Code-Switching By Multilingual Pakistanis On Twitter: A Qualitative Analysis

[1,*]Javeria Jamali; [2]Mishal Rasool; [3]Huma Batool

[1,2,3]Air University, Pakistan

[*]Corresponding author: javeriajamali786@gmail.com

Abstract

Code-switching is practicing two different grammatical systems where multilinguals also move between two languages or between two dialects or registers of the same language. In the current article, code-switching is characterized as the simultaneous use of two or more languages or dialects within a conversation. The current study focused on code-switching practices on the social media website Twitter. While posting on Twitter, multilinguals may use several languages. The aim of this study was to describe and analyze code-switched Tweets for any recurring patterns and practices. The population of this study involved Twitter users living in the Rawalpindi-Islamabad area of Pakistan. The Tweets were collected on the basis of time and location through random cluster sampling method. Qualitative analysis of the individual Tweets was done, and recurring patterns were pointed out. This was purely observational research. It was found that the sampled Tweets only code-switched between Urdu and English. Code-switching at the intra-sentential level was more common than the inter-sentential level. Code-switching at the level of clauses was the most common form of intra-sentential code-switching. Over half of the inter-sentential code-switching had the English sentence(s) preceding the Urdu sentence(s). The findings suggest that code-switching between English and Urdu occurs more commonly at the intra-sentential level. They further imply that the population generally prefers to start inter-sentential code-switching with English before code-switching to Urdu. The results of this study may be useful in demystifying the phenomenon of code-switching in online spaces.

Keywords: Code-switching, Intra-sentential code-switching, Inter-sentential codeswitching, Twitter, Pakistan.

Introduction

The practice of switching between two or more languages or dialects is known as code-switching (Nordquist, 2019). It is a branch of sociolinguistics, and a common phenomenon in non-native English-speaking countries such as Pakistan. This

research article identifies code-switching in Pakistani multilingual Twitter users and analyses the levels at which this code-switching occurs.

The widespread usage of social networks such as Twitter in recent years has resulted in a unique development of languages through code-mixing, which is still not extensively studied in certain contexts. A gap of research is particularly noticed in the case of the Pakistani population and their habits of code-switching on social networks such as Twitter, despite a significant portion of the population regularly using them for communication purposes.

This study attempts to find recurring trends in the code-switching employed on Twitter by people living in specific areas of Pakistan, to shed further light on this gap in research. The code-switching habits of multilingual Pakistanis in an informal context may aid in comparisons with code-switching in other contexts or lead to similar studies with other demographics within Pakistan.

The research objectives are as follows:

a) To determine and describe any occurrences of code-switching among Pakistani Twitter users

b) To find the grammatical and syntactical levels at which the codeswitching takes place

Literature Review

Despite the relatively young age of online platforms, there has been significant research on its effects. Still, as modern technology evolves rapidly, and online cultures along with it, further research of the Internet's effects on language and culture should be conducted. This research aims to contribute in this regard by delving into the use of language and code-switching on online platforms. To achieve this, the current section first describes the relationship between language and social media. Second, it touches on code-switching in general. To conclude, codeswitching in the light of online platforms is discussed.

Language and Social Media

Kamran and Mansoor (2017), using both descriptive and statistical analysis, found that English was the dominantly used language by Pakistani students on Facebook, and the use of English in specific cases indicated the disparity between students of public and private sectors. Ghulam and Soomro's (2019) research, utilizing sentimental analysis, also indicated a similar preference for English by Pakistani

students on Twitter. Language preferences in online spaces can occur for a variety of reasons, and may be indicative of the disparity between students of various backgrounds (Kamran & Mansoor, 2017). The factors resulting in this particular preference of English in Pakistani students have not been discussed extensively; however, some possible factors have been suggested by various researchers.

This preference of English among multilingual Pakistani students may be a result of the lack of technological support for low-resource languages. Urdu is not a supported language script for various technological devices and social media sites, resulting in limited Urdu sentiment analysis in research, as indicated by Ghulam and Soomro (2019).

Alternatively, Abbas et al. (2019) found through an exhaustive examination of previous texts along with a survey using the Likert scale that, among other effects of social media on sustainable education, social media platforms have also "blurred the difference between formal and informal writing" for Pakistani students, resulting in students using online slang in their formal assignments (p. 16). With the greater accessibility of the internet and the popularity of various English-language-based social media sites, students who are active online may not feel the same sense of distance with English as their peers who are not active online, due to their greater consumption of content in English.

Code-Switching

Various definitions of code-switching have been suggested. According to Morrison (n.d.), code-switching can be defined as the "process of shifting from one linguistic code (a language or dialect) to another, depending on the social context or conversational setting" (para. 1). Gumperz (1982) states that code-switching is "the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems" (p.59). Nordquist (2019) defines the phenomenon as "the practice of moving back and forth between two languages or between two dialects or registers of the same language at one time" (para. 1). Code-switching is thus characterized by the simultaneous use of two or more languages or dialects within a conversation in the current article.

Code-switching can be used to perform a variety of functions. According to MyersScotton (1993), code-switching may occur due to the following four reasons: "(i) a lack of knowledge of one language or a lack of facility in it, (ii) the use of another language to exclude some persons from an interaction, (iii) switching into another language in order to introduce a new subject, and (iv) to speak in a different language to impress others" (p.73). The last two functions of code-switching are particularly relevant in the context of online social media platforms. Forming

significant connections with a person in a situation where one can neither see nor hear the other, with written messages being the only form of communication, is quite difficult. People may thus utilize code-switching to create an informal and comfortable social context, and consequently form significant personal connections more easily. Further research on code-switching as a tool for human well-being by transcending the linguistic barrier was conducted by KhudaBukhsh et al. (2020), who utilized a systematic approach to sample code mixed documents, with a polyglot embedding based method that requires minimal moderation.

Despite its various purposes, code-switching has had a negative connotation for society at large. It is discouraged in many formal situations, particularly in academic and professional contexts. Spooner (2017) delves into the convention of teaching only in the target language in academia, and the reasoning behind such a rule, by triangulating the knowledge base of composition theory with that of L2 writing instruction and with the growing literature on translingualism. Spooner describes the limitations of such a singular approach and suggests possible code-switching techniques which can be incorporated in classrooms effectively. The article forms an effective basis upon which academic syllabi can be restructured, to harness codeswitching as a tool in the language learning process.

The purposes of code-switching can be further researched if new datasets can be utilized. Hussain and Arshad (2021) presented an evaluation of various models combined with a Recurrent Neural Network, as well as a new dataset of Roman Urdu and English. They found that using the Attention Mechanism Model provides accuracy and precision. However, the dataset of the low-resource language, Roman Urdu, was overwhelmed with English language data, which is a well-researched and high-resource language. This can be combated by collecting data from platforms with a primary Urdu-speaking population, or by considering the conversational data between people who communicate in Urdu with each other, as may be the case of native Urdu-speaking families. Alternatively, a modern Deep Neural Model, using datasets based on theme, was used by Bansal et al. (2020) to improve upon the accuracy achieved by other models. The research presented a set of nine features, which could be further improved by the use of switching features in the final layer of the Deep Network.

A study of the correlations between code-switching with gender and pragmatics in an Anglophone context was conducted by Arora (2019). The data collection was through group elicitation and stimulated recall conducted in an informal Anglophone context, and was based on three broad research questions on gender, CS patterns and pragmatics. The findings indicated that gender may impact the pragmatic markers, but does not seem involved in code-switching patterns. As the sample size for the study was quite small, the findings are not conclusive. Further analysis can be

conducted with a larger sample size, or by considering other demographic variables, to reach a more complete understanding of the correlations between the mentioned variables.

Code-Switching in Online Spaces

Various studies have been conducted on code-switching in online social networks; however, with the rapid development of the internet as a whole, and the rise and fall of various social media platforms, further research is always significant.

The lack of a word level language detection system is a major barrier in the analysis of online code-mixed content. To combat this, Rijhwani et al. (2017) tried to create a language detection technique to find code-switching between an arbitrary number of languages, with the help of the Hidden Markov Model and Baum Welch reestimation. Their contribution is important for future researchers as it was the first research of its kind and filled an important research gap. A Hindi-English codemixed dataset for hate speech detection, using a classification system consisting of Character N-Grams, Word N-Grams, punctuation, negation words and lexicon, was constructed by Bohra et al. (2018) for the use of future researchers. Its features of detection at character and word levels, along with other lexicon-based features, have made it significant in its relevant fields.

Samhan (2017), using a random cluster sampling method along with analytical analysis, analyzed the written content of active multilingual Twitter users to find instances of code-switching and explore code-switching patterns in light of various demographic factors. The findings indicated a potential correlation between codeswitching patterns and age or educational background, while gender did not seem to have any effect. The research was objective in its data collection by considering all Tweets made within a certain timeframe. Similar research can be further conducted to find a more conclusive result, however.

Agarwal et al. (2017) examined the code-switching preferences of multilingual Indian Twitter users when swearing online. For the detection of swear words; language labelling of words, lexicon match, lexicon match with masked character, consecutive repeated character removal and phonetic match were used. The findings suggest a strong preference for swearing in the population's dominant language, along with possible correlations between topic, gender, and language preferences while swearing. Their findings are important when discussing the reasons behind code-switching. A case study of Pakistani Facebook users was also conducted by Sultana et al. (2020) to find instances of code-switching and the possible causes behind them. SPSS was used for data collection and the data was presented in the

form of accumulated tables. Their findings suggested the reasoning behind codeswitching was a lack of proficiency in the English language.

A qualitative study was conducted by Al-Qaysi and Al-Emran (2017) to find the attitude of language learners and language teachers towards code-switching through two questionnaires. It was found that code-switching is not very common in the Gulf regions; however, most of the students and some of the educators do code-switch among different social media platforms. A very small number of language teachers participated in the survey, so the results were inconclusive. Another significant study, on the code-switching in a private social network context by Panhwar (2020), also took a qualitative approach to the research. By applying the Code-Switching Theory of Bloom and Gumperz, the researcher found many instances of codeswitching in the private chat of a newly-married couple, and pointed out various reasons that caused the code-switching in each case. The researcher could have considered the factors that led to the code-switching for further analysis. In any case, the sample being only two people, the findings are not conclusive; however, the article paves the way for future research of similar situations.

Methodology

In this section of the present research article, the main focus is on the type of research conducted, the processes included in the collection of data, the analysis of the collected data, and the reasoning behind the usage of particular methods.

Data

Qualitative data was used instead of quantitative data. The data collected was primary i.e., the data was collected directly from the public platform, not from any secondary sources. This research considers a population of Pakistani Twitter users in the Rawalpindi-Islamabad area. The sampling frame was Tweets using codeswitching between any local language and English. The random cluster sampling method was used, with only the time of the Tweet's posting as a reference framework.

Collection Methods

This research uses the cluster sampling method of data collection to gather data through direct and unobtrusive observation, without getting involved with the subjects at all. The researchers used the 'Advanced Search' feature of the Twitter website to find Tweets made in specific time frames and from within the locations under research. A designated time period of one hour on three separate days, at a different and random time for each day, was utilized to collect Tweets, in order to

gather data representative of the population. Tweets made in only one language and Tweets containing personal or identifying information were excluded from data analysis, leaving a total of thirty-five Tweets to be discussed and analyzed.

Procedure

The data was first interpreted qualitatively using individual descriptive analysis. Afterward, content analysis was conducted to find any trends permeating the data in relation to code-switching. Various nominal variables on the nature of codeswitching were considered as a reference to organize the data. This data was also represented in tabular and graphical forms where possible for ease of understanding.

The stated data collection methods were used to find unobtrusive data not affected by the knowledge of being observed. To avoid violating research ethics, the public social networking platform of Twitter was used. Twitter users agree to its Privacy Policy, which states that "Twitter is public and Tweets are immediately viewable and searchable by anyone around the world" (Twitter, n.d.) when creating an account on the platform and by continuing their use of the platform. The stated data analysis methods were used in order to gain a better understanding of the population by the complete qualitative analysis of the sampled content.

Data Analysis

The individual Tweets have been qualitatively analyzed using the setup-quotecomment or quote-comment techniques of description.

Descriptive Analysis

Tweet 1

An oft-suggested reason for code-switching is the utilization of noun words not commonly used in the main language. Similarly, short phrases or clauses more easily expressed in the secondary language can be used in addition to content in the main language, as shown below:

> 'smuggling of cigarettes k bais pakistan ko salana #77billionrevenueloss hota hy, jiska market share barh kr 40% hogaya hai even though a few NGOs have claimed it to be 9%. Prime minister of pakistan ny b iss masly ka tazkira kiya or iska solution zaroori hy' This Tweet has code-switching between the languages of Urdu and English. The participant used English nouns and noun phrases within Urdu sentence structures, along with a dependent clause in one

sentence. The participant code-switched because they wanted to convey this message to as many people as they could. This Tweet was posted to tell the people about the effects of smuggling of cigarettes.

Tweet 2

Code switching also often occurs at a sentence level. In the following example, the Twitter user expresses their thoughts using complete sentences in each language that they speak:

> 'Also when I ask my government to cancel the exams this year, I also ask them to activate strict lockdowns throughout the country. Aisa nahi hai ke bas exams cancel ho jayen or baki mulk me party chaly. I hope we all agree on this as well'

The Tweet is related to the current issue of cancellation of exams due to the ongoing Covid-19 pandemic. In this Tweet, the participant code switched between Urdu and English. Firstly, they talked about their stance on the issue in English, then elaborated on it in Urdu, and finally employed English for the conclusion.

Tweet 3

Another common instance of code-switching occurs when quoting someone. For example:

> 'Parents use 'tum bache ho' and 'kya tum bache ho?' As per their convenience'

This Tweet talks about certain common sentences used by parents in Urdu. The languages used for code-switching in this Tweet are Urdu and English. The Twitter user code-switches for the sake of quoting a sentence with full accuracy, so that the readers can have a clear understanding of the context, and subsequently relate more to the subject.

Tweet 4

Another cause of code-switching found was emphasis. To exemplify:

> 'I honestly think what having a stable team feels like, matlab win predictor keh raha hy k 63% chances LQ k hen pr dil ko tasali nhi ho rhi'

Code-switching is used at the level of clauses in this Tweet, between Urdu and English. The Twitter user first wrote in English, and then restated the idea in Urdu to emphasize its meaning.

Tweet 5

Code-switching can also be employed to ask direct questions with the intention of being rude, which may seem awkward in one's non-native language.

> 'Everyone on Twitter wanted to die till a few days ago, now they're complaining about the smog, ab masla kya hy?? Slowly he sahi tum sab ki khwahish toh poori ho rhi hy'

The user started off with English and ended their Tweet using Urdu. This Tweet addressed the issue of smog in a metropolitan city. The use of English when discussing mental health issues may be an indicator of the lack of awareness of such issues in their native tongue. However, the Twitter user asks a rhetorical question in a rude manner by using Urdu.

Tweet 6

Somewhat similar to the previous Tweet is the following case.

> 'I can literally count on one hand k sirf kitny (established) artists hain that are investing in the craft and the music industry continuously. Baki sab?? Bas koi brand ajaye, koi drama krlen, koi controversy ajaye. Aise nhi hota dost. Kuch doge toh kuch milega'

The Tweet in English and Urdu is about the music industry and the artists who invested in the industry. The Twitter user employs a vernacular phrase of English, then proceeds to elaborate further in Urdu. They employ a more familiar way of speaking (i.e., a lower formality register) in Urdu compared to English, and conclude by using a common saying of Urdu, indicating their proficiency in both languages.

Tweet 7

Another instance of code switching for elaboration, this time at a sentence level, is the following.

> 'Yeh toh nhi pata k psl kon jeety ga lekin pindi k crowd nai dil jeet lia. Amazing energy, very positive display of sportsmanship for both the teams. Although their home team lost the game, the crowd was rooting for Karachi as well. Pakistan zindabad'

This Tweet code-switches between Urdu and English. The participant uses codeswitching to explain and emphasize their reasoning for the first statement. They use colloquial expressions from both languages. In the case of an untranslatable slogan that could only be raised in Urdu, Roman Urdu was used again. This Tweet

is posted by an enthusiastic cricket lover who is giving their opinion on the cricket league going on at the time.

Tweet 8

More code-switching can be seen for various purposes in the subsequent Tweet.

> 'Congratulations lifebuoy for being PCBs official partners. Good news kyunky is tarha door door log jo bhi cricket dekhtay hon woh jaan sakenge ke hath dhona kitna zaroori hy. Another step towards healthy pakistan ab rahega pakistan gerasim se door'

It was noticed that code mixing is used in this Tweet because more than one language is being used i.e., Urdu and English. The participant gave the statement in English, then explained in Urdu its reasoning. In this Tweet, the participant is congratulating a toiletries brand for their success in partnering with a popular event in Pakistan, resulting in widespread knowledge of the importance of cleanliness in the current pandemic. The easy switching between two languages while maintaining the grammatical rules of both suggests a great proficiency of language.

Tweet 9

Another case of using nouns from one language in sentences of another is exemplified here.

> 'Ye suicide rate barhta jana hy after results bhi suicides before exams bhi its a mental torture for us'

The Twitter user began with Urdu sentences using an English noun phrase, then switched to English at the end. This Tweet highlights the issue of suicide rate and depression among the students. It is another case of employing code-switching for the discussion of a taboo topic in the local culture i.e., mental health issues.

Tweet 10

Another case of utilizing code-switching for explanatory purposes is shown below. 'More than 90% of the students are not ready for exams kyunk unhy kuch parhaya he nhi gaya tha ab exams kesy den 1 month mai na teacher hen na koi college…. Kren toh kia kren?'

The Twitter user started the Tweet in English by expressing a statement, then, to elaborate on the statement, they used Roman Urdu. The current problem of cancellation of exams is addressed in the Tweet. The user utilizes English for statistical purposes, and further emphasizes their feelings of frustration using Urdu.

Tweet 11

A case of code switching at the level of clauses is observed in the following Tweet.

> 'Political differences aside, I hope the situation in India gets controlled. Prayers with the people of India. We're in this together. Also dear pakistani public thora khayal krlo. Stay home stay safe'

A large majority of the Tweet is in English, with a small portion ("thora khayal krlo") being in Urdu. The Tweet discusses the issues of the Covid-19 pandemic in India and Pakistan. While expressing sympathy with a distant population, the user employs English. However, when advising the people of their own country, they utilize Urdu. This may be for the benefit of an Indian reading the Tweet, or it may indicate the varying levels of closeness with the use of the non-native language in contrast to the native language.

Tweet 12

Another case of utilizing code-switching for the purposes of quotation is the following Tweet.

> ''Gaza needs actions not sympathy' Yeh bolny walo ghar beth k posts kr dena or 2 4 logon ko tag krdena asaan hy pr jab baat ati hy AC waly rooms sy nikl k bahir anay ki tou phr sabhi sympathy k words use krny lag jaty'

In this Tweet, code-switching is used between Roman Urdu and English. The participant primarily used Roman Urdu, only employing English used to quote a statement from a particular group of people. Certain words of English are also used throughout, which may indicate a lack of vocabulary in the Urdu language. This Tweet is addressing the current situation in Palestine.

Tweet 13

Even in the topic of spirituality, code-switching can be utilized.

> 'When you realise the invisible hand of God is with you you never lose hope; you stand upright to the face of oppressors; your valour, your strength and your courage increases with each passing day of cruelty!! 'Zalim' is afraid of your 'bahaduri''

This Twitter user Tweeted primarily in English, and Urdu was used to emphasize the meanings of important words. Only two words of the Urdu language were used.

There can be many reasons behind it, such as a lacking vocabulary, or to emphasize on the words. This Tweet advises people not to lose hope in difficult situations.

Tweet 14

Code-switching at both the word level and the sentence level is employed here without sacrificing the grammar of either language.

> 'How do I explain Facebook who has commissioned me to cover ' The essence of Ramadan in Karachi' in a video that was due within the first ashra to be out. Yahan essence he nahi ho raha roz hangamay ho rhy hen'

Roman Urdu and English were the languages in which code-switching was observed in this Tweet. The user started with English and ended with Urdu. The participant is saying that during Ramadan, instead of respecting the holy month, new conflicts are rising. Specific religiously-significant words were used within English sentences to communicate accurately. In the last sentence, a critique of society is presented in Urdu, while using an English noun for the subject.

Tweet 15

Some cases are quite ambiguous on the reasoning of code-switching, or may utilize code-switching for multiple purposes equally.

> 'Karachi ki fast lanes mai kachre ka dabba rakhny ka idea kis genius ka hai or iski logic kya hy? Yeh practice mukhtalif towns mai prevalent hai tou yaani yeh baqaida town planning ka hissa hai. Main us mahaan shakhsiyat sy milna chahta hoon'

Two different languages are used in this Tweet. Code mixing is observed between Roman Urdu and English. The user primarily Tweeted in Roman Urdu but, due to a lacking vocabulary, they used various words of English language as well. It is observed by the tone of the user that they are criticizing the idea of putting trash cans in certain roads of Karachi.

Tweet 16

> 'If you and your friends don't have a 'yeh ham apni zindagi k sath kya kr rhy hen' moment 3 times a day, then you're doing it wrong'

Code switching is observed between Roman Urdu and English in the Tweet given above. The Tweet is in English but, to quote the sentence, the participant used Roman Urdu. The user utilizes a commonly-used online English posting format along with

a quotation in Urdu to increase relatability with their peers. The post is about an important aspect of close friendships.

Tweet 17

Another case of a commonly-used online posting format, intended for humour, is utilized here with a quotation unique to Urdu speakers, to increase relatability.

> 'Urdu is such a beautiful language like 'yaar dekhta hon' simply means 'no''

The Twitter user employed Urdu to quote something relevant to the Urdu language. The user stated a common sentence in Urdu, and gave its equivalent meaning in English, which is very contradicting to the literal meaning. The phrase 'yaar dekhta hoon' in English means 'I'll see' but here it is said to mean 'no'. The user wanted to emphasize that when people cannot find a valid reason for denying something, they use this sentence.

Tweet 18

> 'Pakistani athletes making us proud yet again. Aisee khabar parh k dil khush ho jata hy knowing k kitna talent bhara hoa hy is mulk mai, sports, music, food, drama, tv… list goes on. Bohat mubarak. Hoping to see you win more accolades. Rooting for you'

Observing this Tweet, the researchers found out that code mixing is used between English and Urdu languages. The user started with English and, throughout the post, they code-switched between the two languages, when congratulating someone's successes. *Tweet 19*

'

Please please please invest in health and education. Kab tak larogy kitna
larogy #India #pakistan #covid'

Code-switching was observed in the Tweet given above. The user started with
English, but switched to Urdu to make their plea. The Tweet discusses the effects of
the Covid-19 pandemic.

Tweet 20

'The government banning youtube randomly in midst of a pandemic is
like jab ammi bore hori hon toh bilawaja router band krden, k boht
internet istemal krlia'

The researchers found out that this is a multilingual Tweet as more than one language
is being used. The user first wrote in English, then, for the sake of exemplifying, they
used Roman Urdu. The participant used Urdu for the example so that people could
understand it more clearly. It also increases the relatability of the post to describe a
local experience in the local language.

Tweet 21

'In pakistan we don't say 'can i have some of your food' we say 'akele akele'
and I think that's beautiful'

Roman Urdu and English were used for the purpose of code mixing in this Tweet.
The poster utilized a common online post format to explain a local colloquialism to
non-native speakers. The user wanted to tell what they say in Urdu in contrast to the
English equivalent to express the unique beauty of the Urdu language.

Tweet 22

Code-switching may also be used for supplications.

'Seeing images and videos of my city drowning. So heartbreaking to
see no one owns karachi. The city that gives so much, has lost when it
needed us yet again. I pray for everybody in my city. Please stay safe.
YA ALLAH KARACHI ko is mushkil waqt sy bachayen Aameen.
#karachi rains'

'

The whole Tweet was in English but at the end the participant used Urdu. The person used Urdu for the sake of spiritual and religious prayer. Code mixing was done in Roman Urdu and English.

Tweet 23

> Exams need preparation, preparation needs notes, notes need time and ab
> time toh baccha hi nahi hy'

The participant started with English and then switched to Urdu to express their personal frustrations with the exams schedule. Code switching was done to express those negative feelings more genuinely. Simple statements can be made in any language one understands but expressing one's feelings is often in the native language.

Tweet 24

> 'These kids are so cute yr. Saray kam kharab kr k itna innocent face le kr
> aty hain k daantny ka bhi dil nhi krta'

Two languages are used in this Tweet. Roman Urdu and English were the languages. The participant gave a stance in the first part in English and then gave reasoning for it in Urdu. The participant switched from English to Urdu to emphasize the meaning of the statement that was given.

Tweet 25

> 'This one time we were reading an Urdu passage about some dead guy
> and they stated that they brought the guy's body to islamabad and the
> question was why. i confidently raised my hand and was like 'kyun ke
> usse islamabad ki scenery dikhani hogi' LIKE BRO WHO IS THAT
> DUMB'

In this Tweet, the participant used Urdu language for the purpose of referring to a past incident. The context that was said by the participant was put as is in the Tweet. The rest of the Tweet was in English. Code mixing was done in Roman Urdu and English for the purpose of quotation.

'

Tweet 26

> 'Everyone's love for Shafqat Mehmood rn [chart increasing emoji] [chart increasing emoji] (Fair decision tha yaar) khush ho jao'

The researchers observed that in this Tweet, code mixing was used between Roman Urdu and English. The participant started off with a positive statement in English, and ended with a plea in Urdu.

Tweet 27

> when ur best friend asks u)'yaar ek photo tou lelo)' what they are actually saying is 'yaar at least 221 photos lo from different angles w different poses and different backgrounds until i have that one photo)' and u gotta do it'

Observing this Tweet, the researchers found out that code-switching was used between Roman Urdu and English. The post is primarily in English but quotations of the dialogues were made in Urdu for an authentic and relatable feeling.

Tweet 28

> 'Even if you truly love the person and you're sincere, you still need some personal space... Har waqt sar pe chadhna b nahi acha lagta'

In this Tweet, the participant used two languages. In the first half of the Tweet, a statement was given, and in the second half, that statement was elaborated on by using another language's colloquialisms, to emphasize the meaning. If someone did not understand it in English, it would be easy for them to understand in Urdu. The languages used for code switching were Roman Urdu and English.

Tweet 29

> 'how do you feel when you see your so-called friends enjoying WITHOUT YOU? Mera to tarah nikal jata he! [confused face emoji]'

Observing the Tweet, it was noticed that English was used to start, but Urdu was used afterwards to put emphasis on the Tweet. The user expresses their frustration in being excluded from events by using their local language of Urdu. Code-switching was used between Roman Urdu and English.

'

Tweet 30

> 'Yaar we had no light for about 9 hours (from 0530 PM till 0230 AM)
> aur us k baad 1 ghanta light, 1 ghanta load-shedding which is still going
> on so yeah, will have breakfast now'

The participant code switched between Roman Urdu and English. The Tweet started
with English, then, in the middle, Urdu was used, and it ended with English again.
English was used to describe the time, while Urdu was used to describe the intervals
of time. There was also some code-switching at the phrasal level.

Tweet 31

> 'No one:

Me in the middle of the night: Ameer hona ha yaar business shuru karna hai koi'

In this Tweet, a commonly-used online post format was utilized with Urdu to increase its relatability. The user gave an example in Urdu to make more people understand. Code-switching was done in Roman Urdu and English.

Tweet 32

'Or thoray arsey tak css ki tayyari karni hai. Filhaal toh Gulrez Akhtar R Nait waghaira k gaaney sunta hun mein. Illegal ko Unlegal boltay hein yeh log apney gaanon mein. rip hai angrezi abhi'

Despite the Tweet primarily being in Urdu, some English words were also used. The user criticizes others by quoting their mistakes when using English. They also employ a vernacular abbreviation of English in a lighthearted and humorous manner, to make fun of those who are not proficient in English.

Tweet 33

'Serious Tweet ko bhi mazak samajhte hai yaar, i hate it here'

The Tweet was in Urdu but some common phrases of online English were used in the Tweet i.e., 'i hate it here'. A trademarked noun along with its descriptor (i.e., 'serious Tweet') is also in English as there is no direct Urdu translation. Code mixing was done between English and Roman Urdu.

Tweet 34

'YAAR SAB KUCH AIK SAATH HO RAHA HAI GIVE ME SOME TIME TO PROCESS IT [loud crying emoji]'

Code mixing was used between English and Urdu. The Tweet started with Roman Urdu and ended with English. The user expressed their frustrations by describing the situation in Urdu, then going on to make a plea in English, along with an emoticon to fully express their emotions.

Tweet 35

'bhai yaar dil sa dua ha ap ka liya Allah ap k saht da ta ka ap hamain yu hi entertain karta raho luv you brother, take care.'

In this Tweet, code mixing was done between Roman Urdu and English. The Tweet was mainly in Urdu but some English words were used due to their multiple meanings. The verb 'entertain' can be just the verb's literal meaning or it can be related to the job of an entertainer. The Twitter user also employed English for a declaration of affection. It may indicate a greater comfort in English or a discomfort of expressing affection in Urdu.

Content Analysis

The data was divided into those showing inter-sentential code-switching and those showing intra-sentential code-switching. The instances of inter-sentential codeswitching were further divided into various nominal categories representing the order of the languages the inter-sentential code-switching turned out to be employed with. The data is represented in the pie chart below.

A majority of 55.6% of the inter-sentential code-switching occurred in the structure of an English sentence preceding an Urdu sentence (English-Urdu order), while the least common inter-sentential structure was the English-Urdu-English order.

The intra-sentential instances of code-switching were also code marked according to the levels at which the code-switching occurred; particularly, the level of words, the level of phrases, and the level of clauses. The cumulative instances for each level are indicated in their respective bar charts below.

Of the twenty-six total instances of intra-sentential code-switching, fifteen evidenced code-switching at the word level, as shown.
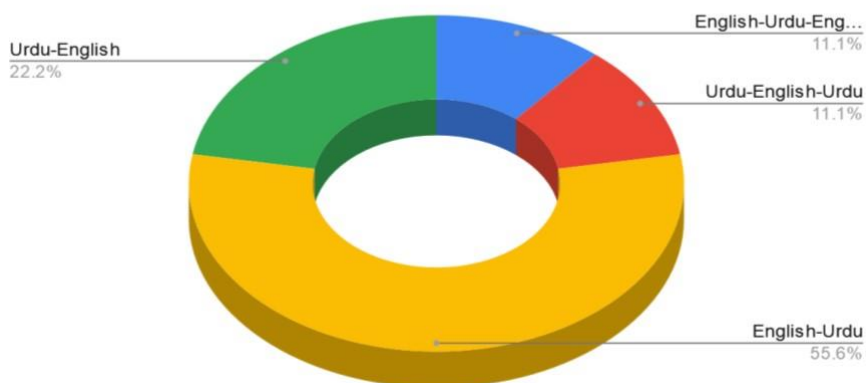
Cumulative Instances

Figure 1.1. *Total Instances of Code-Switching*
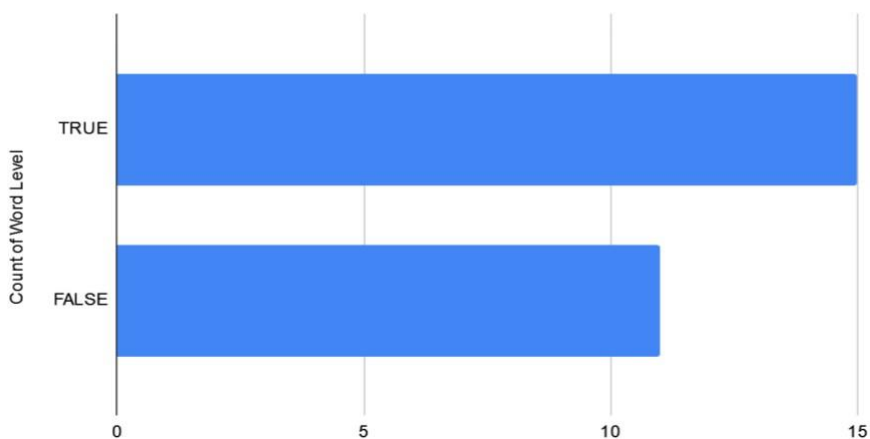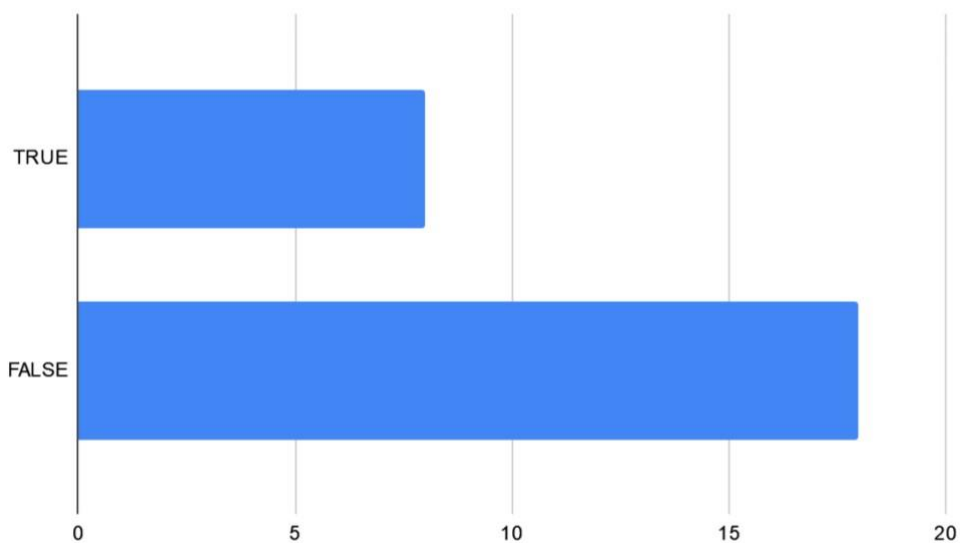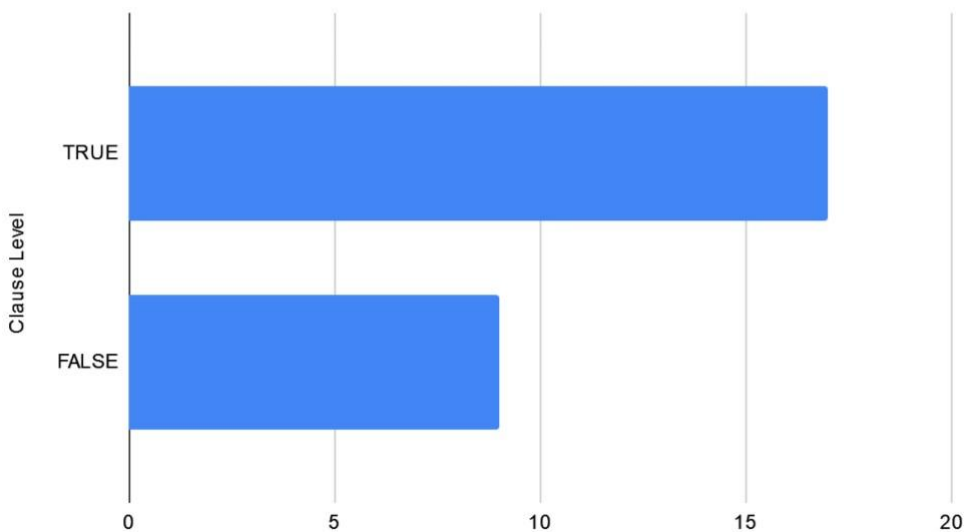
Count of Word Level

Figure 2.1. *Instances of Word Level Code-Switching*

Similarly, the instances of the phrasal level of code-switching are eight out of the twenty-six total.

Figure 2.2. *Instances of Phrase Level Code-Switching*

Figure 2.3. *Instances of Phrase Level Code-Switching*

General Discussion

The researchers observed that the main languages used by the Twitter users were English and Urdu. Tweets utilizing code-switching of regional languages did not appear in the raw data collected, indicating that it may be used less in the online platform of Twitter. Furthermore, Urdu in the form of the Romanized alphabet was widely preferred over other scripts. While several Tweets did exist with Urdu in the form of the Persian script, they did not employ code-switching and thus could not be analyzed within this research paper.

A wide variety of functions was performed by the code-switching of the collected data, implying various reasons for code-switching. Some proposed reasons that may explain a majority of the Tweets are:

a) Lacking proficiency in the main language

b) Expressions and colloquialisms which work better in a particular language

c) Ease of communication and connection with others in similar situations

d) Use of only certain formality registers of a language

These lines of reasoning have been previously discussed in the light of each individual Tweet. Furthermore, it is clear that intra-sentential code-switching is used more commonly than inter-sentential code-switching.

Conclusion

After the researchers collected and analyzed the data, they reached a number of different conclusions. It was demonstrated that code-switching is often used by Pakistani Twitter users. Some of the Tweets code-switched with a clear purpose, such as quoting a reference, while others did not show such clarity and intention. It was observed that the regional languages of Pakistan such as Punjabi, Pushto, Balochi and Sindhi were not used for the purpose of code-switching. The languages used for code-switching were Urdu and English. Urdu in the Roman script was preferred over Urdu in the Persian script. A number of reasons were found behind this pattern of code-switching use emerging.

Code-switching at the intra-sentential level was significantly more common than code-switching at the inter-sentential level. The users showed a strong preference in

starting inter-sentential code-switching with a sentence in English, then using a sentence in Urdu. Meanwhile, intra-sentential code-switching was most common at the level of clauses, followed by the level of words and the level of phrases. It can be concluded that code-switching is a common linguistic strategy of Pakistanis using Twitter.

References

Abbas, J., Aman, J., Nurunnabi, M., & Bano, S. (2019). The impact of social media on learning behavior for sustainable education: Evidence of students from selected universities in Pakistan. *Sustainability*, *11*(6), 1683. https://doi.org/10.3390/su11061683

Agarwal, P., Sharma, A., Grover, J., Sikka, M., Rudra, K., & Choudhury, M. (2017). I may talk in English but gaali toh Hindi mein hi denge: A study of English-Hindi codeswitching and swearing pattern on social networks. *2017 9th International Conference on Communication Systems and Networks*, 554–557. https://doi.org/10.1109/comsnets.2017.7945452

Al-Qaysi, N., & Al-Emran, M. (2017). Code-switching usage in social media: A case study from Oman. *International Journal of Information Technology and Language Studies*, *1*(1), 25–38. https://journals.sfu.ca/ijitls/index.php/ijitls/article/view/5

Arora, A. (2019). *Qualitative analysis of code-switching with reference to gender and pragmatic functions in Indian students at Oxford*. University of Oxford. https://ora.ox.ac.uk/objects/uuid:269d7675-4c91-4601-b926-82a600d7411c

Bansal, S., Garimella, V., Suhane, A., Patro, J., & Mukherjee, A. (2020). Code-switching patterns can be an effective route to improve performance of downstream NLP applications: A case study of humour, sarcasm and hate speech detection. https://arxiv.org/abs/2005.02295

Bohra, A., Vijay, D., Singh, V., Akhtar, S. S., & Shrivastava, M. (2018). A dataset of HindiEnglish code-mixed social media text for hate speech detection. *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, 36–41. https://doi.org/10.1145/3297001.3297048

Ghulam, S. M. & Soomro, T. R. (2019). Current status of Urdu on Twitter. *Sukkur IBA Journal of Computing and Mathematical Sciences*, *3*(1), 28–33. https://doi.org/10.30537/sjcms.v3i1.397

Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge University Press.

Hussain, A., & Arshad, M. U. (2021). *An attention based neural network for code switching detection: English & Roman Urdu*. National University of Computer and Emerging Sciences, Pakistan. https://arxiv.org/abs/2103.02252

Kamran, S., & Mansoor, S. (2017). Globalization and language use on social media in Pakistan. *European Journal of Language and Literature*, *3*(2), 79–84. https://doi.org/10.26417/ejls.v8i1.p79-84

KhudaBukhsh, A. R., Palakodety, S., & Carbonell, J. G. (2020). *Harnessing code switching to transcend the linguistic barrier*. Carnegie Mellon University. https://dl.acm.org/doi/abs/10.5555/3491440.3492042

Morrison, C. D. (n.d.). *Code-switching*. Encyclopedia Britannica. Retrieved June 20, 2021, from https://www.britannica.com/topic/code-switching

Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Clarendon Press.

Nordquist, R. (2019). *What is code switching?* ThoughtCo. Retrieved June 20, 2021, from https://www.thoughtco.com/code-switching-language-1689858

Panhwar, F. Y. (2020). Functions of code-switching in a private chat on Facebook. *Journal of English Language Literature and Education*, *1*(04), 1–16. http://jelle.lgu.edu.pk/index.php/jelle/article/view/6/6

Samhan, A. H. (2017). Social aspects in social media: Code switching and code mixing in Twitter. *Research on Humanities and Social Sciences*, *7*(18), 110–115.

Spooner, M. (2017). Code-Switching and its challenges: Perspectives on translanguaging in the EFL/ESL Classroom. *All Graduate Plan B and Other Reports*. 1126. https://digitalcommons.usu.edu/gradreports/1126

Sultana, I., Adnan, M., & Nawaz, M. B. (2020). Code-switching in social media: A case study of Facebook in Pakistan. *Pakistan Journal of Social Sciences*, *40*(2), 1075–1084. http://pjss.bzu.edu.pk/website/journal/article/5efa50d784e91/page

Rijhwani, S., Sequiera, R., Choudhury, M., Bali, K., & Maddila, C. S. (2017). Estimating code-switching on Twitter with a novel generalized word-level language detection technique. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, *1*, 1971–1982. https://doi.org/10.18653/v1/P17-1180

Twitter. (n.d.). *Privacy Policy*. Retrieved June 20, 2021, from https://twitter.com/en/privacy

Appendix

Tables referenced in figures:
https://docs.google.com/spreadsheets/d/1JK9Vt_vrWKvkPlnZDz4GZV4N0aA6BUSkMCU
cc_BKw6k/edit?usp=sharing