International Journal of Computing Sciences and Technology Studies
https://journals.hu.edu.so/index.php/ijcsts

Orthographic Variability Among Somali People in Writing Somali Script

Fahad Mohamed Abdullahi
Faculty of Computer Science and Information Technology, Hormuud University, Mogadishu, Somalia

Email: fahad@albasrawie.com

Abstract

This study investigates the orthographic variability in the Somali language, examining how inconsistent educational practices, the influence of foreign languages, and outdated linguistic resources contribute to divergent spelling and writing conventions among Somali speakers. The purpose of the study is to identify the underlying causes of this variability and to propose practical measures for standardization. A multifaceted, field-based methodology was employed, combining visits to academic institutions, engagement with media outlets, consultations with Somali scholars and linguists, and an empirical dictation exercise administered to a group of students to capture real-world writing variation. The findings confirm that orthographic variability is a pervasive and persistent issue, stemming from limited advancement beyond basic literacy, weak institutional incentives for mastering standardized Somali writing, and the comparatively stronger emphasis placed on foreign-language grammar instruction. Based on these findings, the study proposes the development of a unified Somali language curriculum together with a Natural Language Processing (NLP)-based platform that offers automatic correction and suggestions, educational modules, a standardized vocabulary database, and tools for community engagement. In conclusion, standardizing Somali orthography is essential not only for clearer communication but also for preserving the cultural richness, linguistic unity, and long-term relevance of the Somali language, and the study recommends coordinated educational reform and technological innovation to achieve this goal.

Keywords: Somali language, orthographic variability, standardization, Natural Language Processing (NLP), educational practices, cultural preservation

1. Introduction

The Somali language, rich in history and cultural significance, faces a unique challenge: a lack of standardized orthographic practices. This issue manifests in the varied ways in which Somali words are written, leading to confusion and inconsistency. For example, common words such as "gabar" and "gabadh" or "waxay" and "wexey" are often written differently by different people, highlighting a significant orthographic divergence. This variability not only complicates communication but also undermines the linguistic unity of the Somali people.

The problem is further exacerbated by the widespread use of foreign languages, such as English, which many Somalis learn in formal educational settings. The proficiency in

English grammar among Somali speakers contrasts sharply with the inconsistencies in Somali script writing. This phenomenon raises important questions about the factors driving these orthographic practices and their implications for the Somali language. Traditionally, Somali people vary significantly in their writing styles, often influenced by regional dialects and the integration of foreign language terms. This fragmentation has led to a communication barrier that undermines the collective linguistic identity.

The Somali language, integral to the cultural and national identity of the Somali people, faces significant challenges in both written and spoken forms within the community. Many individuals encounter difficulties either in writing, speaking, or both. A prevalent mindset among Somali speakers is encapsulated in the belief, "The important thing is to understand each other—the one who is speaking and the one who is being spoken to." While this pragmatic approach prioritizes mutual comprehension, it has inadvertently contributed to the erosion of standardized linguistic practices, posing a serious problem for the preservation and consistency of the Somali language. In the context of linguistic practices, a clear disparity exists among Somali speakers when it comes to writing their mother tongue versus foreign languages. If a group of Somalis were assembled and tasked with a Somali dictation exercise, their writing methods would vary significantly, reflecting the phrase, "Somali language, as it is said, is written." Conversely, the same group would likely write with accuracy and consistency if given a dictation exercise in a foreign language such as English. This proficiency can be attributed to formal education and the emphasis on learning English grammar, reinforced by the notion, "Learn English; it is a universal language." [1]

This linguistic phenomenon has led to the incorporation of foreign words into everyday Somali usage, despite the existence of equivalent Somali terms. Alarmingly, many Somali speakers are familiar with certain English words but are unaware of their Somali counterparts, further highlighting the extent of this issue. This trend underscores the need for focused efforts to standardize and promote the correct usage of the Somali language.

1.1 Disparity in Writing Practices

The inconsistency in writing Somali arises from a lack of standardized orthographic practices and formal education focused on Somali grammar. Many individuals learn the basics of reading and writing Somali during their early education but do not progress to more advanced levels. This results in a superficial understanding of the language, where people are comfortable with basic words and phrases but struggle with more complex constructions. Consequently, when asked to write in Somali, individuals often employ varied orthographic conventions, leading to significant discrepancies in written communication. Every phenomenon has an underlying cause, and the divergence of Somali youth and children from standardized orthographic practices in their mother tongue is no different. Various factors contribute to this issue [2].

1.1.1 Lack of Advanced Learning

When Somali children and individuals initially learn to read and write in Somali, they often cease to advance their knowledge beyond the basic level. They become familiar with simple and frequently used words but do not pursue a deeper understanding of the language. As they progress through their education, from high school to university, they fail to expand their vocabulary or grasp the orthographic rules governing long and short vowels and double consonants. This lack of advanced learning results in individuals who, when confronted with unfamiliar words, cannot correctly apply orthographic conventions [3].

1.1.2 Absence of Motivation and Necessity

The Somali youth and children do not face any significant repercussions for not mastering their language's standardized writing system. There is no societal or professional pressure, such as job loss or lack of respect, associated with improper Somali writing. Consequently, there is no

intrinsic motivation driving them to learn more about their language. Moreover, there is a prevalent attitude among Somali students that proficiency in their mother tongue is not an obligation. This lack of perceived necessity further diminishes their incentive to pursue linguistic accuracy and consistency.

1.2 Influence of Foreign Languages

The proficiency in foreign languages, particularly English, among Somali speakers is a testament to the effectiveness of structured language education. English is taught with a strong emphasis on grammar, spelling, and syntax, providing learners with a comprehensive understanding of the language. This rigorous approach is often lacking in Somali language education, resulting in a weaker grasp of Somali orthographic rules. The societal perception that English is a "universal language" further motivates individuals to prioritize its learning, often at the expense of their mother tongue [2].

2. Literature Review

The Somali language, a vital component of the Somali people's identity, plays a crucial role in unifying the nation through its rich history and vibrant culture. However, the modern challenges of linguistic inconsistency and the influence of foreign languages necessitate a renewed focus on standardizing and preserving the Somali script. The scholarly discourse surrounding the Somali language recognizes its profound potential for scientific and literary applications, as well as the importance of leveraging contemporary technological advancements to ensure its continued relevance and integrity.

A people can become one nation because of its language, history, and culture. Scholars who have researched the Somali language, both foreign and local, have observed and reported that Somali is sufficiently equipped to be used for science and literature. Somali is a copious language, with a great number of words; in every respect it is a rich language, its words and expressions, its poetry and proverbs, its odes and elegies, its poems and plays, its songs and ballads, its stories and tales, its written and oral genres, whatever aspect one considers, it is a great language, full of wisdom and sayings that convey profound ideas [1].

The proliferation of computers and the internet has greatly amplified the necessity for standardizing written language. When written language is standardized, computers can provide various linguistic services, such as spell checking, semantic analysis, machine translation, and more. The convergence of interests between computer scientists and linguists in leveraging computers for linguistic research and language services is increasingly evident. This interdisciplinary collaboration has led to the emergence of a new field known as "Natural Language Processing (NLP)" or "Computational Linguistics." This field, which has been evolving since the 1950s, continues to expand and develop, driven by advancements in both computational and linguistic sciences [1].

In the more recent work "Unity, Development, and Modernization of Professor Abdalla Omar Mansur" (2024), Professor Abdalla Omar Mansur offers an insightful exploration into the evolution and standardization of the Somali language. The first chapter of this seminal book tackles the debate over the potential "literacy" or "voices" left behind during the script development process. Professor Abdalla asserts that creating the Somali script was not merely about linguistic knowledge but required a systematic and scientific approach to achieve a unified writing system. His work stands as the first and most systematic effort towards standardizing Somali writing, making it a crucial resource for scholars and educators dedicated to Somali literacy [2].

The standardization and orthographic stability of the Somali language have been major subjects of scholarly inquiry since the adoption of the official Latin script in 1972 [3]. The official script creation marked a historic transition, shifting Somali from an unwritten national language with competing scripts (including Arabic and Osmanya) into a unified medium for

education and governance [4]. However, scholars have long noted that formal script adoption did not eliminate orthographic and dialectal variations [5]. The selection of the central/northern *Maxaa* dialect as the standard written form inevitably created friction with southern and regional dialectal variants [6], leading to persistent divergence in spelling and word choice across different Somali-speaking populations.

From a descriptive linguistic standpoint, the complexity of Somali grammar and phonology presents inherent challenges for orthographic consistency [7]. Key features such as vowel length distinctions (short vs. long vowels), consonant gemination (double consonants), and tonal accents are frequently sources of orthographic variation among writers [8]. For example, common functional words and lexical items display widespread spelling fluctuations—such as *"gabar"* vs. *"gabadh"* or *"waxay"* vs. *"wexey"*—primarily due to regional phonetic shifts and a lack of standardized advanced grammar instruction in formal schooling [9]. Furthermore, historical analyses show that while early standardization policies established basic literacy, the lack of continuous, advanced language curricula in secondary and tertiary education has caused written proficiency to stall at elementary levels [10].

Beyond early national decrees, the structural expansion of written Somali required intensive lexical modernization to support public administration, education, and modern press [11]. Early foundational work on transcribing Somali narratives demonstrated that creating a Latin-based writing system required reconciling complex acoustic phonetics with visual graphemes [12]. As functional language usage expanded in government and media, sociolinguistic research revealed that nominal compounding and morphological derivation became primary drivers of vocabulary growth, though often resulting in competing spelling choices across technical registers [13].

In educational and grammatical contexts, pedagogical grammars have sought to systematize Somali for both native speakers and academic researchers [14]. However, specific grammatical structures continue to create orthographic ambiguity. For instance, noun inflectional patterns—specifically gender polarity, tone placement, and irregular pluralization—frequently produce written errors among authors who lack formal instruction in advanced Somali morphology [15]. To address these inconsistencies, national lexicographical projects culminated in comprehensive monolingual reference works aimed at codifying standard headwords and orthographic conventions [16]. Furthermore, transcribing the vast body of Somali oral literature and proverbs highlighted the ongoing tension between capturing regional acoustic nuances and adhering to a single standardized written script [17].

3. Methodology

To investigate the orthographic variability and linguistic practices in Somali script, I employed a multifaceted approach involving field visits, academic consultations, and empirical studies. My methodology is outlined as follows:

3.1 Field Visits to Academic Institutions

I conducted extensive visits to various academic institutions, including universities, where Somali is taught and used as a medium of instruction. These visits enabled me to observe the current teaching practices and gather insights from educators on the challenges faced in maintaining orthographic consistency.

3.2 Engagement with Media Outlets

Recognizing the influence of media on language use, I visited radio and television stations where Somali is widely used. These platforms play a crucial role in shaping linguistic practices among the general population. Interviews with broadcasters and journalists provided valuable perspectives on the practical use of Somali in daily communication and the prevalence of orthographic variations.

3.3 Consultations with Somali Scholars and Linguists

I engaged with prominent Somali scholars, linguists, and literary figures to gain a deeper understanding of the historical and contemporary issues surrounding Somali orthography. These experts highlighted the reliance on traditional manual dictionaries for looking up words, a method they described as outdated and ineffective for modern linguistic needs. They emphasized the lack of contextually rich sentences and meaningful phrases in current reference materials.

3.4 Empirical Study: Somali Dictation Exercise

To empirically assess the orthographic variability among Somali speakers, I assembled a group of students and administered a Somali dictation exercise. The dictation was designed to capture common variations in writing practices. The resulting samples, shown in Figures 1–3, revealed significant differences in the students' writing methods. These variations underscored the necessity for standardized orthographic practices and highlighted the challenges faced in achieving uniformity in written Somali.

Figure 1: Dictation sample from Student 1, illustrating variation in writing style.

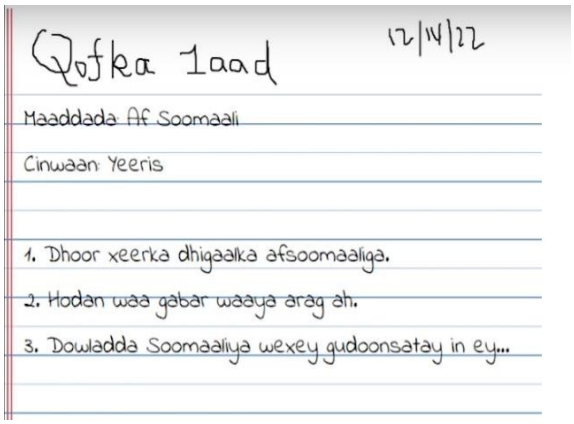


Figure 2: Dictation sample from Student 2, illustrating variation in writing style.

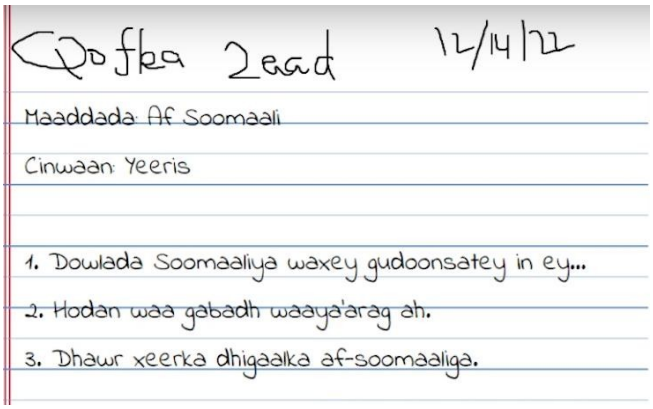
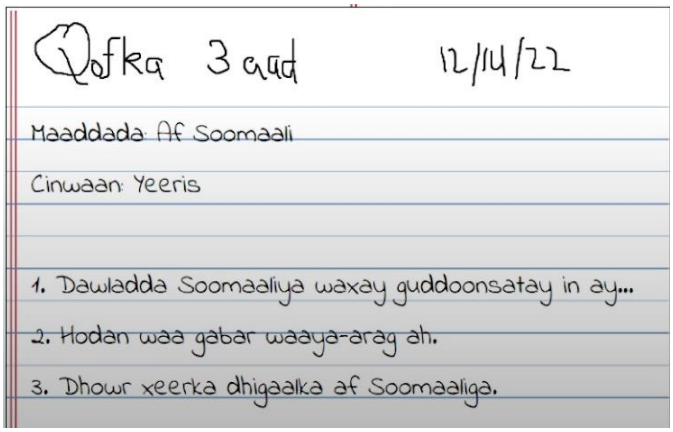


Figure 3: Dictation sample from Student 3, illustrating variation in writing style.



3.5 Data Analysis

The collected data from academic institutions, media outlets, and the dictation exercise were meticulously analyzed to identify common patterns and deviations in orthographic practices. This analysis provided a comprehensive overview of the current state of Somali writing and informed the development of strategies for promoting standardization.

4. Results

The results of my study confirm that orthographic variability is a pervasive issue among Somali speakers. The variability stems from inconsistent educational practices, the influence of bilingualism, outdated linguistic resources, and individual writing habits. Addressing these challenges through standardized education, updated resources, and technological solutions such as NLP-based platforms is essential for preserving the integrity and richness of the Somali language.

To address these challenges, this paper proposes the creation of a unified linguistic platform utilizing Natural Language Processing (NLP). This initiative aims not only to standardize Somali script but also to foster a deeper connection among Somali speakers worldwide by enhancing readability and comprehension.

4.1 Platform Description

The proposed NLP-based platform will serve as a central tool for standardizing the Somali script. It will leverage advanced computational linguistics techniques to analyze and reconcile variations in Somali orthography. Key features of the platform include:

- **Automatic Correction and Suggestions:** Utilizing machine learning algorithms, the platform will offer real-time orthographic corrections and suggestions, promoting consistency in writing.
- **Educational Modules:** Interactive modules will educate users on standardized practices, enhancing their writing skills and linguistic knowledge.
- **Language Database:** A comprehensive database of standardized Somali vocabulary, including technical and everyday terms, to reduce the reliance on foreign words.
- **Community Engagement:** Tools for community feedback and contributions, allowing continuous improvement and adaptation of the language model to real-world usage.

5. Discussion

5.1 Cultural and Educational Impacts

The widespread use of foreign words in Somali conversations is a direct outcome of the emphasis on foreign language proficiency. Phrases such as "Learn English; it is a universal language" promote the integration of English into daily communication. This bilingual or even multilingual environment can lead to a dilution of the native language, where speakers increasingly rely on foreign words due to familiarity and ease of use. This trend is concerning, as it can lead to a gradual erosion of the Somali language's richness and uniqueness.

The educational system plays a pivotal role in this linguistic shift. In many Somali schools, English and other foreign languages are often prioritized over Somali, both in terms of instructional time and perceived importance. This focus results in students developing a stronger command of foreign languages while their proficiency in Somali remains basic and underdeveloped. The lack of comprehensive Somali language education contributes to the variability in writing practices observed among Somali speakers.

5.2 The Urgency for Standardization

Addressing this issue requires a concerted effort to standardize Somali orthographic practices and enhance the quality of Somali language education. Developing and implementing a unified Somali language curriculum that emphasizes advanced grammar, orthography, and

vocabulary is crucial. Additionally, leveraging technology, such as NLP-based platforms, can facilitate the learning and standardization process, providing real-time feedback and corrections to learners.

Promoting the importance of linguistic purity and cultural heritage is equally vital. Encouraging Somali speakers to value and use their native language in all aspects of life, including professional and educational contexts, can help mitigate the encroachment of foreign words. Highlighting the cultural significance of Somali and its role in preserving national identity can foster a sense of pride and responsibility among speakers to maintain the integrity of their language.

6. Conclusion

The research highlights a critical issue within the Somali-speaking community: the significant orthographic variability and the lack of standardized writing practices. This study identified several contributing factors, including inconsistent educational approaches, the influence of bilingualism and multilingualism, outdated linguistic resources, and varied individual writing habits. These factors collectively undermine the linguistic unity and integrity of the Somali language.

My findings underscore the urgent need for a concerted effort to standardize Somali orthography. A unified curriculum emphasizing advanced grammar, orthography, and vocabulary, supported by NLP-based tools that provide real-time feedback, is essential. These technological tools can bridge the gap between traditional linguistic practices and contemporary needs, ensuring that the Somali language remains relevant and robust in the digital age. Promoting pride in Somali as a marker of cultural heritage and national identity is equally vital. In conclusion, addressing the orthographic variability in Somali requires a holistic approach that includes educational reform, technological innovation, and cultural advocacy. These steps can ensure the preservation and enrichment of the Somali language for future generations, reinforcing its role as a cornerstone of Somali identity and culture.

Acknowledgments

I would like to express my sincere gratitude to all the individuals and organizations that contributed to this research. My heartfelt thanks go to the academic institutions and their staff for providing invaluable insights during my field visits. I am also deeply grateful to the journalists and broadcasters at Somali radio and television stations for their time and perspectives. Special thanks to the Somali scholars, linguists, and literary figures whose expertise and guidance were instrumental in shaping this study. Your dedication to preserving the Somali language is truly inspiring. I would also like to acknowledge the students who participated in my dictation exercises; your contributions were essential to my findings. Thank you to everyone who supported this research, directly or indirectly, and to those who continue to champion the cause of linguistic and cultural preservation.

Conflict of Interests

The author declares no conflict of interests.

References

[1] Intergovernmental Academy of Somali Language (AGA), Afmaal, Djibouti, Dec. 17–21.

[2] A. A. Fiqi, Haldhigaalka Af Soomaaliga Guud, Mogadishu: https://www.facebook.com/share/p/bXtyuy5JyiQnEyWA/, 2024.

[3] B. W. Andrzejewski, The Introduction of a National Orthography for Somali, London: School of Oriental and African Studies, 1974.

[4] D. D. Laitin, Politics, Language, and Thought: The Somali Experience, Chicago: University of Chicago Press, 1977.

[5] A. Hussen, M. Abdullahi, and H. Mohamed, From Linguistic Imperialism to Language Standardization: The Dynamics of the Somali Orthography Project, Mogadishu: Journal of Somali Studies, 2016.

[6] M. Lamberti, Map of the Somali Dialects in the Somali Democratic Republic, Hamburg: Helmut Buske Verlag, 1986.

[7] J. I. Saeed, Somali, Amsterdam: John Benjamins Publishing Company, 1999.

[8] A. Puglielli and A. O. Mansur, Grammatica della Lingua Somala, Rome: Università degli Studi Roma Tre, 1999.

[9] A. M. Ismail, G. Banti, and M. D. Abdullahi, Building and Evaluating Somali Language Corpora: Asaas and Wargeys, Gothenburg: Association for Computational Linguistics, 2014.

[10] A. O. Mansur, Taariikhda iyo Sahanka Af-Soomaaliga, Rome: Roma TrE-Press, 2009.

[11] J. C. Caney, The Modernisation of Somali Vocabulary, with Particular Reference to the Period from 1972 to the Present, Hamburg: Helmut Buske Verlag, 1984.

[12] M. H. I. Galaal and B. W. Andrzejewski, Hikmad Soomaali, London: Oxford University Press, 1956.

[13] M. M. Hared, Elaboration of Somali Vocabulary: A Macro-Sociolinguistic Study of Somali Functional Usage, Los Angeles: University of Southern California, 1992.

[14] M. Orwin, Colloquial Somali: A Complete Language Course, London: Routledge, 1995.

[15] M. Nilsson, Grammatical Gender and Number in Somali Nouns, Pisa: Redsea Online / University of Gothenburg, 2018.

[16] A. Puglielli and A. O. Mansur, Qaamuuska Af-Soomaaliga, Rome: Roma TrE-Press, 2012.

[17] G. Kapchits, Hubsiino Hal Baa La Siistaa: Somali Proverbs, Aachen: Shaker Verlag, 2002.