media interacts and writes about political parties and their leaders. Delving deeper, the top negative and positive words were extracted for political leaders as detailed in table 4.1. Notably, from table 4.1, is the significantly large gap between the top three negative terms and the top three positive terms for Ramaphosa. The top negative terms associated with Ramaphosa depict him as a crooked leader with terms such as “corruption”, “crisis” and “allegations” outpacing his top five positive terms. In the same vein, the leader of the EFF, also has strongly negative top terms such as “violence”, “kill”, and “hate”. Steenhuisen receives more favourable online media representation and characterisation but lacks overall visibility.

As per table 4.1, Ramaphosa has the highest visibility which is in line with expectations as the head of the ruling party and president of the republic. However, the leader of the second largest opposition party, Malema, has greater visibility than that of the leader of the leading opposition party, Steenhuisen. This, however, does not aid his political aspirations as the tonality is negatively skewed against the leader. As per theory, strong negative tonality towards the leader will heighten negative leader perceptions. Thus, the negative tonality likely impacts the president and the leader of the EFF most significantly.

The analysis of net sentiment depicts a picture that is pro the opposition and decidedly free markets or neoliberal. The critique of tonality bias begins to gain credence and validity, with the prominence

of skewed positive sentiment for the DA versus the two “populist” and African led parties. However, there is margin for contention with the leader of the EFF having less negative tonality than that of his political party.

Table 4.1: Political leader top term frequency

Ramaphosa				Steenhuisen				Malema			
Positive	Frequency	Negative	Frequency	Positive	Frequency	Negative	Frequency	Positive	Frequency	Negative	Frequency
support	288	against	810	like	29	opposition	35	supporters	81	against	151
well	164	corruption	293	support	18	against	31	like	51	accused	55
special	163	crisis	253	freedom	15	challenge	16	support	50	violence	49
confirmed	151	allegations	237	right	15	crisis	16	open	29	kill	41
like	136	criminal	193	well	14	corruption	15	well	24	warned	37
ensure	124	disaster	161	confidence	13	failed	13	freedom	20	assault	37
confidence	124	suspension	156	better	13	disaster	13	free	17	hate	30
protector	120	violence	144	good	11	lost	12	right	17	alleged	27
help	91	accused	131	open	11	seat	11	apologise	17	charges	26
peace	91	alleged	124	accountability	11	accused	10	good	15	denied	26

4.2. Lexical-based sentiment analysis by media publisher

To further the discourse on tonality bias, the Lexicoder Sentiment Dictionary is applied to select online media publishers. The 50 000 articles are subset by publisher, and the lexical-based sentiment analysis is evaluated between the top commercial publishers versus the state-owned publisher.

Utilising the distribution of articles available within the sample and internet traffic by site, the section focuses on the leading online media publishers. These were earlier identified as News24, IOL, Times Live, The Daily Maverick and The Citizen. They are contrasted against the state-owned online media publisher, the SABC, in ascertaining whether the state utilises their own media publisher to drive a pro state agenda and bias. Each online media publisher underwent lexical-based sentiment analysis by political party and political leader. The analysis separated political leaders from their political parties given the diverse results from the above analysis.

4.2.1. Online news media and the ANC

The ruling party had the highest visibility as per the frequency of mentions across media publishers. The *a priori* expectation was for the party to be polarising and have negative sentiment across many of the publishers given its history with the media. As per figure 4.5, the results varied across media publishers. The state-owned publisher, the SABC, stood apart and exhibited a consistent trend of positive tonality. The party received low and negative patterns in sentiment across multiple publishers with the Citizen, and News24 leading negative tonality. Net sentiment preceding the 2021 Municipal Elections remained net negative for the ruling party among many online media publishers. This is in direct contrast to tonality post the elections across all media publishers. Overall, there was little joy to be found in the ruling party across many commercial media publishers and the state-owned media publisher played a mitigating role. Although the sentiment graph for the SABC is more muted than those of other publishers, it has the lowest levels of net negative sentiment, and many periods of positive tonality. The highest spike in net positive sentiment was registered by the Times Live in mid-2022, whilst many of the commercial online media publishers trended close to the zero threshold. The extent of negative tonality was inconsistent amongst online media publishers.

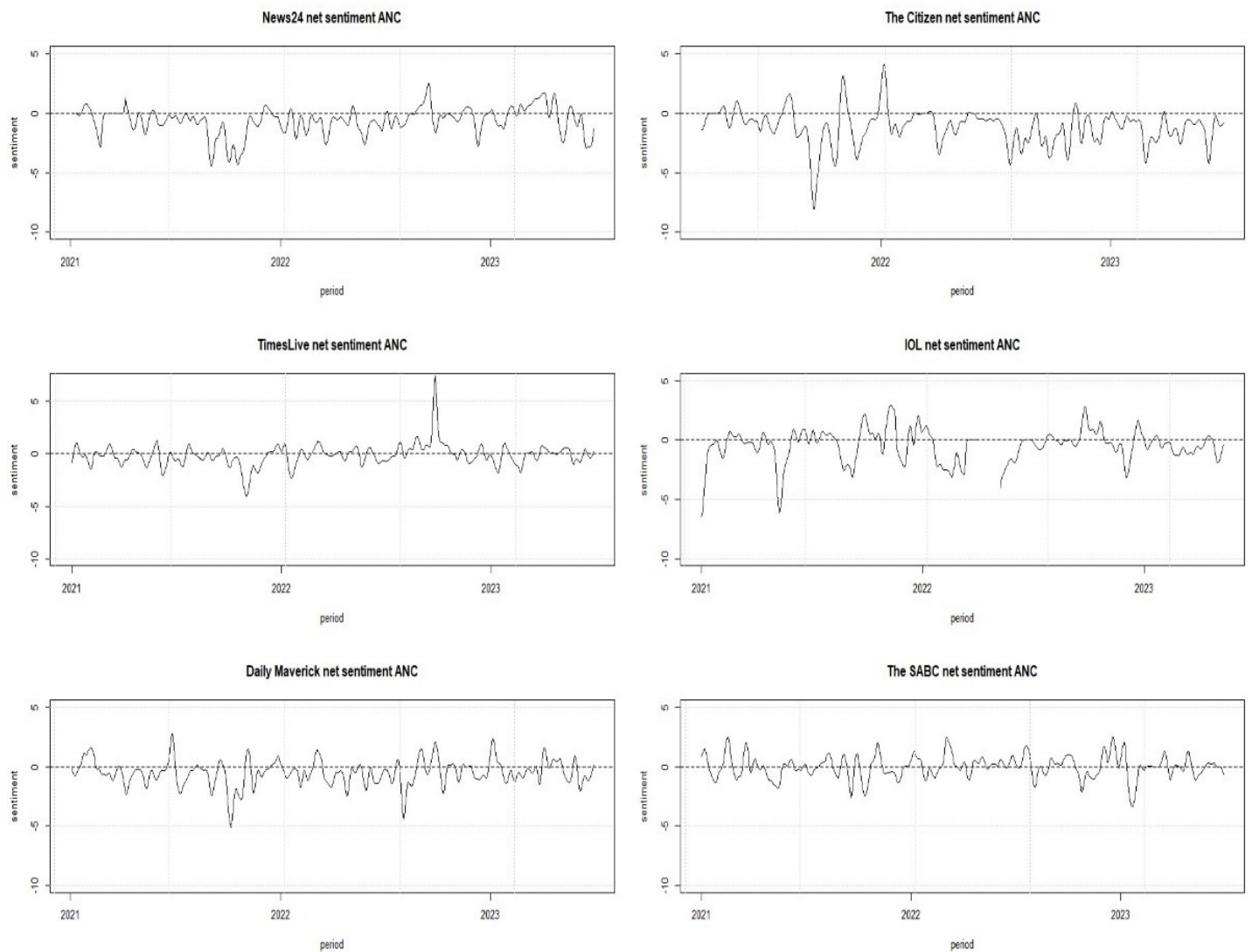


Figure 4.5: LSD Sentiment Chart by publisher: ANC

4.2.2. Online news media and the DA

The sentiment analysis outcomes of the DA are largely consistent across commercial online media publishers and congruent with the state-owned online media publisher as per figure 4.6. Overall tonality is muted in comparison to that of the ANC. The party is marginally represented in early 2021 by commercial online media publishers with a largely flat sentiment chart. Party sentiment gains momentum in the latter half of 2021 as the party is positively depicted by many commercial news media publishers, led by the IOL and the Daily Maverick. The uptick in sentiment is in time with electioneering for the 2021 Municipal Elections. There are, however, distinct disparities amongst online commercial media publishers in their tonality towards the DA over the period of analysis. The Citizen exhibits a net negative sentiment towards the party whilst The Daily Maverick and Times Live appear to have a total net positive tonality towards the party. The state-owned media publisher sentiment chart goes against *a priori* expectations and is net positive over the period of analysis. Overall, sentiment towards the party trends close to zero and neutrality. Although the charts exhibit significant volatility, net sentiment peaks and lows are an improvement in comparison to opposition parties.

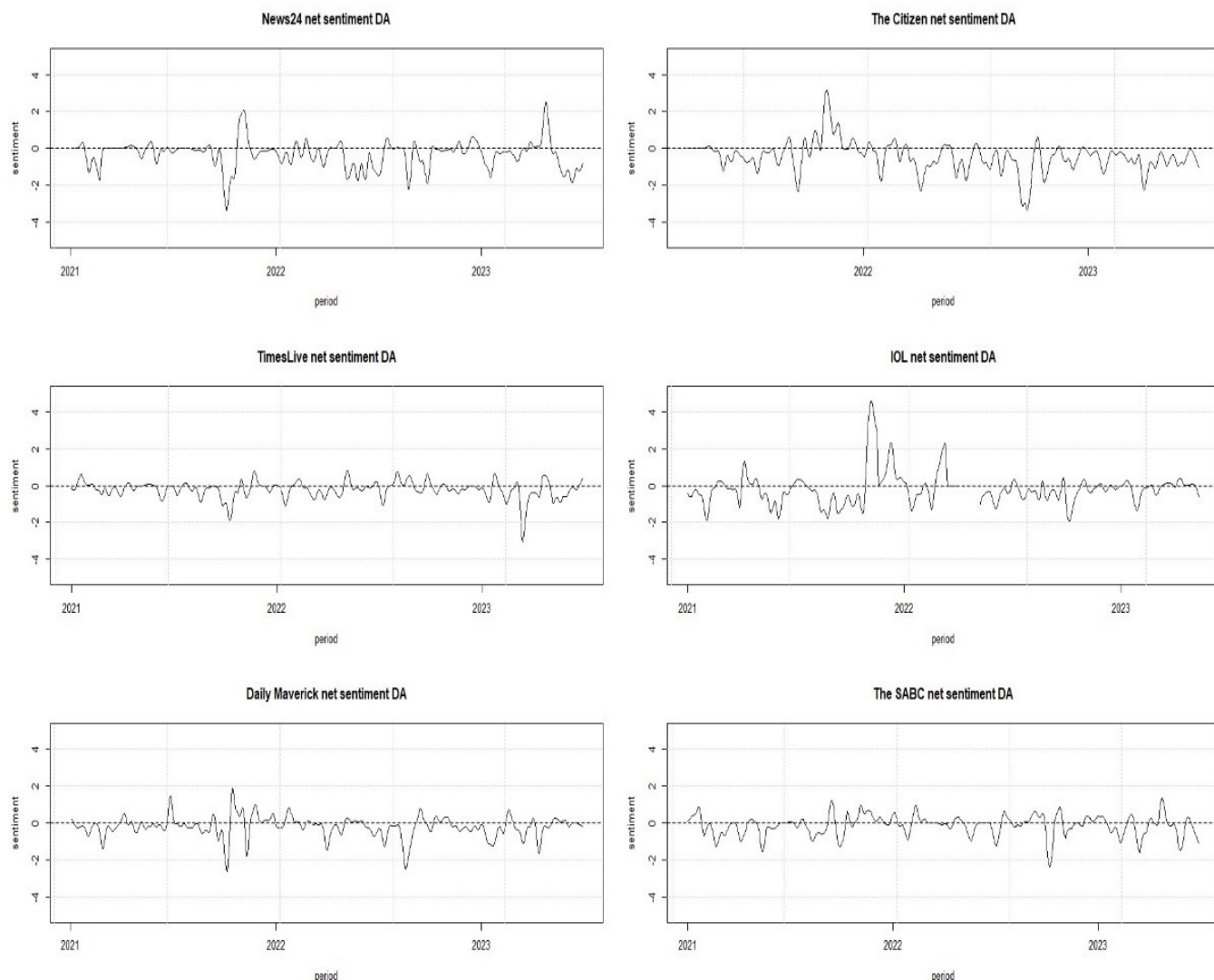


Figure 4.6: LSD Sentiment Chart by publisher: DA

4.2.3. Online news media and the EFF

The EFF sentiment charts buck the trend and attract consistently negative sentiment from most online commercial publishers as depicted in figure 4.7. The SABC, the state-owned media publisher, is more vocal in its negative sentiment towards the EFF unlike towards the DA. The Citizen and the IOL are the biggest detractors with significant net negative sentiment charts. News24 and the Daily Maverick evidence the least negative tonality bias with instances of peaks in positive sentiment geared at the EFF. The largest dip in tonality for the party occurs in 2023 and is aligned with the timing of the National Shutdown on 20 March 2023. This is indicative of the misalignment of the politics of the party and the political opinions of commercial media publishers, hence the consistent trend in net negative tonality.

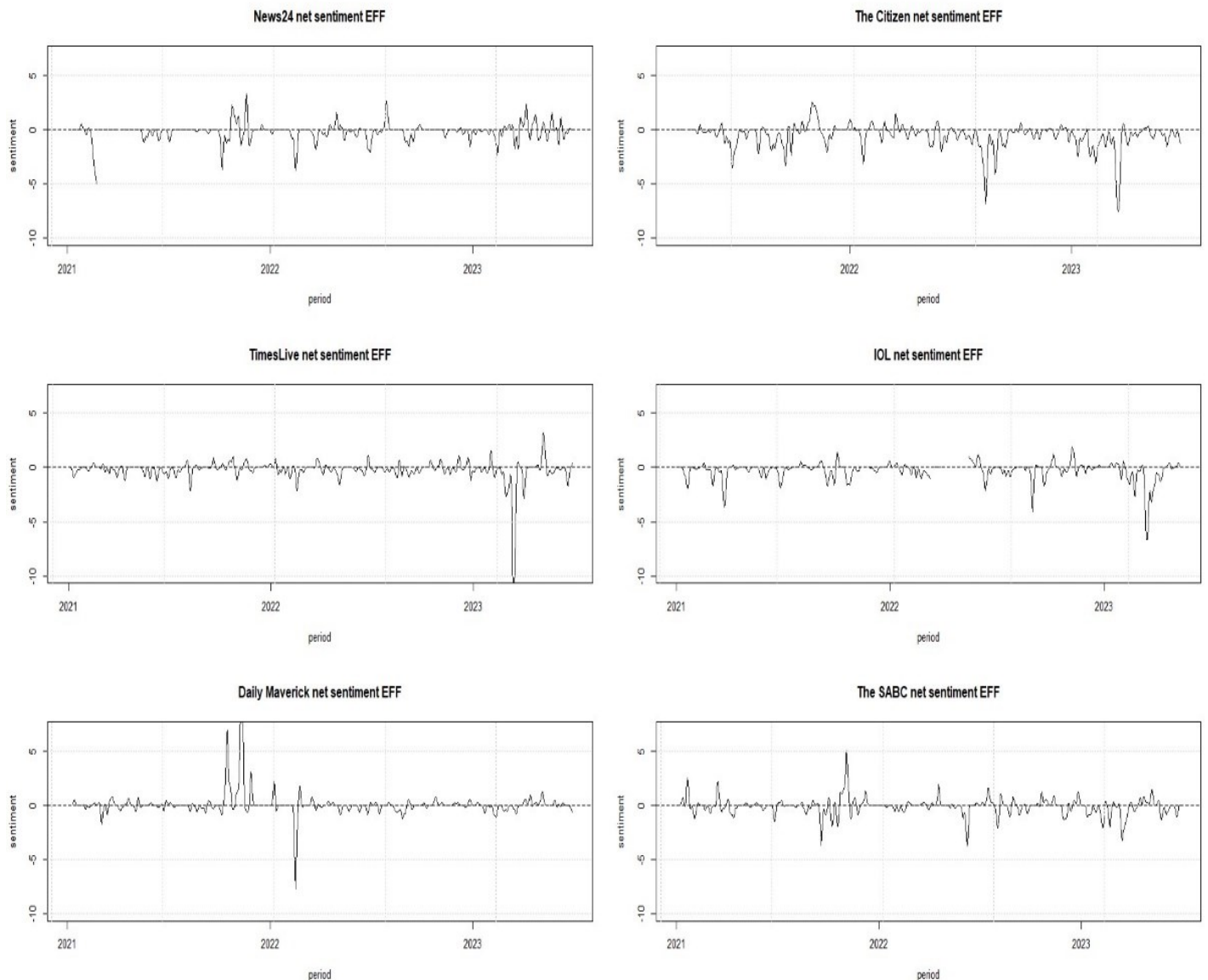


Figure 4.7: LSD Sentiment Chart by publisher: EFF

4.2.4. Online news media and political party leaders

The above sentiment analysis examined tonality by publication for the three leading political parties in South Africa. The search included only the mentions of the political parties and not their leaders. In this section, each political party head is analysed to determine the existence of tonality bias and comparison is made across selected online publications. In addition, by separating the political leaders from their party, the analysis attempts to determine the presence of leader effects. The results are organised as that of Ramaphosa of the ANC; Steenhuisen, leader of the DA; and Malema, leader of the EFF. Once again, the Lexicoder Sentiment Dictionary is utilised as the basis for the lexical-based sentiment approach.

4.2.5. Online news media and President Ramaphosa

The net sentiment charts for President Ramaphosa are by in large consistent across online media publishers as depicted by figure 4.8. The net sentiment charts are also disparate from those of his

political home, the ANC. The online media publisher, Times Live, exhibits numerous periods of net positive tonality alongside that of the state publisher, the SABC. Nonetheless, net sentiment declined consistently across all commercial media publishers, towards the end of 2022, with the most significant drop in sentiment evidenced by the Citizen. This is likely related to the Phala Phala game farm robbery and the failed attempt at impeachment. The theft of a significant amount of undeclared dollars at Ramaphosa's Phala Phala game farm came to light in June 2022 and the subsequent abuse of power, having occurred in February 2020, began the process towards the calls for the impeachment of the president ("South Africa's Ramaphosa avoids impeachment, 2022). A special parliamentary panel would find that Ramaphosa violated his oath of office which led to a failed impeachment on the 13th of December 2022 ("South Africa's Ramaphosa avoids impeachment, 2022). This sentiment towards the president struggled to recover and was bolstered by events relating to the National Shutdown with an upturn evidenced in the second quarter of 2023. The President's sentiment charts detail a period prior to mid-2022 where commercial media publishers were marginally more enamoured, but with recent scandals, the lustre appears to have worn off. Ramaphosa has benefitted from individual leader effects but is now facing the same challenges alongside the ANC. This is likely exacerbated by his level of visibility as president of the republic.

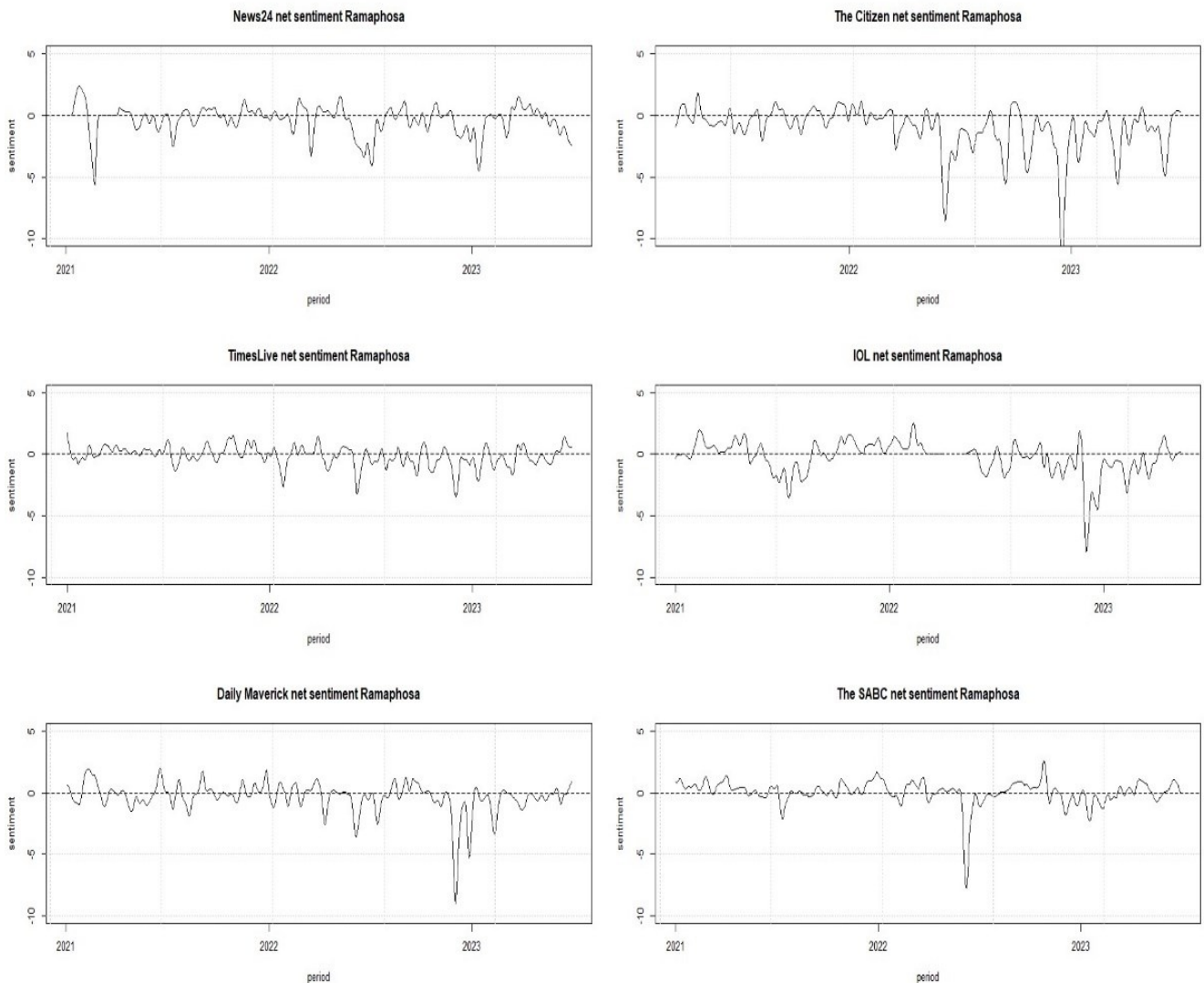


Figure 4.8: LSD Sentiment Chart by publisher: Ramaphosa

4.2.6. Online news media and John Steenhuisen

Steenhuisen, leader of the DA, has lower visibility than that of his political home as was evidenced with the frequency of his mentions. Further, his sentiment charts are muted and inconsistent across media publishers (figure 4.9). Steenhuisen experienced a period of net positive sentiment, followed by approximate neutrality in the period leading up to the 2021 Municipal Elections. His sentiment was marginally worse than that of his political party, the DA. The rise in positive tonality is most evidenced by the Daily Maverick and the Citizen. Examining Steenhuisen's charts alongside that of the DA does not indicate for separate leader effects.

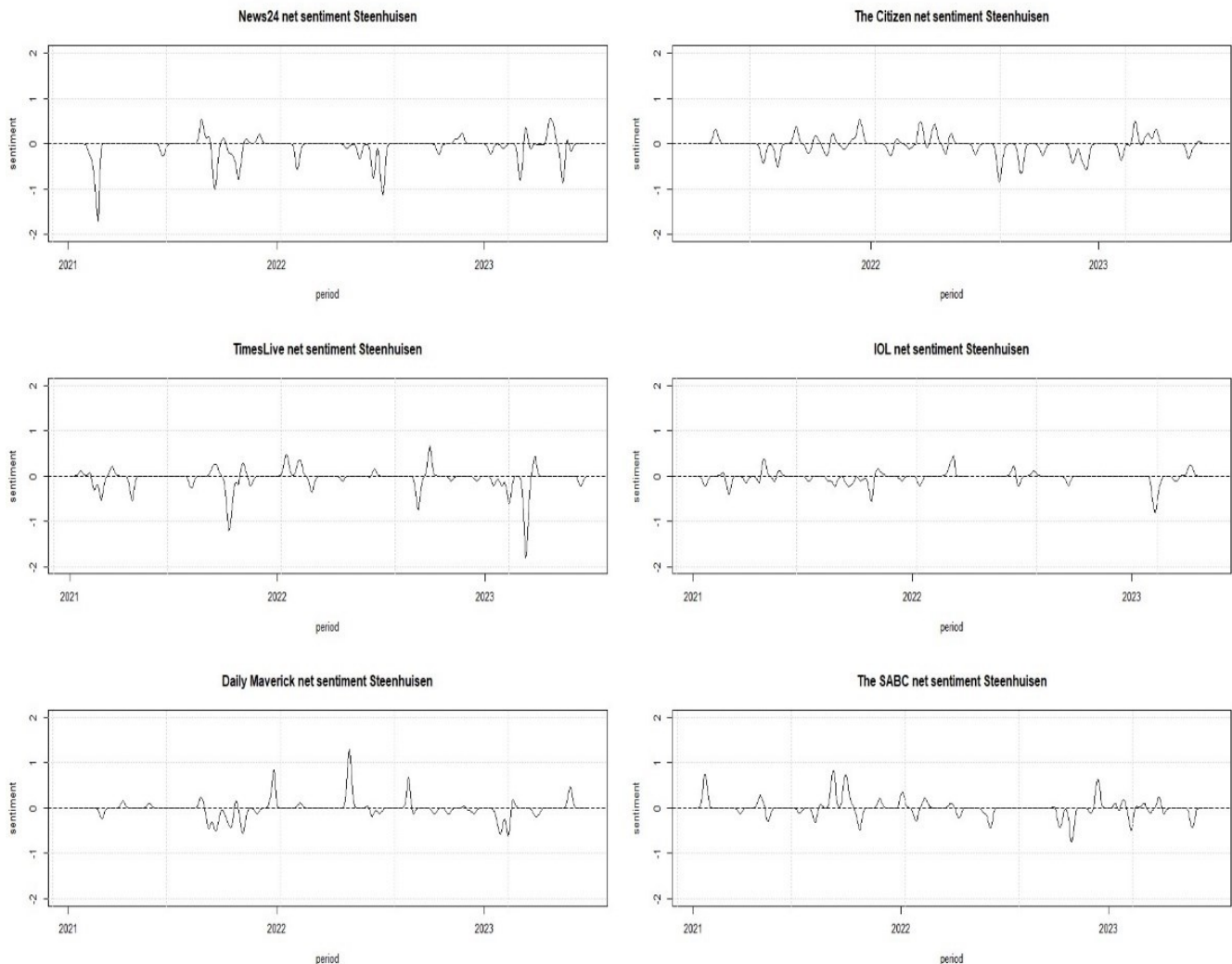


Figure 4.9: LSD Sentiment Chart by publisher: Steenhuisen

4.2.7. Online news media and Julius Malema

Malema, leader of the EFF has had an acrimonious political relationship with South African media publishers as per the sentiment charts in figure 4.10. All sentiment charts by commercial media publishers were largely net negative over the period of analysis. The leader of the EFF's net sentiment did, however, exhibit a rise linked with the 2021 Municipal Elections from the Times Live and the Daily Maverick. This was once again followed by negative tonality after the elections. The Citizen exhibited the most markedly net negative sentiment towards Malema, followed by IOL and News24 over the period of analysis.

The results highlight the ability of commercial media publishers to vie for political parties and to detract from an undesirable candidate vis a vis the ANC. Amongst the political leaders, Malema and Ramaphosa are the most impacted by negative tonality. However, Malema's net sentiment is an improvement from the tonality exhibited towards his political party.

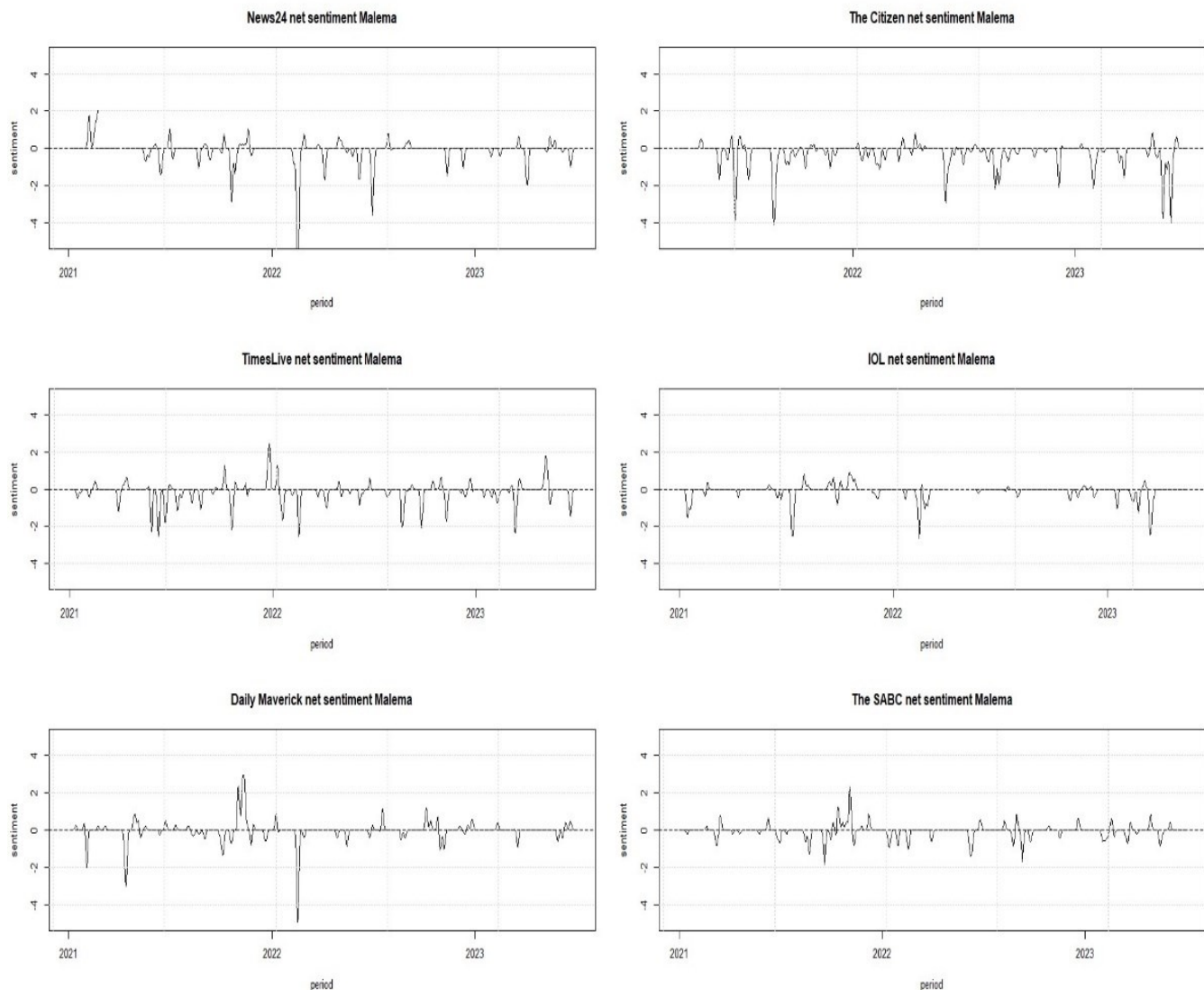


Figure 4.10: LSD Sentiment Chart by publisher: Malema

4.3. Discussion

The Lexicoder Sentiment Dictionary (LSD) was chosen for its domain specificity for political text to examine tonality bias in online political news reporting. Tonality is an important indicator for bias as it conveys meaning and has the ability to shape the views of the electorate, thus impacting decision making. This is of importance in the electoral process as this can sway and influence voting habits. The lexical-based sentiment analysis was conducted across online media publishers to determine tonality towards political parties and political leaders. This was limited to the top three political parties in South Africa, namely the ANC, DA, and EFF.

The descriptive analysis revealed a high level of text correlation amongst online media publishers which indicated a lack of diversity in reporting. Important to this was to examine the role of tonality

bias in an election cycle. A consistent pattern amongst the commercial online media publishers was the uptick in positive tonality prior to the 2021 Municipal Elections for the DA. This was also evidenced marginally for the EFF, although they were not a media friendly party at other times. The alignment across media publishers was a possible indication of endorsement to their readers for the opposition. This was in direct contrast to the negative tonality bias for the ruling party. Commercial publishers thus partisanship in favour of the market friendly, DA. The president of the republic, the ANC, Malema, and the EFF received the highest level of negative tonality across media publishers.

Utilising the LSD, allowed for establishing the presence of leader effects. Ramaphosa and the ANC were on the receiving end of negative tonality bias; however, the negative bias was inconsistent with one another, with the president losing significant media favour from 2022 onwards. Ramaphosa's sentiment chart was indicative of having individual leader effects. Whereas Steenhuisen benefitted from marginal positive representation in the media, the DA, by in large had its own party identity with greater visibility and positive tonality than its leader. Malema evidenced marginal leader effects with improved sentiment to his political party.

Key to the analysis was ascertaining the intersection between tonality and visibility. A rudimentary measure of visibility examined the frequency of online mentions for political leaders. This was led by the ruling party and its leader, president Ramaphosa. Theory posits that visibility only further serves to heighten negative or positive tonality bias. Thus, visibility is only desirable if media representation is positively skewed. With negative tonality bias skewed towards Ramaphosa and his political party, increased visibility likely proved detrimental to his electoral aspirations. This was the similar case for Malema whose public initiatives garnered significant negative tonality bias.

The state-owned media publisher, the SABC, appeared to take a more moderate approach to sentiment. It was more positively biased towards the ruling party but did not appear to have negative tonality bias towards the opposition party, the DA. The analysis of the state-owned media publisher showed it failed to deliver in agenda-setting for the state as it also amplified the opposition.

The Lexicoder Sentiment Dictionary led analysis revealed a pattern in negative tonality bias with commercial media positively biased towards market friendly and neoliberal political partisanship with the DA benefitting from a muted and positive tone. This differed to the negative stance taken towards the opposition and the president of the republic. There is merit in the call to diversify media voices that are not necessarily state or commercially owned to ensure balanced reporting is provided to better inform voting choices. This is of particular importance with the upcoming National Elections in 2024.

The subsequent chapter will examine tonality bias by utilising a semi-supervised machine learning approach, Latent Semantic Scaling. It will also serve as the basis for comparison to the Lexicoder Sentiment Dictionary findings.

5. CHAPTER FIVE- LATENT SEMANTIC SCALING ANALYSIS

5. Introduction

Chapter 5 presents the results and data analysis stemming from the application of the Latent Semantic Scaling (Watanabe, 2021), a semi-supervised machine learning method, as detailed in the methods chapter. The chapter begins by examining for tonality bias as an aggregate function across the South African online news publishers included in the data sample. This provides for a comparative analysis of overall tonality bias across political parties and their respective leaders. The analysis is subsequently expanded to examine tonality by select individual media publishers, as to provide for a comparative analysis within South African political news reporting. Lastly, an ordinary least squares regression is estimated utilising the sentiment scores predicted by the LSS. The regression is utilised to determine the presence of tonality bias, specifically at the time of the 2021 Municipal Elections. This serves to deepen understanding of tonality bias during an election cycle in a South African context.

The subsequent sections replicate the themes explored in the lexical-based analysis of chapter four, through the lens of Latent Semantic Scaling. This is to ensure comparability across methods. The analysis is undertaken in the programming language R (R Core Team, 2022), utilising the package *quanteda* (Benoit et al., 2018).

5.1. Testing for tonality bias across South African media

The LSS relies on a small set of seed words to estimate polarity as a function of the weighted proximity to words within a corpus. The study utilised a small set of generic dictionary seed words for the estimation of positive-negative sentiment. A small set of generic seed words were chosen as not to introduce any *a priori* bias.

The LSS as estimated within the *quanteda* package (Benoit et al., 2018) allows for the visualisation of terms associated with both negative and positive polarity within the corpus of news text. The outcomes of the LSS estimation are depicted in a text plot which gives tonal evidence to the relationship between South African online media publishers and political actors. The X axis depicts the polarity of the pattern of words used to estimate the LSS, whilst the Y axis represents the logged frequency of the terms within the corpus. The text plot is further compared to the lexical-based method through the application of the LSD to the LSS outcomes. The subsequent section thus examines the outcomes of LSS estimates for political parties and their respective leaders.

5.1.1. South African media and the ANC

In the estimation of the LSS, the pattern “anc” was utilised to determine the tonality associated with the ruling party within the corpus of news text. This required the initial substitution of the phrase “African National Congress” with “anc” as to not skew the sample. Figure 5.1 depicts the text plot of the terms associated with the political party. The plot appears to be balanced between negative and positive polarity with a marginal skew towards negative tonality. The issues identified by the LSS point to a party in decline and in crisis. This is evidenced with terms such as “factional”, “weakened” and “struggle” amongst others.

The LSS also highlights ideological differences from South African online media publishers as policy documents, which are deemed socialist such as the National Democratic Revolution (NDR), receive negative tonality. This is indicative of the media playing a role in agenda-setting by negatively representing policy not in alignment with commercial or editorial convictions. This is line with the earlier assertion of a commercial media that is neoliberal and pro capital.

The ability of the LSS to assign sentiment to terms within the corpus enables the analysis of the topical issues affecting media sentiment. The ruling party received negative media sentiment during the Municipal Elections and in their dealings with the opposition. The ruling party does, however, receive positive sentiment for leadership changes as evidenced in Kwa Zulu Natal. The LSS thus provides clarity on the sentiment on political decisions enacted by the ANC.

When the LSD dictionary is applied to the LSS outcomes, the topic of a weakened party remains, and terms such as “scandal” and “chaos” further confirm the negative tonality findings of the LSS as per figure 5.2. Whereas the LSS utilises the terms within the corpus and assigns polarity and the Lexicoder Sentiment Dictionary (LSD) utilises a predefined lexical dictionary, the application of the LSD on the LSS confirms the polarity and tonality towards the ANC.

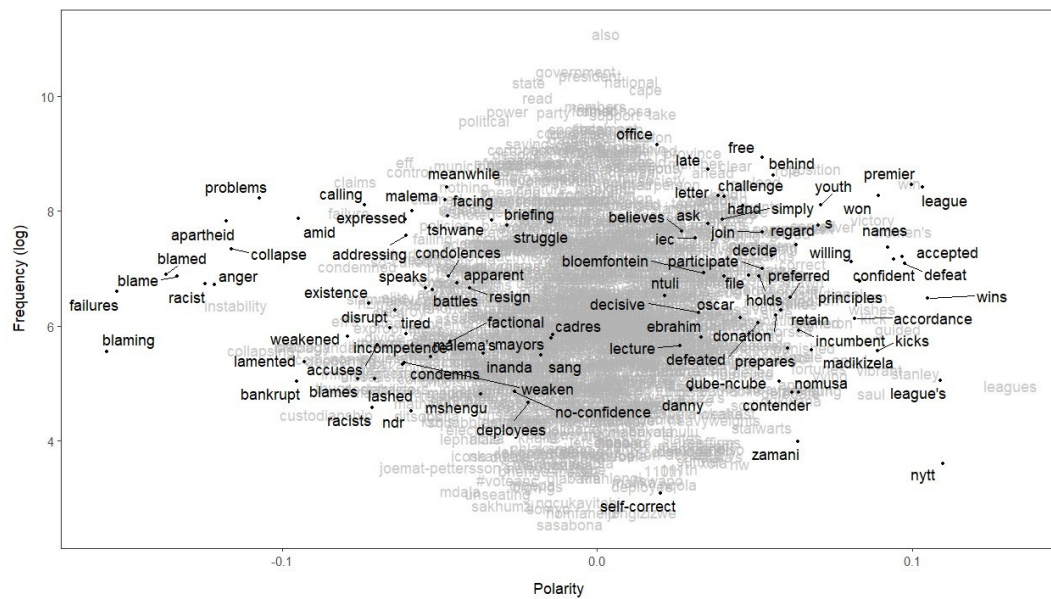


Figure 5.1: LSS Text plot terms: ANC

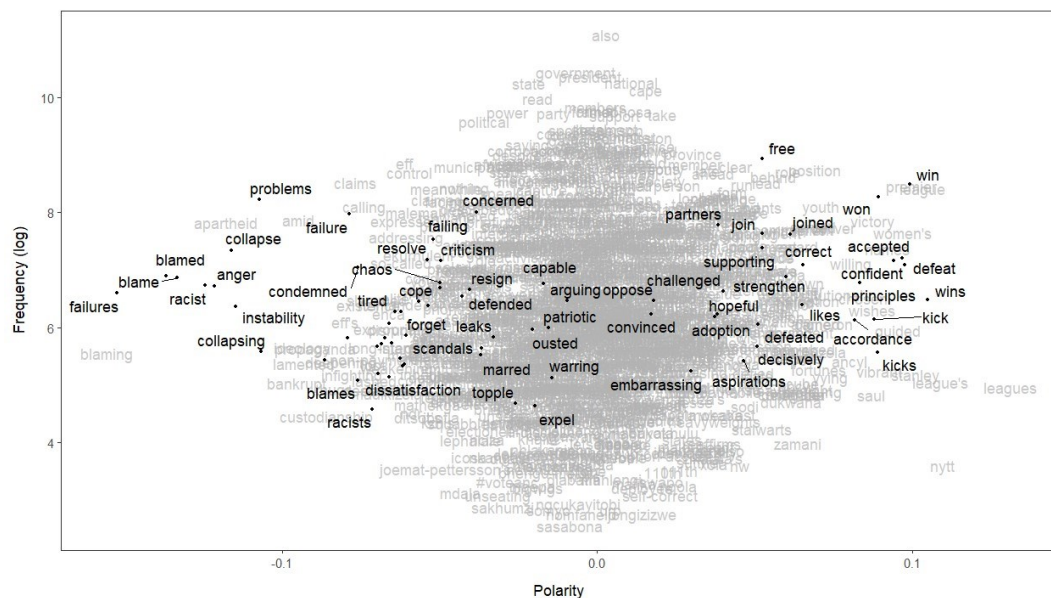


Figure 5.2: LSS-LSD Text plot terms: ANC

5.1.2. South African media and the DA

The LSS model outcome for the DA indicate greater negative tonality in contrast with the previously estimated approximate neutrality by the LSD. Figure 5.3 illustrates the text plot for the party following the estimation of the LSS as per the pattern “da” and the replacement of the term “democratic alliance”. Examining the terms in the text plot, a pattern emerges of positive tonality bias towards the electioneering efforts of the DA. Terms such as “principles”, “win”, “premier”, and “partners” point to overall positive sentiment aligned with the victory of the party and associating party victory to positive political leadership. The party scores with terms such as “confident” and “clarity” which confirm positive tonality bias towards its leadership. Whilst the overall LSS appears to be negative, the terms appear to exhibit positive bias towards the party’s electoral efforts.

Examining the negative polarity outcomes, evidences the party's continued racial representation in electioneering with terms such as "coloured" and "racism" rising to the fore. An additional area of contention is the negative tonality between the party and its opposition, the EFF, and the ruling party.

Turning to the LSD as applied to the LSS (figure 5.4), confirms the findings of the LSS text plot with the party largely portrayed positively in its electoral efforts. The negative tonality of the terms “arrogant”, “arrogance” and the negative association of the party with its former leader “Maimane” provides detail on possible deterring factors from the party having broader appeal.

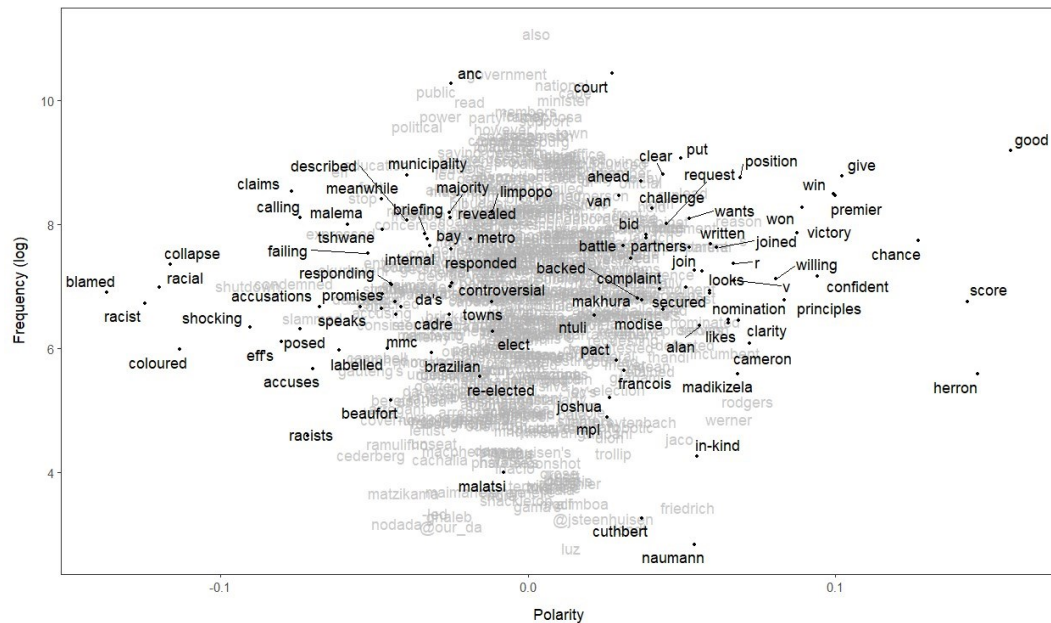


Figure 5.3: LSS Text plot terms: DA

5.1.4. South African media and President Ramaphosa (ANC)

The LSS text plot outcomes for the president of the ANC and the republic, Ramaphosa, are depicted in figure 5.6. The pattern used for the president's LSS was "Ramaphosa". Unlike the LSS outcomes for the ANC, the LSS for Ramaphosa suggests a more positively skewed text plot. The areas of negative sentiment identified by the LSS look to be policy failings such as load shedding and his ability to deal with civil unrest. In contrast, the annual presidential address attracts marginal positive tonality from online media publishers. This is in contrast with the negative tonality attached to ANC political policy. This is likely evidence of Ramaphosa's separate leader effects and his distinction in delivering government planning that is market friendly unlike the policies of his political home. The LSS does well in identifying areas of discord for the president which include parastatal failure as sore points that likely impact his electoral aspirations.

The LSD as applied to the LSS outcomes, provides greater detail on tonality towards the president (figure 5.7). As per the findings of the lexical-based analysis, the burglary at the president's farm had a significant negative impact on media sentiment as it indicted the president of possible money laundering. The LSS unlike the LSD depicts a more balanced outcome in sentiment towards the president.

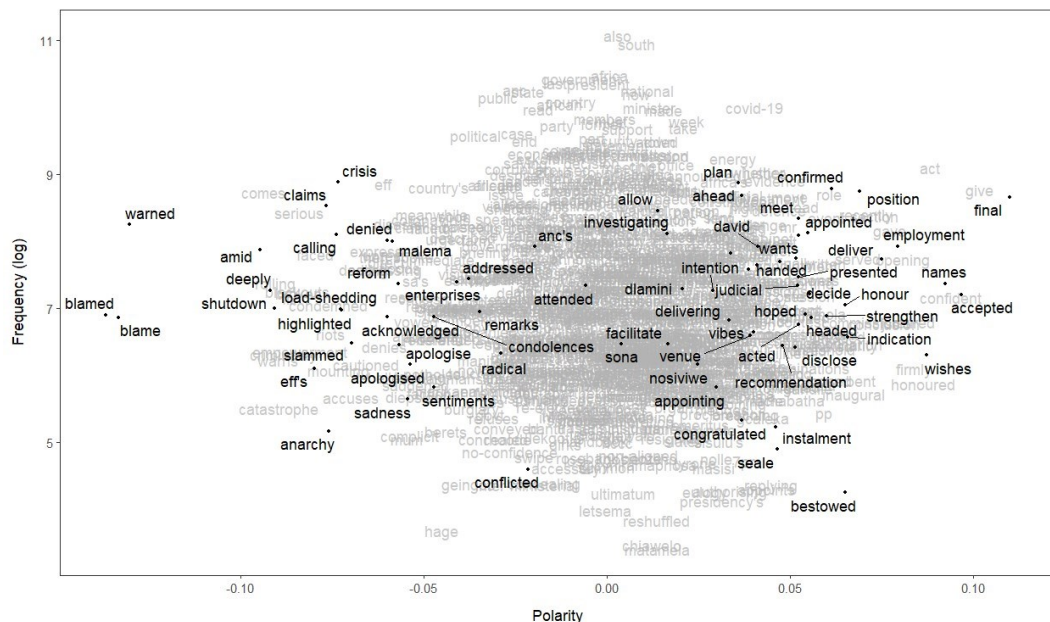


Figure 5.6: LSS Text plot terms: Ramaphosa

5.1.6. South African media and Julius Malema (EFF)

The EFF and its leader have had a trying time with South African media publishers. This raises concern for the upcoming political contestations, particularly if the media is found to play a role in setting the political agenda. Figure 5.9 illustrates the outcomes of the LSS for the leader of the EFF, Malema, when the pattern “malema” and “julius malema” were utilised. Unlike that of the political party, the text plot for Malema is more evenly distributed in polarity. The focus of the LSS outcomes assigns negative polarity to Malema’s political and policy statements. Once again, the tonality is negative regarding senior leadership within the party and towards the party itself. This is a signal of South African media’s fractious relationship with the party. Similar to the issue of race relations raised within the LSS outcomes of the DA, they are also contentious for the leader of the EFF. Terms such as “white” and “AfriForum” receive strong negative polarity. Race dynamics in South Africa appear to be a sore spot across the board for South African media publishers on aggregate.

In alignment with the findings of the party, the term “land” is strongly negative indicating continued adverse sentiment for the policy stance of the EFF as delivered by its leader. Key to the above findings is the ability of South African media publishers to negatively influence and set an anti-party policy agenda.

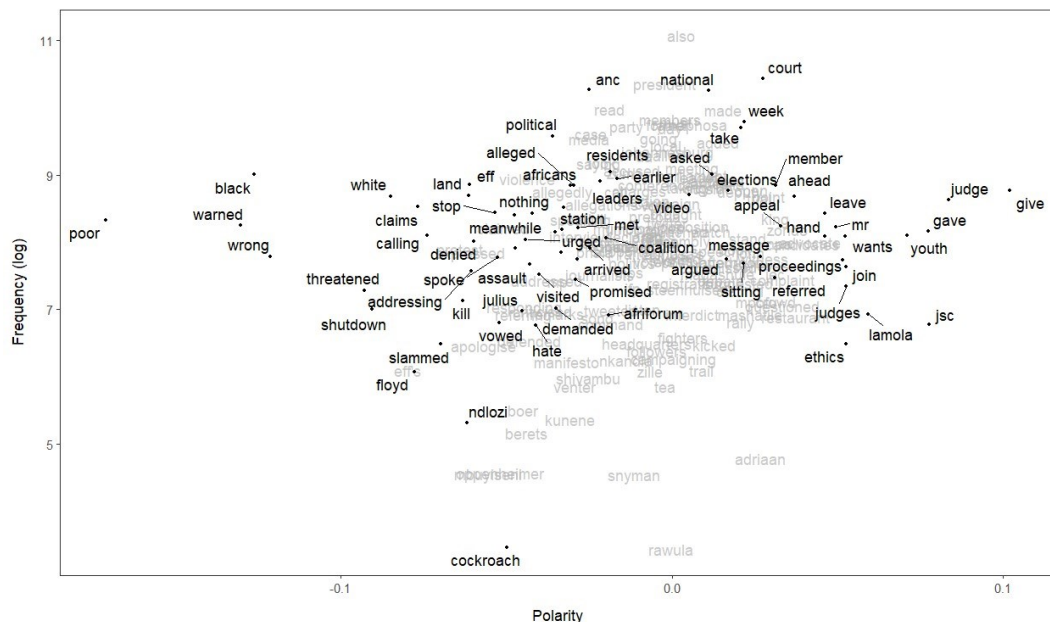


Figure 5.9: LSS Text plot terms: Malema

5.2. Tonality bias by publication: Applying regression to Latent Semantic Scaling outcomes.

The study set out to examine tonality bias of South African online media publishers in political news reporting. The analysis of Latent Semantic Scoring text plot terms provided a view of skew and the presence of tonality bias. It identified significant areas of concern, particularly those relating to electioneering and leader effects. A prominent determinant of tonality across South African online media publishers were the 2021 Municipal Elections, as evidenced by the findings of the LSD, and confirmed by analysis of the LSS. Thus, the subsequent regression analysis is focused on estimating tonality bias for political parties and their respective leaders during an election cycle, as proxied by the 2021 Municipal Elections.

The method of analysis is analogous to a difference in differences design proposed by Watanabe (2017). The method examines the differences in tonality following a pivotal event by performing an ordinary least squares (OLS) regression on continuous sentiment scores, as detailed in the methods chapter.

In line with the analysis of the lexicon-based approach and the requirements of the LSS, the group of online media publishers were reduced due to the availability of articles and online media presence by publisher. The media publisher, the Eyewitness News (EWN) was included due to the quantity of articles available for use in the estimation of the LSS. The regression results are presented and discussed below.

5.2.1. Bias and the 2021 Municipal Elections

The Municipal elections of 2021 were hotly contested amongst political parties. Prior to the political melee, debate centred on the likelihood of the elections taking place. The lead up to the elections was mired by the ongoing world health crisis presented by Covid-19. Concerns over an election occurring during a health crisis were disputed to be a delaying tactic by the ruling party, as South Africa had been in the midst of the declining pandemic for approximately two years. Facing possible censure, government announced the date of the 2021 Municipal Elections on the 20th of September 2021 with the elections taking place on the 01st of November 2021. These seminal dates formed the basis for formulating the regression models.

In estimating the regressions for tonality bias associated with the 2021 Municipal Elections, three significant date points were identified in the data. The date of the elections was established as the change point for the LSS sentiment scores, with the date of announcement as the commencement of the dataset. The dataset was subset by the selected South African media publishers.

5.2.2. Bias and the 2021 Municipal Elections: political party news

In analysing the relationship between tonality bias aimed at political parties during an election period, three regression models were estimated as detailed in table 5.1. Model one estimates the regression outcomes of aggregate media tonality bias for each political party following the elections. For all political parties, tonality was significantly negative on average over the election period. The most significant negative tonality was evidenced towards the EFF. The regression analysis revealed that the 2021 Municipal Elections were an overall negative and significant event that resulted in negative tonality bias by online media publishers, attaining significance at the 10 percent level. In contrast, the DA, attained significant positive tonality bias following the 2021 Municipal Elections, an indication of aggregate media support of party victories. The subsequent model estimations looked to establish the linkages between select online media publisher's news reporting and political parties.

Model two explores the relationship between media publishers and the 2021 Municipal Elections. Of the South African media publishers selected, only two media publishers exhibited positive tonality over the period, namely the Citizen and EWN. Two online news media publishers exhibited significant negative sentiment on average over the period, namely News24 and SABC news. News sentiment declined significantly at the 5 percent level for the Citizen and the 10 percent level for SABC news following the elections, indicative of negative tonality bias. Though both publishers exhibited tonality bias, this is likely due to conflicting rationale. The significant negative tonality bias exhibited by the SABC is likely a reaction to the poor performance of the ruling party as a state broadcaster. A third model is formulated to examine the relationship between select online media publishers and political parties.

Model three provides for the interaction between political parties and the media over the course of the 2021 Municipal Elections. In this permutation, the ANC is now the only party that on average has negative tonality over the period. Commercial media publishers, the Citizen and News24 are now the only significantly negative media publishers over the period of the elections. Additionally,

EWN is shown to be on average significantly tonally biased against the political party, the EFF at the 5 percent level.

The regression analysis evidenced that not all South African media publishers displayed negative tonality over the period of the 2021 Municipal Elections. On average, the elections were significantly negatively reported on across all parties whereas analysis of individual media publishers revealed that some media publishers were more biased than others, with News24 and the SABC both taking a significant negative tonal bias following the elections. Lastly, from the final permutation, only the ANC had negative tonal bias on average across the period. However, this was not attributable to any specific media publisher whilst the EFF received significant tonality bias from EWN over the period. The DA benefitted following the 2021 Municipal Elections from positive tonality bias.

The sentiment changes over the period of the elections are plotted, to give visual effect to the interactions between the media and political parties as per figure 5.10. As per the regression results, the 2021 Municipal Elections were not on aggregate a significant event across online media publishers. Examining the first chart in figure 5.10, News24 and the SABC had significantly low negative sentiment scores in both periods. The most significant change in sentiment is confirmed to be that of the SABC and the Citizen post the elections with markedly negative tonality bias. The individual party charts in figure 5.10 evidence the discrepancies in sentiment in online news reporting. The Daily Maverick consistently exhibited the highest sentiment across political parties prior to the elections whilst News24 exhibited the lowest sentiment values across parties. The ANC experienced a decline in tonality across all media publishers, following the elections. Subsequent to the elections, negative tonality bias was exhibited by EWN for the EFF, though the EWN improved its tonality for the DA. The DA received aggregate positive tonality bias scores but none that were significant from an individual media publisher.

These findings somewhat confirm those of the LSD with tonality negatively skewed towards the ANC. However, none could be attributed to specific media publishers. Only the EFF was found to have a negatively biased relationship with a specific media publisher, the EWN.

Table 5.1: 2021 Municipal Elections regression models: political party news

	Model 1	Model 2	Model 3
(Intercept)	0.064**		0.222***
ANC	-0.108+	0.068	-0.358*
DA	-0.189**		-0.199
EFF	-0.440***		-0.103
Municipal Election	-0.073*	0.100	
ANC: Municipal Election	-0.107		
DA: Municipal Election	0.216+		
EFF: Municipal Election	0.198		
EWN		0.054	-0.072
IOL		-0.031	-0.096
News24		-0.346***	-0.457***
SABC		-0.197*	-0.382***
Citizen		0.016	-0.124*
EWN: Municipal Election		-0.166	
IOL: Municipal Election		0.004	
News24: Municipal Election		-0.088	
SABC: Municipal Election		-0.200+	
Citizen: Municipal Election		-0.240*	
ANC: Eyewitness News			0.145
ANC: IOL			0.265
ANC: News24			0.310
ANC: SABC Online News			0.274
ANC: The Citizen			0.212
EFF: Eyewitness News			-0.600*
EFF: IOL			-0.300
EFF: News24			-0.026
EFF: SABC Online News			-0.087
EFF: The Citizen			-0.300
DA: Eyewitness News			0.234
DA: IOL			0.103
DA: News24			0.109
DA: SABC Online News			0.194
DA: The Citizen			-0.082
Num.Obs.	4592	4592	4592
R2	0.023	0.026	0.048
R2 Adj.	0.022	0.023	0.043
AIC	12 378.7	12 376.0	12 294.4
BIC	12 436.6	12 459.6	12 455.2
Log.Lik.	-6180.367	-6174.977	-6122.206
F	15.666	10.963	9.969
RMSE	0.93	0.93	0.92

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

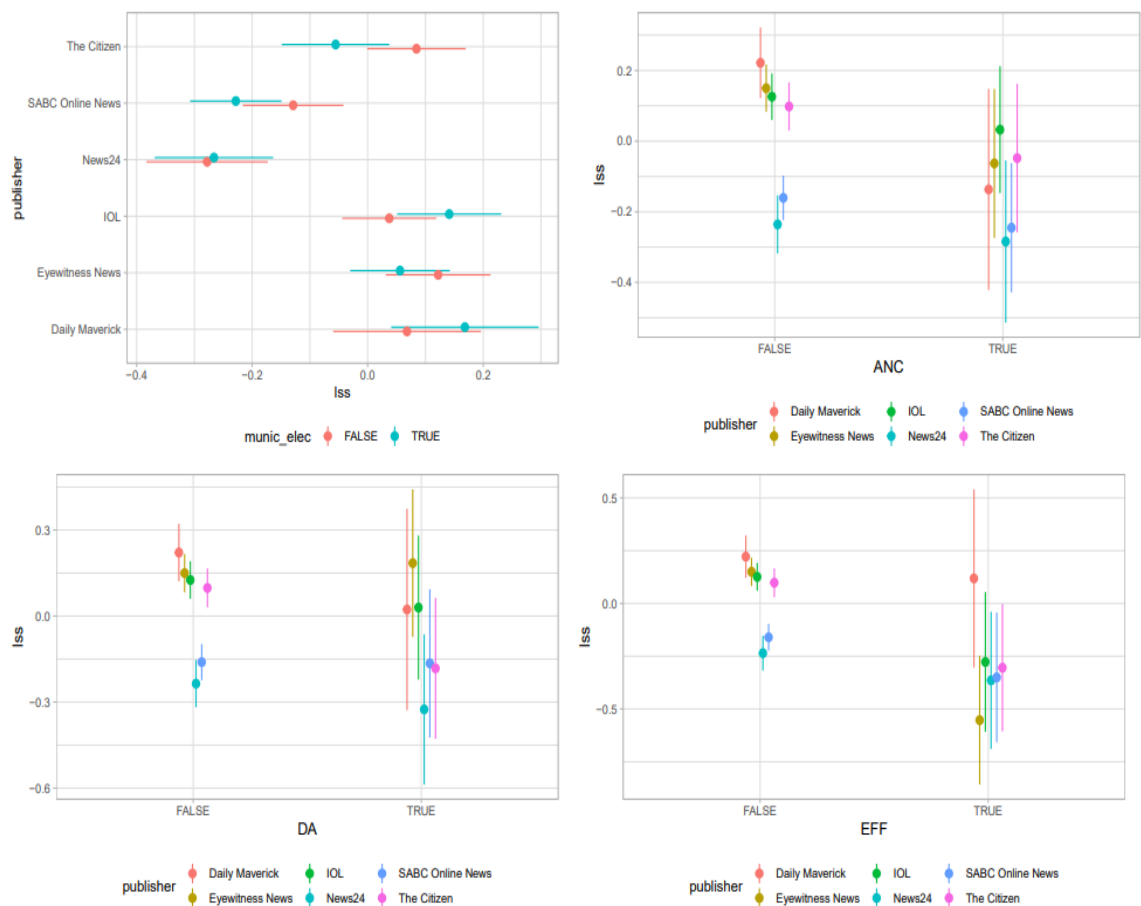


Figure 5.10: Regression interaction effect plots: 2021 Municipal Election party news

5.2.3. Bias and the 2021 Municipal Elections: political leader news

In turn, a comparative analysis was applied amongst the three political leaders in respect to the 2021 Municipal Elections. The estimation of the regression models replicates that of the political parties and the pivotal dates went unchanged to ensure alignment with the political party analysis. The results were informed by the LSS outcomes estimated on revised patterns as per the leaders of each political party. The regression output is summarised in table 5.2 below.

Model one estimates the relationship between political leaders and the 2021 Municipal Elections through an aggregated media lens. The elections were found to be a significant change point in tonality for political news reporting on political leaders at the 10 percent significance level. The negative tonality bias exhibited post the elections, did not impact political leader sentiment. On average, negative tonality in reporting was hardest on Malema of the EFF, followed by Steenhuisen of the DA. In contrast, Ramaphosa received positive sentiment on average across the period of reporting, however it was not significant.

Model two explored the sentiment espoused by select media publishers on the 2021 Municipal Elections. All online media publisher's political news reporting was negative on average across the period. However, News24 and the SABC evidenced significant negative tonality in reporting over the period. In contrast, post the elections, only the SABC evidenced significant negative tonality bias whilst all commercial media publisher's sentiment was insignificant.

The third regression estimation provides a breakdown of the interaction between select media publishers and political leader sentiment over the election period. In this permutation of the model,

the media is only significantly negative about the leader of the DA on aggregate over the election period. An increasing number of media publishers are found to exhibit significant negative tonality over the election period, and now include EWN, News24, the Citizen and the SABC. The interactions between media publishers and political leaders evidence once again that not only was the EWN significantly negative against the EFF party, but they were also significantly negative against the party leader, Malema.

The LSS sentiment scores are plotted to give visual effect to the interactions between the media and political leader news reporting as per figure 5.11. The first plot gives visual effect to the changes in tonality of media publishers during on the elections. News24 led negative sentiment overall whilst the results of the elections resulted in a significant dip in sentiment for the state publisher, the SABC. News24 and the SABC consistently espoused negative tonality during the period of the 2021 Municipal Elections for both political parties and their respective leaders. The SABC, although negative about the elections overall, exhibited positive tonality bias towards the president. President Ramaphosa appeared to have leader effects as his plot evidenced improved tonality across media publishers in comparison to the ANC. Conversely, Steenhuisen performed poorly in comparison to the DA and had negative tonality from media publishers. The EFF and Malema evidenced little to no change between the party and its leader, with the party attracting significant tonality bias from the EWN.

The regression results for political leaders evidenced the presence of leader effects during an election cycle. The state media publisher played a pro state agenda role as it was significantly biased in favour of the president over the period but not for the ruling party. South African online media publishers consistently displayed negative tonality in reporting with only the EFF and Malema attracting the ire of the EWN.

Table 5.2: 2021 Municipal Elections regression models: party leader news

	Model 1	Model 2	Model 3
(Intercept)	0.026	0.118+	0.237***
Ramaphosa	0.074		-0.185
Steenhuisen	-0.424***		-0.740*
Malema	-0.805***		-0.446
Municipal Election	-0.060+	0.132	
Ramaphosa: Municipal Election	0.040		
Election			
Steenhuisen: Municipal Election	0.117		
Election			
Malema: Municipal Election	-0.041		
EWN		-0.058	-0.161**
IOL		-0.027	-0.089
News24		-0.341***	-0.448***
SABC		-0.212*	-0.438***
Citizen		-0.125	-0.228***
EWN: Municipal Election		-0.173	
IOL: Municipal Election		-0.041	
News24: Municipal Election		-0.121	
SABC: Municipal Election		-0.318**	
Citizen: Municipal Election		-0.186	
Ramaphosa: Eyewitness			0.178
News			
Ramaphosa: IOL			0.285
Ramaphosa: News24			0.326+
Ramaphosa: SABC Online			0.494**
News			
Ramaphosa: The Citizen			0.240
Steenhuisen: Eyewitness			0.382
News			
Steenhuisen: IOL			0.531
Steenhuisen: News24			0.335
Steenhuisen: SABC Online			0.606
News			
Steenhuisen: The Citizen			0.258
Malema: Eyewitness News			-0.805*
Malema: IOL			-0.568
Malema: News24			0.215
Malema: SABC Online			-0.316
News			
Malema: The Citizen			-0.510
Num.Obs.	4592	4592	4592
R2	0.024	0.024	0.049
R2 Adj.	0.022	0.021	0.044
AIC	12 848.6	12 857.2	12 762.8
BIC	12 906.5	12 940.8	12 923.6
Log.Lik.	-6415.284	-6415.608	-6356.422
F	16.089	10.169	10.164
RMSE	0.98	0.98	0.97

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

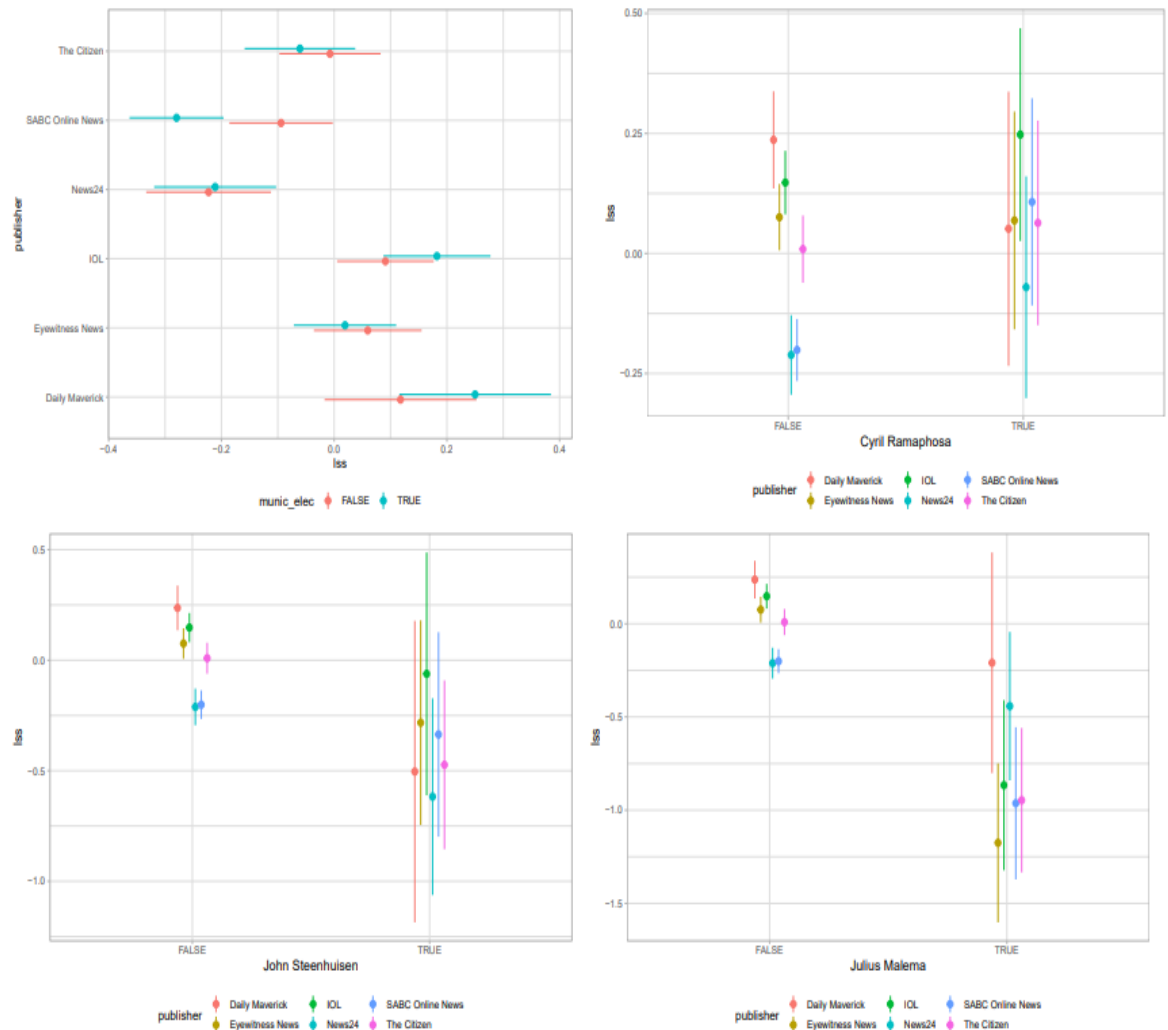


Figure 5.11: Regression interaction effect plots: 2021 Municipal Elections party leader news

5.3. Discussion

The Latent Semantic Scaling approach provided a nuanced understanding of the relationship between South African media publishers and political parties and their respective leaders during an election cycle. The analysis revealed that while the media by and large exhibits negative tonality to political news reporting, there are winners and losers, and these vary by theme. News24, SABC and the Citizen were found to be the most significantly negative media publishers in the South African media landscape as per the sample. The biggest losers were the ANC, EFF, Malema and the leader of the DA, Steenhuisen. There was little evidence individualism in the South African media landscape as media publishers largely echoed the same tonality. In this vein, some were more notably anti the EFF as per EWN with the EFF and party leader, Malema.

A possible compelling rationale would be the need for South African media publishers to act as agenda setters and direct political policy. This was evidenced in the recurring outcomes of a decidedly anti ruling party and EFF tonality biased media. The media took further ire with policy documents that were not in line with neoliberal policy such as the National Democratic Revolution of the ANC, and the proposal of land expropriation by the EFF.

Notably the president of the republic fared better with the LSS than with the LSD with pockets of positive tonality. The regression analysis revealed that the state broadcaster, the SABC, played a

pivotal role in driving positive tonality bias for the President and to a lesser extent, so did the commercial media publisher, News24 during the 2021 Municipal Elections.

Although the LSS was not conclusive in confirming the outcomes of the LSD, the regression results were indicative of negative tonality bias regarding the ruling party, the ANC during an election cycle. Conversely, the DA benefitted from positive tonality bias during an election cycle. This does not abstract away from the role played by the state broadcaster, the SABC, in delivering more favourable sentiment for the President.

A significant outcome of the LSS regression analysis was the evidence of leader effects in South African electioneering. This likely served the political aspirations of the ruling party due to the positive tonality for the president which is further reinforced by his visibility in the media. Though the leader of the DA was on the receiving end of negative tonality during the 2021 Municipal Elections, his lack of visibility and the higher visibility of the DA, likely dampened his impact. This will play a role in future electioneering prospects of political parties as they contest the 2024 National Elections.

6. CHAPTER SIX- PROTEST ACTION IN THE NEWS

6. Introduction

Protest action has been theorised to receive significant negative reporting from South African media publishers with protesters characterised as violent and destructive mobs. Equally, protest action is often the only tool available for the marginalised majority to call attention to their socioeconomic challenges, and failure to contextualise unrest serves to further polarise the already disadvantaged. This chapter utilises the Lexicoder Sentiment Dictionary, alongside Latent Semantic Scoring to examine the subject of news reporting on social unrest.

6.1. Lexicoder Sentiment Dictionary: Are the marginalised allowed to protest?

The looting of July 2021 and the planned National Shutdown of 2023 were identified as the two seminal events of protest action to be examined for tonality bias utilising a lexicon-based method. These events were separated through the term “loot*”. This was done to exclude the extraordinary events of July 2021, which led to loss of life and significant national disruption (Davis et al, 2021). In a three-day period of civil unrest in July 2021, over 300 lives were lost, with significant racial tensions and protests led by the jailing of former president Jacob Zuma (Davis et al, 2021). Yet no subsequent research reports link the significant economic losses and loss of life to a specific event (Davis, 2024). The estimations included the terms “protest” and “shutdown”.

Figure 6.1 depicts the results of the LSD with the chart on the left representing net sentiment excluding the term “loot”. Net sentiment is negative throughout the period of analysis, and it is a clear indication that protest action is viewed as destructive on aggregate with significant tonality bias. This is further evidenced by the significant negative decline in sentiment in the period approximating the occurrence of the National Shutdown, March 2023. This is in alignment with the negative tonality bias impacting the opposition party, the EFF.

The net sentiment graph accounting for the inclusion of looting depicts a sobering portrait of the impact of a substantive negative event and its representation by the media. The inclusion outweighs all other protest action in net negative sentiment. The LSD thus demonstrates that South African media publishers have a strong negative view of protest action even with the exclusion of the extraordinary events of 2021. This gives the poor marginal room to express their discontent. The analysis is decomposed by the top terms to examine the interaction between frequency and tonality bias in table 6.1.

The frequency of terms detailed in table 6.1 examines the interaction between visibility and tonality. As per table 6.1, the top negative terms associated with protest action outnumber the frequency of positive terms tenfold. The term “protest” garners strong negative emphasis and is repeatedly utilised to capture the media’s strong anti-protest stance. The term “violence” is utilised more frequently than that of “right” and “violent” also appears more frequently than “peaceful”. This gives credence to the assertion that the media has a negative tonality bias towards protests and depicts what is a tool for many to call attention to their plight, as criminality, and a threat to society.

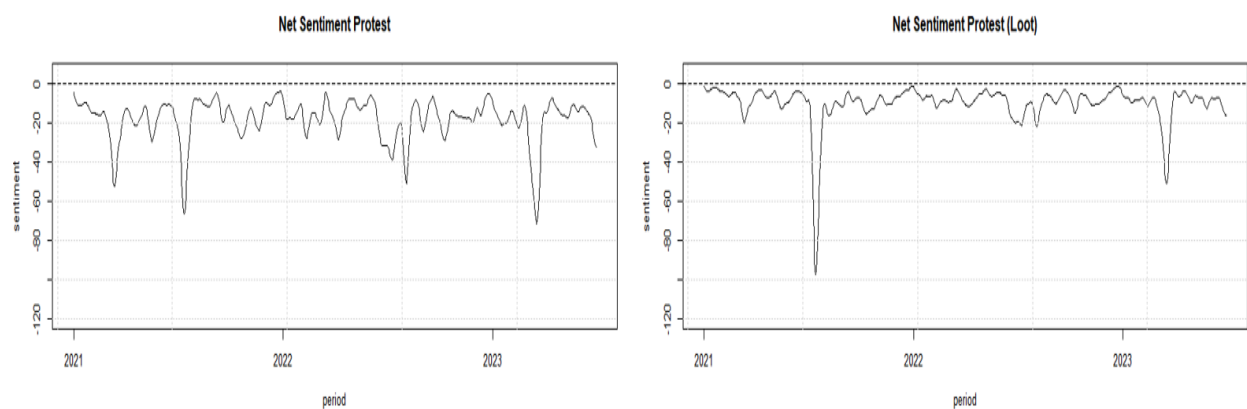


Figure 6.1: LSD Sentiment Chart: Protests

Table 6.1: Term frequency: Protests

Protest			
[Negative]	Frequency	[Positive]	Frequency
protest	2709	right	194
against	652	well	109
violence	448	peaceful	106
unrest	281	support	103
killed	233	civil	96
protests	231	like	95
violent	199	join	92
affected	139	rights	86
destruction	136	supporters	66
protesters	108	protect	65

6.1.1. Decomposing online news media publishers and protest action

The LSD is estimated for selected media publishers to measure individual tonality towards protest action. Decomposing sentiment analysis by each online media publisher does not yield any significant changes as depicted in figure 6.2. The net sentiment charts are negative across all commercial online media publishers. The state-owned media publisher also assumes negative tonality towards protest action, and this can be seen as a pro government stance, given the ire raised by protesters for government failings in service delivery. The IOL, the Citizen and the SABC take the strongest stance on protest action with the lowest troughs in tonality. The 2023 National Shutdown received largely negative sentiment across all online news media publishers, in line with the negative tonality bias on protest action evidenced in South Africa.

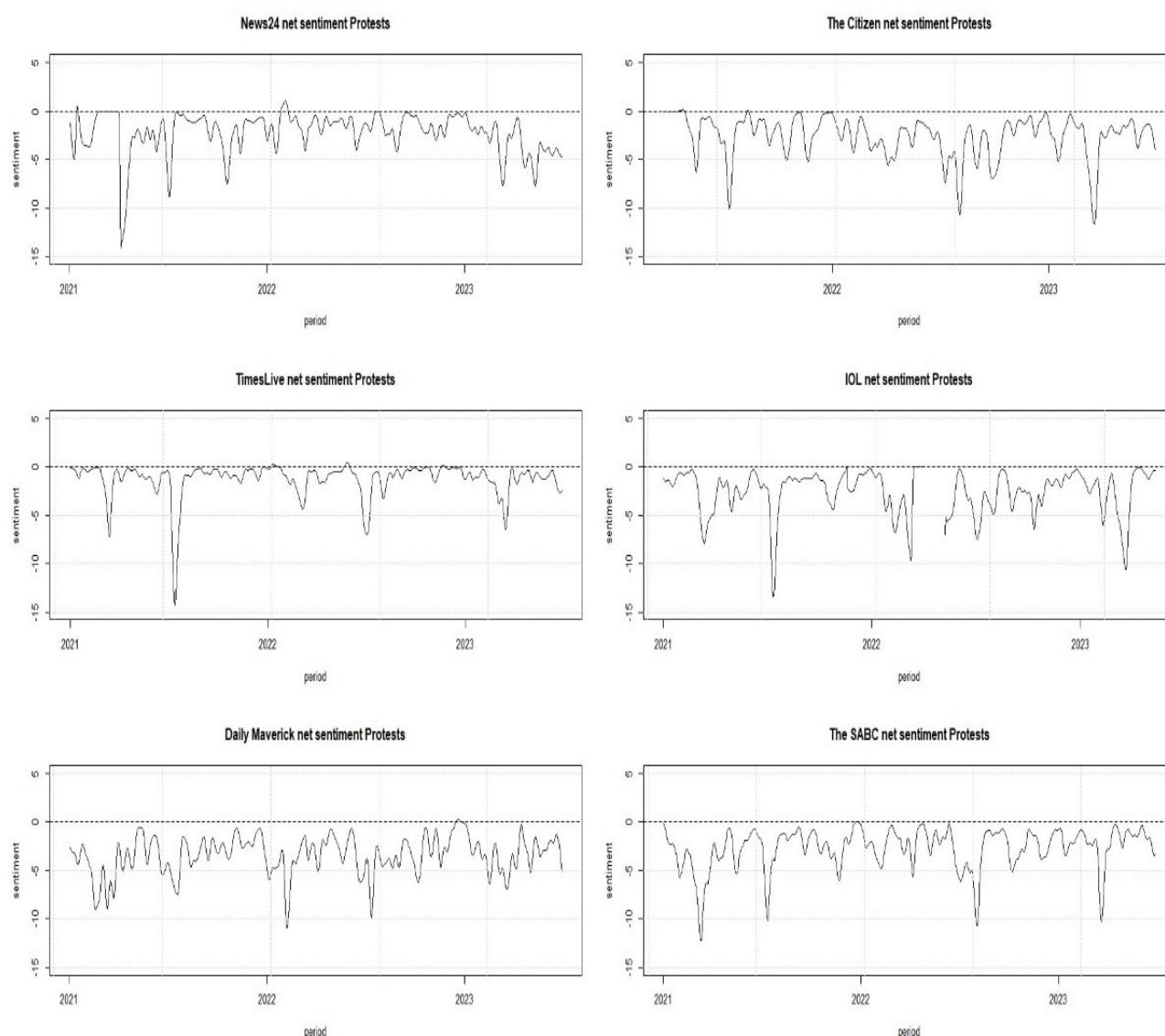


Figure 6.2: LSD Sentiment Chart by publisher: Protests

6.2. Latent Semantic Scoring: South African media and Protest action

Protest action has received significant negative tonality bias from South African media publishers whether commercial or state owned as per the findings of the Lexicoder Sentiment Dictionary method. Applying Latent Semantic Scaling methodology provides greater detail into the tonality associated with protest action through the visualisation of terms and polarity scores. As per figure 6.3, the text plot is significantly negatively skewed. Terms such as “criminality”, “disrupting” and “violent” are highlighted through the application of the LSS as characterisation of protest action by South African online media publishers. This is in alignment with accusations of the negatively skewed representation of protest action and its delegitimisation.

The text plot outcomes also include the identification of leaders within the EFF who receive negative tonality for their role in agitating for protest action from online media publishers. The negative tonality associated with the EFF as per the LSS method, is in alignment with the LSD findings, which evidenced a decline in sentiment due to the party’s leading role in the 2023 National Shutdown.

Applying the LSD dictionary to the LSS outcomes provides further emphasis on the negative tonal bias associated with protest action as depicted in figure 6.4. Significant to both outcomes were the negative sentiment of violence and disruption. The LSS method confirms that South African media publishers represent protest action with negative tonality bias to their audience rather than examining the underlying endemic issues.

A focal point of the analysis remains the analysis of the protests and how these impact political leaders given their prominence in the LSS text plots.

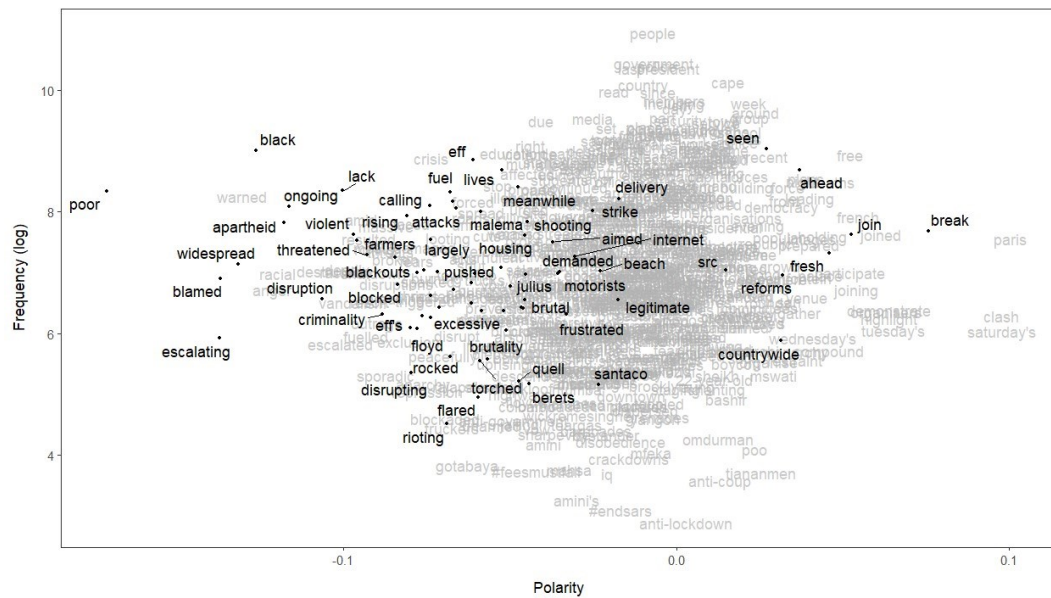


Figure 6.3: LSS Text plot terms: Protests

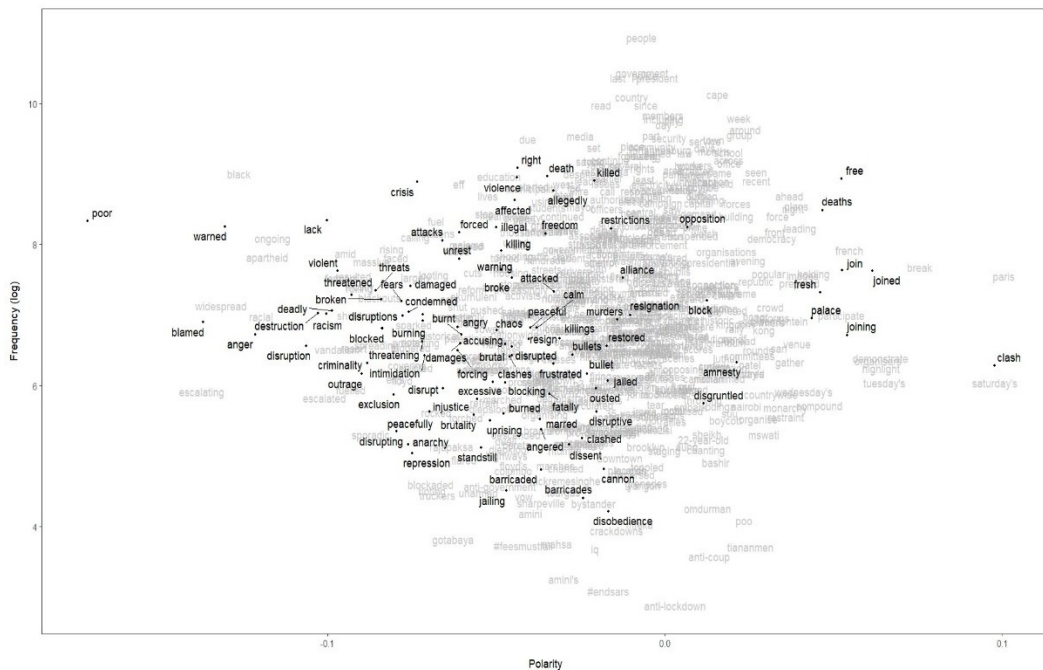


Figure 6.4: LSS with LSD Text plot terms: Protests

6.2.1. Bias and the 2023 National Shutdown: political leader news

As South Africa approaches the 2024 National Election, political parties and their leaders have begun positioning to curry favour. The media will serve a crucial function in the deliberations of the public in the lead up to election day. It is in this vein that the issue of agenda-setting, and media bias is examined through the lens of the 2023 National Shutdown. The shutdown was selected as it is a preceding event signalling the start of electioneering and a highly contested approach to political action.

In estimating the regression, three date points were chosen to ensure relevance to the shutdown period. The key date points identified included the announcement of the shutdown, 30 January 2023, and the date of the shutdown, 20 March 2023. The regression analysis was informed by the LSS outcomes estimated on revised patterns as per the leaders of each political party.

The outcomes of the regression model estimations are depicted in table 6.2. The first column represents the outcomes of the tonality change following the shutdown aggregated across all South African media publishers for each political leader. The South African media has on average significant negative tonality for Ramaphosa of the ruling party and Malema of the EFF. Although Steenhuisen is on the receiving end of negative tonality, this is insignificant. The shutdown was not perceived as successful by online media publishers with the event not achieving significant tonality change. The media's tonality bias against protest action is further evidenced with the significant positive tonality bias exhibited towards the president, Ramaphosa, following the National Shutdown. The president is the only political leader to gain significant positive sentiment, an indication of his ability to have quelled the unrest.

Model two provides the outcomes of the interaction between the media and the shutdown. With the perceived lack of impact, there is no significant tonality bias changes exhibited following the shutdown. All online media publishers evidenced significant negative sentiment on average across the period of the shutdown. This is line with previous outcomes of media tonality on protest action. Model three evidenced that overall media sentiment was significantly negative over the period, positive tonality bias is limited to specific political leaders and their role in the shutdown, whereby Ramaphosa benefits from the perceived failure of the shutdown. The state broadcaster worked to benefit the president; however, he also benefited from significant positive sentiment from News24 and to a lesser extent by EWN.

The LSS sentiment scores are plotted to give visual effect to the interactions between the media, and political leader news reporting on the shutdown as per figure 6.5. The first plot gives visual effect to the tonality of select media publisher's reporting on the shutdown. The SABC as the state broadcaster led overall negative sentiment on the shutdown, likely driving a state agenda. Of the commercial media publishers, negative sentiment was evenly distributed between EWN, News24 and the Citizen. The Daily Maverick evidenced the highest positive tonality amongst all media publishers. Figure 6.5 confirms the biggest gains for Ramaphosa were led by News24 and the SABC post the shutdown. As depicted in figure 6.5, Julius Malema experienced negative tonality bias as net sentiment declined across all media publishers, following his leading role in the shutdown.

The analysis revealed that South African media publishers have an adverse relationship with protest action and the role played by political leaders greatly influenced tonality bias, with the shutdown perceived as a failure by commercial and state media publishers.

Table 6.2: 2023 National Shutdown regression models: party leader news

	Model 1	Model 2	Model 3
(Intercept)	−0.099***	0.098*	0.173***
Ramaphosa	−0.150**		−0.335*
Steenhuisen	−0.090		−0.647*
Malema	−1.030***		−1.293***
shutdownTRUE	0.007	0.033	
Ramaphosa:shutdownTRUE	0.380***		
Steenhuisen:shutdownTRUE	−0.193		
Malema:shutdownTRUE	−0.239		
EWN		−0.324***	−0.293***
IOL		−0.203***	−0.204***
News24		−0.295***	−0.363***
SABC		−0.394***	−0.488***
Citizen		−0.361***	−0.343***
EWN:shutdownTRUE		0.115	
IOL:shutdownTRUE		0.147	
News24:shutdownTRUE		0.016	
SABC:shutdownTRUE		−0.067	
Citizen:shutdownTRUE		0.064	
Ramaphosa: Eyewitness			0.344+
News			
Ramaphosa: IOL			0.195
Ramaphosa: News24			0.640**
Ramaphosa: SABC Online			0.599***
News			
Ramaphosa: The Citizen			0.163
Steenhuisen: Eyewitness			0.593
News			
Steenhuisen: IOL			0.180
Steenhuisen: News24			0.674+
Steenhuisen: SABC Online			0.623
News			
Steenhuisen: The Citizen			0.392
Malema: Eyewitness News			−0.162
Malema: IOL			0.714+
Malema: News24			0.272
Malema: SABC Online			0.251
News			
Malema: The Citizen			0.258
Num.Obs.	3662	3662	3662
R2	0.036	0.020	0.057
R2 Adj.	0.034	0.018	0.051
AIC	10 041.3	10 109.0	9992.8
BIC	10 097.2	10 189.7	10 147.9
Log.Lik.	−5011.664	−5041.525	−4971.378
F	19.681	6.941	9.615
RMSE	0.95	0.96	0.94

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

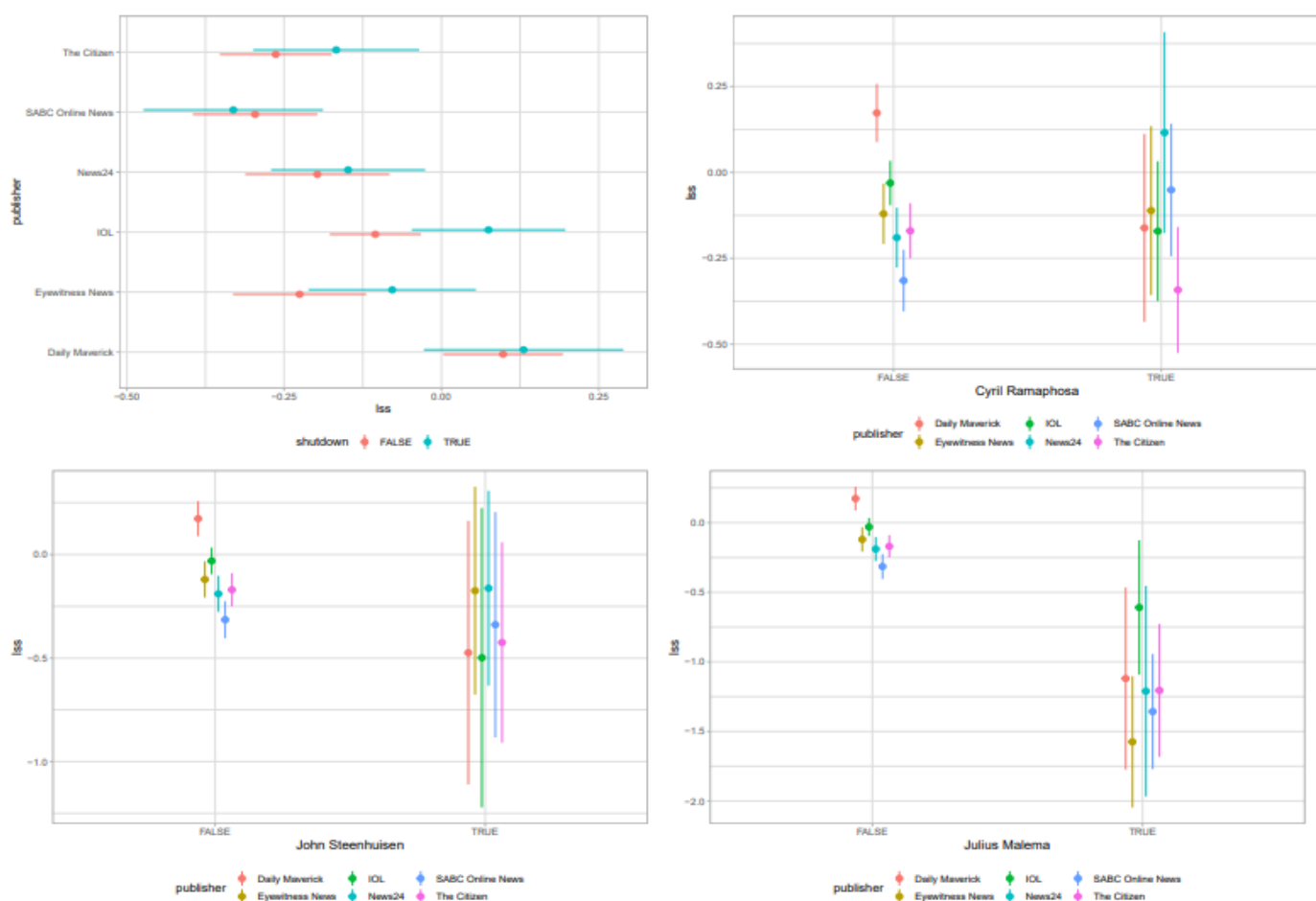


Figure 6.5: Regression interaction effect plots: 2023 National Shutdown party leader news

6.3. Discussion

In a fractured country such as South Africa, the media is critical in bridging or widening the divide through the narratives created in reporting. It is glaringly apparent that protest action is viewed through a narrow, pro market lens that leaves little room for engaging with the issues of socioeconomic imbalance and inequality. A significant example of negative tonality bias is evidenced in the media's response to the National Shutdown of March 2023, which likely had negative leakage effects for opposition party, the EFF.

The negative tonality bias in reporting further marginalises the majority. The tough stance taken on protest action likely has an impact on the deliberations of South Africans in understanding the socioeconomic divides which the media serves to entrench.

7. CHAPTER SEVEN- CONCLUSION

Thirty years on, South Africa has ostensibly lost the sheen of transitioning to a democracy and within an increasingly fractured socioeconomic environment, the tumultuous political landscape has become a dominant topic of contention. South African media publishers have been accused of not just reporting the news but fanning the flames. The progression in political news reporting since the attainment of democracy has been criticised to be untransformed, dominated by the white minority elite and not reflective of the socioeconomic environment of South Africa. South Africa, like many other nations negotiating trying socioeconomic times, coupled with a society built on structural racism, remains vulnerable to threats of greater divisive action.

With political actors having direct access to their constituents, often in fragmented societies, political news media publishers have asserted the need for continued media independence in their quest to keep leaders accountable and society informed. Media publishers have stressed that political actors have used new technological platforms to undermine the credibility of reporting. Opposingly, political hopefuls and incumbents maintain that their political aspirations are susceptible to the political leanings of news media publishers that cloud reporting. It is in this vein that political scientists have applied their research, to determine the hidden meaning in political text and the potential adverse influence on society, particularly during electoral cycles. With the 2024 National Elections rapidly approaching, the role of online news media publishers is an important facet to uncover within the South African political landscape, and how it impacts voting choices.

The study was thus initiated as an opportunity to marry the study of political communication and the use of data science techniques in answering topical issues currently being debated to have a central role in shaping South Africa. The study focused on tonality bias, defined as the framing of a political actor in comparison to other candidates. The sentiment methods utilised for the study included the Lexicoder Sentiment Dictionary (LSD), as the lexical-based methodology. This was followed by the utilisation of the Latent Semantic Scaling (LSS), a semi-supervised machine learning methodology.

In the lexical-based analysis, the LSD findings evidenced that the overall tonality of South African media was negative in online political news reporting with a significant focus on governance and ruling party oversight. There was indication of text similarity between South African online media publishers as a function of text correlation. This has the implication of limiting diversity of thought and only serving as a voice for the white minority elite of society.

The EFF fared poorly with South African media publishers with significant negative tonality bias, followed by the ruling party to a lesser degree with both the LSD and LSS. This differed with the improved tonality for the DA political party nearing the 2021 Municipal Elections. The LSD evidenced alignment in negative tonality bias towards protest action and increased negative sentiment for the EFF and the 2023 National Shutdown. This was in alignment with the lack of diversity in commercial and state reporting as evidenced in the negative bias towards social justice actions such as protests.

An overarching theme consistent among South African online media publishers evidenced with both the LSD and the LSS was the negative tonality towards populist policies and protest action. This is cause for pause in the ability of the media to drive policy and agenda-setting rather than reporting on political actors. Herman and Chomsky (2008) advanced that propaganda can remain a function of a competitive media regime when centralised within the elite. This occurs even when the media maintains objectivity as its central doctrine. Thus, agenda-setting becomes a greater concern when contextualised within the socioeconomic environment of South Africa. Commercial media publishers are likely to only represent the concerns of the white minority elite but have influencing power over the marginalised majority in policy priority setting. This leaves the marginalised majority voiceless with little room to shape policy priorities.

Whilst there was evidence of tonality bias across South African media publishers against the ruling party, it was not consistent, and the party did garner favour with a few commercial news media publishers, even though many were significantly negatively biased. The frontrunner across both methods would prove to be the EFF in garnering significant negative tonality bias.

The findings from the LSS further proved to be more nuanced and whilst they evidenced significant negative tonality bias, it was unequal across media publishers and political actors. Paradoxically, the LSS was more pronounced in establishing the role of the state-owned media publisher, the SABC, in narrating a pro government agenda. Together with the limited views provided by commercial media and the role evidenced by the state publisher, there is merit in the proposal to diversify media.

An important outcome of the study was the identification of leader effects as evidenced by President Ramaphosa and to a lesser extent, Malema of the EFF. Steenhuisen lacked visibility and the ability to differentiate himself from his political home, the DA. However, the relationship between visibility only serves to amplify the leader effects if tonality is positive.

The utilisation of sentiment methods worked well in establishing tonality bias and the comparative analysis emphasised the importance of ensuring the veracity of outcomes from any methodology. The lexical-based method and semi-supervised machine learning method improved on each set of findings but could have been further improved with a locally developed political lexicon.

7. Future study considerations

The results garnered from the study are important to the discourse on the methodologies utilised to assess political text, particularly in an African context. The study evidenced the relevance of sentiment analysis tools as evaluation techniques within the political sciences. There remains a significant gap in South African computational led political science research.

A secondary consideration is the need to confirm a South African lexicon for political text. Future work could thus look towards building a South African political lexicon likened to that of Young and Soroka (2012). This could be developed to further include diverse South African languages.

The study could have also benefited from an extension into content analysis and testing against human coding. Whilst computational methods allow for statistical results and greater volumes of data to be analysed, the inclusion of human coding and content analysis would have enriched the outcomes. This is indicative of the role that both qualitative and quantitative analysis can improve upon each other when applied simultaneously.

A consideration is the further application of the study to electoral political news reporting prior to the onset of the 2024 South African National Elections, with the inclusion of visibility and agenda-setting measures as forward indicators, rather than regressively. This could form the basis of electoral outcome predictions and a test for media accountability.

Lastly, the study has evidenced that South Africa still has significant opportunity to expand the use of sentiment tools in the political sciences, and to marry these with other forms of data such as polling data, economic data amongst a plethora of sources, to greater inform the policy space. Key to this, is expanding the reach and impact of political science deliberations beyond research shelves and into the forefront of political communication deliberations.

REFERENCES

- Barclay, F. P., Venkat, A., & Pichandy, C. (2015). Media effect: correlation between press trends and election results. *Media Asia*, 42(3-4), 192-208.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774-774.
- Brändle, V. K., & Eisele, O. (2023). A Thin Line: Governmental Border Communication in Times of European Crises. *JCMS: Journal of Common Market Studies*, 61(3), 597-615.
- Chiang, C. F., & Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of economic studies*, 78(3), 795-820.
- Chiumbu, S. (2016). Media, race and capital: A decolonial analysis of representation of miners' strikes in South Africa. *African Studies*, 75(3), 417-435.
- Chiumbu, S. H., Reddy, V., Bohler-Muller, N., Gumede, N. A., & Mtshengu, A. (2018). Social justice for the poor: The framing of socioeconomic rights in selected South African newspapers. *Journalism*, 19(7), 959-975.
- Cornelissen, L. A., Daly, L. I., Sinandile, Q., de Lange, H., & Barnett, R. J. (2019). A computational analysis of news media bias: A South African case study. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists* 2019, 1-10.
- Daniels, G. The relationship between the media and democracy. (2013). *Fight for Democracy: The ANC and the Media in South Africa*. Wits University Press, 22-40.
- Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707-714.
- Duncan, J. (2014). South African journalism and the Marikana massacre: A case study of an editorial failure. *The Political Economy of Communication*, 1(2).
- Economic Freedom Fighters[@EFFSouthAfrica]. (2023, January 29). *CIC@Julius_S_Malema addressing #EFFPresser. The EFF has resolved to hold a national shutdown to demand the return of electricity and for the resignation of Cyril Ramaphosa on the 20th of March 2023*. X. <https://x.com/EFFSouthAfrica/status/1619742535261503490>.
- Eberl, J. M., Boomgaarden, H. G., & Wagner, M. (2017). One bias fits all? Three types of media bias and their effects on party preferences. *Communication Research*, 44(8), 1125-1148.
- Endersby, J. W. (2011). Fair and balanced? News media bias and influence on voters. *News Media Bias and Influence on Voters*, August 1, 2011.
- Ferreira, E. (2022, 30 March) . ANC defeats motion of no confidence in Ramaphosa's cabinet by 100 votes. *Mail&Guardian*. <https://mg.co.za/politics/2022-03-30-anc-defeats-motion-of-no-confidence-in-ramaphosas-cabinet-by-100-votes/>.
- Davis, R. (2024, 30 January). How the reports into the July 2021 unrest let South Africa down. *Daily Maverick*. <https://www.dailymaverick.co.za/article/2024-01-30-how-the-reports-into-the-july-2021-unrest-let-south-africa-down/>.
- Davis, R., Nicolson, G., & Simelane, B. (2021, 08 August). South Africa's three bloodiest days: 342 dead and we are still in the dark. *Daily Maverick*. <https://www.dailymaverick.co.za/article/2021-08-08-south-africas-three-bloodiest-days-342-dead-and-we-are-still-in-the-dark/>.

- Friedman, S. (2011). Whose freedom? South Africa's press, middle-class bias and the threat of control. *Ecquid Novi: African Journalism Studies*, 32(2), 106-121.
- Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of political Economy*, 114(2), 280-316.
- Hackett, R. A. (1984). Decline of a paradigm? Bias and objectivity in news media studies. *Critical Studies in Media Communication*, 1(3), 229-259.
- Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & quantity*, 51, 2623-2646.
- Haselmayer, M., Wagner, M., & Meyer, T. M. (2017). Partisan bias in message selection: Media gatekeeping of party press releases. *Political communication*, 34(3), 367-384.
- Haselmayer, M. (2021). Candidates rather than context shape campaign sentiment in French Presidential Elections (1965–2017). *French Politics*, 19(4), 394-420.
- Herman, E. S., & Chomsky, N. (2008). *Manufacturing consent: The political economy of the mass media*. Random House.
- Hopmann, D. N., Vliegenthart, R., De Vreese, C., & Albæk, E. (2010). Effects of election news coverage: How visibility and tone influence party choice. *Political communication*, 27(4), 389-405.
- Hossain, A., Karimuzzaman, M., Hossain, M. M., & Rahman, A. (2021). Text mining and sentiment analysis of newspaper headlines. *Information*, 12(10), 414.
- Lippmann, W. (1922). The world outside and the pictures in our heads. *Public opinion*.
- Masson, E. (2022, 14 February). DA tables motion of no confidence in Ramaphosa's cabinet. *Mail&Guardian*. <https://mg.co.za/politics/2022-02-14-da-tables-motion-of-no-confidence-in-ramaphosas-cabinet/>.
- McCarthy, K. J., & Dolfsma, W. (2014). Neutral media? Evidence of media bias and its economic impact. *Review of Social Economy*, 72(1), 42-54.
- Merkley, E. (2019). Partisan bias in economic news content: New evidence. *American Politics Research*, 47(6), 1303-1323.
- Media Monitoring Africa (n.d.). About MMA. <https://www.mediamonitoringafrica.org/about-us/>.
- Mullainathan, S., & Shleifer, A. (2005). The market for news. *American economic review*, 95(4), 1031-1053.
- Plaut, M. (2018). Media freedom in South Africa. *The Round Table*, 107(2), 151-162.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Radebe, K. (2019). Emergence of an authoritarian democracy: the ghost of Nat Nakasa. *Acta Academica*, 51(1), 125-143.
- Rauh, Christian. "Supranational emergency politics? What executives' public crisis communication may tell us." *Journal of European Public Policy* 29, no. 6 (2022): 966-978.
- Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5, 1-29.
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. "O'Reilly Media, Inc."
- Similar web. (n.d.). Traffic and engagement. <https://www.similarweb.com/topwebsites/south-africa/news-and-media/>.

- Soontjens, K., Van Remoortere, A., & Walgrave, S. (2021). The hostile media: Politicians' perceptions of coverage bias. *West European Politics*, 44(4), 991-1002.
- South African National Editors' Forum. (2020, August). *Constitution of the South African National Editors' Forum Constitution amended 22 August 2020*. <https://sanef.org.za/wp-content/uploads/2020/12/Signed-Sanef-Constitution-August-2020.pdf>
- South African National Editors' Forum. (2022, June). *The state of news media*. <https://sanef.org.za/wp-content/uploads/2022/08/State-of-the-Media-June-2022-SANEF.pdf>
- South Africa's Ramaphosa avoids impeachment: A timeline. (2022. 13 December). *Aljazeera*. <https://www.aljazeera.com/news/2022/12/13/south-africas-ramaphosa-faces-impeachment-debate-a-timeline>.
- Strömbäck, J. (2008). Four phases of mediatization: An analysis of the mediatization of politics. *The international journal of press/politics*, 13(3), 228-246.
- Takens, J., Kleinnijenhuis, J., Van Hoof, A., & Van Atteveldt, W. (2015). Party leaders in the media and voting behavior: Priming rather than learning or projection. *Political Communication*, 32(2), 249-267.
- Wasserman, H. (2020). The state of South African media: A space to contest democracy. *Publizistik*, 65(3), 451-465.
- Watanabe, K. (2017). Measuring news bias: Russia's official news agency ITAR-TASS' coverage of the Ukraine crisis. *European Journal of Communication*, 32(3), 224-241.
- Watanabe, K. (2021). Latent semantic scaling: A semisupervised text analysis technique for new domains and languages. *Communication Methods and Measures*, 15(2), 81-102.
- Wasserman, H., Chuma, W., & Bosch, T. (2018). Print media coverage of service delivery protests in South Africa: A content analysis. *African Studies*, 77(1), 145-156.
- Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29(2), 205-231.