



Media Discourse and Society



Volume 2, Issue 1, Summer 2026

ISSN (P): 3106-6453 & ISSN (E): 3106-6461

Homepage: <https://journal.vu.edu.pk/MDS>

A publication of

Department of Mass Communication

Virtual University of Pakistan

Title: Patterns of Roman Urdu Usage among Pakistani Social Media Users: A Content Analysis Study

Author (s): Dr. Shahid Minhas, Lecturer, Digital Communication, School of Communication & Design, Hanoi Campus Vietnam

Corresponding Author: Dr. Shahid Minhas

Citation: Minhas, S. (2026). Patterns of Roman Urdu usage among Pakistani social media users: A content analysis study. *Media Discourse and Society*, 2(1). Department of Mass Communication, Virtual University of Pakistan. ISSN (Print): 3106-6453; ISSN (Online): 3106-6461.

Copyright: © The Authors

Date of Publication: 24-06-2026

Licensing: This article is open access and is distributed under the terms of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Conflict of Interest: Author (s) declared no conflict of interest.

Virtual University of Pakistan

Sir Syed Memorial Society Building, 19-Ataturk Avenue, G-5/1, Islamabad

Abstract

This study has analyzed the forms and patterns in which roman Urdu is used in the comments by the social media users. In this study the Facebook posts were considered for the analysis in which the various comments were studied. For this purpose, the content analysis method of data analysis has been used. This method used both qualitative and quantitative approach to study the content. The time period for this study was 1st Oct 2022 to 31st Oct 2022. All those posts have been considered in this study that were posted in connection to a certain issues and events of immense importance for example, issues concerning ousting of PTI govt., Elections, Pak-America Relations, T20 world Cup etc. The various comments have been categorized in to the following categories such as numeric homophones, character repetition, spelling consistency, frequency of comments on a post, Pattern of roman Urdu, Short words, Sentence structure, Usage of English and Roman Urdu collectively and understanding of messages/posts content. The results show that, Pakistani social media users are constantly and regularly using the roman Urdu in the comments, they use it for any type of purpose to communicate their messages, the messages are clear to understand and they follow the same pattern of sentence structure as used in proper Urdu literature. However, this trend is causing to ignore the usage of proper Urdu on social media, which might be not favorable in the context of a country this has Urdu as national language.

Keywords: Roman Urdu, forms and patterns, social media, users, Pakistan.

Introduction

Urdu is the official language of Pakistan, it is spoken throughout the country especially in the Punjab and Sindh. Since the 6th century, Urdu language has emerged from other languages like “Arabic, Hindi, Persian, Prakrit, and Sanskrit”. However, in many Government offices, schools, colleges and Universities across Khyber Pakhtunkhwa and Baluchistan it is widely spoken language. Pakistan is a multilingual country with more than 70 languages are used (Shah, Riaz & Malik, 2019).

Statistics indicate that approximately 60% of the 1.65 billion people in the Indian subcontinent speak Hindi/Urdu (Bilal, Et al., 2017). The majority of Pakistan's more than 200 million inhabitants speak and write Urdu fluently (Moazam & Shekhani, 2018). Urdu is the country's official language. Nonetheless, as a result of technological development, Roman Urdu is now

widely used in everyday communication. The majority of people use roman Urdu due to its convenience. As not everyone can type Urdu perfectly on tabs, mobiles, therefore they opt. to use Roman Urdu for their ease. The problem at which this study focus on is that, users are continuously shifting towards Roman Urdu, thus decreasing the use of proper Urdu in their communication, hence this study highlights the forms and usage pattern in which the Roman Urdu is used. Thus, it identifies the areas and put recommendation for users to use proper Urdu in communication.

Besides Low-cost mobile phones have facilitated the widespread adoption of smart phones in developing nations; for instance, South Asian users represent one of the world's largest mobile phone user bases. Researchers have referred to user groups from developing countries as emerging users in recent years. These user groups are either less educated, economically disadvantaged, geographically dispersed, or have a background that is culturally diverse. According to the Pakistan Telecommunication Authority, there were 133.24 million mobile subscribers in 2016 and over 301.7 billion text messages were sent in 2014 (Bilal, Rextin, Kakakhel & Nasim, 2018). We observed that Pakistani users of text messaging predominantly employ Roman Urdu. Roman Urdu refers to the practice of writing Urdu (the national language of Pakistan) using the Roman alphabet (Mushtaque & Zaki, 2019). Local users are of the opinion that this practice originated when non-native English speakers began to communicate via text messages (Bilal, Rextin, Kakakhel & Nasim, 2018). Looking to the trend of using roman Urdu, this study has been planned to analyze the forms and pattern in which it is used on social media in Pakistan.

Research Objectives

1. To examine the patterns and linguistic features of Roman Urdu used in Facebook comments related to significant political, social, and international events.
2. To analyze the use of specific Roman Urdu writing practices, including numeric homophones, character repetition, spelling variations, and shortened words in online communication.
3. To investigate the occurrence of Urdu-English code-mixing and sentence structure patterns in Roman Urdu Facebook comments.
4. To assess the comprehensibility and communicative effectiveness of Roman Urdu messages used in social media interactions.

Research Questions

RQ1: What are the predominant patterns and linguistic features of Roman Urdu used in Facebook comments related to significant political, social, and international events?

RQ2: How are numeric homophones, character repetition, spelling variations, and shortened words used in Roman Urdu communication on Facebook?

RQ3: To what extent do Facebook users employ Urdu-English code-mixing and different sentence structures in Roman Urdu comments?

RQ4: How does the use of Roman Urdu affect the comprehensibility and communicative effectiveness of messages on Facebook?

Literature Review

As part of the global trend, local users in south Asia especially in Pakistan the users use text messages for a variety of purposes, including maintaining social bindings, expressing of emotions, and coordinating their daily events. (Ting-Toomey & Dorjee, 2018). It has been argued that emerging users, particularly illiterate users, find it challenging to use text-based phone features (Daud, Khan & Che, 2017).

But text communications work as well – they're private and asynchronous, in some social situation more convenient. The users in Pakistan have taken an advantage of the messages and their complexity and altered the code into Roman script to convey in Urdu. This type of writing is known as Roman Urdu (Cramer, De Juan & Tetreault, 2016). As the Roman Urdu is gaining popularity in the digital communication system in Pakistan, this is a worthy field of study. Urdu is written with Roman Urdu, which is popular in informal contexts, e.g., social media, text messages and online forums, and uses Latin script. It has come to be associated strongly with technological accessibility, ease of use and computer mediated communication (CMC) changes.

According to Bilal, Rextin, Kakakhel, and Nasim (2017), a large collection of Urdu text messages from the Romans was analysed in one of the basic studies in this area. Their research is the first to show that Roman Urdu is one of the most-widely-used modes of communication among Pakistani users, particularly among those who are not confident in English or lack convenient access to Urdu script keyboards. The study highlights that Roman Urdu is more user-friendly than traditional

Nastaliq script and is more suited to the digital devices, thus making it a viable alternative.

An important feature of the Roman Urdu is that it lacks a standard orthography, unlike formal written languages. Khan et al. (2020) noted that the Roman Urdu has a wide range of spelling variations, with users using phonetic approximations instead of rules. This means that the word can be spelt many ways depending on the individual perception, regional pronunciation, and typing style. This variability can cause problems for computational processing, but also illuminates the flexible and user driven nature of the language.

Roman Urdu is an important concept for comprehension, which is called phonetic spelling. Digital communication is more about speed and efficiency than language; users tend to prefer being understood over being linguistically correct. This mirrors the trend in CMC with the change of language in response to the need for quick interaction (Crystal, 2011). In this context Roman Urdu is a phonetic system; a system of writing the words according to their sound, not how they are normally spelt.

There is another significant aspect of Roman Urdu communication and that is code-mixing, where both English and Urdu are mixed in a single sentence or discourse. Code-mixing is a common phenomenon in multilingual societies like Pakistan as it reflects cultural and social phenomenon. Research shows that the language a user switches to or from is context, audience and communicative purpose dependent (Androutsopoulos, 2015). Roman Urdu is therefore a hybrid medium which enables the easy incorporation of English words in the Urdu conversation.

The status of Roman Urdu as an instrument of social identity and interpersonal relations has also been extensively discussed. The context of digital communication can affect the choice of language, including due to intimacy, social distance, and group identity. Bilal et al. (2017) propose that Roman Urdu helps people to preserve their cultural identity while communicating in a contemporary fashion, using technology. This is an example of the sociolinguistic value of Roman Urdu as a language of communication and identity.

Roman Urdu is also very useful in the expression of emotions and attitudes apart from its use in the expression of identity. Because digital communication doesn't have the non-verbal elements

that make up tone, facial expressions, or gestures, users find a way to communicate their emotions instead. These can be character repetitions, informal spellings and variations in style. The features that are found in online communication are not uncommon, as users find their own ways to write language to convey the meaning they can convey in speech (Crystal, 2011). Urdu is very flexible in its structure, and makes a lot of expressive practices possible, in particular, in the context of Roman Urdu.

On the technological front, Roman Urdu has become a subject of interest in the domain of natural language processing (NLP) and machine learning. It has been studied for sentiment analysis, opinion mining and text classification. For example, Qutab et al. (2020) have done comparative study on sentiment analysis for Roman Urdu text, which reveals that the problems of variation in spelling and lack of standardisation in Urdu language are the problems involved. Likewise, Khwaja et al. (2020) explored the usage of machine learning methods for sentiment analysis of Roman Urdu, highlighting the necessity of having specialised datasets and models.

With the recent developments in AI, the research on Roman Urdu has also grown in importance, especially in the field of hate speech detection and content moderation. The study by Ashiq et al. (2024) highlights the need for identifying harmful content in Roman Urdu, particularly as it has become increasingly significant in online communication. The findings of these studies highlight the importance of Roman Urdu in the present digital context and the need for further research to tackle its challenges.

Educational technology can also be a factor explaining the prevalence of Roman Urdu. Historically, the Urdu script has been hard to learn, and there have been few user-friendly keyboards available. Consequently, Urdu in Roman script is used as an alternative, which is easier for the users. This is reminiscent of the notion of technological determinism, which is a view that communication technologies influence language use (Thurlow, Lengel, & Tomic, 2004).

Moreover, Roman Urdu is an emergent language behaviour which is developed by the users of the language itself in terms of the norms of communication, given the social and technological changes. However, Minhas and Salawu (2025) suggest that Urdu, as spoken by the Romans, is an

adaptive system that is continually in a state of evolution, due to the interaction of the user with the language. This view focuses on the user as an agent in the process of shaping linguistic practices.

The use of Roman Urdu has its merits but the expansion of this usage has raised concern about its effect on the Urdu language and script. The control of Urdu language by the Roman script may result in the reduction in the use of Nastaliq Urdu script, especially among the younger generation, some scholars believe. This may affect literacy, cultural heritage and the preservation of language. Some other researchers, however, propose that Roman Urdu should be considered as a system which complementizes the standard Urdu system.

A functional differentiation concept can be helpful to understand this relationship. The standard Urdu is used predominantly in formal communication, education and literature, while the Roman Urdu is also used in formal communication, but mainly in informal communication. The language is flexible to meet changing communicative needs in this co-existence. However, Minhas and Salawu (2025) has claimed digital communication frequently results in diversification of language and not its demise.

In the light of globalization, the development of Roman Urdu can be viewed as a part of a larger trend of using the Latin script for non-English languages. English as a global language has had an impact on the use of Romanized forms in many linguistic communities. Roman Urdu, as such, is a localised version of global communication practices, with a mixture of Urdu and English.

While a large number of books have been written there are some gaps in research. Previous research tends to be limited to text messaging or computational analysis, rather than social media platforms and differences in content. The lack of examination of the use of Urdu among the Romans in different situations, e.g., political, social and entertainment, has created a need for more research in this area. Furthermore, more research is needed to investigate the sociolinguistic aspect of Roman Urdu that is its influence on the identity, education and language policy.

There is no standardised spelling of Roman Urdu and sometimes individuals use different spellings for the same word (Döring, Hellwig & Klimsa, 2005). Text message communication is increasing in popularity and has thus attracted the interest of researchers for study. Some studies show that text messages can be segmented into two different types. In other studies, the structure and function of informational messages and relational messages, sent by users with different age, gender, experience, etc. are compared. Studies are even available that investigate the difference between the language(s) used on Facebook, WhatsApp, etc. and the standard language(s) (Bilal, et al., 2017) and studies that investigate the functions of emojis in text-based one-to-one communication (Bilal, et al., 2017).

Lots of research on European language text messages is available for analysis of misspellings, text prediction, spam detection (Batinca, & Treleaven, 2015). While developing countries are estimated to have a large population of mobile phone users, little research has been carried out on this population. For example, India has more mobile phone subscribers than any other country. Its neighbouring country Pakistan is in the eighth position in the world list (Donner, 2008).

Several studies indicate that the Roman Urdu text messages have some unique features. There are only few empirical studies that have analyzed the language of text messages from Pakistan. (Bilal, Rextin, Kakakhel, & Nasim, 2018).

According to (Safdar, Et. al., 2020) Roman Urdu is mostly used by those who are not skilled in English language, therefore their needs are met with Roman Urdu.

However, the true Urdu has been rapidly appropriated by the Roman Urdu and this change is observed in various forms and with varying accentuation by (Atiq, 2019).

Methodology

Research Design

The mixed method content analysis methodology is used in this study, that is combined both quantitative and qualitative approach for studying the pattern of use of Roman Urdu in the social

media. Content Analysis technique in research is a systematic method in analysing the content of communication in which the content of communication is classified and interpreted based on the text information. In the quantitative aspect of the study, the frequency of the specific linguistic features of Roman Urdu and their occurrence were measured while the qualitative aspect involved examination of the patterns and meanings of the use of these Roman Urdu linguistic features in online communication.

Population

This study comprised of all Facebook posts and comments on major political, social and international issues that were discussed between the 1st October 2022 and 31st October 2022. These were the subject of the ousting of the PTI government, elections, Pakistan-USA relations, ICC T20 World Cup and other important events of the country. The sample consisted of the user generated contents posted on Facebook between the given time frame and written in Roman Urdu language.

Sampling Technique

The purposive sampling technique was used in selecting the data to be analyzed. The researcher opted for this non-probability sampling technique because he wanted to be able to identify and select the posts on Facebook that are directly related to the issue and events. To allow analysis of the content of the selected posts, only a few posts which were highly relevant to the subject and all comments were selected. The sample was determined appropriate to study Urdu usage patterns in the context of the online discourses about the important event of contemporary times in the Roman period.

Unit of Analysis

Single comment in Roman Urdu on Facebook was the unit of analysis in this study. In addition, the texts of Facebook posts were selected that were relevant to the context where the Facebook comments occurred to discuss the texts. All comments were analysed based on certain pre-defined categories such as numeric homophones, repetition of characters, spelling consistency, frequency of comments, Roman Urdu writing style, usage of short words, sentence structure, combined usage of English & Roman Urdu, and message comprehensibility. These categories were used to identify

the linguistic pattern and interpretation of Roman Urdu communication in social media was carried out according to this pattern.

Coding Scheme

A coding scheme was created to analyze the use of Roman Urdu in Facebook comments, systematically. The coding categories were developed through a review of literature on the topic of computer mediated communication, and preliminary analysis of the data gathered. They were homophones, repetition of characters, consistency of spelling, comments referring to each other, Urdu writing patterns in Roman script, short words, structure of sentences, code mixing (Urdu and English) and understanding of the messages. To ensure consistency of analysis all comments were analysed and categorised into the following pre-determined categories.

Coding Procedures

All the Facebook comments were manually reviewed and coded. One or more coding categories were assigned to each comment and all comments were read and coded based on the linguistic features of the comments. A coding sheet has been created to tally the number and percentage of each category. To facilitate coding process, the patterns of the Roman Urdu were identified which were repetitive in nature and to assist in qualitative interpretation and quantitative analysis of the data.

Reliability and Validity

The coding categories were well-defined before the complete analysis was conducted for a more reliable coding. The coding scheme was pilot coded using a small sample of Facebook comments to test the clarity and applicability of the coding scheme. Some minor changes were made in the definitions of categories and some edits were made to reduce ambiguity, as a result of the pilot test. The consistency of these categories with the study objective and literature survey on Roman Urdu and online communication was used to check their validity. Also, the categories were constantly revisited throughout the coding process, making sure that they represent the phenomenon of interest in the linguistic domain.

Inter-Coder Reliability

No inter-coder reliability testing was conducted as it was a study of one researcher's experience. The coding categories were well defined and the coding procedures were systematic, therefore the coding was consistent. In future research, reliability of coding could be further increased by using several coders and by ensuring it by using statistical analysis such as Cohen's Kappa.

Results

Based on the results of the numeric homophone analysis, it is observed that the Roman Urdu users start using numbers as a “shortcut” and this is observed at different levels in the various forms. The most common substitution is “1” for “one” (35.7%) which has a high degree of congruence with the universally accepted digital shorthand in the internet, suggesting that Roman Urdu users are influenced by the general norms of internet communication.

However, the “7” in “together” (31.0%) is a sign of local phonetic adaption, and illustrates how users can link between the pronunciation of Urdu and local context of Arabic, by showing creative ways in which the Urdu pronunciation is adapted. The percentages of the "13" (19.0%) and "2" (14.3%) are rather low, suggesting that the forms are probably not standardized and will need more skill to utilize in context.

In sum, the results reflect a hybrid language usage, in which an Urdu language uses Roman Urdu in conjunction with local phonetic innovation. This suggests that numeric homophones are not only efficiency devices but also represent the emerging digital identity and informal communication patterns in the social media scene in Pakistan.

Numeric Form	Intended Meaning	Frequency (f)	Percentage (%)	Usage Pattern
1	One	15	35.7%	Standard digital shorthand; widely recognized
7	Together	13	31.0%	Phonetic adaptation influenced by Urdu pronunciation
13	Yours	8	19.0%	Creative, less standardized

Numeric Form	Intended Meaning	Frequency (f)	Percentage (%)	Usage Pattern
2	Give	6	14.3%	substitution Limited usage; context-dependent
Total	—	42	100%	—

The repeated use of letters (Orthographic elongation): From the analysis of the data, it was noted that Orthographic elongation as a repeated use of letters is one of the important stylistic features in the Roman Urdu communication in social media. The most common elongation is called “Yaarrrr” (33.3%) which is used for emphasizing elongation for purposes of emotional intimacy and casual bonding in oral communication where it is thought that tone is not perceived in the digital form.

Similarly, elongation is used to show emphasis and intensity, e.g. “Haaaann” (24.1%) and “Na karrrr” (23.4%) to signify agreement, or insistence, particularly in the expression of agreement or insistence. The other elongated words are not as common as “Nahiiiiii” (11.3%) and “Afsoossssss” (7.1%) words used for expressing negative emotions which are used less in the friendly/Persuasive context.

It is overall concluded that the elongation of orthographic forms in Roman Urdu is not random, rather it is a paralinguistic feature which mimics the speech-like features like tone, stress and emotional features. This illustrates how the written language can be flexibly adapted in the computer mediated communication context, to achieve the desired richness of expression.

Word Form	Standard Form	Frequency (f)	Percentage (%)	Functional Use
Yaarrrr	Yaar	47	33.3%	Emotional emphasis / intimacy
Haaaannnn	Haan	34	24.1%	Agreement with emphasis

Word Form	Standard Form	Frequency (f)	Percentage (%)	Functional Use
Na karrrr	Na kar	33	23.4%	(request/insistence) with intensity
Nahiiii	Nahi	16	11.3%	Strong negation
Afsoosssss	Afsoos	10	7.1%	Expression of regret / sympathy
Total	—	140	100%	—

Spelling consistency: In Roman Urdu there are no consistent spelling rules given so there is a possibility of different spellings for the same word. Nearly half of our participants indicated they use consistent spellings and nearly the other half indicated that they do not use consistent spellings while. The data shows that, 33% of the comments had consistent words usage while the remaining had used different words with the same meaning.

Spelling Variations: The results show large orthographic variations in the spelling of Roman Urdu which can be attributed mainly to the lack of standardisation in the spelling rules. The most common variation is in the representation of the vowel “Zaleel/Zalil/Zalel” (41.8%), where users indiscriminately mix the different isochores without the risk of affecting each other's understanding. This implies that it's more important to approximate than to be correct.

Likewise, the different forms of “Kiya/Kia/Kiyea” (27.0%) and “Batein/Batay/Baateain” (18.0%), show variation in the representation of long vowels and plural forms, indicating regional pronunciation and typing preferences. There is partial stabilization of some commonly used words, with the relatively lower frequency of “Ziada/Zyada/Zyaada,” which indicates a partial stabilization of these words.

In summary, the data has illustrated that Roman Urdu is a flexible and personal writing system, which has been influenced by the phonetics, convenience and individual preferences. This

flexibility makes Urdu communication easier, but at the same time indicates the lack of uniformity in digital communication in Urdu.

Word Variants	Standard Form	Frequency (f)	Percentage (%)	Linguistic Feature
Kiya / Kia / Kiyea	Kiya	33	27.0%	Vowel variation / phonetic spelling
Bateain / Batay / Baateain	Baatain	22	18.0%	Morphological & vowel inconsistency
Ziada / Zyada / Zyaada	Zyada	16	13.1%	Orthographic simplification
Zaleel / Zalil / Zalel	Zaleel	51	41.8%	Consonant-vowel variation
Total	—	122	100%	—

Slang Words on Different Posts in Roman Urdu

Content Category	Slang Frequency (f)	Total Posts (N)	Percentage (%)	Relative Usage Level
Political	87	556	15.6%	Moderate
Sports	53	700	7.6%	Low
Crime	67	482	13.9%	Moderate
Miscellaneous	204	1100	18.5%	High
Total	411	2838	14.5%	—

The slang distribution by content categories shows distinct differences in verbal attitudes based on type of content. Miscellaneous posts (18.5%) contain the highest percentage of slang, indicating that in non-formal and personal posts, linguistic creativity and relaxed communication styles seem to be preferred.

Political posts (15.6%) also have a relatively high rate of slang use, suggesting that users use informal language to convey opinions, criticism, or emotional reactions to political language.

Likewise, crime-related posts (13.9%) demonstrate moderate slang, which may be a sign of expressive communication – shock, anger or sarcasm.

Sports posts (7.6%), however, have the lowest usage of slang, which may indicate a comparatively more standardized or commentary-driven style, likely due to the nature of discussions on matches and teams, and performance.

Overall, the results show that slang is context sensitive in Roman Urdu, with more informal and emotional content showing more slang and more structured/informational content showing less slang. This demonstrates the importance of social media as a language context in which language selections are determined by topic and communicative purpose.

Pattern of roman Urdu

Linguistic Feature	Frequency (f)	Total (N)	Percentage (%)	Interpretation Level
Use of Short Words	96	992	9.7%	Limited
Structured Sentence Formation	888	992	89.5%	High
Code-Mixing (English + Roman Urdu)	399	992	40.2%	Moderate
Message Comprehensibility	992	992	100%	Very High

Examining the usage patterns of Roman Urdu shows that in digital communication clarity and structural coherence are dominant and there is no formal standardization. 89.5% of the comments contained a recognisable sentence structure and thus users' grammar was maintained even in the informal situation.

There is not a great deal of use of short words (9.7%) either which suggests that the reliance on abbreviation in Roman Urdu communication is not a dominant characteristic in Roman Urdu communication in this dataset. In contrast, code-mixing (40.2%) appears as a salient pattern which reflects the mixed language setting in Pakistani social media, where English and Roman Urdu

language are present together and have become a part of their lives.

Message comprehensibility is very high (100%), indicating that no differences in spelling, structure, language mixing, etc. inhibits comprehension. This further supports the concept that Roman Urdu is an adaptive and efficient method of communication which emphasises content over form.

The results were not entirely uniform; however, it can be concluded that Roman Urdu communication is not completely formal; it is semi-structured communication in which most of the users are able to express themselves clearly and use sentence structure, but they use informal elements like code-mixing and, sometimes, abbreviations selectively. This balance brings expressiveness and mutual intelligibility in computer-mediated communication.

Discussion

This study has tried to investigate the form and pattern of Roman Urdu in various kind of social media (SM) content with a special focus on Urdu use in user-generated commentaries (UGC). Linguistic features were categorised numerically using a content analysis approach, with spelling variations, repetition of characters (orthographic elongation), short word usage, sentence structure, numeric homophones, short word usage and overall message comprehensibility.

The results indicate that the Roman Urdu is a flexible and adaptive language system that is mostly user convenient, phonetically interpreted and digital communication influenced. The use of numeric homophones is one of the significant characteristics of the data – that is, when numerals take the place of sound-alike words. For example, the word ‘one’ was written out 15 times as 1 and 13 times as together. Together- 13 times, one- 15 times. Although these are not the dominant forms of the genre in all the comments, the forms found indicate the influence of global digital shorthand usage on Roman Urdu.

Another noticeable feature noted is also the repetition of the character or orthographic elongation, which is then expressed as a paralinguistic means to indicate the intensity of the emotions. There are terms like ‘yaarrrr’ (47 times used), ‘haaannnn’ (34 times) and ‘na karrrr’ (33 times) that

showcase how users cope with the deficit in voice tone in written communication. These forms simulate the flow of speech and the user has the ability to emphasis, intimacy or urgency in delivery. Other less familiar, are “nahiiii” (16) which represents elongation and positive emotions, and “afsoosssss” (10) which represents elongation and negative emotions.

The analysis further reveals that there is no standardisation of spelling in Roman Urdu where there are variations in spelling. The variants are "Kiya/Kia/Kiyea" (33 times), "Bateain/Batay/Baateain" (22 times), "Ziada/Zyada/Zzaida" (16 times) and "Zaleel/Zalil/Zalel" (51 times, all of which are examples of phonetic approximations rather than spellings that follow fixed rules. There was about a 33 per cent consistency rate in the percentage of comments that were spelt correctly. There are a number of variations in the spelling of many of the comments. This indicates that the orthographic uniformity of Roman Urdu has been made available and more emphasis has been given to the phonological representation and typing.

Language informality in terms of slang use shows that there are variations in content categories. Miscellaneous (204/1100), political (87/556), crime (67/482) and sports (53/700) content were the most frequently occurring categories of the slang. This difference would indicate that context is important to language use and informal emotional conversations would influence the frequency of slang use, whereas more structured information-oriented conversations such as sports would have an opposite effect.

The study also explored some broader trends that emerged with respect to how the Romans used Urdu: it was found that sentence structure did not change much, as 888 comments out of 992 had coherent construction. It means that the users' underlying grammar structure remains almost similar to the standard Urdu, even though the spelling and stylistic variations are not standard. Moreover, a code-mixing phenomenon between English and Roman Urdu was noticed in 399 comments, showcasing the bilingual nature of digital communication in Pakistan. The low percentage of short words (96/992) indicates that there is not a high number of short words as is believed in the realm of online language.

Perhaps the most important finding is that none of the analyzed comments (992/992) were

unintelligible, indicating that differences in spelling, structure and mixing language do not impede communication. This helps to strengthen the view that Roman Urdu is efficient and functional medium of communication as it can be used effectively to communicate even in an informal and non-standardized environment.

The overall results are in line with the previous studies (e.g. Ansari et al., 2020) that argued that Roman Urdu has irregular but systematic structure, which is determined by user practices in the context of formal language rules. The language seems to be a mix of the local and the global, as it incorporates aspects of both local languages and global digital conventions.

Conclusion

The study is concluded that "Roman Urdu" is a prevailing and accepted method of communication amongst the social media users of Pakistan. It is consistently used in a variety of settings, political, social and informal, for the effective and efficient communication of messages. Though not standardized, Roman Urdu has a high degree of clarity and comprehensibility, partly because the sentence structure and basic meaning is preserved.

The findings prove that Roman Urdu is not a "sloppy" or solely a defective version of Urdu but a living and growing linguistic system. It involves a variety of phonetic spellings, code-mixing, slang and stylistic aspects of writing including character elongation, in order to meet the communicative goals of digital interaction. These features allow users to convey emotion, stress, and to participate in conversational discourse in a manner similar to oral communication.

Meanwhile, the growing trend in the use of Roman Urdu also has some significance. The results established that the usage of non-standard forms may be attributed to preference for Roman script and it can be speculated that this is a contributing factor to the diminishing usage of standard Urdu script (Nastaliq) in online spaces. This change could impact linguistic preservation, literacy and cultural identity in a nation with Urdu as a national language.

The current study is in consistency with the previous one (Bilal, Rextin, Kakakhel, & Nasim, 2017), which has concluded that the Roman Urdu language is well rooted in the digital communication. It can, however, be considered as a complementary linguistic mode and means of

technological adaptation and change in the Urdu sociolinguistic context, not only as a threat to standard Urdu.

The long-term effect of Roman Urdu on language development, literacy and identity formation can be further investigated in future research and differences between platforms, age and education level can also be explored.

Limitations of the Study

This study has some limitations, which are discussed here, the major of which is that this study has only examined one set of data on the pattern of Roman Urdu. The results provide some interesting elements related to the current use of digital communication, however it is not possible to capture the use of Urdu in various regions, age and platform. In addition, since Urdu is not a standard language which is written in a standard way, some variations in spelling of the same word exist and this affects the uniformity and confidence of linguistic analysis. Besides, Urdu code mixing with English is also common and the identification of the pattern of language in Urdu is challenging. Additionally, the study focused on a particular time and context of use and online language use can be in constant change. Therefore, it is important to consider the results within the context and data.

Recommendations for Future Research

Further studies are required to collect more data from bigger and more varied Roman Urdu corpus from different social media and communication types. Standardisation of Roman Urdu should also be done with the purpose of representing the variations in Urdu spelling and improve analytical precision. Moreover, other research could focus on sociolinguistic issues impacting the utilization of Roman Urdu, such as age, schooling, gender, geographical area, and so on. State-of-the-art machine learning and natural language processing may provide new insights into how language patterns and language development. Secondly, comparison between Roman Urdu and Urdu script would help in understanding the adaptation of the language in the digital communication and help in developing the language technology of Urdu in low resource languages.

References

- Ansari, Z., Ali, S., & Khan, F. (2020). Use of Roman script for writing Urdu language. *International Journal of Linguistics and Culture*, 1(2), 165-178.
- Ashiq, W., Kanwal, S., Rafique, A., et al. (2024). Roman Urdu hate speech detection using hybrid machine learning models. *Scientific Reports*, 14, 1–15.
- Atiq, S. (2019). The Roman Conquest of Urdu in Pakistani Movie Posters: A Study of the Cannibalistic Politics of Transliteration. *Pakistan Journal of Languages and Translation Studies*.
- Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, 30(1), 89-116.
- Bilal, A., Rextin, A., Kakakhel, A., & Nasim, M. (2017). Roman-txt: forms and functions of roman urdu texting. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-9).
- Bilal, A., Rextin, A., Kakakhel, A., & Nasim, M. (2017). Roman-txt: Forms and functions of Roman Urdu texting. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*.
- Bilal, A., Rextin, A., Kakakhel, A., & Nasim, M. (2018). Analysing emergent users' text messages data and exploring its benefits. *arXiv preprint arXiv:1811.11983*.
- Bilal, A., Rextin, A., Kakakhel, A., & Nasim, M. (2018). Analyzing emergent users' text messages data and exploring its benefits. *IEEE Access*, 7, 2870-2879.
- Cramer, H., De Juan, P., & Tetreault, J. (2016, September). Sender-intended functions of emojis in US messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 504-509).
- Crystal, D. (2011). *Internet linguistics: A student guide*. Routledge.
- Daud, A., Khan, W., & Che, D. (2017). Urdu language processing: a survey. *Artificial Intelligence Review*, 47(3), 279-311.
- Donner, J. (2008). Research approaches to mobile use in the developing world: A review of the literature. *The information society*, 24(3), 140-159.
- Doring, N., Hellwig, K., & Klimsa, P. (2005). Mobile communication among youth in

- Germany. *Wien: A Sense of Place: The Global and the Local in Mobile Communication*, ed. K. Nyiri (Vienna, 2005), 209-20.
- Forman, J., & Damschroder, L. (2007). Qualitative content analysis. In *Empirical methods for bioethics: A primer* (pp. 39-62). Emerald Group Publishing Limited.
- Khan, A. R., Karim, A., Sajjad, H., Kamiran, F., & Xu, J. (2020). A clustering framework for lexical normalization of Roman Urdu. *arXiv preprint arXiv:2004.00088*.
- Khwaja, S. A., Khwaja, S. U., Mughal, B., Mushtaq, F., & Zahra, S. (2020). Roman Urdu sentiment analysis using machine learning. *Pakistan Journal of Engineering and Technology*, 3(2), 45–52.
- Minhas, S., & Salawu, A. (2025). Preserving and promoting indigenous languages: social media analysis of Punjabi and Setswana languages. *Journal of Asian and African Studies*, 60(7), 4109-4126.
- Minhas, S., & Salawu, A. (2025). Indigenous Languages Activism on Social Media: A Comparative Analysis of Setswana and Punjabi Activism. *Journal of African Languages & Literary Studies*, 6(1), 57.
- Moazam, F., & Shekhani, S. (2018). Why women go to medical college but fail to practise medicine: perspectives from the Islamic Republic of Pakistan. *Medical education*, 52(7), 705-715.
- Mushtaque, S., & Zaki, S. (2019). Language identity of tertiary ESL learners: Understanding Urdu and English language identities. *International Research Journal of Arts and Humanities*, 47(47), 19.
- Qutab, I., Malik, K. I., & Arooj, H. (2020). Sentiment analysis for Roman Urdu text over social media: A comparative study. *arXiv preprint arXiv:2010.16408*.
- Safdar, Z., Bajwa, R. S., Hussain, S., Abdullah, H. B., Safdar, K., & Draz, U. (2020). The role of Roman Urdu in multilingual information retrieval: A regional study. *The Journal of Academic Librarianship*, 46(6), 102258.
- Shah, M., Riaz, W., & Malik, S. (2019). AN EXPLORATION OF THE FUNCTIONS OF BI-MULTILINGUAL PRACTICES INSIDE AND OUTSIDE CLASSROOM. *ABOUT AUTHORS*, 57.
- Soomro, T. R., & Ghulam, S. M. (2019). Current status of urdu on Twitter. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 3(1), 28-33.

Thurlow, C., Lengel, L., & Tomic, A. (2004). *Computer mediated communication: Social interaction and the Internet*. Sage Publications.

Ting-Toomey, S., & Dorjee, T. (2018). *Communicating across cultures*. Guilford Publications.