A Forensic Interpretation of Hateful Micro-Speech Acts and Performative Modality in Facebook and Twitter during 2017 Election-Kenya

Fredrick Meeme Irimba [1] 👤 📧 , **Jacinta Ndambuki** [2] 👤 **and Florence Mwithi** [3] 👤

[123]*Department of Literary and Communication Studies, Laikipia University, Nyahururu, Kenya*

📧 **Corresponding Author**: Fredrick Meeme Irimba, **E-mail**: fredrickmeeme@yahoo.com

ARTICLE INFORMATION	ABSTRACT
	The increasing shift of human activities to online spaces in Kenya has resulted in the new behaviours among internet consumers. One such behaviour is the growing online public journalism phenomenon amid legal and regulatory gaps permeating expression of online hate speech rhetoric disguised as 'politically correct talk' which often goes unquestioned despite its injurious force and the potential to precipitate physical violence in the long run. To judge content as hateful, Kenya's judicial processes rely the establishment of speech intention to hurt a legally protected entity. However, hate speech law enforcers lack skill and capacity to accurately determine the pragmatic force of hateful language. This article, which is a part of broad study that examined the discursive construction of online hate rhetoric, examines the injurious potential of online micro-speech acts and performative modality of selected Facebook posts and tweets constituting the day-to-day communicative practices online during the 2017 general election in Kenya. Working within forensic-based Computer Mediated Discourse Analysis (CMDA) framework, we analyse a purposive sample of 160 posts; FB (120) and Twitter (40) collected through online observation of Facebook groups and hashtags trending in Kenya between July and November 2017. The findings show how micro-speech acts and performative modality worked in service of aggressive ideology in the form of overt and covert appeals for collective prejudice against marked ethno-political out-groups. These insights are relevant for policy makers such as NCIC, KHR and CAK as well as the hate speech law enforcers especially National Police Service and prosecutors in understanding how certain commonsensical day to day online communicative practices yield pragmatic potential to propagate ideologically rooted culture of hate and violence in multi-ethnic cultural contexts such as Kenya.

1. Introduction

Facebook (FB) and Twitter in Kenya are perceived to afford previously silenced individuals and groups a rare chance for free participation in controversial socio-political debates that often degenerate into extreme sentiments in the form of hateful online verbal flames. We focus on FB and Twitter communication in Kenya during the electioneering period; the time believed to herald more online hate speech than other times. This work is located within Computer Mediated Discourse Analysis (CMDA), a multidisciplinary approach dealing with communication produced when human beings interact via networked ICT devices. We draw from KNCHR (2018) and NCIC Act (2017) to coin the term online hate rhetoric which describes any speech action in form of posts or tweets intended to hurt, intimidate, threaten, degrade and embarrass others or promote hatred and violence against groups based on their ethnic and or political affiliations. We argue that although the majority of Facebook and Twitter users employed a fairly neutral language, a significant number of users reproduced targeted speech activities whose design and context comprised overt and covert appeals for collective prejudice against marked ethno-political out-groups. The article interrogates the discursive mechanisms underlying the linguistic choices made, potential audience meaning as well as their social implication

2. Overview of Research on OnlineHate Speech Discourse

Hate speech online has drawn a lot of scholarly attention over recent years although this has not been quite the case in Kenya despite the apparent shift of hate speech from the mainstream media to the online platforms. Online platforms are often perceived to be less regulated, rich in diverse audience and communicative modes which users often leverage to achieve deliberate speech intentions (Anat & Matamoros, 2016) while at the same time, feigning compliance with the existing user policy and legal provisions.

Siegel's (2008) prediction in the early days of web 2.0 that internet would be another communicative arsenal for 'racist' and 'hate mongers' to spread hate messages is witnessed in the upsurge of online hate speech which Anat and Matamoros (2016), Eltis (2013), Perry and Olson (2009) and Duffy, (2003) attribute to the lack of the editorial oversight that is mandatory in the mainstream media. Perry and Olson (2009) found out that the Web creates new common spaces that foster 'collective identity' for previously fractured hate groups resulting to 'web anarchy'. Concurring with Perry and Olson (2009), Commaerts (2009), cited in Christoforou (2014), demonstrates how the internet hosts racial hatred and discrimination talks. Mullen and Leader (2005) on the other hand, examined the nature of derogatory language in the American perspective and came up with contextualized "nouns that cut slices" and promote prejudice and exclusion of the target. Kenya comprises a uniquely different context thus making it interesting to find out how this alters the manifestation forms and patterns of online hate speech.

2.1 Power, Ideology and Discourse Practice in the Web

Simply defined, discourse is spoken or written language above the sentence (Cameron, 2001 p.11). Fairclough (1995) expands the concept 'discourse' to include the whole array of other semiotic activities such as visual images whose type and patterns can reveal important social relations and suggest the implied and nuanced audience meaning in regard to the portrayal of sensitive topics. Discourse, power and ideology are three social forces that are in intimate relation in any society (Fairclough, 2001, p. 2). According to Bourdieu (1988) cited in Ndambuki (2010, p. 77), there is a link between linguistic practice and forms of power evident in pervasive features of society such as inequality. As construed in this study, hate speech is a form of identity-based unequal and discriminative representation of an individual or group in deficit discourse based on real or imagined characteristics. For example, van Dijk (1993) argues that assertions and micro speech activities evident in the contributions made to national debates by those in authority can serve to confirm and legitimize certain constructed realities, including those of discriminative nature.

In conceptualising the discourse practice, Bourdieu (1988) uses the term 'habitus' to define the set of predisposes which incline agents to act and react in certain predisposed ways that generate practices, perceptions and attitudes as basis of social representation (Dijk, 1998) which is traceable in discourse as ways of speaking, writing, doing, being, believing and valuing. Other theorists such as Hodge and Kress (1979), Thompson (1997), Wodak (2015) and van Dijk (2005) explicate the complex relationship among discourse, power and ideology in contemporary society by identifying strategies by which language practice is employed not only in communication but also as discursive tool for social-political control in service of power and ideology. This idea is further expounded by Wodak (1997), Yieke (2002), Herring (2007) and Ndambuki (2010) who use Discourse Analysis Approaches to demonstrate how the apparently harmless daily conversational practices such as turn taking, interruptions, choice of words and topic structure appear to exercise subtle forms of power and symbolic linguistic violence within various socio-political and institutional contexts. However, Hannah (2002) cited in Posset (2017) attempts to delink language as a system from violence by warning that if language is violent, it is not its inherent attribute. Rather, it depicts the inner disposition of the language user and the wider society.

According to Herring (2004), computer networks, including Facebook and Twitter, do not normally guarantee a democratic equal opportunity interaction. Rather, the pre-existing discriminative social arrangements, bias, and power symmetries carry over into cyber space as users leverage its multiple semiotic ensembles by deliberately choosing those that suit their purpose to perform violence and discrimination online. Therefore, we view online hate speech as a form of extreme oppressive power struggle performed not only against virtual individuals but also their offline networks through persistent use of profiling expressions and speech acts that overtly or covertly classify and negatively characterise a social entity often with undertones of appeals and attempts to justify collective discrimination of the perceived 'threat'. However, not much research work known by the researcher has focussed on how twitter and Facebook communication practices produce, propagate and shape the online hate discourses in the Kenyan context where hate speech, despite being a proscribed misdemeanour, keeps incarnating and persisting in the social media platforms while successfully circumventing various monitoring systems. It is against this backdrop that it becomes not only interesting but also necessary to investigate how aggressive ideologies are embedded in the performative features of discourse in the light of the new media while evaluating its implication on the text consumers and the society as a whole

2.2 Inferential Intentionality and Expressions of Hate Speech

Although the judgement of content as hate speech in Kenya depends on proof that the speaker's speech intention is indeed to hurt, Lotta *et al.* (2016) posit that there is no easy way to getting into the speaker's mind. Nevertheless, Lotta *et al.* (2016) impute

that certain linguistic clues can point out to the intention as resides in the speaker's topic and a myriad of other discourse strategies. Levinson (2006, p. 87) warns of potentially ambiguous utterances, which may sometimes result in differences between the original intention of the author and the intention as revealed by the clues (Lotta *et al.*, 2016). Austin (1962) identifies a class of utterances he called performatives because when uttered in certain context, they perform an action in and of themselves. For Searle (1992b), the meaning of the acts of speech can be analysed in terms of intentional states such as beliefs or intentions. Both Searle (1983) and Lotta *et al.* (2016) suggest that the mind 'imposes' intentionality on the linguistic expression, in that the basic intention to represent influences the choice of meaning-making resources which can be interpreted in context to shed light on the overall subject positioning and the agency of participants in tokens of communication. Searle (1983) contends that linguistic output can shed light on the intentionality of mental states that underlay the specific act of communication and bestow on such act the so called conditions of satisfaction captured.

In brief, Searle (1983) believes that mental states such as beliefs, attitudes, aversions, wishes, motives or resolves impose satisfaction on the expressions. Searle summarises John Austin's speech acts theory into five acts representing various intentions of acts speech. These include: representative acts which commit to the truth of the propositions in question and they include claiming, reciting, describing, concluding, suggesting, predicting and so on); declarative speech acts which seek to change the reality in accordance with the propositions of the speakers and these include baptising, labelling, pronouncements, etc; directive speech acts which are meant to cause the hearer to take some specified action, e.g. requesting, commanding, warning etc; expressive speech acts which express speaker's attitude and emotions towards the proposition, e.g., congratulating, complaining, scolding, apologising and so on; and finally commissive acts which denote committing the speaker to particular action such as promising, threatening, vowing, betting and challenging. This study seeks to carry out a CMDA of selected micro-speech acts with a view to unravelling the underlying speech intention and ideological motivation of their production as evident in the expressive linguistic clues.

2.3 Cyber Forensics and Linguistics Approaches to Hate Speech

Hate speech has been characterised as a social misdemeanour perpetrated primarily through language hence, according to Bardici (2012), it is difficult to analyse contemporary online hate speech independent of the language structure, conventions and the contexts that make it a possible instance of linguistic violence. To buttress Bardici's views, Wafula (2016) admit to the existence of a glaring limitation in attempting an interpretation of hate speech without a basis in a linguistic theory to reveal whether what meets the criteria of the monitoring software as hate speech is indeed hurtful or not. These observations underline the significance of linguistic analysis as a way of gaining understanding into the attributive, performative and interpretive role of language which is central not only in understanding how hate speech plays out but also in explaining how the target is able to perceive it as hateful. Various attempts have been made in applying linguistic insights to the analysis and investigation into what entails language crimes (Olsson, 2013) thus giving rise to Forensic linguistics as a field of inquiry that focuses on investigating authorship attribution, authenticity of witness statements and more importantly for this paper, the meaning imputation and disputation in the potential hate speech messages.

Whereas Cyber forensics in general, entails an application of scientific methods and tools to identify, extract, document and interpret digital data for digital evidence or for a root cause analysis (Wafula, 2016, p. 49), the forensic linguistics approach applies ethnographic approaches to investigate interaction in order to understand how hate speech is constructed in the context of in-groups and out-groups. Terence (2014) and Wafula (2016) characterise cyber hate forensics as monitoring web forums such as the FB and Twitter and identifying the potential hate speech. On the other hand, forensic language analysis involves a determination of meaning through objective analysis of key speech acts, key words, phrases, and other linguistic materials documenting them and giving a comment on their potential interpretation in context.

Olsson's (2013) argues that analysing social media postings can reveal whether they are truly hateful or not in relation to an existing legal framework. To this end scholars have applied discourse analysis techniques for forensic purposes. For example, Prentice *et al.* (2011) uses forensic discourse analysis approaches on the language of extremist hate group with a view to identifying beliefs, motivations and justification of popular extremism hate discourse while Watt *et al.* (2013) uses discourse analysis approach to document inferences of threats in apparently neutrally worded utterances.

This article interrogates how specific speech acts and contextual resources are leveraged to create group entities as worthy targets of discrimination, which in Brink's (2001) terms results to exclusion and emotional or physical violence.

3. Theoretical Framework

This paper applies Herring's (2004, 2007) Computer Mediated Discourse Analysis hereafter known as CMDA to carry out contextualised observation of the day to day online communication identifying potentially hateful posts and tweets; distinguishing them from neutral ones, describing and interpreting them while inferring their potential audience meaning in their specific contexts of use. The approach provides theoretical lenses for analysing an aversive online communicative behaviour-hate

speech- which is evident in micro level online textual traces. Herring (2019) grounds CMDA approach around four linguistic levels showing how micro level analysis can be used to shed light on macro level online phenomenon such as hate speech. These levels include 1) structure- dealing with typography, morphology and syntax; 2) meaning which deals with contextual meaning of words, utterances, speech acts and exchanges; 3) interactional management which entails interactivity, turns and coherence and finally, 4) social phenomena which entails linguistic expressions of power play, conflict, identity, group membership and cultural differences as evident in discourse and rhetoric styles (Herring, 2019, p. 27). This paper focuses on meaning level to do a contextual interpretation of speech acts by means of pragmatic analysis techniques borrowing from Austin's (1962) Speech Acts Theory as reviewed by Searle (1996) to explicate how the illocutionary force and the potential perlocutionary effect of the illocutionary speech were leveraged by Facebook and Twitter users to achieve the desired hateful speech intention. Systemic Functional Grammar postulated by Halliday (1985) offered descriptive categories for analysing performative modality as a tool of understanding the aggressive attitudes/emotions and the stance of the hate speech author towards the objectified predicate of clauses.

Although Herring's (2007) Faceted Classification Scheme for contextual interpretation of social situational and medium factors provided the general context for the broad study, the socio-situation factors, and not the medium factors, were used in interpreting online socio-political context which is the point of focus in this paper.

4. Research Methodology

This study employs a qualitative case study design using purposive sampling to observe and collect the materials which had characteristics relevant to the objective of this paper (Strauss & Corbin, 1998). One hundred and twenty posts from Facebook and forty tweets drawn from the initial sample of three hundred and sixty posts were collected by observing politically affiliated Facebook groups and trending hashtags on topics related to key political and social events over the period between July and November 2017. Public groups, rather than private groups, were preferred since they offered the diversity in terms of membership from different socio-political affiliations and cross-cutting themes which captured the typical context for identity-based discourse. The posts and tweets were actually read within their conversational context while paying attention to the specific posts and tweets whose apparent communicative content and intent appeared to hurt, threaten, intimidate, degrade, embarrass others or promote feeling of hatred, enmity and encourage violence against individuals and groups based on the perceived ethnic and political associations (KNCHR, 2018, p. 7; NCIC Manual 2017, pp 10-12).

Internet enabled and memory enhanced tablet was used to monitor the unfolding digital communication in Facebook and Twitter asynchronous conversations while taking screenshots of posts and tweets of interest before saving the data in retrievable Joint Photographic Experts Group (JPEG) format. According to Shuy (2014), the intention and the actual meaning of any speech must be substantiated through a careful examination of such speech in its context. Therefore, Herrings' (2007) social situation factors of CMDA were drawn on in analysing the context of specific text by posing relevant questions about the data in order to capture essential aspect of socio-situation context such as the participants **(S1)** ,conversational structured (**S2**), overriding goal of the studied online sites as well as purpose of the specific speech events (**S3**), topics (**S4**), tone of participation (**S5**), speech activities involved(**S6**), norms of communication (**S7**) and finally the codes and modes used in communication (**S8**). The analysis involved identifying, classifying and counting actual instances hate speech acts and modality across data sets from Facebook and twitter.

5.Results and Discussion

5.1 Micro Speech Acts in Service of Aggressive Ideology

Most of the online hate messages analysed occurred in the context of internet flaming which is an online verbal exchange characterised by aggressive, impolite and violent language which results from discrepancy of ideas and opinions between interlocutors (Fracchiolla, 2013) (Cited in Lotta *et al.* (2016). The Findings show how specific micro speech acts and aspects of modality ranging from simple lexicon to more complex pragmatic structures worked together in the service of aggressive ideology by positioning individuals and groups as 'deserving' victims of violent speech acts that contained propositions ranging from subtle to extreme forms of discrimination (Irimba, 2014, pp. 129-130), micro-aggression (Kassia, 2015, p. 28) and explicit threats as well as calls to harm the perceived 'Other'.

Relevant speech acts associated with the verbal aggression and hate speech from both the Facebook and Twitter were categorised as shown in table 5.1

Table 5.1 Hateful Speech Acts by Type and Intention

Speech acts by type	Speech Activities	Frequencies of occurrence			
		Facebook	Twitter	Total	Percentage (%)
Representative acts	Describing	31	5	36	12.7
	Claiming/asserting	6	2	8	4.6
	Predicting	7	3	10	5.8 {23.1%}
Directive acts	Commanding	10	5	15	8.8
	Warning	25	7	31	15 {23.8%}
Commissive acts	Threatening	23	3	26	15
	Vowing	7	-	7	4
	Challenging to act	4	1	5	2.9 {21.9%}
Expressive acts	Complaining	7	2	9	5.2
	Scolding	5	2	7	4 {9.2%}
Declarative acts	Pronouncements	8	6	14	8.1
	Labelling	19	5	24	13.9 {22%}
Totals		**120**	**40**	**160**	**100%**

Source: Researcher's analysis of Facebook groups' walls and Twitter hashtags, 2020

Overall, Violent speech acts featured 160 times out of 950 texts analysed for linguistic features. Generally, speech acts observed were classified into different illocution types (Searle, 1996) but just for analytical convenience and for in-depth understanding of specific speaker's intentions but in practice, all instances of hate speech acts observed tended to combine various types of illocutionary acts in the same proposition in what appeared to be participants' effort to produce what they believed to be the most effective perlocutionary force to suit specific communicative intentions ranging from biased representation, issuing instruction to violate or cause to violate those who are not 'Us', influencing to believe, think or inciting to act in prejudicial manner against the objectified target individuals and groups.

The 160 speech acts correspond to Searle (1996) classification of speech acts as follows: negatively representative acts, 23.1 % (comprising negatively describing, asserting & predicting micro speech acts); directive acts, 23.8 % (comprising commanding, instructing to harm & warning); declarative acts, 22% (comprising violent pronouncements and negative labelling); commissive acts, 21.9 % (including threatening to hurt, vowing to hurt and challenging others to act with prejudice toward the perceived enemy) and finally expressive acts at 9.2 % (complaining and scolding). Austin (1962) and Searle (1996) submit that language is a tool for performing discursive actions and that the meaning ascribable to a linguistic expression such as Facebook and Twitter posts indicates the producer's denotative as well as connotative communicative intentions rather than just a presentation of the

sum total of the meanings of words and expressions used. In relating language use and violence, Posset (2017) maintains that although language is not violent in and of itself, it gives an expression to the underlying state of the actual source. Evidence tabulated above also points out to deliberate use of aversive impoliteness and aggressive performative modality as constitutive part of violent speech acts and which aggravate their injurious perlocutionary force

Following are the examples showing the speech acts analysis of specific posts in order to reveal their violent locutionary, illocutionary and perlocutionary acts.

Text F1

1.Locution: <>*Islam is a religion of murdering, intimidating, fighting others, demeaning others, terrorism, jins, pirates, Jihads, suicidal missions to kill innocent….etc*

2. Illocutionary act: -direct (negatively describing and claiming)

-Indirect (warning)

3. Intended perlocutionary effect: causing fear and anxiety that can lead to isolation and revengeful aggression against the named target.

Text F1 was a response to an apparently misplaced post inviting participants in a political FB group to join Islam. The response is an Islamophobic remark in which the writer who takes the stance of a non-Muslim (Us) assumes a conclusive tone in describing what he believes Islam (Them) to be. The post comprises a locution that corresponds to the representative speech acts types advanced by (Searle, 1969). The locution pragmatically yields illocutionary act at two levels; directly (negatively describing) and indirectly (sending a warning). Both direct and indirect illocutionary acts were considered in this analysis. Whereas direct illocutionary acts contain a performative verb indicating the action, the indirect speech acts manifest the speaker's illocution act by relying mainly on the ability of the hearer to draw inference from what is being said and act accordingly. The direct illocutionary force in text F1 above entails attributing the subject by ascribing negative characteristics. Therefore, this representation further yields the indirect (unspoken) illocution act of warning the in-group about the adversary (out-group) who has the potential to cause harm to them. Lotta *et al.* (2016) suggest that the mind 'imposes' intentionality on the linguistic expression. In this view, at yet another level, the direct and indirect illocution acts are intended to wield the perlocutionary effect of provoking the feelings of terror, fear and anxiety on the part of the in-group participants and depicting any engagement between the in-group members and the Islam and by extension with Muslim faithful (out-group) as an extremely dangerous affair which should be avoided at all cost. Out-group participants reading such a post may also experience a different but equally disharmonious perlocutionary effect of feeling misperceived, isolated and threatened which according to FRA (2012) and Posset (2017) causes fear and anxiety on their part. Indeed, whatever the direction the perlocutionary effect takes, it is generally a recipe for disharmony, mutual suspicion and emotional pain that can also escalate to physical acts of violence.

Text T2

1. Locution

<>*… for the Kikuyus to respect U (you) they must be slaughtered to the ground [.] for Raila to be the president I think a particular tribe must be multiplied by zero* (Twitter)

2. Illocution acts: -direct (representative act of asserting, directive act of ordering, commissive act of vowing and betting

-indirect illocution (threatening)

3. Expected perlocutionary effect: (inciting violence against identified target)

The Twitter text above occurred under the hashtag #kenyapoll2017. The conversation contained views from across the political divide with participants drumming up for support of their preferred candidate. The tweet was a response to a previous one which stated that Raila (NASA) will never lead Kenya. T2 which combines representative illocutionary acts, directive acts and acts of commission for maximum perlocutionary effect that seemingly intended to incite genocide attacks on a particular Kenyan ethnic group depicted as a stumbling block and hindrance for their preferred presidential candidates from other ethnic groups perceived as a minority to ascend to power. This captures Glowinski (2015) characterization of hate speech as a highly contextual phenomenon where the considered object of hate must be destroyed- the sooner the better. Zero symbol (0), a mathematical figure that reduces its multiples to nil is combined with the imperative modal verb 'must' of obligation to make a strong call to violently eliminate the already marked and labelled out-group.

Austin (1962) outlines the felicity conditions that must be fulfilled for the perlocutionary force to be realised. Some participants attempted infusing authority, authenticity and other favourable felicity conditions to their online illocutionary hate speech through impersonation of popular personalities. Texts T3 and T4 clearly demonstrate this strategy.

Text T3

Text T4

Texts T3 and T4 above bear what appears to be retaliatory incitement to gender-based violence. Tuiyot (2013), in her comparative analysis of electoral gender-based violence in Africa maintains that violence against women is widespread in Africa and manifest in different context such as conflicts, humanitarian situations, electioneering periods and many other situations. Whether the alleged authors (Charity Ngilu and Nimrod Mbai actually uttered the quoted words exactly as they are or not, is not immediately clear but the two posts appear to have gone through deliberate rigorous multimodal message designing efforts that cannot be taken for granted. First, the tweets were not collected from the alleged authors' accounts but rather from two separate twitter hashtags. Secondly, the appendage of the supposed authors' names and their coloured photos sought to both legitimize and rationalize (Thompson 1997 & Wodak and van Leeuwen 2015) the violence contained thereof by adding credibility to the aggressive propositions and perlocutionary effects of the posts. The perlocutionary effects of appending Ngilu's authorship to Text T3 is tarnishing her reputation as a credible candidate for gubernatorial seat especially among women who were believed to constitute the majority of her supporters while the two texts generally wield perlocutionary effect that include threatening, inciting to hurt and causing anxiety as well as intimidating female voters.

The findings further corroborate Tuiyott's (2013) claim that in Africa, GBV is often used by the opposite gender as a tool to intimidate and exclude women from active political participation. Text T3 above points out two possibilities; firstly, if indeed the text is from Ngilu, a female politician, then it would point at the attempts by women politicians to use coercive force to intimidate their women folk into supporting them. on the flipside though, the post could actually be a case of gender violence in form of political attacks against the purported author by depicting her in bad light to other women who were perceived as constituting majority supporters in her gubernatorial candidature in 2017 general election which she won becoming one of a paltry three female governors against forty four male counterparts in Kenya.

Text F 5

1. Locution<>*ONYO!!! ONYO!!! ONYO!!!Wakikuyu, Wakisii na Wameru wote mwatakiwa kuhama Kajiado county kuamukia usiku wa 7/8/2017 vile alivyoamrisha Rais mtarajiwa… RAILA AMOLO ODINGA. PUUZA ILANI HII NA UTAJUTA CHA MTEMA KUNI*

(WARNING!!! WARNING!!! WARNING!!!All the kikuyus, Kisiis, and Merus are required to vacate Kajiado County on the night before 7/8/2017 as the president to be, RAILA AMOLO ODINGA directed. Ignore this notice and you are going to regret).

2. Illocutionary act: direct-(naming, commanding/directing, vowing and challenging)

                                Indirect-(warning/threatening and predicting violence)

3. Intended perlocutionary effect (defamation, stir fear& anxiety, incite eviction)

Text F5 was collected from 'ODM youth' Facebook group which was affiliated to NASA coalition, one of the two major political factions in 2017 election. The post was uploaded a few days after a campaign tour to Kajiado; a perceived NASA stronghold, by

the NASA flag bearer who allegedly asked the locals of Maasai origin to vote wisely or else they would remain poor and marginalised only selling their native land to people from other ethnic groups mentioned in the text and who were perceived to be ardent supporters of the rival Jubilee party. The text combines both the aspects of commissive and directive speech acts of violence packaged in form of a notice that begins by drawing the reader's attention to the salient issue through the use of capitalisation and nonconventional use of multiple exclamatory punctuation marks; a feature of Facebook language (Mwithi, 2016, p. 196) that in this case serves to enhance the textual tone of the urgency. The locution itself contains an illocutionary act of commanding or directing the perceived out-group to vacate the area named. The user then goes further to vow on the intended commission of unknown violence (prediction) against those out-group members who may choose to defy the proposition thereby contained in the notice. Although the illocution act derives authority from its tone of urgency, it is made more authoritative by deliberately appending the name of a popular political party leader with profound political influence whose inclusion is seen as a key factor for the perlocutionary effect to be achieved successfully. According to van Dijk (2005), this is a form of a*uthentication* of personal argument making it appear binding and worthy of support (Thompson, 1985, p. 96). The intended perlocutionary effect appears to be causing fear, anxiety on the target groups and also stimulating collective action of prejudice against the perceived enemy and possible forceful eviction of the marked out-group from the region in question.

Text F6

<>*I will beat anyone wearing clothes nasa, odm, wiper, anc cloths on Tuesday watajua hii Nairobi iko na wenyewe* ..(they will know this Nairobi has the owners)*(Starehe M.P Charles Jagua)*

F6 was posted in Jubilee party supporters FB group two days before the return of Raila Odinga from a tour of Britain amid claims that NASA supporters would congregate at the airport to welcome him back against state prohibition of such gathering. F6 was posted by individual user who appeared to attribute the claims to other sources by appending the names of a politician in order to create deliberate ambiguity (Sindoni, 2018, p. 74), as a way of tactfully avoiding the responsibility of being sought after as the originator the prejudicial information or as a way of citing credible and authoritative sources which build up on the discursive force of such statements.

A. *5.3 Modality as a Persuasive and Manipulative Device*

Following Fowler (1985) and Lillian (2008), the researcher read through the Facebook and Twitter sample highlighting the occurrence of modal forms of interest before classifying them in accordance with Fowler's (1985) and Palmer (1986) categorisation paying attention to the context of use. Perfomative modality appeared to reinforce various violent speech acts by imposing conditions of desirability- suitability, moral acceptability and obligation where threats and violent instructions are presented as obligatory information that needs to be followed without the privilege of the addressees deciding on their own hence manipulative. Generally, modality was mainly expressed by modal verbs 'ought to'; adverbs such as possibly, certainly, probably, etc; adjectives such as necessary, probable and finally by use of primary auxiliary verbs 'have' and be' forms such as "have to' 'be able to'. A total of 88 modal forms were identified in relation to hate speech propositions and their occurrence and distribution by type across data sets was tallied and tabulated in table 5.2.

Table 5.2: Modal form in Facebook and Twitter by Category

Modality			Frequency of occurrence			
Major categories	**Modal subcategories**	**Examples**	**FB**	**Twr**	**total**	**%**
Epistemic modality	Validity	Be, certain, may	5	3	8	9.6
	Prediction	Will, would	10	5	15	18.1
Deontic modality	desirability	Should, advisable	11	6	17	20.5
	Obligation	Must, have to, should	18	4	22	21.5
	Permission	May, can	3	1	4	4.8
	Hedging statements	Arguably,	7	5	12	14.5
	Habitual verb tense	Belongs, hates	4	2	6	7.2

	Null modality	-	1	3	4	4.8
Total			54	29	88	100

Source: Researcher's analysis of Facebook group walls and Twitter hashtags, 2020

From table 1.2, it is evident that certain categories of modality manifested across data sets at a relatively higher frequency than others. Obligatory and desirability modalities which entailed the use of modal verb forms that give directives ordering or prohibiting people to do were the most frequently used each with the frequency of 20.5 percent. This is indicative of the persuasively manipulative nature of the online hate speech discourse in Kenya as illustrated in the texts below which show how Facebook and Twitter discourse participants linguistically coerced or cleverly persuaded the readers to act in specific prejudicial and discriminative ways.

Text F7

><..Lazima warudi kwao. Uhuru atupee kibarua ya kurudisha refugee na wachaina kwao ka ameshindwa!!(they must go back to their countries. Uhuru should give us the job of sending them back home if he is unable to)

Text F8

><... Acheni na wakongo ni mabro, wasudu??? Niwabaya na wachafu jo (let alone the Congolese, they are our brothers, Sudanese??? They are bad and dirty)

The participants in F7 and F8 above use the obligatory modality marker '*must*' and a subjectless imperative clause *'let alone the Congolese'* and a desirability modality marker *'ought to/should* to impose their prejudicial personal desires to hurt or cause to hurt the target by way of causing fear and anxiety of being deported on the identified object of the proposition, i.e. refugees (Congolese and Sudanese) as well the Chinese. By using obligatory modal verbs 'must and should', the writer explicitly attempts to impose his discriminative influence on the free will of the in-group readers which in effect assigns those who live in close proximity with the target out-group a duty to act discriminatively against them. According to van Dijk (1993), one of the most potent sources of collective social power in influencing perceptions of the contemporary society is having a privileged access to popular media that has a wide circulation, which is essentially what Facebook and twitter offer.

Text F 9

*><Tutawatupa kwa lake (*we shall throw them into the lake*)*

Text T10

><Luos will be the last tribe to rule this nation, never (negative Prediction)

Text F9 and T10 above employ modal verbs 'shall' and 'will' to achieve two apparently discriminative discursive functions: threatening and negative prediction. In text F9 the participant contributes to the conversation about the refugees and other foreigners living in Kenya and their influence on local politics. The self appointed representative of his in-group members (*we*) makes targeted threat (to throw foreigners in the lake). Similarly, text T10 contains a proposition that in the Kenyan multi-ethno-political context may be interpreted as provocative, demeaning and an outright discriminative prediction touching on the limitation of the rights of a particular ethnic group to produce successful presidential candidate material. Although the text above presents personal opinions of their contributors, they also serve to set excitable agenda for the online conversation which ends up drawing support from other like-minded participants to create ideologically skewed and highly polarised discussions that may negatively influence general attitudes and perceptions towards the target groups. The pragmatic force of such use of modality is explored by Palmer (1986) who relates the deontic modality to performatives when she points out that by uttering a deontic modal form the speaker/writer may actually grant permission, prohibit, promise or threaten as in the cases illustrated above.

Instances of null modality were also observed in declarative clause structures stating the author's personal and often skewed opinion as facts. Five percent (4.8%) of such declarative clauses omitted the modal markers of possibility or probability even when the propositions contained therein clearly entailed such grammatical moods since they were statements of personal evaluation of the truth value about the subject.

Text F11

<>This Mount Kenya region???are the core problem of this country from 1963!!!!! That they behave as if this country belongs to them alone yet we are 44 tribes always defending their stinking thiefs [thieves]

Text F12

<>Heheheeee...UR [you are] cursed...fools. you never seem to understand that jaluo will be the last community to rule this country...@ a free prophesy!! TAKE IT!! OR LEAVE IT!! Hata heshima hamna

Text F11 and F12 are two turns in a conversation in a pro NASA youth group. The two emotional participants use modality of certainty marked by primary auxiliary verb 'be' in simple present tense to represent particular negative predication and attribution associated with the objectified targets as relatively habitual and stable states that define their nature in no uncertain terms. This constitutes deficit descriptive discourse that seeks to justify various speech acts of discrimination, prejudice or violence against objects represented as such.

While the first representation depicts the mount Kenya region as the common historic problem that all other kenyan communities should unite to deal with, the second representation justifies the exclusion of the objectified target from country's leadership depicting them as jynxed, cursed and unworthy of such a high level resposibility as the presidency. This representation corroborates Kress and van Leeuwen (1996, 2006) and Herring (2007) argument that the process of assigning linguistic description to the individual or group targets is by no means a neutral activity but a well thought out discursive process that enhances particular narratives and attitudes meant to position the subjects in particular ways. It is common in Kenya that the community the president hails from is itself perceived to be in power thus making presidency a communal responsibility. When such clauses were used in negative description, their propositions were meant to be interpreted as the inherent attributes of the object of description, thus justifying the course of the author. Hannah (2002) cited in Posset (2017) delinks language as a system from violence by positing that if language is seen to be violent, it is not its inherent attribute. Rather, it depicts the inner disposition and engraved discriminative attitudes of the language user and the wider society.

6. Conclusions

Indeed, evidence points out that the hateful dispositions of the user's mind 'impose' on the online linguistic expression; hence it is possible to distinguish potentially hateful speech acts from the neutral ones by use of contextually nuanced linguistic criteria. Violent speech acts yielded potential pragmatic force in the service of aggressive ideology by leveraging the performative and action-like property of language whose utterance may be construed to actually perform an act of linguistic violence with potential to cause mental, emotional and or physical pain. Carefully selected violent micro speech acts expressed and performed repressed hostility online through implicitly or explicitly encoding targeted threats, making declaration of violence and directly appealing to virtual masses and their offline networks to discriminate the 'bad other' or support and amplify the existing offline discriminative order against those who are not 'us' ethno-politically as well as along other social identity lines such as gender, nationality and race. Aspects of performative deontic and epistemic modality served to re-affirm the discursive commitment of the authors to their violent declarations thus intensifying their pragmatic force in order to persuade and manipulate audience by imposing conditions of desirability, suitability, obligation and moral acceptability of prejudice and discrimination where threats and violent instructions were presented as necessary and obligatory information that need to be followed and acted on in the interest of the in-group members

7. Recommendations

Online hate speech is primarily a linguistic phenomenon that is observable, predictable and can be arrested at its earliest stages by offering counter narratives to minimize or forestall its adverse effect in the long run. Stakeholders such as NCIC, CAK and National Intelligence Service (NIS) can greatly benefit from the data and insights offered by this study in coming up with objective criteria for identifying early symptoms of hate speech to develop early sign-early response framework. Moreover, the findings speak to the social media users and policy makers as well as law enforcers to be mindful of how the seemingly harmless tokens of aggressive online content could yield valid interpretation as hate speech often with a build up effect on emotional states of online audience.

The findings further point out the futility of artificial intelligent-based approaches in monitoring, flagging out and classifying online hate speech in Kenya. Such software cannot account for various contextual factors and the use of highly symbolic local idioms, figurative language, and metaphors that collectively inform the audience's overall interpretations of online content as being hateful. So, this study suggests advocacy of more nuanced human criteria for hate speech detection to complement the automated techniques by considering the multiplicity of contextual factors and dynamic meaning resources at play in the production and interpretation of contemporary online hate speech.

Both the Kenya Evidence Act cap 80 of (2014) and NCIC Act (2008) do not clearly advocate for evidence from linguists' citing indeterminacy of the direction of motive, the actual originality and the susceptibility of electronic evidence to digital manipulation. Nevertheless, these evidential forms largely define the contemporary online content in light of the new media. Insight of this article could form the basis for amendment of existing hate speech laws to allow admission of linguists' evidence

and induction of hate speech law enforcers, especially NCIC investigators, prosecutors, lawyers and judges to equip them with basic linguistic skill since hate speech is largely a linguistic misdemeanour

8. Suggestions for Future Research

This research establishes a number of theoretical and methodological gaps which the future researchers ought to note in order to make internet research more effective and informative.

There is no consensus on how to distinguish hate speech from free expression of extreme sentiments. Whereas in legal definitions hate speech must result in actual occurrence of acts of violation in real life, some scholarly definitions focus on the potential inherent in an expression to injure the target emotionally or cause to harm in physical acts of discrimination or actual violence. Attempts to link online tokens of hate speech directly to specific acts of violence in the offline world remain largely speculative rather than empirical. Studying the cycle of events between hate speech stage to widespread negativity and finally, commissions or omissions that amount to the actual violence is yet to happen. The apparent vagueness of the term hate speech makes it necessary for the researchers to review the existing definitions before drawing the contextualised definition that applies in their studies.

Hate speech is one of numerous social identity issues arising in the digital environment that also hosts myriads of other repressive discursive activities that arouse intellectual curiosity in diverse perspectives especially regarding how this dynamic CMC context influences contemporary discourses. It would be worthwhile paying attention to other forms of representation such as social class relations among virtual membership and other discursive situations that accrue from this rich context.

This research notes a significant deviation from the traditional interactional norms guided by principles of cooperative, informative and politeness conversational approach aimed at ideally privileging the social and emotional needs and 'safety' of the addressee. The conversation appeared to shift towards a dehumanised, depersonalised and carefree online interaction in which users appear to mind less about social and moral expectation of them to be empathetic to their addressee. This raises the pertinent question of whether Facebook and Twitter create or facilitate anti-social attitudes and perceptions, or whether the manifest hate content in them is simply an innocent reflection of changing attitudes and perceptions of Kenyan users towards an ideologically rooted culture of hate and violence. This points out the possibility of other factors not explored in the current study and can provide entry point for new inquiry.

References

[1] Anat, B. N. and Matamoros, A. F. (2016) Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of the Extreme Right Political Parties in Spain. *International Journal of Communication, 10(2016) 1167-1193*

[2] Austin, J. L. (1962) *How to Do Things with Words*. Clarendon Press.

[3] Bardici, M. V. (2012) A Discourse Analysis of the Media Representation of Social Media for Social Change- the Case of Egyptian Revolution and Political Changes. [Unpublished MA Thesis], Malmo University.

[4] Brenner, S. W. (2007) At Light Speed: Attribution and Response to Cyber Crime/ Terrorism/ Warfare. *Journal of Criminal Law and Criminology*, 97 (2), 379- 476.

[5] Brink, D. O. (2001) Millian Principles, Freedom of expression and Hate speech, *Legal Theory 7*(2001), 119-157.

[6] CAK. (2017) Second Quarter Sector Statistics Report for the Year 2016/2017. http:// www.ca.go.ke.

[7] Cameron, D. (2001) *Working with Spoken Discourse*. Sage Publications Ltd.

[8] Christoforou, S. (2014) 'Social Media and Political Communication'. Un published [M.A Thesis], *Erasmus University*. Danet, B. (2013) Flaming and Linguistic Impoliteness on a Listserv. In C. S, Herring and D. S, Virtanen (Eds.) *Pragmatics of Computer Mediated Communication* (639-664). De Gruyter.

[9] Eltis, K. (2013) Hate Speech, Genocide and Revisiting the 'Market Place of Ideas' in the *Digital Age, Law Journal*, 43 (2), pp. 267- 328.

[10] Eronen, M. (2015) *Rhetoric of Self- expression in Online Celebrity Gossip*. University of Vaasa.

[11] Fairclough, N. (1995) *Media Discourse*. Oxford University Press.

[12] Fairclough, N. (2001) *Language and Power*. Longmann.

[13] Fowler, R. (1985) Power. In T. A. van Dijk (Ed) *Handbook of Discourse Analysis. Vol*. 4. Academic Press, Inc. Pp.61-82.

[14] Glowinski, B, D. (2015) Sociological Norming of Members' Concept of Violent Music. Unpublished [M.A. Thesis], Texas State University.

[15] Halliday, M. (1985) *Introduction to Functional Grammar*. Edward Arnold ltd.

[16] Herring, C. S. (1999a) Interactional Coherence in Computer Mediated Discourse. *Journal of Computer Mediated Communication*, 4(4). Retrieved 6/11/2016 at http:/jcmc.indiana.educ/vol4/issue4/.

[17] Herring, S.C. (2004) 'Computer Mediated Discourse Analysis: An Approach to Research in Online Behaviour'. In S.A. Barab, R. Kling and J. Gray (Eds.) *Designing for Virtual Communities in the Service of Learning*, pp. 338-376). Cambridge University Press.

[18] Herring, S. C. (2007) *A Faceted Classification Scheme for Computer Mediated Discourse*. Language@ Internet (4). http:/www.languageatinternet.org./articles 7/12/ 2018.

[19] Herring, S. C. (2011) *Grammar and Electronic Communication*. Http://ela.slis.indiana.edu.

[20] Herring, S. C. (2019) Grammar and Electronic Communication. In C. Chapelle (Ed.), the Concise Encyclopedia of Applied Linguistics. Wilie-Blakwell.

[21] Hodge, R. and Kress, G. (1979) *Critical Linguistics*. Polity Press.

[22] Irimba, F. (2014) Discursive Construction of Social Class in Education Placement Practice: A Case of the New Quota System Debate in Isiolo Township, Kenya, Unpublished [M.A. Thesis], Laikipia University.

[23] Kassia, E. (2015) Racism and Media: A Textual Analysis. Research Paper. St. Catherine University

[24] KNCHR (2006) 'Need for Hate Speech Legislation' KNCHR Report to the Parliament. http://www.knchr.org.

[25] KNCHR (2008) Annual Report for the Year 2008-2009 Financial Year. http://www.knchr.org.

[26] Kress, G., and van Leeuwen, T. (1996, 2006) *Reading Images: the Grammar of Visual Design.* Routledge.

[27] Levinson, S. (2006) Cognition at the Heart of Human Interaction. *Journal of Discourse Studies*, 8 (1) pp.85- 93.

[28] Lillian, D. L. (2008) Modality, Persuasion and Manipulation in Canadian Conservative Discourse. *Critical Approaches to Discourse Analysis across Disciplines*, 2 (10), pp.1- 16.

[29] Lotta, L., Johanna, I., Veronika, L., and Matti, L. (2016) 'Linguistic Analysis of Online Conflicts: A Case Study of Flaming in the Smokahontas Comment Threads on You Tube'. http://wiedescreen-fi/numerot/2016-1-2/.

[30] Mullen, B., and Leader, T. (2005) Linguistic Factors: Antillocutions, Ethnonyms, Ethnophaulisms and other Varieties of Hate Speech. In J. F, Dovidio, P, Glick and L. A, Rudman (Ed.) *On the Nature of Prejudice: Fifty years after Allport*, 192-207: Blackwell.

[31] Mwithi, F.M. (2016) Social Media Discourse: Linguistic Study of Facebook Among a Selected Group of Kenyan internet Users, Unpublished [PhD. Thesis], Laikipia University

[32] NCIC (2006) *National Cohesion and Integration* Act. Council for Law Reporting.

[33] NCIC (2008) *National Cohesion and Integration* Act. Council for Law Reporting.

[34] NCIC (2017) Police Training Manual: Enforcing the Law on Hate Speech. NCIC.

[35] Ndambuki, J. (2010) Discursive Representation of Women's Interests and Needs in Makueni District-Kenya. Unpublished [PhD. Thesis], University of Witwatersrand.

[36] Olsson, J. (2013) *Word Crime: Solving Crime through Forensic Linguistics*. http:/www.googlebooks.co.ke.

[37] Palmer, F. R. (1986) *Mood and Modality, Second Edition*. Cambridge University Press.

[38] Perry, B., and Olson, P (2009) Cyberhate: The Globalization of Hate. *Journal of Information and Technology Law*, 18 (2), 185- 199. http://doi:10.1080/13600830902814984.

[39] Posset, G. (2017) Can Hatred Speak? On the Linguistic Dimension of Hate Crime. *Linguistic Online* 82 (3) 17.

[40] Prentice, S., Taylor, P. J., Rayson, P., Hoskins, A., O'Loughlin, B. (2011). 'Anlysing the Semantic Content and Persuasive Composition of Extremist Media: A Case Study of Text Produced During the Gaza Conflict.' *Information System Frontiers,* 13, 61-73. Accessed on 23/2017, http://dx.doi.org/10.1007/s10796-010-9272-y.

[41] Searle, J. R. (1992b) *The Rediscovery of the Mind*. Mass MIT Press.

[42] Siegel, L. (2008) *Against the Machine: Being Human in the Age of Electronic Media*. Spiegel.

[43] Shuy, W. R. (2014) *The Language of Murder Cases: Intentionality, Predisposition and Voluntariness*. Oxford University Press.

[44] Strauss, A. L., and Corbin, J. M. (1998) *Basics of Qualitative Research- Techniques and Procedures for Developing Grounded Theory*. Sage.

[45] Terence, C. (2014) Being Im(polite): A forensic Linguistic Approach to Interpreting Hate Speech Case. *Language Matters*, 45 (3). Pp. 37-51.

[46] Thompson, J, B. (1997) *Ideology and Modern Culture*. Polity Press.

[47] Tuiyott, P. K. (2013) Comparative Analysis of Electoral Gender-Based Violence: A Case Study of Kenya and Uganda. Unpublished [MA Thesis], Nairobi University.

[48] van Dijk, T. (1993) *Elite Discourse and Racism*. CA, Sage.

[49] van Dijk, T. (1998) *Ideology: A Multidisciplinary Approach*. Sage.

[50] van Dijk, T. (2005) *Racism and Discourse in Spain and Latin America*. John Benjamins Pub.

[51] Wafula, W.G. (2016) Social Media Forensics for Hate Speech Opinion Mining. Unpublished [M. Sc. Thesis], University of Nairobi.

[52] Watt, D., Kelly, S., and Llamas, C. (2013) Analysis of Inference of Threats in Neutrally Worded Utterances of Familiar and Unfamiliar Language. *York Papers in Linguistics* 2 (3).

[53] Wodak, R. (1997) *Gender and Discourse*. Sage.

[54] Wodak, R., and van Leeuwen, T, J. (2015) 'Legitimizing Immigration Control: A Discourse Historical Analysis'. *Discourse Studies*, Vol 1 (83). http://doi.org.10.1177/1461445699001001005/

[55] Wodak, R. (2015) The Discursive Construction of Strangers: Analysing Discourses about Migrants and Migration from a Discourse-historical Perspective. Migration and Citizenship. *Newsletter of American Political Science*, p. 6-105.

[56] Yieke, F. (2002) Language and Discrimination*:* A Study of Gender and Discourse in Work places in Kenya, Unpublished [PhD. Thesis], University of Vienna.