



The Role of Language in Digital Marginalization: A Critical Discourse Analysis of Linguistic Exclusion in Online Platforms and Policies

Asad Khan¹, Adil Hussain²

¹ Research Scholar, Department of English, Hazara University Mansehra, Pakistan.

Email: mrasadkhan0112@gmail.com

² MS Scholar, Applied Linguistics, Higher School of Economics, Nizhny Novgorod Russia. Email: adilhh194@gmail.com

ARTICLE INFO

Article History:

Received: June 28, 2025

Revised: September 04, 2025

Accepted: September 05, 2025

Available Online: September 06, 2025

Keywords:

Cultural Identity

Digital Citizenship

Intertextuality

Linguistic Hierarchies

Multilingual Digital Norms

Qualitative Method

Funding:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Digital technologies have revolutionized the access to communications, information and governance but language is an essential axis of inequality in digital transition. The supremacy of English and other international languages in the online space has established a status in the sense that those who speak marginalized or minor languages tend to have difficulties in their access to participation. The present study examines how the language used in the context of digital marginalization is examined with a Critical Discourse Analysis (CDA) of such policies, user interfaces, and communication on large Internet platforms such as Meta, YouTube, and Twitter. The theoretical focus involves finding answers to the question of how language policies and interface designs help exclude the speakers of minority and indigenous languages in their full access to the digital world. Based on three research objectives, the research study critically examines linguistic patterns, visual semiotics, framing as well as ideological constructions found in digital governance texts. Following a qualitative research design, the research adopts textual analysis, discourse coding, intertextuality, and contextual reading to unlock the prioritization of dominant global languages and vague modality, depersonalized pronouns as expressions of linguistic hierarchies. The findings shows linguistic exclusions are not accidental as they are ideologically embedded and thus define user access, user engagement, and cultural identity. Future researchers can further enhance the demand of a legally binding multilingual digital norms and user-centred designing policy and linguistic right to operate the basics of equitable digital citizenship.

© 2025 The Authors, Published by IRASD. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License

Corresponding Author's Email: mrasadkhan0112@gmail.com

1. Introduction

The introduction explains why language is a key factor of influence to the access and experience in digital spaces and therefore there is a necessity to explore linguistic exclusion as a variant of digital marginalization. It puts the study into the wider perspective of social justice and digital rights. Emphasis is laid on the way in which dominant languages are being promoted over minor ones in the Internet. This part constructs the argument of the perception of language as a barrier as well as a tool of digital involvement.

1.1. Background of the Study

Digital technologies have revolutionized the access to communications, information and governance but language is an essential axis of inequality in digital transition. The supremacy of English and other international languages in the online space has established a status quo in the sense that those who speak marginalized or minor languages tend to have difficulties in their access to participation. Warschauer (2003) points out that hardware access will soon not be the biggest issue about digital divide as the element of symbolic resources will matter, such as language. In the same way, Deumert (2014) observes that linguistic hierarchies present in

the real world are replicated online through digital communication, which oppresses the voice of minorities. Such exclusion plays out in web design, in content delivery along with algorithmic suggestions that favor the use of hegemonic languages, a tendency that has been also highlighted by Blommaert (2010), who addresses the politics of voice in discourses of the Internet. Environments such as Google, Facebook, and Twitter are spaces available on line, which are international in character, but which seldom accommodate a linguistic diversity through meaningfulness. The user interface and the automated content moderation systems have a language option that favors the dominance languages without considering the facts about multicultural societies. According to the position of Bender and Friedman (2018), the assumptions underlying the NLP technologies favor normalized forms of language and exclude speakers of non-dominant dialects. This argument is factored in by Van Dijk (2005), who recommends that unequal meaning on the Internet is structure exclusion. Further, Canagarajah (2006) states that today where the digital world seems to strengthen the role of English as the gatekeeping tool, particularly, in postcolonial settings in which the local languages are frequently stigmatised or non-existent in digital infrastructure.

The issue of linguistic exclusion cannot be described as purely technical but highly political and ideological. The policies which guide the way of developing digital content, translation services, and governance on the platform hardly contain any promises of linguistic justice. Tollefson and Tsui (2017) say that language policies are parenting, control mechanisms laid down with respect to greater power relationships, especially in multilingual communities. Through Critical Discourse Analysis, Fairclough (2013) explains the way language constructs but at the same time reflects social inequality. Besides, Shohamy (2006) affirms the oppression of language rights in covert manipulation of language in policy and planning. These two authors emphasize that digital marginalization is a continuation of the historical linguistic imperialism when the digital space becomes one more realm of exclusion. This is not a neutral process since the choice of languages to be given or omitted depends on ideological positions (deep-rooted ideologies) that favor the powerful cultures. Through the digital hierarchy, therefore, we have a contemporary place of linguistic gatekeeping, which continues to deepen the inequalities rooted in the colonial and post-colonial states of formation. Language in this regard acts as a metaphorical battle field whereby power is retained by silencing the voices of minorities. By limiting themselves to one or few languages to publish information, platforms are enforcing systematic exclusion since the right to expression is given to only specific ways of expression. Consequently, the digital space, being affected by linguistic exclusion, contributes to the larger tendencies of social and political marginalization.

Structural inequality has been reproduced in online language use, and Critical Discourse Analysis (CDA) is a fruitful framework to study this process. Under the CDA, scholars can examine how language decisions in terms of platform guidelines and content are used to support the interests of the dominant groups and make others unseen. Wodak (2004) believe that CDA can be very instrumental in ensuring that instances of power relations which can be obscured in institutional and public discourses are revealed. To this Fairclough (2013) further adds that discourse is not only something that reflects reality but functions as a constitutive force that reflects social structures. Further, according to Blommaert (2005), the language in digital texts which may include terms of service or algorithmic instructions are presented as a type of gatekeeper of linguistic identity, either included or excluded. Such a masking of inequality as the expression of neutrality is a factor that leads to a normalized view of exclusionary processes in algorithmic decision-making and platform moderation. CDA makes possible the interrogative questioning of how digital design and governance codifies ideologies that secure hegemonic forms. Consider, e.g., the moderation policies that target some kind of vernacular or dialect in the name of inappropriate language policy that demonstrates the bias of even objective systems. These illustrations prove that discourse cannot exist outside the context of the sociopolitical processes; it needs to be critically deconstructed so that the power asymmetries could become revealed. Consequently, CDA not only enlightens such inequalities but is a means of combating and redefining them, as well.

The Consequences of such exclusion are significant to the users of the underrated linguistic backgrounds. Digital accessibility is associated with the capacities of running operations in a dominant language which influence civic participation, the access to a range of public services, and the opportunities in the surrounding world conversations. According to Piller and Takahashi (2011), digital inclusion is just a myth without linguistic accessibility.

Pennycook (2010) goes further to say that linguistic justice is any treatise on equality in the information age. In the same vein, Skutnabb-Kangas and Phillipson (1995) emphasize that linguistic human rights are to be safeguarded online to avoid disappearance of linguistic diversity. The digital divide is contributed to by exclusionary language, thus, making speakers of minoritized languages an invisible entity in the online civic life. Consequently, they can hardly take part in the e-governance, online education, and medical services as well as rely on other people representing the dominant languages. This does not only support dependency but also derives communities of their power to speak linguistically in their digital communication. Thereby, the digital world turns out to be the environment of cultural oblivion to which specific identities are condemned to be voiceless. A situation is especially severe in the multilingual countries like India, Nigeria, and Pakistan. Health apps and educational and government websites use only official or colonial languages, leaving behind the major part of the population. According to UNESCO (2016), the availability of online content in African languages is situated below 5 percent, although Africa uses a diversity of languages. Another similarity between Tufte and Hemer is that they point out the restricted linguistic scope of development communication on the Internet (2005). The imbalance provincializes whole communities and naturalizes pre-existing socioeconomic differences by gatekeeping language. Digital disenfranchisement is especially destructive in nations with enormous linguistic variety that is limitedly represented in formal or internet infrastructure. Major languages like Pashto, Sindhi, Balochi, which are also predominant in Pakistan are not represented in government portals at all in Pakistan. This brings out cases of digital illiteracy, not because people have no connections to the internet, but because of the linguistic inaccessibility. The giving of more privileged position to the colonial or official languages also means downgrading of indigenous languages practice and self-complexes of cultural inferiority. Such a digital exclusion is not only restrictive but also demeaning linguistic culture as it will be left in the backstage of technological innovation (Dhivya et al., 2023; Wided & Alfalih, 2023).

The argument that businesses should satisfy dominant languages as an efficiency or global impacts issue is common among companies and digital policy designers. Nevertheless, these explanations fail to address inclusive social responsibility and equal access. Due to the exclusionary practices of better-off AI and NLP developers, Bender et al. (2021) state that these developers do not consider any language variation and sociolinguistic diversity. Shohamy (2006) laments such instrumental approaches to language and she insists that language planning should be carried out under a rights-based perspective. Technocratic language approaches that fail to consider history and culture, two very important elements in the digital world, are also discouraged by Tollefson and Tsui (2017). The privilege ranking of the major languages at the pretext of scalability or efficiency induces structural hindrance to the linguistic minorities. These reasons are used to justify the status quo at the expense of ethical imperative to enhance linguistic diversity and equity. Besides, language is a carrier of identity, memory and worldview as well as means of communication which can not be quantified down to the rules of the market. Bender et al. (2021) hold the idea of language-inclusive datasets and warn against the development of digital tools that incorporate only linguistic standards of the elite. When language is marginalized to a functional commodity, the questioning of digital inclusion in a wider cultural and ethical sense is completely imaginary. Exclusive language practice on the Internet is indicative of the larger body of linguistic imperialism and neoliberal regulation. The propagation of English as a global language is, in fact, much related to economic and technological dominance as Pennycook (2006) asserts. According to May (2013), true multilingualism in the digital era cannot be implemented with tokenism representation, but there should be a structural transformation. Wodak (2004) reaches the conclusion that the digital marginalization is a facility of systemic discrimination that will not disappear until institutions, governments, and digital platforms recognize language as a source of power. Such imposition of English as a de facto one on the internet carries on the tradition of colonial linguistic hegemony in a neoliberal form. It also deepens a mono-culture in the digital area in which fluency with English is a prerequisite to access, credibility, and visibility. May (2013) cautions not to act deliberately to change digital language policy; otherwise the unity will continue excluding platforms that favor the already empowered groups. To realize both language as a right and language as a Resource, the belief of linguistic neutrality in the digital governance should be confronted by the institutions (Ahmed, Azhar, & Mohammad, 2024; Mohammad & Ahmed, 2017).

1.2. Problem Statement

This growing digitalization of our society has illuminated an ongoing challenge of linguistic exclusion on the Internet, which is that many platforms prioritize certain languages, especially English, in their platforms based on interface, algorithms, and policies. Consequently, linguistic minorities and indigenous language users take up the lower end of marginality and, thus, have restricted access to digital information, services, and engagement in internet-based discourse. Language has been a clear and yet little-discussed obstacle that supports the dynamics of power imbalance and erasure of cultures brought up in the discourse of digital inclusion. Despite the fact that in previous literature (Bender & Friedman, 2018; Canagarajah, 2006; Tollefson & Tsui, 2017), an emphasis on the importance of equal language policy has been stressed, little focus has been dedicated to the idea of sourcing digital marginalization in discourse as embedded in platform design and governance.

1.3. Research Objectives

1. To develop a critical analysis of how language policies and interface designs of the platforms lead to the digital marginalization of minority and indigenous language speakers.
2. To examine the expression and appearance of various languages in the discussion of online platforms and their official policy on communication with the people with the help of Critical Discourse Analysis.
3. To discuss the consequences of linguistic exclusion in the online environments on user engagement, accessibility to services and cultural identification maintenance.

1.4. Research Questions

1. What are the roles of platform language policies and interface designs towards digital marginalization of minority and indigenous language users?
2. How the various languages are incorporated or omitted in the speech of online sites and their communication strategies?
3. How does linguistic exclusion in an online environment impact the participation of the user, access to services and cultural identity?

1.5. Significance of the Research

The study offer valuable insights about the phenomenon of language and digital inequality as it will be possible to find out how linguistic exclusion works in the system of online platforms and policies. In critical discourse perspective, the study point out the unseen manner in which language decisions underpin social practices, thereby defining who may view, take part, and be listened to in online domains. In addition, the study serve to end the discussion of digital equity, multilingualism, and linguistic justice by providing empirical data and analytical approaches in telling the contribution of language to the formation of digital participation and identity.

2. Literature Review

2.1. Overview

The literature review examines the most important works on digital inequality, linguistic discrimination, and critical discourse analysis and forms a basis of knowledge of how the language can be an obstacle in the online spheres. It employs theories of language/identities/powers intersection. Warschauer, Fairclough, and Blommaert are the authors who give the essential information about digital language hierarchies. In the review, the gaps in the existing research are identified, in particular, the linguistic exclusion in the platform policies. The concept of digital inequality is very complex and cannot be used to cover aspects of hardware and spatial access, but rather all parts of access such as access to language and symbolic capital. Warschauer (2003) suggests that the digital inclusion does not only entail a technical infrastructure, but it also has more to do with linguistic and cultural accessibility. Van Dijk (2005) backs this up by stressing the fact that inequality of information magnifies itself when the digital content is inaccessible to all users with regards to language. Likewise, Hargittai (2018) proves the strong linkage between digital literacy and comprehension of languages, and that a speaker of a non-dominant language usually suffers in having access to online environment. It is this dynamism of digital inequality, which brings to light that symbolic access, e.g. access to dominating discourses and linguistic representation, is the doorman to complete digital participation (Van Dijk, 2005; Warschauer, 2003). Thus, it can be concluded that digital marginalization involves more than just a material obstacle; it is

systematically grounded in semiotic stratifications, which exclude the non-hegemonic linguistic areas (Hargittai, 2018). Through these omissions, new layers of exclusion are constructed, and due to the repeated nature of the process of exclusion, language is a component of this hierarchical structure that is reflected and reinforced by it.

Online space usually reproduces and emphasises the hierarchies of languages globally where English and other dominant languages dominate. Pennycook (2006) demonstrates how English can be traced as a complete component of the communication technologies all over the world, and the people who do not speak English are marginalized accordingly. The power of English in the digital spaces of education and learning has also been denounced by Canagarajah (2006), as it relegates any local knowledge and expression. Crystal (2011) continues to say that as much as internet offers some linguistic diversity, in reality, the internet reinforces linguistic centralism to the few major languages. As Pennycook (2006) stresses, the linguistic imperialism of English in the digital communication does not only replicate the relations of domination and subordination, but also plays an active role in determining the valid and accessible knowledge. This means that linguistic resources do not act in digital spaces neutrally, instead serving as an instrument of homogenization across the world. Canagarajah (2006) goes further to add that the digital privilege given to the dominant languages also discredits local and indigenous epistemologies that are founded on local and indigenous cultures. Platform designs often fail to meet the demands of diversity, encompassing multilingual requirements, which causes implicit linguistic discrimination. According to Shohamy (2006), language policy is the form of control, and in most cases, digital space exerts monolingual ideologies. Bender and Friedman (2018) demonstrate that language models and AI-based platforms are discriminating towards homogeneous varieties of the powerful languages. According to Tollefson and Tsui (2017), the loss of minority languages in digital settings is one of the ways of language oppression of institutions, which is not even recognized or told. Crystal (2011) observes that the internet has the capacity of encouraging multilingualism, but the data algorithmic and structural modelling of sites is more inclined towards encouraging dominant tongues, in which case there is a false appearance of diversity. Referring to the notion of digital monolingualism, Shohamy (2006) postulates that in many cases this phenomenon is masked under the tag of neutrality, though its domineering nature is indeed a top-down control which shuts out other linguistic behaviors. This makes most of the users invisible in the discourse communities of the digital world and further patterns a chain of language disenfranchisement (Bender & Friedman, 2018).

The Critical Discourse Analysis (CDA) remains a useful methodological method of exposing power relationships inherent in the use of language in the internet. According to Fairclough (2013), discourse structures continue to dominate the social structures by repeating ideologies. Wodak (2004) also stress that CDA reveals institutional regime preserving the linguistic inequality. Blommaert (2005) supplements this by stating that the discourse of digital texts is a manifestation of the wider asymmetries in the society particularly vis-à-vis the language representation at the discourse level and access. According to Tollefson and Tsui (2017), such an invisibility is not accidental, but it is institutionalized and realized through the routine process of normalizing linguistic homogeneity in online content. Fairclough (2013) demonstrates that the language pattern used in online texts give preference to dominant ideology by supporting which speaking and subjectivities are justified. Institutional power implied in a discourse therefore expresses to push down other linguistic identities (Blommaert, 2005; Wodak, 2004). The internet is a recreation of the colonial ideology of language stratifications in which the local and indigenous languages are systematically locked out. The use of the dominant language in the way the world communicates first led to the concept of linguistic imperialism that was coined by Phillipson (1992). The studies by Rickford (2016) state that online language policy can usually isolate non-standard and racialized dialects. To this Pennycook (2006) add that the global digital flows recreate the colonial power structure by adopting language choice in media and platforms. Such language stratification resembles the colonialism systems, in which the standard language was colonized by the use of Western and elite languages, which are currently reproduced digitally (Phillipson, 1992). According to Rickford (2016) the raciolinguistic ideologies remain to oppress non-standard dialect speakers and usually depict them as unfortunate linguistically weak. In online space this has led to algorithmic discrimination in which language has been the tool of proxy exclusion of racial and cultural groups (Pennycook, 2006).

According to Skutnabb-Kangas and Phillipson (1995), the concept of linguistic human rights must form part of the digital development agendas, particularly because the internet has become an intermediary to complete civic engagement. According to Piller and Takahashi (2011), it is reasonable to state that digital unevenness brought about by linguistic isolation not only transforms into technical divide but also results to civic and democratic white spot. May (2013) appends that the ideal of digital inclusion would only be achieved when linguistic diversity would no longer be viewed as an appendage to public digital infrastructure but rather on its core. Language is not only a means to communicate but it is a right which guarantees digital citizenship. The situation in the online environment forces (Skutnabb-Kangas & Phillipson, 1995) to note the necessity to integrate linguistic human rights in order to guarantee equal participation. According to Piller and Takahashi (2011), inability to access linguistically the online sphere hinders the possibilities of participating in the overall online democracy. May (2013) goes further to argue that inclusive digital societies particularly those where digital tools are involved in provision of public services require multilingualism. Blodgett et al. (2020) emphasize that most AI systems do not generalize beyond linguistic norms, which results in the systematic exclusion of the minoritized dialect in natural language processing. As Birhane et al. (2022) state, this shows a more profound problem because technological design does not take a reflexive standpoint and tends to construct social hierarchies in an unconscious manner. As Bender et al. (2021) remark, rather than reducing bias, the existing language models could increase it because they accept assumptions that conform to linguistic power hierarchies. The ability of the AI to relate to non-standard and minority dialects is not sufficient, especially in language models. Blodgett et al. (2020) prove that NLP models are systematically unable to do well with underrepresented dialects, leaving them excluded digitally. Birhane et al. (2022) believe that ethical models of AI do not pay enough attention to linguistic variability and focus more on the efficiency of technology rather than inclusion. Bender et al. (2021) stressed that Big language models carry out societal biasness by intensifying language ideologies of power at the expense of linguistic diversity.

According to Gillespie (2019), algorithmic moderation is developed on the top of existing language hierarchies and thus, it is inapplicable to decipher or maintain the individuality of marginalized speech. In his work, Roberts (2019) shows that such systems do not work in non-English environments, where content can be incorrectly labeled or deleted because of its unfamiliarity to the language. (Noble, 2018) criticizes market reasoning that supports search and classification systems to show how the latter is aggravating inequalities by obscurity with black box-relied logic that punishes linguistic diversity. The tendency to exclude minority language and dialects in social media platforms becomes common as a result of content moderation practice. According to Gillespie (2019), the dominant language assumptions in the creation of algorithmic filtering do not take into account the different linguistic expressions. According to Roberts (2019), there is no transparency in content moderation especially in the non-English settings where content removal decisions are outsourced or automated. Noble (2018) demonstrates that the practice of indexing and classification of language in search engines and other digital devices propagates structural inequalities. Although the linguistic rights are promoted, the policy gap in multilingual internet governance is still high. According to UNESCO (2016), there is also the lack of representation of indigenous languages online, and it requires welcoming language policies. Hemer and Tufte (2005) maintain that, development communication is still hegemonic in nature, it does not give local voices much airplay. In her article titled Language policy in digital governance, Busch (2012) stresses that language policy in digital governance cannot afford to be passive in terms of pursuing linguistic diversity but should instead take the initiative to agree on the neutrality of English language. UNESCO (2016) appeals to implement practical actions to conserve and develop native languages in the digital reality since the existing policies are not supportive enough to promote effective integration. According to Hemer and Tufte (2005) the historical popularity of Northern paradigm in development communication can be seen as a mirror of the overall postcolonial power structure. According to Busch (2012) digital language policy should be shifted or rather pushed to become active promotion rather than the complacent and passive tolerance so that we may be able to unlock and define equitable futures with no nasty digital surprises. In such countries as India, Nigeria, and Pakistan the digital divide is legitimized by the use of languages. Kothari (2003) reveal that there is a lack of linguistic diversity, in the Global South given by the digital infrastructures that are propagated by the international development agencies. According to Omoniyi and White (2006), the local language are not usually included in the design of ICT and thus the majority of the population in the countryside have no access

to the digital services. Banda (2009) considers that the digital language marginalization is a factor that perpetuates larger shifts in socio-political exclusion in multilingual postcolonial societies.

Language and identity on the internet have been the object of more and more scholars, who cite the linguistic marginalization of online cultural and individual identities as a relevant point of concern. According to Bucholtz and Hall (2005), language is one of the main modes of identity construction and because when diverse languages cannot be accommodated in the digital platform, it results in disconnection of many people and their ability to express real selves. Androutsopoulos (2006) answers the question of online communities and demonstrates that the language used in the digital interactions largely defines the participation and belonging to certain groups of people. Lee and Chau (2018) goes further to state how the ideologies of language on platforms influence the behavior of the user and the language they use, and how people align themselves or run away with certain online communities. According to Kothari (2003), structural designs in most digital initiatives by the global development actors have conformed to Western linguistic assumptions making the locals invisible through their expressions. Instead, Omoniyi and White (2006) note that such lack is not coincidental because the understanding of local knowledge systems when it pertains to technological planning was long overdue. According to Banda (2009), this kind of marginalization furthers the social cleavages, language is a symptom and a mode of extended reclusive behaviors within the postcolonial societies. This has led to impoverishment of indigenous languages on an online platform which is now a major issue among linguists and those who advocate digital rights. Grenoble and Whaley (2005) mention the threat caused by the lack of indigenous language support in the digital world, and the lack of such support only leads to endangered languages. Davis (2010) reviews the lack of the aboriginal representation in the web conversation and emphasizes that saving the language would also involve internet technology and interface. Perley (2012) criticizes the tokenistic display of language coverage on online media and presents a demand to rethink the representation of indigenous languages in the digital media and use the rights-based, community-centered perspective of the language integration process.

According to Grenoble and Whaley (2005), the demise of indigenous languages is being fast-tracked by digital neglect since digital presence has become the critical factor that defines the language vitality. According to Davis (2010), incorporation of indigenous language in to web infrastructure is not only a technical need but a cultural necessity. According to Perley (2012), shallow representations of indigenous languages in vogue should be changed with more profound, community-driven modes of encounters that lay importance on the cultural genuineness and agency. Digital learning currently lacks multilingualism because majority of the e-learning environment and online curriculums focus on giving prominence to standardized global languages. Hornberger (2009) emphasizes the proficiency of the additive multilingualism in schools but digital education tools tend to support subtractive practices that marginalize mother tongues. Benson (2013) indicates that local language is not mostly incorporated in the e-learning systems thus affecting both learner engagement and their understanding negatively. In the case of bilingual and multilingual situations, Cummins (2000) says that language inclusivity in online pedagogy is important to boost academic excellence and culture affirmation.

2.2. Research Gap

An increased understanding of the place of language in digital inequality, especially concerning how dominant languages are favored online and in AI-based tools, appears in the literature. Nevertheless, one would notice that a big gap hedges in analyzing Critical Discourse Analysis (CDA) use on issues of ideological programs and policy discourse of controlling linguistic exclusion in online settings. Although the research on digital colonialism, algorithmic bias and language rights has been conducted in the past, there are limited studies that have taken a systematic examination of how platforms have engaged in communicative practices and governance of technologies in the manner that is CDA shaped. The proposed research will address the gap by examining the construction of exclusion and marginality through language used in digital platforms and policy documents, thus, offering a better perception of the opportunity to cognize language as a tool of digital discrimination.

3. Research Methodology

3.1. Overview

In this paper, using a qualitative research method, Critical Discourse Analysis (CDA), is used to discuss the role of language policy and platform design in digital marginalization. CDA is used to explore the concealed ideology that lies in the digital texts and interfaces. Data gathered through online platforms, official language policies and user-generated contents.

3.2. Research Method

The research methodology used in this study is qualitative method, the reason being that it is the most appropriate to investigate the use of language as a maintenance or resistance tool of power in the digital world. Qualitative methods aim at finding meanings, ideologies, and patterns in the block of text/discourse or discursive material instead of quantifying variables. As this study investigate language policies, interface discourses and governance texts related to online platforms using the Critical Discourse Analysis (CDA) perspective, a qualitative research method can be used due to the possibility of reading the text and its interpretation critically and thoroughly. Creswell and Poth (2016) recommends qualitative research when examining the study of complex social phenomenon that is internalized in language and culture. In the same manner, Denzin and Lincoln (2011) claim that clarity of symbolic interactions and ideologies can be achieved through the means of qualitative methods proving the notion that qualitative methods work best when analyzing the digital marginalization typified by a language.

3.3. Theoretical Framework

Critical Discourse Analysis (CDA) be used as the theoretical base of this research due to the works of Fairclough (2013) especially. The world view of CDA represents a social practice with language and concerns the way discourse reveals and reinforces social power, inequality, and ideologies. The Linguistic Human Rights Theory developed by Skutnabb-Kangas and Phillipson (1995) is also considered the part of the conceptual framework, and it help to explore language as a human right within the online environment. The implementation structure is the use of CDA to analyze a chosen corpus of the policies and interfaces of digital platforms and official correspondence with a view towards exposing covert power structures and ideology that discriminates against non-dominant-language speakers. The key notions of the study that is operationalized in the course of the investigation of texts and documents is the digital marginalization, linguistic exclusion, multilingual access, and platform governance.

3.4. Research Population

This investigation rely on the official discourse materials created by digital platforms and ruling institutions as the research population, including terms of service, language policy, the language menu in the interface, the user guidance, etc. One can find such materials created by international tech companies, national agencies of digital governance, the stakeholders involved with the creation of language access in the digital space. This conforms to the qualitative approach that aims at interpreting and analyzing use of language in the field.

3.5. Sample and Sampling Technique

A representative sample of digital platforms and policy texts is chosen on the basis of a purposive sampling technique in order to conduct discourse analysis. These involve large international sites like Facebook, Twitter, YouTube and Google among others and national e-governance sites in multilingual nations (e.g. India, South Africa, Pakistan). Such arenas are chosen according to their ability to appeal to worldwide audiences and the capacity to shape the linguistic representation on the online platform. As Patton (2002) indicates, purposeful sampling is suitable in qualitative research in which cases with the rich information are chosen to be analyzed in further detail.

3.6. Data Sources

Policies on digital platform, terms of service, documentation in help centers, interface language settings, language accessibility policies, and governmental guidelines on communication shall be listed as the main sources of data. These written materials is gained through internet official websites. The interpretation of the context rely on secondary data that includes previous academic studies of platform voice (e.g. Gillespie, Dunbar-Hester) and cross-nation data on digital inclusion (e.g. UNESCO, ITU).

3.7. Procedure of collecting Data

Data retrieval consist of systematic retrieval of the documents and texts on the chosen digital environments and official digital platforms of administration. This include the downloading of terms and conditions, language policy section, user interface documentation and other areas of texts. The process of search is done based on certain preset criteria concerning the representation and inclusion of language. Archival of documents would be made and thematically organized according to the method of coding. The information gathered in favorable, non-threatening as well as publicly visible ways; ways that are in line with ethical principles that apply to document-based qualitative studies (Bowen, 2009).

3.8. Theoretical Framework

Critical Discourse Analysis (CDA) or its formulation by Norman Fairclough is the main theoretical background of this research. CDA considers language as a social practice and as such, it is a product of social structures of power as well as an agency, which impacts the structures of social power (Fairclough, 2013). It does not just practice the use of language in isolation but also how it constructs social realities and legitimises institutional practices and maintains domination of ideologies. On a digital platform, CDA can help the researcher to unearth the ways in which language policies, interface guidelines and the user guide act as a representation of where power resides between the privileged and the non-privileged lingual groups. According to Fairclough (2003), discourse exists on three planes, which include: text, discourse practice and social practice, which enables an explanatory analysis of what has been said, how and why it has been said in addition to a detailed coverage of the same. Wodak (2004) also state that CDA best fits the analysis of institutional texts because such arrangements of power are frequently hidden behind several seemingly neutral or technical terms. To this end, CDA can offer a very powerful perspective through which the strategic application of language by online governmental control regimes and corporate entities can be criticised on the grounds of promoting linguistic exclusivity and digital marginalisation.

4. Analysis and discussion

4.1. Overview

The analysis shows the trends in the exclusionary language and power among online platforms, indicating how specific linguistic practices drive inequality and restrict the online exposure of minority groups. It brings out how linguistic diversity is oppressed by English-dominant algorithms and the pattern of normalcy. The They are provided through illustrations on platform terms of service, content moderation rules, and user language interface settings. The discussion links the linguistic marginalization to other concerns of representation, equity and access.

4.2. Textual Selection

The first step of Critical Discourse Analysis (CDA) is called as textual selection, in which mostly desirable and representative texts were sought in relation to the discourses of linguistic exclusion in cyberspaces. The most relevant policy documents belonging to such platforms as Facebook, Twitter (now X), and Youtube were chosen in this research, as well as official interface guidelines and community standards. These documents were downloaded based on publicly available areas of the Help Center and Transparency Report collections of respective platforms. As an example, the text about the rules and policies in Twitter (2023) highlights the importance of following them but provides rules on how to use the interface in only mainstream languages that are also used worldwide, i.e., ones like English, Spanish, and Arabic, and only a little consideration is given to other languages, particularly the native ones. In the same way, policy documents of YouTube despite being provided in more than 40 languages leave out significant number of minority dialects. Discourse does not just reproduce power according to Van Dijk (2006) rather, it generates discourse through power. The platform texts act as power tools, as they make some languages visible and others invisible. Their marginalization turns into the process of marginalization when the access and participation of minority language speakers become restricted. In this way, textual selection targets at discourse which is an expression of system incorporation of the dominant at the expense of the local and minoritized voices. Moreover, the language access policies of such platforms tend to act as a projection of the corporate strategies rather than an intention to linguistic equity. You can see this, e.g., in the Community Standards published by Facebook (Meta, 2023), which is available in various

languages, whereas most technical and safety features are almost never localized outside of the most high-traffic languages. These texts were not only chosen due to their availability factor, but since they embody what (Fairclough, 2003) referred to as language as a form of social practice. In short, this impairs the notion that languages on platforms are neutral, they are filtered through ideological, economic and cultural identity. As indicated by Bourdieu (1991), language is a power vehicle through which people invoke their interests.

4.3. Contextual Reading

To generate insights on the ideologies behind the linguistic exclusion on online platforms, it is important to grasp the underlying socio-politics of platform language policies. The majority of governance documents on platforms are built on a Western-centric neoliberal model where more emphasis is put on global market coverage than on language-cultural integration. As an example, the Community Guidelines of YouTube (2023) are based on the compliance with the norms of the U.S. law and political-cultural values is the highest priority, and in most cases, they do not consider regional traditions of communication between people or regional peculiarities of the language used in countries of Asia, Africa, or South America, which have a rather low level of political and cultural representation. These writings usually arise in situations where the importance of linguistic diversity will be lowered to the aim of homogenized, expandable platform management. According to Pennycook (2010), contexts of linguistic practices are always in the greater social, political, and historical forces. Reducing the multilingual support is not a neutral action, but it is a form of supporting the digital inequity that is reinforced within the wider political-economic context. Moreover, historical elimination of the native or indigenous language by the colonial rule is now playing out in digital space as a result of linguistic homogenization. A lot of postcolonial cultures as well include downloadable digital administrative papers, which are color-coded linguistically on a colonial scale, say Facebook India policies, which are all written in Hindi and English instead of the Dravidian or tribal language, such as Tamil, Telugu or Gondi. This trend repeats what Ngungi wa Thiong (1986) termed as linguistic imperialism because the prevailing languages will be used as means of suppression and culture destruction. The production of Facebook online is not highly supportive in African languages like Swahili or Yoruba although they had many speakers. According to Fairclough (2013), discourse context contains institutional and societal structures which are informed and form the language used. With this, through the contextual reading, we are able to judge harshly on how the socio-political forces give new meaning to the design and implementation of language policies on the internet- maintaining the inequalities rather than breaking it.

4.4. Discourse Coding

In the exercise of discourse coding, both open and axial coding methods were used in isolating the repetitive linguistic patterns and ideological frameworks ingrained in policy statements and policy platforms. As an example, the term user safety, phrase maintain integrity, or variations of preventing harm was used many times on the Twitter page called the Enforcement Philosophy (2023). Nevertheless, these have never been translated or put into the context of indigenous people, implying an implicit target audience of people native speakers of the mainstream languages of the world. These repetitive wordings were coded as a reflection of the universalist framing that obscures language exclusion through the alleged single-size-fits all model of communication. Titscher et al. (2000) indicate that discussion coding in CDA aids to determine the institution of discursive structures that serve as a facade of ideologies. As a result of open coding, it was evident that the sites constantly use ambiguous and depersonalized terminologies like the terms users, community and people and avoid identifying a particular lingual or cultural group. Such neutralizing language though has hidden asymmetrical power relations and it systematically marginalizes non dominant users. The surface-level expressions could then be connected to the ideological commitment on a deeper level through axial coding. In the case of Transparency Reports (2022) by Meta, such key terms as harmful content, misinformation, and trust and safety circulated and were linked to such peripheral ones as language moderation or localization. But such terms were hardly used in regional or linguistic contexts. The inference is that the majority cultural-linguistic norms guide on what should be identified as either harmful or trustworthy, marginalizing the voices of the minority. The themes are coded to the extent that their communicative utility operates as means of symbolic reproduction, as Van Dijk (2005) puts it, these themes are regulated by discourse produced by elites in a sense that they are used as the means of power. With the help of discourse coding, it was determined in this research how abstract language of policy

may be effectively used to exclude people speaking linguistic minorities and construct a normative environment in the digital world around dominant languages and Western epistemologies. Not only this coding process reveals how the visibility of languages is selective but also highlights the ideological machineries operating within content governance systems.

4.5. Intertextual Analysis

Intertextual analysis assists in clarifying how ideologies are embedded, echoed, and replicated on the online platforms in various text and papers and a coherent discursive cross which perpetuates lingual exclusion. Where community standards are concerned, in the case of Meta, language uses are reflective of similar standards of global regulation to that of online life on platforms as articulated by the OECD (2020) in the principles of artificial intelligence and digital cooperation roadmap (UN, 2020) that ensure globalization of a wide scope and neutrality of platforms. Nevertheless, linguistic diversity is not a priority in these international documents, and all these texts are silent on this topic, and Meta takes this silence with her. It is an ideological borrowing that corresponds to the idea of Fairclough (2003) who suggests the concept of intertextual chains according to which discourses achieve legitimacy and authority through their integration with already developed texts. In the case, online platforms legitimize their policies by appealing to global governance norms, norms that, in turn, tend to be Anglophone and elitist. This intertextual harmony has the effect of entrenching the linguistic status quo in a circular tautology that sanctifies exclusion by appeal to other more exclusive texts. A third illustration of intertextuality is possible to find in the policy enforcement guidelines on YouTube which refers to some specific terms, e.g., to local laws, language support, and community expectations, being only selectively invoked. Such appeals sound syntactically closer to legal texts in the U.S. of communications law and international copyright conventions such as the Berne Convention but at best occur intertextually with indigenous or community initiatives in local linguistic justice. According to Blommaert (2005), intertextuality in the globalized discourse is most often borrowing of authorities and innocence of not taking cognizance of the subaltern narrative. When platforms invoke the concept of international norms they tend to invoke documents that were developed without the consultations of linguistically diverse or digitally marginalized groups. A consequence that is produced is a digital governance ecosystem where language representation has been both constructed and limited to a small collection of dominant-source texts. This process of intertextual silence on the part of minority languages introduces an additional level of marginalization of these groups, displaying how even neutral policies can create ideological bias by the means of discursive borrowing.

4.6. Lexical Analysis

The present concern of lexical analysis involved the targeted vocabulary use and euphemism in the platform governance texts as well as on interface texts that contribute majorly in the building of stories of inclusion and exclusion. As an example, Meta policies on Safety and Integrity work with generic, cleaned up words such as violation, user risk, harmful behavior, etc. without explaining how those translate across language boundaries. Employing these types of abstract words can be used to create an illusion of universality; a fact that neglects the fact that there are specific needs which minority language speakers have. In accordance with what Halliday and Hasan (2014) emphasize, lexical cohesion is a one of the main sources to develop the meaning and demonstrate the ideological intentions. An evasion of exact linguistic markers, either using the term, non-English speakers, or indigenous language communities, platforms hide the phenomenon of digital lingual inequality. It is the aspect of dominance of world hegemonic languages since they make the use of their languages to be a given in a way that leaves other users invisible. Moreover, so called content moderation or language preferences can frequently imply the agency of a user, but does not cover instances of structural constraints since some languages are not even presented as an eligible choice in the first place. Lexical patterns are also there and imply implicit prejudices and power differentials. The authoritarian tone of the policies, use of assertive verbs and nominalization (enforce, restrict, protect, flag) – tend to imply power and order in the policy discourse, yet they convey linguistic exclusivity. As an example, in the web-site of YouTube, Hate Speech Policy (2023), there is only one case of the word minority, and the word language is used only to mean profanity or hate terms in English. Such selective use of words identifies whose experiences matter when defining discourse of the safety of the platforms. It has been indicated in the writing of Fairclough (2003) that there is never a neutrality in lexical choices,

rather it constitutes and signifies relations of power. In this regard, lexical analysis demystifies the fact that the language (even in its vocabulary) applied in the digital governance excludes and marginalizes users whereas it codes dominant ideologies of digital communication even to the values towards standardized English and Eurocentric values.

4.7. Modality and Hedging

As the discussion of modality and hedging shows, digital platforms create forms of power, accountability, and inclusion via degrees of certainty and obligation at the expense of minorities of language speakers. Modality markers such as *may* or *can*, *should*, or *where possible* are most common in the documents of Community Standard and Transparency Center in Facebook, when citing localization services or translations. To give an example, there is the following clause, used in the documentation relating to Meta Language Support: We can deliver translated content when possible (Meta, 2023). The implications of such hedging are that linguistic accessibility is not a given right but is a conditional offering: it is given at the whim of the platform, or whether it is viewed to be possible. Halliday and Hasan (2014) tells us that modality records the judgments of the speaker in terms of probabilities, obligations, and so on, and in this instance the digital gatekeepers employ low-modality hedging in order to avoid the responsibility of achieving full language inclusiveness. The outcome is a watered-down multilingual access characterized by disenfranchisement of non-dominant user groups linguistically. Besides, high-modality expressions are strikingly understated in dealing with content control rather than linguistic fairness. According to YouTube policy on hate speech, it is permissible because it transfers the duties and responsibilities, as we can see: "Content that incites violence or hate against a protected group will not be allowed," with a declarative modality of prohibition and certainty. Conversely, in the sections concerned with language availability of the interface, the same policies tend to read, "We would like to serve as ubiquitous as possible in terms of language choices," demonstrating future desire and not strong intent. Modality, according to Fairclough (2003), is one of the important linguistic properties of the analysis of the correspondence of power and responsibility in texts. This tactical difference in modality illustrates how platforms have arrived at a ferocious grasp of their power over the rule as they are ambiguous and indifferent about linguistic access. Marginalization of assertiveness will help solidify the inequality of the digital through the over-cautious promises of support, rather than outright assurances of inclusion, a dynamic that will manifest in a systemic way as the policies of systems that provide support to language will be made optional instead of part of a guaranteed practice.

4.8. Pronouns Usage

The use of pronouns in and on platform policy documents is a very conspicuous but very strong factor of defining group identities, inclusion behaviour, and institutional power. In official papers such as Community Standards by Meta or Terms of Service by Google, the collective pronouns which refer to a specific entity, whether that be *we*, *our* or similar, distinguish the group entity as the platform itself and symbolically create an entity with perceived power and good heart. Considering our example of the ideas behind the platforms, such as the "We work so that everyone could feel safe and respected" (Meta, 2023), the platform can be deemed as the central and protective force. However, the same user pronouns are depersonalized and generalized the type of pronouns used to refer to a user is generally the pronouns at the user level such as *users* or *they* and thus there is a discursive gap between the user and the user. Van Dijk (1998) observed that, pronouns are a strategy of discursive which can constitute or eliminate, make human or depersonalize. This de-personalizing aspect usually removes the uniqueness of the identity of the minority language speakers and homogenized them into the abstract label of users and masks their need of serving their particular language or culture. What is also interesting is that an inclusive, collective pronoun, *we* is employed to establish a sense of moral concurrence and in the same breath it establishes control over institutional authority. YouTube, for instance, has had the following message on its page How youtube works: We take down the content that breaks our policies in order to guard our community. The phrase the community in this case means inclusion, while concealing the lack of authentic multicultural or multilingual representation in policymaking. It is actually a fantasy homogeneity, linguistically positively associated with the prevailing cultural norms and disassociated with linguistic minorities as the object of the created *we*. Reisigl and Wodak (2005) say that Pronouns have an ideological usage to form in-groups, and out-groups. Such polarizing approach perpetuates digital marginalization because the digital medium will be symbolized as the group bearing the majority identity, and, in the process, marginalize others

as peripheral others. Overall, the appearance of neutrality in terms of pronouns within digital governance texts can be discussed as achieving much of the work of boundary-building, which is a type of ideologically situated action with linguistic roots and foundations.

4.9. Framing Analysis

The framing analysis can assist in revealing the ways digital inclusion and multilingualism are framed (or marginalized) in the form of the official language of online platforms. Upon a close reading of the statements of Facebook on the topic of accessibility and inclusion, it is possible to note that the prevalent frame is the techno optimism. Through building technology, the company believes that it is building something to provide people with power to generate a community and make the world closer together (Meta, 2023). Although this frame appears to be all-embracing, it focuses on technological progress as the major force of equality leaving behind structural and cultural aspect, such as language rights. By framing it as an effect of innovation as opposed to a basic user right, the platform makes the issue of multilingual accessibility optional, not a requirement. Entman (1993) states that a frame means to choose some things of the perceived reality and increase their salience in the text of communication. Here, platforms prefer to emphasize their technological successes at the expense of the socio-linguistic impossibility of many users to work and interact with them, including the lack of the indigenous or regional language in such important sections as help center, content monetization portal, or AI products. Likewise, in the public documents of YouTube, the issue of multilingualism is presented with the perspective of scaling the business and expanding the audience. Such statements as "We are committed to growing our language services as part of our mission to serve the world" (YouTube Policy, 2023) indicate that the expansion of the number of languages is based on the market volume and not on the linguistic equality. This penny-pinching frame puts languages with low resources in a disadvantaged position, in seeing such languages as failing to be scalable investments. Such is the case that frames organize meaning by specifying what is going on here according to Goffman (1974). In the case of Youtube and the likes of it, the frame works towards eliminating the understanding of language access as a social need and sets it as a market commodity. Alternatively, the language policy, as it is, is not merely about language, rather, it is a matter of ideology, economics and control (Shohamy, 2006). Framing analysis thus reveals the insidious but inveterate means through which platform discourse persuasively edifies users of its concepts of inclusion.

4.10. Ideological Analysis

Under the critical discourse analysis (CDA) emphasis, ideological analysis brings to the fore that the policies of online platforms are saturated with powerful neoliberal and technocentric ideologies to the detriment of English and other dominating languages on the one hand to the disadvantage and marginalization of the indigenous and other minority language speakers. To give just one example, phrases, such as, the global community, scalable safety, and universal standards, are being commonly used in the official documentation conducted by Meta (Meta, 2023), seemingly neutral phrases that nevertheless show a world order that assumes Western digital hegemony. These slogans suppose some linguistic and cultural universalism which is absolutely mistaken. According to Fairclough (2003) who also explains that ideology is most powerful when it is naturalized which implies that prevailing views are naturalized in such a manner that they appear to be common sense or non-ideological. The usage of dominant languages and abstract globalization ideals does the trick of keeping the status quo at least, or absolutely. Excluding those with other languages and not fitting in within the boundaries of the digital participatory society. The ideology of efficiency, which is another influential value against these texts, pressurizes the platforms to cut off costs and streamline language services at the expense of cultural and linguistic equity.

4.11. Comparative Analysis

A comparative study of an analysis of language policies on platforms and national digital frameworks has shown a certain continuity, such as language inclusion principles are minimalistic in online platforms compared to the multilingual requirements provided in the national constitutions or linguistic diverse nations in education. India notably has 22 official languages in its Constitution, but Facebook India offers the interface in only a minor number of them, only Hindi, Bengali, and Tamil, whereas other ones, including Maithili or Santali, are left out. Conversely, national digital policy in Estonia implements the availability of the entire

Estonian language in every e-governance portal. This detour shows that language access through platforms is not free because of democratic or ethical requirements but can be determined by the number of users and the advertising market. As was observed, writing and speaking policy at the ground level is frequently inconsistent with what the nation on the podium suggests, and the same applies to the cyberspace.

4.12. Strategies of Recurrent Discourse Grouping

These repeated discursive patterns across the documents of the platform can be summarized under more general themes to preconceptualize the same exclusionary logic of linguistic governance of the Internet. Three key discourse groups are identified which include techno-optimism, content regulation, and abstract inclusivity. Techno-optimism presents innovation as the universal equalizer, as seen in the repeated assertion of Meta to bring people closer together without taking into account a linguistic barrier. The second category which is content regulation speaks of trust and security and adopts a legalistic and disciplinary language such as enforcement, violation, and injury which is frequently used as an overarching approach independent of cultural and linguistic context. Abstraction of commitments are noted in the third category, abstract inclusivity, whereby commitments such as the need to support language diversity where feasible are made, but in abstract forms; rhetorically this diffuses responsibilities as has been demonstrated by Pennycook (2010). Such groupings supports the stance of Bourdieu (1991) that language can be viewed as a symbolic power that African Americans in the lower-level dialectic categories accomplish visibility, inequality which is naturalized through selective and repetitive themes. Through systematic reinforcement of the recurring themes, platforms establish a comprehensive approach to normalization of a digitally sanitized space in which multilingualism is something marginal, optional, or contingent on, rather than the core of the equitable participation in the digital citizenship.

4.13. Cross-Checking Validation Secondary Data

Secondary data on linguistic diversity cross-validation showed that there is a giant gap between the promises made in various international bodies (UNESCO, ITU, and the UN Broadband Commission, among others) and action on the platform. As one example, the 2022 report by UNESCO titled *ReShaping Policies for Creativity* stresses the necessity of language as one significant pillar of digital cultural rights and suggests that we should promote inclusive digital ecosystems that take into account the linguistic diversity of the world. However, these principles of globalization seem to be entirely idealistic, when considering practices of platforms. According to the ITU, most global platforms received low ranks in the Digital Inclusion Index of 2023 which includes linguistic accessibility among other factors. The evidence-base justifies the claim by Fairclough (2013) that discourse interpretation should be performed through the lens of the institution since platforms purport to be inclusive, and yet, they inherently exclude those same groups the global institutions should be protecting.

4.14. Researcher Reflexivity

One of the biggest questions that a researcher critically analysing digital language policies should answer is to address the interpretive lens that texts have been researched within. The academic focus of linguistic justice and postcolonial language theory defined my belief regarding the identification of exclusionary discourse, particularly the constructions of platforms in the way they do or do not represent non-dominant language speakers. Since I am a bilingual speaker of the same language myself, and sit in a context of linguistic diversity that is not only rich but also touchy as it relates to questions of politics (e.g. Pakistan), my positionality is sensitive to the forms of erasure that occurs in spaces where communities discussing the world have their voices formalized online. As articulated by Denzin and Lincoln (2011), qualitative research is political and moral in origin and this analysis has been completed in the realization that discourse analysis cannot be expected or desired to be neutral.

5. Conclusion

This study was a critical reflection of how language policy and the interface design of web-based platforms enact digital marginalization of the minority and indigenous language speakers. The analysis of dominant digital discourses through Critical Discourse Analysis (CDA) in the form of a qualitative study revealed that dominant discourses in the digital sphere focus on such languages as Arabic or English, disregarding other, non-popular spoken languages, and thus, they shape online communication processes and functions and systematically lock out

other languages historically left behind. Analysis of the excerpts taken in textual and visual form discussions of documents on websites and platforms, including Meta, YouTube, Twitter, proved how the answers given in it contain abstract, universalist rhetoric that does not imply specific obligations towards multilingual inclusion. Policies that were often adhered to were vague modality (may, where possible) and technocratic vocabulary as a way to avoid that responsibility in linguistic inequity. Categories of depersonalised user identifiers and vague statements of inclusion ducrumansard and mmh97 show the same effect of neutrality of ideology keeping power in the hands of the same as described by Van Dijk (2006) as the elite discourse. The results attest the fact that language is not only a medium of digital communication but a form of gatekeeping that defines who should fully access the digital world. It is concluded that online platforms employ the selective model of multilingualism, which runs in tandem with economic interests but not with cultural or linguistic justice. Comparative and intertextual approaches allowed proving the fact that as the principles of language diversity are propagated by such global mechanisms as UNESCO and ITU, the digital space supports the nullification of these values as the partial or commercial language choice is constantly supported.

Observation of modality, the use of pronouns, and semiotics of looks also demonstrated that platform design cannot be linguistically neutral because it reproduces and propagates cultural stratifications that gain privileges to languages used by dominant groups and silences those of marginalised ones. This trend is not only restricted to user accountability and service delivery but also poses a danger on the cultural preservation and language revival of the less dominant communities. According to Fairclough (2013) and Bourdieu (1991), the functionality of language as symbolic capital in the real world, language is exploited in digital places to exert more control on those who own the platforms, which encourages the creation and perpetration of inequalities on a global scale. These study implications are practical and theoretical. In practice, it involves the need to immediately reform how platforms do and signal their language policies, shifting away from applying hazy promises to enforcement of universal rules where exclusionary results no longer reflect reality. Theoretically it confirms that it is important to employ CDA to the digital spaces so as to reveal the power structures that are manifested in the language and design itself. Future studies ought to be extended to incorporate user-generated material as well as interviews of the speakers of marginalized languages to further comprehend lived experiences of marginalization. Also, there must be partnership among linguistic communities, policy makers and digital companies to create fair digital infrastructures. It is only after understanding language as a right and a resource rather than as a mere technicalisation that the layers of digital marginalisation will start to fall and make way to a globalised, but not digitalised society.

5.1. Suggestions and Recommendation

Based on the results of the critical discourse analysis done in this paper, a number of practical recommendations and proposals can be listed. Online platforms also need to reform their language regulations that are market-inclusive to linguistic fairness by applying complete interface and content moderation of the minority and indigenous languages. It does not imply just the extension of language choices but use of available fonts, scripts, layouts and content of assistance with accessibility and cultural relevance. Policymakers and platform designers are urged to work with the communities to develop language diverse domains in cooperation with local scholars and cultural organizations that establish the usage of linguistic rights as digital rights. Transnational digital governance organizations, including UNESCO and ITU ought to have better engagements with technology giants to develop binding multilingual standards of accessibility. More so, the platforms are advised to use open, bullish modality in their policies text working towards exclusion of ambiguous language such as where possible to concrete and quantifiable assurances. The content moderator and AI developer training programs should be inclusive of linguistic diversity literacy to eliminate bias in automated systems that might be inadvertent. Last but not least, the future work of the scholars should be aimed at including the voices of the marginalized users in the form of qualitative interviews, participatory policy inquiry, and ethnographic research to co-construct an inclusive digital discourse landscape.

References

- Ahmed, D. M., Azhar, Z., & Mohammad, A. J. (2024). The Role of Corporate Governance on Reducing Information Asymmetry: Mediating Role of International Standards for Accounting (IAS, IFRS). *Kurdish Studies*, 12(1).
- Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), 419-438. <https://doi.org/10.1111/j.1467-9841.2006.00286.x>
- Banda, F. (2009). Critical perspectives on language planning and policy in Africa: Accounting for the notion of multilingualism. *Stellenbosch Papers in Linguistics PLUS*, 38, 1-11.
- Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, 6, 587-604. https://doi.org/10.1162/tacl_a_00041
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, 2021-03-03). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ? FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency,
- Benson, C. (2013). Towards adopting a multilingual habitus in educational development. In *Language issues in comparative education: Inclusive teaching and learning in non-dominant languages and cultures* (pp. 283-299). SensePublishers Rotterdam.
- Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2022, 2022-06-21). The Values Encoded in Machine Learning Research. FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency,
- Blodgett, S. L., Barocas, S., Daumé Iii, H., & Wallach, H. (2020, 2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,
- Blommaert, J. (2005). *Discourse: A Critical Introduction* (1 ed.). Cambridge University Press.
- Blommaert, J. (2010). *The Sociolinguistics of Globalization* (1 ed.). Cambridge University Press.
- Bourdieu, P. (1991). *Language and symbolic power*. Harvard university press.
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4-5), 585-614. <https://doi.org/10.1177/1461445605054407>
- Busch, B. (2012). The Linguistic Repertoire Revisited. *Applied Linguistics*, 33(5), 503-523. <https://doi.org/10.1093/applin/ams056>
- Canagarajah, A. S. (2006). NEGOTIATING THE LOCAL IN ENGLISH AS A LINGUA FRANCA. *Annual Review of Applied Linguistics*, 26, 197-218. <https://doi.org/10.1017/S0267190506000109>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Crystal, D. (2011). *Internet Linguistics* (0 ed.). Routledge.
- Cummins, J. (2000). *Language, Power and Pedagogy: Bilingual Children in the Crossfire*. Multilingual Matters.
- Davis, J. L. (2010). Indigenous language cyberactivism: How and why Indigenous people are using the Internet to preserve their languages. *The International Journal of Language Society and Culture*, 3-9.
- Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*. sage.
- Deumert, A. (2014). *Sociolinguistics and Mobile Communication*. Edinburgh University Press.
- Dhivya, D. S., Hariharasudan, A., Ragmoun, W., & Alfalih, A. A. (2023). ELSA as an Education 4.0 Tool for Learning Business English Communication. *Sustainability*, 15(4), 3809. <https://doi.org/10.3390/su15043809>
- Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4), 51-58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Fairclough, N. (2003). *Analysing Discourse: Textual Analysis for Social Research* (1 ed.). Routledge.
- Fairclough, N. (2013). *Critical Discourse Analysis* (0 ed.). Routledge.
- Gillespie, T. (2019). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- Goffman, E. (1974). Frame analysis: An essay on the organization of experience. *Northeastern UP*.
- Grenoble, L. A., & Whaley, L. J. (2005). *Saving Languages: An Introduction to Language Revitalization* (1 ed.). Cambridge University Press.
- Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in english*. Routledge.

- Hargittai, E. (2018). The digital reproduction of inequality. In *The inequality reader* (pp. 660-670). Routledge.
- Hemer, O., & Tufte, T. (2005). Media and Glocal Change. Rethinking Communication for Development. *Göteborg and Buenos Aires: NORDICOM and CLASCO Books*.
- Hornberger, N. H. (2009). Multilingual education policy and practice: Ten certainties (grounded in Indigenous experience). *Language Teaching*, 42(2), 197-211. <https://doi.org/10.1017/S0261444808005491>
- Kothari, U., & Hulme, D. . (2003). Narratives of development: An anthropological critique. Zed Books.
- Lee, C., & Chau, D. (2018). Language as pride, love, and hate: Archiving emotions through multilingual Instagram hashtags. *Discourse, context & media*, 22, 21-29.
- May, S. (2013). *Language and minority rights: Ethnicity, nationalism and the politics of language*. Routledge.
- Mohammad, A. J., & Ahmed, D. M. (2017). The impact of audit committee and external auditor characteristics on financial reporting quality among Malaysian firms. *Research Journal of Finance and Accounting*, 8(13), 9-16.
- Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York university press.
- Omoniyi, T., & White, G. (2006). *The sociolinguistics of identity*. Bloomsbury Publishing.
- Pennycook, A. (2006). *Global Englishes and Transcultural Flows* (0 ed.). Routledge.
- Pennycook, A. (2010). *Language as a Local Practice* (0 ed.). Routledge.
- Perley, B. C. (2012). Zombie Linguistics: Experts, Endangered Languages and the Curse of Undead Voices. *Anthropological Forum*, 22(2), 133-149. <https://doi.org/10.1080/00664677.2012.694170>
- Phillipson, R. (1992). *Linguistic imperialism*. Oxford university press.
- Piller, I., & Takahashi, K. (2011). Linguistic diversity and social inclusion. *International Journal of Bilingual Education and Bilingualism*, 14(4), 371-381. <https://doi.org/10.1080/13670050.2011.573062>
- Reisigl, M., & Wodak, R. (2005). *Discourse and discrimination: Rhetorics of racism and antisemitism*. Routledge.
- Rickford, J. R. (2016). *Raciolinguistics: How language shapes our ideas about race*. Oxford University Press.
- Roberts, S. T. (2019). *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press.
- Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches* (0 ed.). Routledge.
- Skutnabb-Kangas, T., & Phillipson, R. (1995). Linguistic human rights. *Berlin: Mouton de Gruyter*.
- Tollefson, J. W., & Tsui, A. B. M. (2017). Issues in Language Policy, Culture, and Identity. In A. B. M. Tsui & J. W. Tollefson (Eds.), *Language Policy, Culture, and Identity in Asian Contexts* (1 ed., pp. 259-270). Routledge.
- UNESCO. (2016). *If you don't understand, how can you learn? Global Education Monitoring* <https://unesdoc.unesco.org/ark:/48223/pf0000244575>
- Van Dijk, J. (2005). *The Deepening Divide: Inequality in the Information Society*. SAGE Publications, Inc.
- Van Dijk, T. A. (2006). Discourse and manipulation. *Discourse & Society*, 17(3), 359-383. <https://doi.org/10.1177/0957926506060250>
- Warschauer, M. (2003). *Technology and Social Inclusion: Rethinking the Digital Divide*. The MIT Press.
- Wided, R., & Alfalih, A. A. (2023). Extremism immunity through artificial intelligence networks: Extremism awareness and social intelligence. *International Journal of Data and Network Science*, 7(1), 341-356. <https://doi.org/10.5267/j.ijdns.2022.9.013>
- Wodak, R. (2004). Preface: The power of language in political discourse. *Journal of Language and Politics*, 3(3), 381-383. <https://doi.org/10.1075/jlp.3.3.01wod>