The Role of Digital Multilingual Patterns & Functions in Branding and Self-Branding: An Investigation of Practices by Saudi Users on Twitter

Reem Al Madani

Thesis submitted in partial fulfilment of the degree of

Doctor of Philosophy

Centre for Language and Communication Research

School of English, Communication and Philosophy

Cardiff University

August 2022

Abstract

This study is conducted within the multilingual digital communication domain, where it aims to investigate the linguistic resources employed on Twitter in Saudi Arabia. The study acknowledges the differences between the user-types, and thus compares the linguistic behaviours of three distinct groups of Twitter users: corporations, social media influencers, and ordinary users. Focusing on the language choices of Twitter users, the study investigates the relation between multilingualism and its role in presenting users to the virtual audience for (self-)branding.

The study investigates the use of different languages and varieties, and the inclusion of paralinguistic cues, in 13,426 tweets collected from 100 public Twitter accounts. Furthermore, the study explores the tweets' communicative functions to establish a link between the linguistic codes utilised for each respective function. Descriptive and statistical tools were employed to analyse the tweets and establish a general foundation for the different groups' utilisation of the platform's languages and varieties. Following the tweet's analysis, fifteen interviews were conducted to illuminate the motivation for the multilingual practices and its relation to (self-)branding.

The analysis revealed that regardless of the numerous language affordances available to the users, Modern Standard Arabic dominates and was perceived by most interviewees as the ideal variety for use on Saudi Twitter. Furthermore, the analysis found that Arabizi, reportedly very popular in Arab social media, is disappearing with negative attitudes towards it expressed by the interviewees. The analysis also highlighted the emergence of new modes for online communication, namely, reversed Romanisation (English with Arabic script), in addition to a new Arabic variety described by the interviewees as the 'white dialect'. Moreover, the different means of utilising English on Twitter, such as English with Arabic in parallel text bilingualism, were found. Finally, the analysis discusses the general trend for the association of the tweets' languages and communicative functions.

Acknowledgments

This thesis would not have been possible without a great number of people who have provided support throughout my PhD journey.

First and foremost, I would like to thank my supervisor Dr Mercedes Durham for all her insightful feedback, helpful guidance, and thoughtful observations. This work would not have been completed at this level without her considerable expertise and patience. Furthermore, I must acknowledge her encouragement to persevere with my work, especially during the challenges of the Covid-19 pandemic, for which I am most grateful.

I must also extend my gratitude to Prof. Michael Handford for his supervision and support when Dr Durham was on leave, while I am also thankful to Dr Tereza Spilioti, Dr Sara Pons-Sanz, and Dr Frances Rock for all their constructive suggestions and comments at different stages of my research. I must also express my appreciation to Cardiff University and all the amazing staff, especially Rhian Rattray for her advice and encouragement.

My profound appreciation must be offered to my dearest parents and aunt, my adorable sisters, my amazing brother, and my wonderful parents-in-law for their continuous support and prayers. I would not have been able to reach this conclusion without them.

I must express my deepest gratitude to my husband, Ahmed, for being ever-present at my side and supporting me to navigate through all the challenges that have presented. Without his love and belief, I would never have been able to submit this thesis. I must also thank Faisal, Farah, and Adam for all the excitement and happiness they bring into my life. They are my source of motivation and inspiration, and I am eternally blessed to have such wonderful children to light up my life.

I am fortunate to have forged many amazing friendships through this journey, and I must thank them for the wonderful times and great conversations we have shared. Thank you, Banan, Anfal, Alaa, Kate, Kateryna, Zainah, Sarab, and Shahad. I am also grateful to my writing group friends in Jeddah Shahla, Ruaa, and Noor. I hope our friendships last a lifetime.

I would also like to express my sincere gratitude to the Ministry of Education in Saudi Arabia and the Saudi Cultural Bureau in London for sponsoring my PhD research, and for all their support, as well as Effat University in Jeddah for recommending me for the scholarship programme to pursue my PhD.

Table of Contents

List of Illustrations

List of Tables

List of Abbreviations

ASCII	American Standard Code for Information Interchange
CA	Colloquial Arabic
CMC	Computer-Mediated Communication
DCOE	Discourse-Centred Online Ethnography
EFL	English as a Foreign Language
H	High (variety)
IRC	Internet Relay Chat
L	Low (variety)
MSA	Modern Standard Arabic
SMS	Short Message Service

Chapter 1:
Introduction

1.1 Research background

Social media technologies allow people to connect and share different content. With the Internet offering “a home to all the languages” (Crystal, 2006, p. 229), the huge volume of authentic multilingual written data publicly available online makes this a rich source for researchers from different disciplines such as linguistics, media studies, and the social sciences (Lee, 2017). The primary aim of this study is to examine multilingual practices in digital communication by Arabic speakers in Saudi Arabia in order to identify the characteristics and functions of the different languages and varieties that are used, mainly the different varieties of Arabic alongside English. Throughout, the study intends to shed light on how social media users’ choice of language is utilised as a tool for branding and self-branding (cf. Marwick & boyd, 2011; Page, 2012; Davis, 2014; Khamis et al., 2017). According to Androutsopoulos (2013a), the linguistic diversity on the Internet and the co-existence of languages on a website or social media platforms, such as a Twitter timeline, constitute a multilingual space that may not be related dialogically, and thus users alternate between languages in their different posts to communicate with their different groups of followers. Hence, the current research considers the Internet as a multilingual space and seeks to measure the distribution of the different languages and varieties that appear on the timelines of selected Twitter accounts. As the research is concerned with the relation between language, branding, and self-branding, the study offers a detailed analysis of the languages and varieties employed by three main groups on Twitter: corporations, influencers, and ordinary users (cf. Page, 2012). The use of different languages and varieties in digital communication should not be viewed as spontaneous production, particularly on a widely available public platform such as Twitter, since typing a tweet involves awareness and self-consciousness prior to posting it publicly, thus indicating that the choice of language is unlikely to be arbitrary. Moreover, as users have different types of audience among their followers, this may add complexity to their online choice of language as it forms a context collapse, which Marwick and boyd (2011) describe as the process by which various offline networks with different socio-demographics and types of relations are co-presented in a virtual space. Thus, the language choice of corporations, social media influencers, and ordinary users is dictated by the different kinds of followers to whom they wish to present themselves.

Arabic, which is the official language of Saudi Arabia, is characterised by the existence of a number of dialects utilised in different contexts (Albirini, 2016). This is clearly reflected in the difference between the formal standardised language employed in education and governmental communications, and the more informal spoken language such as that used on social media. Among Saudis, the common language varieties and forms utilised to communicate on social media are multilingual, and primarily include Modern Standard Arabic (henceforth, MSA), Colloquial Arabic (henceforth, CA), English, Arabic using Latin script (henceforth, Arabizi), and English with Arabic script. Furthermore, different paralinguistic cues such as emoticons and emojis are employed on social media to express emotions (cf. Seargeant, 2019; Tantawi & Rosson, 2019). The current research investigates how Twitter users in Saudi Arabia incorporate all of these languages and varieties on Twitter to create affiliations with their followers and develop their image. The concept of self-branding was first introduced by Tom Peters' FastCompany article in 1997, who asserts, "We are CEOs of our own companies: Me Inc. To be in business today, our most important job is to be head marketer for the brand called You". Then, in terms of reinforcing a positive image, Evans (2017, p. 271) states that "self-branding denotes how we want to be perceived by others". In the current digital age, self-branding tactics involve creating and maintaining social and networking profiles, personal websites, and blogs (Labrecque et al., 2011). Social media participants on the various platforms imagine other users to be an audience to whom they seek to be appropriately presented (Marwick & boyd, 2011). Hence, the current study explores how multilingualism is employed by Saudi Twitter users to communicate with their audience and brand themselves. The results are expected to guide Twitter users on the most appropriate languages and varieties for different communication purposes, such as to promote the user, maintain their existing followers, and acquire new followers.

This study analyses the different languages and varieties that occur in a data-set of posts from Twitter, which according to Statistica.com is among the most popular of social media platforms utilised by Arabs in general, and by Saudis in particular. In addition, multilingual practices will be explored through interviews with Saudi Twitter users, which are expected to offer insight into the motivations for such practices and how multilingualism is employed by the users as a tool for branding and self-branding.

1.2 Research aim and questions

With this study aiming to examine multilingual practices in digital communication by Arabic speakers in Saudi Arabia, the focus is on examining the different languages and varieties of Saudi users on Twitter employed on Twitter by three different groups of Saudi users: corporations, social media influencers, and ordinary users. Furthermore, the study seeks to

investigate the relation between the function of the tweet and the language and variety employed to fulfil this function. A further focus of the study is considering how the three groups use the patterns of the languages and varieties for branding and self-branding. Therefore, this study seeks answers to the following four research questions:

1. What are the linguistic characteristics of Saudi users' tweets? What languages and varieties are used?
2. What are the functions of the languages and varieties utilised on Twitter by Saudi corporations, social media influencers, and ordinary users?
3. How are the different groups using the different language choices in relation to the functions?
4. What are the patterns of communication employed by the different groups for branding and self-branding?

After establishing the research aim and questions, section 1.3 provides an overview of the methodology used to attain responses to the research questions.

1.3 Methodology overview

This study applies a mixed-method approach to investigate the practices of corporations, social media influencers, and ordinary users on Twitter by analysing the tweets posted on their public profiles, in addition to conducting interviews with a sample from each group to explore different insights that might not be apparent from analysing Twitter data alone, such as the motivations and general attitudes towards the languages and varieties used on Twitter.

In order to answer the research questions, the study adopts two main approaches: Discourse-Centred Online Ethnography (DCOE) and Computer-Mediated Discourse Analysis. The DCOE approach (cf. Androutsopoulos, 2007) combines the systematic observation of selected online discourse with direct contact with its social actors. This approach entails observing patterns of discourse and engaging with language users to elicit their perspectives on language use. Furthermore, the analysis of the tweets follows the methodological framework of Computer-Mediated Discourse Analysis presented by Herring (2004), and adds to the traditional assumptions of discourse analysis in that it is shaped by the technological features of digital communication.

The first data-set includes the tweets generated from Twitter by the public accounts of three main groups of Twitter users (i.e., corporations, social media influencers, and ordinary users) following the same categorisation of groups utilised by Page (2012). The second

data-set is collected through semi-structured interviews with a sample drawn from each of the three groups, that is, corporations, social media influencers, and ordinary users. The interview questions are specifically designed based on the results of the initial quantitative data analysis, and the results will be primarily utilised to offer significant insight not accessible through analysing the first data-set alone, such as the preferred language choice, the motivations for using each language or variety, and the attitudes towards the use of the different languages and varieties.

1.4 Research significance

As mentioned in section 1.1, there is a huge volume of multilingual text available online. Therefore, there is growing scholarly interest in multilingual practices in digital communication, and specifically in various social media platforms (cf. Jaworska, 2014; Seargeant & Tagg, 2014; Androutsopoulos, 2015; Tagg, 2015; Spilioti, 2019). The current research seeks to supplement our understanding of the emerging linguistic trends of written multilingualism employed in digital communication in Saudi Arabia, and reflects on its implications for the study of digital communication in the global domain.

Various studies that explored the multilingual practices of users on different platforms in digital communication (e.g., Danet & Herring, 2007; Georgakopoulou & Spilioti, 2015; Lee, 2017) reveal that participants navigate between different codes, including English as the Internet's lingua franca, the official standard language of the user, and the non-standard variety. These studies mainly investigated the process of alternation between the different codes and the functional reasons why participants engage in code-switching in digital communication. This written code-switching online is facilitated by eliminating the boundaries that separate speech from writing, which Tagliamonte and Denis (2008, p. 8) describe as "the emerging tendency for written genres to be more like speech, a process referred to as colloquialization". Some of these features are utilised to facilitate accessibility to a broader virtual audience through employing different languages and varieties. Furthermore, other features such as emoticons and emotions are employed to convey the paralinguistic features of spoken communication that are usually expressed by facial expressions, tone of voice, and physical gestures. Hence, it is important to examine the different characteristics of language variation employed on Twitter.

A number of studies have also investigated the linguistic practices of Arab users on different social media platforms from different perspectives (e.g., Palfreyman & Khalil, 2003; Salia, 2011; Alothman, 2012; Alfaifi, 2013; Eldin, 2014), with most focused on code-switching between Arabic and other foreign languages, primarily English and French. Furthermore,

researchers have explored the different Arabic varieties utilised online (e.g., Albirini, 2016; Khalil, 2018; Alhejely, 2020). For example, in 2003, Palfreyman and Al-Khalil investigated female university students in the United Arab Emirates through data based on a corpus of instant messenger conversations, where they concluded that Arabic was employed for formulaic phrases and English for university-related topics. Additionally, Eldin (2014) studied the code-switching functions performed by Egyptian Arabic–English bilingual users in their Facebook interactions, while Alfaifi (2013) investigated code-switching on Facebook by focusing on Saudi participants and reporting on the different topics utilised in English and Arabic. Another study was conducted by Alothman (2012), where she investigated the orthographic representation and communicative functions of Najdi Arabic in Internet Relay Chat (IRC), and examined the attitudes and beliefs towards the use of the different language forms. Khalil (2018) explored Egyptian Arabic writing on Facebook during the Arab Spring, arguing that this movement has played an important role in the widespread writing of colloquial Egyptian Arabic. Another recent study was conducted by Alhejely (2020), who investigated the linguistic practices on Twitter of five Arab students in UK using online ethnography, with the findings revealing that the users relied more on CA and that English was not dominant in the data. Nevertheless, despite the research on the multilingual practices of users on different platforms in digital communication, no study has explored this in the context of Saudi Arabia via a mixed-methods study of corporations, social media influencers, and ordinary users, with inter-group comparison of the groups.

This study investigates the broader multilingual use of Arabic speakers to include MSA, CA, English, and Arabizi, as well as paralinguistic cues (i.e., emoticons and emojis). Furthermore, it is of great significance to comprehend the kinds of languages and varieties employed on social media and to shine a light on the forms that support visibility on social media (Page, 2012). The current study also considers the functions and attitudes towards the use of languages/varieties as an important aspect of the analysis, thus aiming to investigate not only the language choice, but also the functions and attitudes towards these choices by the different users. Moreover, by comparing the multilingual practices of three different groups of Twitter users, this study provides a detailed quantitative analysis of the languages and varieties employed, while previous studies in the Arab context focused more on the qualitative facets of the phenomenon.

Finally, this research not only reports on the languages and varieties that exist on Twitter in Saudi Arabia, but also provides insight into the languages and varieties utilised online in the country. Moreover, evidence is gathered on the national status of English and the changes

that are occurring to the Arabic language with the massive development of digital communication.

1.5 Thesis structure

This section presents an overview of the chapter structure of the thesis.

Chapter 1 introduced the topic of the current research. It offered a background to the research phenomenon, in addition to a discussion of the main languages pertinent for this study, namely, Arabic and English. The research questions were established, and an overview of the methodology provided. Finally, the gap in the existing literature was presented that highlights the academic significance of this research.

Chapter 2 provides a discussion of the main languages under study in this research, that is, Arabic and English. First, the chapter offers an overview of Arabic and considers the different varieties that are found on Twitter and utilised by Saudi users. Then, English and its global status is discussed, with a special focus on the status of English in Saudi Arabia.

Chapter 3 conducts a literature review of the relevant research. It is divided into four main sections: (i) digital communication in general, and on Twitter in particular; (ii) multilingualism and multilingual digital communication; (iii) self-presentation and identity in digital communication; and (iv) branding and self-branding on social media and Twitter.

In Chapter 4, the methodology of the thesis is presented. First, the research approach is discussed. Then, the data-collection process is explained, with an account provided for the coding process of the languages, varieties, and functions that appear in the data, in addition to an explanation of the statistical analyses employed to examine the data both quantitatively and qualitatively. After that, the chapter discusses the ethical considerations for the research, and finally explores the pilot study and its advantages in shaping the research design.

Chapter 5 (corporations), Chapter 6 (social media influencers), and Chapter 7 (ordinary users) present a detailed analysis of the languages, varieties, paralinguistic cues, and functions of the three groups' Twitter usage. Furthermore, each chapter presents the findings emerging from the semi-structured interviews conducted with a sample drawn from each of the three groups.

Chapter 8 conducts further analysis by comparing the overall results of the three groups, and reports on the statistical findings to describe the association between the language

mixing, the languages used, the paralinguistic cues, and the functions of the three groups under study in this research. Furthermore, the chapter provides a detailed account of the languages utilised with each function that appeared in the data collected from Twitter.

Chapter 9 presents the responses to the research questions and links the findings emerging from the analysis in chapters 5–7 with the overall conclusions drawn from the recurring patterns. Finally, the chapter concludes by highlighting the research findings, implications, and recommendations for future study.

Chapter 2:
Languages in Saudi Arabia

2.1 Introduction

The current study explores the linguistic practices of Saudi users on one of the most popular digital communication platforms: Twitter. Therefore, it is necessary to consider Arabic, which is the official and main language employed in Saudi Arabia. Furthermore, when coding Twitter data extracted from Saudi users, it is vital to distinguish between the different Arabic varieties, as will be further explored in this chapter. The description of Arabic in section 2.2 is based on the literature on the development of Arabic and its various linguistic features, and supported by examples from the data collected for the current research. Furthermore, as English was found to be popular in Saudi social media (cf. Alothman, 2012; Al Alaslaa, 2018), section 2.3 provides an overview of English as a world language, as well as the status of English in Saudi Arabia and on Twitter in the Kingdom.

It is important to note that the chapter will only review relevant studies on the Arabic language variations, since other aspects of Arabic socio-linguistics such as the historical changes and gender differences are not relevant to the present study.

2.2 Arabic

Arabic is one of the most widely used languages worldwide. Therefore, this section conducts an in-depth examination of the varieties of Arabic and the factors that affect their classification. Section 2.2.1 provides an overview of the Arabic language in light of the linguistic framework of diglossia, in addition to the different local dialects of Arabic, while section 2.2.2 discusses each variety of Arabic and provides examples from the data collected for this study.

2.2.1 Aspects of the Arabic language

Arabic is the (co-)official language of 25 countries, and represents one of the six official languages of the United Nations. The number of Arabic speakers depends on the source utilised. Ethnologue (n.d.), which is widely considered as a reliable language data source, asserts that Arabic is the fifth most spoken language globally with approximately 274 million speakers in 2021. Another recent report published by the World Bank states that Arabic is spoken by 467 million speakers worldwide (Gregory et al., 2021). It should be noted that

some sources, such as Ethnologue, have recently begun to count different varieties of Arabic separately (e.g., Standard Arabic, Egyptian Arabic, and Saudi Arabic), which explains why its statistic is significantly lower than that of the World Bank. However, regardless of these discrepancies, Arabic still tends to present as one of the most frequently employed languages globally. Furthermore, uncertainty regarding the number of Arabic speakers and different varieties raises an important question: What is Arabic?

Research interest into the social aspects of Arabic commenced in the mid-twentieth century with Ferguson's (1959) pioneering work, where he described the social and linguistic variables of Arabic, and used the term 'diglossia' to refer to a specific relationship between two or more varieties of the same language in use in a speech community with different functions. Ferguson (1959, p. 336) defines diglossia as:

> a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) super-imposed variety, the vehicle of a large and respected body of written literature, either in an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.

Ferguson investigated the linguistic situation in Arabic-speaking countries, in addition to Greece, German-speaking Switzerland, and Haiti that are characterised by similar functional distributions between two varieties of the same language. The superposed variety is referred to as the high (H) variety that is acquired through formal education, with the low (L) variety acquired informally in the home. In the Saudi Arabia context, Ferguson's H variety can be compared with MSA as it is the formal written variety, while the L variety can be compared with CA. It can also be argued that while MSA enjoys a higher status, it is the less familiar variety due to being learned formally and not acquired naturally at home. The most important characteristic of diglossia is the functional specialisation of the H and L varieties. Primarily, the H is employed in formal settings, writing tasks at school, religious settings, books, newspapers, and magazines. On the other hand, the L variety is utilised for daily conversation.

As discussed above, diglossia views the language as having two varieties, each with its own features and uses. However, there are different overlapping features between the H and L varieties of diglossic languages. Hence, Ferguson's model of H and L varieties in Arabic has been refined to acknowledge the existence of an intermediate level, with many of the recent studies on the Arabic language identifying three main varieties: Classical Arabic, MSA, and CA (cf. Versteegh, 2014; Albirini, 2016; Holes, 2018). Note that each of these varieties will be further explained in section 2.2.2.

Based on the discussion above, Arabic can be described as a continuum of forms including Classical Arabic, MSA, and the many informal CA varieties utilised for daily communication. A learner of Arabic therefore needs to master MSA and one of the regional dialects to be a speaker of Arabic (Gregory et al., 2021). MSA is acquired through formal education, unlike dialects that children acquire from an early age at home (Al-Huri, 2015). As soon as children enter school, they are expected to start using MSA for reading, writing, listening, and speaking. This presents a challenge in defining native speakers of Arabic, since the exposure of young Arabic speakers to MSA prior to commencing their schooling is limited to reading the Quran and stories, or watching certain, mostly dubbed, children's programmes. Therefore, MSA cannot be considered a native language for any Arabic speaker, because their mother tongue is the dialect related to their particular region (Alghamdi, 1998; Biadsy et al., 2009). Furthermore, many parents in Arab countries are incentivised to introduce foreign languages (primarily English and French) to their children, which obviously competes with the available time and prioritisation for Arabic (Gregory et al., 2021). This raises a challenge in terms of mastering MSA that could result in errors, including when writing a tweet.

2.2.2 Varieties of Arabic

This section presents a brief description of the languages and varieties appearing in the data collected from Twitter, in addition to examples. It is important to discuss the varieties of Arabic, and how they appear in the data, because the coding of the tweets will be based on the linguistic features of each variety, as described in the sub-sections below.

2.2.2.1 Classical Arabic

Classical Arabic is often identified as the literary language employed in the Quran and the medieval pre-Islamic era (Albirini, 2016). Furthermore, Classical Arabic is associated with the Hadith (i.e., the reported words of the Prophet Mohammed), which is considered the second-most important literary source in Islam. The strong link between Classical Arabic and Islam has contributed to its wider distribution in many countries around the world. Moreover, Classical Arabic is not solely used in Muslim religious settings, but also in Christian ceremonies and rituals in Arab countries that have Christian populations such as Egypt, Lebanon, and Syria. Nowadays, Classical Arabic is primarily employed in religious settings and literary poems. Similarly, in the data collected from Twitter, Classical Arabic is used when quoting verses from the Quran, the Prophet Mohammed's words, and poems. The following example in Figure 2.1 shows a tweet with a verse from the Quran.

[O ye who believe! If a wicked person comes to you with any news, ascertain the truth, lest ye harm people unwittingly, and afterwards become full of repentance for what ye have done.]

Figure 2.1: A tweet with a verse from the Quran

2.2.2.2 Modern Standard Arabic

As a result of the contact between Arab countries and Europe in the nineteenth century, linguistic changes started to appear in Classical Arabic (Versteegh, 2014), which began with translation from European languages, especially English and French, into Arabic. As many scientific concepts in the source languages had no equivalents in Arabic, new words, expressions, and stylistic features started to dilute Arabic. With the spread of many novel terms and expressions, the language gradually diverged from Classical Arabic, thus leading to the emergence of MSA (Abdulaziz, 1986). The new form's title was coined to reflect the changes that Classical Arabic was undergoing. The relationship between Classical Arabic and MSA can be viewed as MSA being the development of Classical Arabic, which reflects the developments that all languages typically undergo. According to Albirini (2016), the new words and forms found in MSA comprise part of the evolutionary process, just as many of the new words found in contemporary English did not exist in the language several decades ago. However, speakers of Arabic and those familiar with the language would agree that the difference between Classical Arabic and MSA is not as striking as the difference between Old and Modern English. Researchers of Arabic have indicated that the variances between Classical Arabic and MSA are primarily lexical and stylistic, as opposed to being morphosytactic or phonological (cf. Bateson, 1967; Versteegh, 2001; Holes, 2014).

MSA is generally utilised in education and business, as well as for official televised programmes (e.g. news and documentaries), along with newspapers and magazines, letters, and official speeches. As MSA is the official variety of Arabic, with a recognised writing system and taught in schools, it is widely employed in the data, as well as in various digital communication platforms in general. The following example in Figure 2.2 shows a tweet in MSA by a Saudi corporation.

[Footage of #KingFaisal, then Emir, during his official visit to #Britain on behalf of his father. His visit was at the official invitation of King George V, to participate in the celebrations of Britain's victory in #WorldWarI. This was his first visit to Europe in 1919 at the age of 13.]

Figure 2.2: A tweet in Modern Standard Arabic

2.2.2.3 Colloquial Arabic

CA refers to the different regional dialects spoken by Arabic speakers in daily conversation and in informal settings to discuss different topics such as sports and music, and as used in most television programmes other than the news. According to Mitchell and El-Hassan (1994), although these varieties differ in a number of ways in terms of their lexicon and

phonology, they share a broad range of lexical, syntactic, phonological, and morphological features. Arab academics attempt to maintain the distinction between MSA and CA through various means such as Arabising foreign words and maintaining MSA dictionaries. However, CA is developing due to the use of the variety on television and in digital communication through social media, where it is now even written rather than merely being spoken.

CA is spoken across a diverse geographical area. Therefore, there has been a tendency to classify Arabic dialects according to their geographical locations. For example, Biadsy et al. (2009, p. 55) classify Arabic dialects geo-linguistically into the following regions:

1. Gulf Arabic: including the dialects of Kuwait, Saudi Arabia, Bahrain, Qatar, the United Arab Emirates, and Oman.
2. Iraqi Arabic: the dialect of Iraq—in some dialect classifications, Iraqi Arabic is considered a sub-dialect of Gulf Arabic.
3. Levantine Arabic: including the dialects of Lebanon, Syria, Jordan, and Palestine.
4. Egyptian Arabic: featuring the dialects of the Nile Valley, namely, Egypt and Sudan.
5. Maghrebi Arabic: including the dialects of Morocco, Algeria, Tunisia, and Mauritania, with Libya sometimes also included.

According to the above classification, Saudi Arabia has been classified under Gulf Arabic. However, there are many regional colloquial dialects within Saudi Arabia. Few studies have investigated the linguistic patterns in Saudi regions and hypothesised their linguistic boundaries (Ingham, 1994). Aldarsoni (2011) investigated the linguistic features in Saudi dialects and divided them into five main dialect regions: the Najdi dialect (Central), the Hijazi dialect (Western), the Southern dialect, the Northern dialect, and the Eastern dialect. Nevertheless, the current study does not differentiate between the different local varieties and regards them all as CA.

One of the challenges during the analysis of tweets concerns the different varieties of Arabic utilised, and in particular differentiating between MSA and CA, which can overlap in many aspects of their morphological and syntactic rules, in addition to some lexical items. The following examples illustrate the similarities and differences between MSA and CA:

a. MSA: *al-walad**u** katab**a** fii ad-daftar*

   English: The boy wrote in the notebook

b. CA (Saudi dialect): *al-walad katab fi l-daftar*

English: The boy wrote in the notebook

Clearly, these examples are almost identical apart from minor phonological and morphological differences, such as the presence of the case markers in MSA (highlighted in bold above). Therefore, in addition to relying on the morphological and syntactic differences between MSA and CA, the researcher consulted native speakers of Arabic when rating some of the more ambiguous tweets.

Arabic colloquial dialects share the following common features:

1) They are all acquired at home from family.
2) They are the main language varieties employed for everyday interactions.
3) Arabic colloquial varieties are not typically written.
4) CA does not have an official status in any of the Arabic-speaking countries (as discussed regarding Ferguson's L variety above).

It is worth noting that changes to CA occur faster than in the case of MSA. Firstly, new concepts and expressions can be easily adapted and used. Furthermore, in similarity to other languages, some speakers, especially those from the younger generations, deliberately attempt to change the rules of their dialects by introducing new borrowed or modern terms to indicate their sophistication, intelligence, modernity, or socio-economic class (Albirini, 2016). Albirini (2016) highlights the main differences between MSA and CA as follows. First, MSA is characterised by an elaborate morphological system that involves inflections for such features as the number, gender, person, case, and definiteness. On the other hand, CA has a more simplified morphological system that lacks many of these inflections such as the representation of dual and plural-feminine categories, case, and infinite endings. Second, sentential negation in MSA is realised by five main negative particles—*laa*, *maa*, *lam*, *lan*, and *laysa*—while in CA this is realised by two negative particles, which are *maa* and *mahu/mahi* in the Saudi dialect. There are a number of structures that are expressed uniquely in MSA or CA. For example, affirmative existential constructions in MSA are introduced by the locative form *hunaaka* (there) or the verb *yuujad* (exists). However, in CA existential sentences are introduced by the preposition *fi* (in). In terms of phonology, some sounds in MSA words such as /q/, /ð/, /θ/, and /j/ are utilised differently in colloquial words. For example, most CA dialects change the MSA /q/ to /ʔ/, /g/, /y/, or /k/; and the /ð/ and /θ/ to /d/, /z/, /s/, or /t/. According to Albirini (2016), MSA has a richer vocabulary and fewer foreign words in comparison to CA. Borrowed words from different languages such as English, French, and Turkish are shared by more than one Arabic dialect, such as *asˁansˁeir* (lift), *boliis* (police), and *ṭarabeiza* (table). It should be

noted that these words have established counterparts in MSA, namely, *ma sˁ ʕad* (lift), *ʃor tˁa* (police), and *tˁawla* (table), but they are commonly utilised in CA.

One of the observations from the data of the current study is that there is a variety that shares features of MSA and CA. Generally, this hybrid adopts the syntax and morphology of CA and lexical items from MSA. Hence, there is no regional marking for this variety. The general purpose for using this hybrid variety is to be both formal and congenial, simultaneously. Furthermore, as it adopts lexical items from MSA, local colloquial terms are not used that may obscure the origin of the user while rendering the post more easily read by all Arabic speakers. As was also apparent from the interviews, the users were aware of this variety, whereby they consciously use it and edit their posts to represent this variety. They referred to it by different terms such as the 'white dialect' and 'plain dialect'. The term white or plain are used as a metaphor to indicate purity and clarity to the recipients as colloquial regional terms are not used. It should be noted that some researchers of Arabic identified and explained different levels of Arabic between the H and L varieties presented by Ferguson. Badawi (1973) explained varying degrees of overlap between the H and L, and he described a specific spoken colloquial variety of the educated that he named *ʕaːmmiyyat ʔal-muθaqqafiːn.* Furthermore, Meiseles (1980) and Ryding (1991) emphasised an intermediate variety of Arabic that they described as formal spoken and referred to as Educated Spoken Arabic, which is an informal variety utilised among educated Arabs for its "intercomprehensibility among speakers of different vernaculars, arising mainly from the speaker's incentive to share a common language with his interlocutors" (Meiseles, 1980, p. 126). Ryding (1991, p. 216) lists a number of distinctions that differentiate Educated Spoken Arabic from MSA:

1. Omission of inflection (i.e., final short vowels in all parts of speech).
2. Consequent metathesis of vowels on pronoun suffixes.
3. Reduction of inflectional endings in dual and masculine plural sounds to the oblique or non-nominative form.
4. Elimination of the separate feminine plural categories in verbs and pronouns, and reduction to one non-gender-specific plural.
5. Elimination of the dual category in verbs and pronouns (both second and third person) and merging of this category with the plural.
6. Omission of the final 'nuun' in inflectional suffixes for second-person feminine singular and second- and third-person plural in the imperfect.
7. Generalisation of the defective suffixable stem to geminate verbs in the past tense.

8. Conversion of final nunation in indefinite defective nouns to a long vowel.
9. Creation of a category of verbs with an embedded indirect object.

Educated Spoken Arabic is also employed in television programmes, particularly on pan-Arab channels that are watched by Arabs from different countries and use different dialects. It is also used in business meetings, as MSA is not utilised and such meetings are not an appropriate setting for CA. Hence, a middle variety is also employed. However, no recent research has been conducted in this area to investigate the contemporary use of this variety. In the data of the current research, it was clearly noticed that the users from the three different groups tended to employ a hybrid between MSA and CA, which somehow resembles Educated Spoken Arabic but in a written form.

As stated earlier, CA was not typically used in written form. However, digital communication has given rise to its use on such platforms. In particular, family members and friends find it uncomfortable to communicate in MSA when discussing personal matters, and therefore they tend to use CA (cf. Albirini, 2016). On Twitter, which is a more public platform, CA is employed to convey congeniality to the audience. The following example in Figure 2.3 is from Careem, a transportation network company that tends to use CA more frequently than MSA to communicate with its clients. It should be noted that this tweet was determined as being written in CA, due to the inclusion of the word *alħein* (now), which is CA in several dialects utilised in Saudi Arabia.

[Your options with Careem are increasing, now you can add credit through Mada cards.] 👍

[Picture: After the transaction, you will see the balance in your Wallet account.]

Figure 2.3: A tweet in Colloquial Arabic

2.2.2.4 Arabizi

Arabizi is a term employed to describe the use of Roman characters as an alternative orthographic form of the Arabic language, which is normally written in Arabic script (Alothman, 2012). The widespread use of writing Arabic words using English characters among online users in the early use of the internet led to several new terms that describe this phenomenon such as Romanized Arabic, Arabish, and Arabizi (Albirini, 2016). Section 3.3.4.1 describes the use of Arabizi in more detail as creative orthography in digital communication.

One of the main areas of research motivating studies of language in digital communication is a desire to determine why Arabic speakers tend to employ Arabizi rather than Arabic script in their online communications (cf. Palfreyman & Al Khalil 2003; Yaghan, 2008). However, the data from the current study revealed that Arabizi is now less extensively employed than previously, with only a few examples having been identified. Figure 2.4 shows an example of Arabizi usage by one of the social media influencers.

[Hello]

Figure 2.4: A tweet in Arabizi

2.3 English

People from around the world with diverse cultural and linguistic backgrounds employ English for communication, diplomatic relations, education, trade, and business (Jenkins, 2003). Furthermore, English is recognised as a global language and there has never been a language with such a wide spread of usage (Jenkins, 2003; Schneider, 2007; Durham, 2014). The globalisation of English is relevant to the present study as English represents

one of the important languages used for Twitter communication in Saudi Arabia. Here, section 2.3.1 explores English as a world language, while section 2.3.2 discusses the situation of English in Saudi Arabia, and finally section 2.3.3 displays how English is utilised by Saudi users on Twitter.

2.3.1 English as a global language

English has been considered as having a global language status since it was spread throughout the British Empire between the sixteenth and twentieth centuries. After that, the United States became the world's leading economic power following the conclusion of World War II (Crystal, 2003). Furthermore, in addition to the influential political status of those English-speaking countries, one of the important factors playing a prominent role in the popularity of English during the twentieth and twenty-first centuries has been globalisation and the need for global communication. English is the global language for many sectors including business, science, technology, and electronics (McArthur, 2002), with multinational companies worldwide mandating English as the corporate language in order to facilitate communication between the diverse businesses globally. Furthermore, in the domain of digital communication, it has become possible for those with different linguistic and cultural backgrounds, and from various geographical locations, to communicate. Therefore, there has been an increasing need for a common language to ensure more efficient communication. Hence, English is employed for facilitating worldwide communication (Crystal, 2003; Jenkins, 2003).

The globalisation of English resulted in its grouping into three broad categories: English as a Native Language, English as a Second Language, and English as a Foreign Language (EFL). One of the most influential models for classifying English varieties is Kachru's (1982) model of the three circles of English, whereby the first circle is the inner circle that includes countries such as the United Kingdom, the United States, and Australia where English is the primary language; the second circle is the outer circle that includes countries like India and Singapore where English is widely used for internal purposes; and the third circle is the expanding circle where English is treated as an international language that the society learns (i.e., EFL), featuring countries such as China and Russia. Based on this classification, Saudi Arabia belongs to the third, expanding circle, since English is considered an international language that is restricted to specific contexts such as business, trade, and education, and is learned as a foreign language. Kachru's model has been widely cited by socio-linguists to address the varieties of English, including the Arabic setting. However, the situation of English nowadays shows that classifying the types of English speakers in different countries does not reflect the entire situation, since there are myriad

ways of categorising world Englishes (Durham, 2016). For instance, although Kachru clearly defines inner circle countries, the distinction between the outer circle and the expanding circle is blurred, and has thus received further analysis over time with the technological advancements. Furthermore, Kachru's classification focuses on the countries rather than the citizens and how they use English. Nevertheless, Kashru's model remains one of the most cited classifications to situate the status of English in the majority of socio-demographic studies, addressing issues related to variations in terms of the use of English or its communicative functions worldwide. Hence, section 2.3.2 uses the classification to contextualise how English is utilised in the Kingdom of Saudi Arabia.

2.3.2 English in Saudi Arabia

As explained in section 2.2 above, Arabic is the dominant and official language in Saudi Arabia. However, English also represents an important language in many sectors in the Kingdom, such as business, education, tourism, and healthcare. Moreover, English can be seen alongside Arabic on road signs, in restaurants, and on advertising billboards, in addition to within most hospitals and universities.

The rapid economic growth in Saudi Arabia has played a major role in the spread of English among the society. The use of English, especially in business, became vital in the early 1940s following the discovery of oil reserves with the assistance of American oil companies (Zuhur, 2011). In recent decades, the economic openness has provided myriad opportunities for investment in different fields, thus attracting international companies to launch or franchise their business in Saudi Arabia, which has enhanced the status of English as the official language of business in the Kingdom. Furthermore, English has become an essential requirement for recruitment by many national and multinational corporations. One of the key focus points for the Saudi Vision 2030 is diversifying the Saudi economy away from a dependence on oil revenues (Moshashai et al., 2020), with one of the goals being to strengthen the economy by expanding business and broadening the economic base. Therefore, English usage is expected to increase further in Saudi Arabia in the light of this initiative and its aims.

Furthermore, according to the latest official census in Saudi Arabia, more than 30% of the population are non-Saudis, and even though a large proportion are from other Arab countries such as Egypt, Yemen, Syria, and Lebanon, many of the expatriates are from non-Arabic speaking countries such as India, Pakistan, and Philippines. This has resulted in English, or an English pidgin, being employed as a lingua franca in Saudi Arabia by non-Arabic and Arabic speakers alike (Al-Rawi, 2012).

In order to better comprehend the situation of English usage in Saudi Arabia, it is useful to refer to Kachru's (1982) three circles model presented in section 2.3.1 above. Based on this classification, although English is widely employed in different sectors of Saudi Arabia, the country belongs to the expanding circle since English is considered an international language that is restricted to specific contexts such as business, trade, and education, and is learned as a foreign language. However, according to Kachru (1982), the transition from the expanding circle to the outer circle occurs for different reasons, such as the length of time in use, the functional importance, and the socio-linguistic status. The rapid economic and technological advances unfolding globally have thus resulted in situating certain countries between the expanding and outer circles. Hence, the current situation of English in Saudi Arabia can be described as being positioned between the outer and the expanding circles of Kachru's (1982) model.

English is officially taught as a second language in public and private schools in Saudi Arabia, where the school system comprises of 12 years: 6 years at the primary stage, followed by 3 years at the intermediate stage, and finally 3 years at the secondary stage. English was previously taught from the intermediate stage onwards, but since 2003 it commenced from Grade 4 of primary school. Moreover, in 2021 the Ministry of Education announced that English tuition would begin from Grade 1 of primary school, resulting in 12 years of English instruction in Saudi public and private schools. The economic growth and the expansion of English in Saudi Arabia has thus underscored the need to improve its proficiency among Saudi students from a younger age. At university, many subjects are taught through the medium of English in a range of fields such as medicine, engineering, and science. In other academic fields taught in Arabic, students take English language courses throughout the duration of their studies in government universities. On the other hand, private universities teach the majority of the courses that populate the different majors in English. In addition, there are several scholarship opportunities to study in English-speaking countries that aim to qualify the students to be able to meet the requirements and demands of the domestic employment market. In fact, one of the most important requirements for most positions in Saudi Arabia is the ability to communicate in English, which is essential in the globalised present and will ultimately affect the overall economic success of the country. Following on from this discussion about English as a global language and in Saudi Arabia, the next section introduces how English is utilised on Saudi Twitter accounts.

2.3.3 English in Saudi Twitter

As discussed in section 2.3.2, English in Saudi Arabia is taught and utilised as a foreign language and is increasingly employed for teaching various majors in higher education (e.g., science, engineering, and medicine), along with business interactions such as emails and meetings (Rahman & Alhaisoni, 2013). Therefore, it is to be expected that English represents the most used international language in the data. Researchers have also identified a coexistence of English and Arabic in digital communications on websites, as well as for Facebook comments, text messages, and emails (cf. Warschauer et al., 2002; Palfreyman & Al Khalil, 2003; Al-Khatib and Sabbah, 2008; Alfaifi, 2013). The Twitter data collected for the current study reveals different uses of English by the three groups (i.e., the corporations, social media influencers, and ordinary users). Furthermore, the data highlights three main ways of communicating in English, as summarised below and further explained in greater detail in the analysis chapters.

1) A post in English and Arabic with similar content:

Figure 2.5: A tweet in Arabic and English

2) A post in English, followed by another in Arabic featuring the same content (or vice versa):

Figure 2.6: A tweet with an English translation

3) A post solely in English:

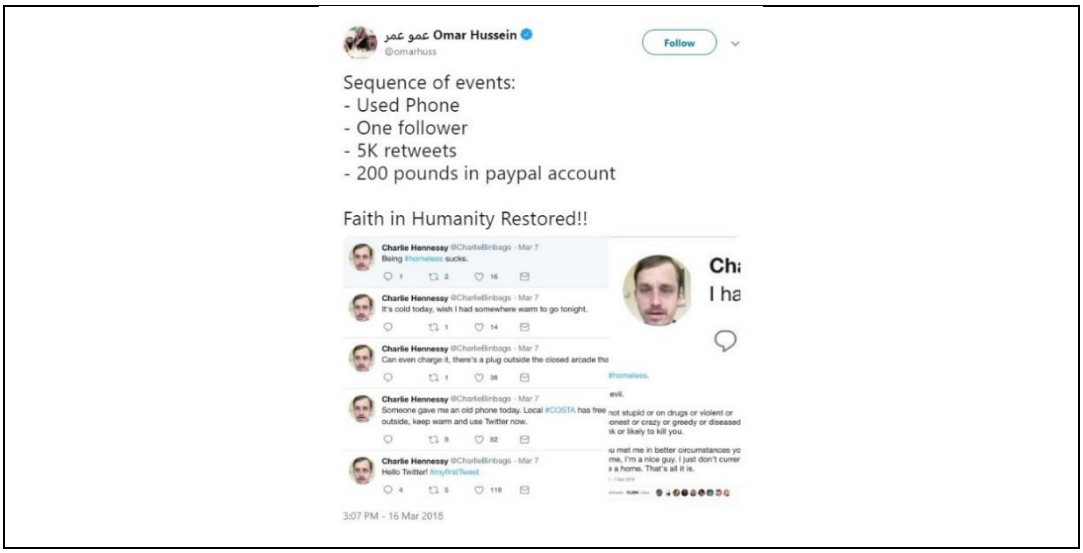
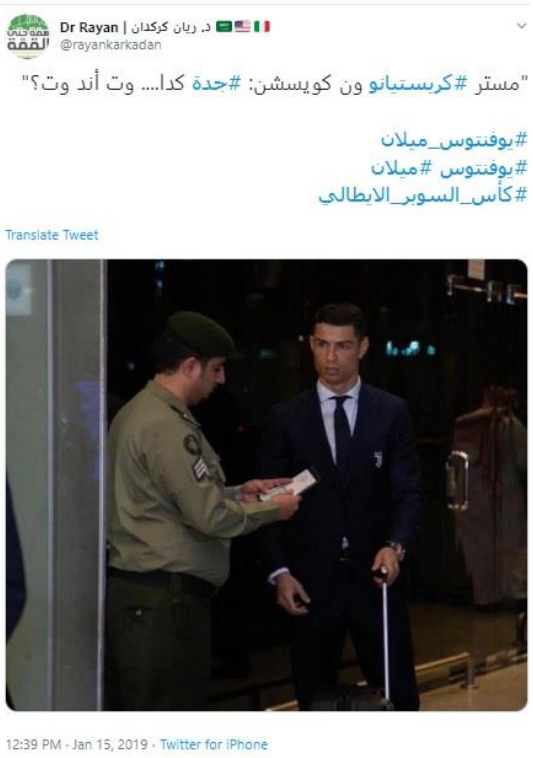


Figure 2.7: A tweet in English only

Another observation arising from the data is the use of English with an Arabic script (i.e., the opposite of Arabizi), as exemplified in the following example from one of the social media influencers who poses an open question in English (but in Arabic script) to the footballer Cristiano Ronaldo and his followers:



“Mr. #Cristiano one question: #Jeddah kida (like that) .. what and what?”

#Juventus_Milan

#Juventus

#Italian_Super Cup

Figure 2.8: A tweet in English with Arabic script

2.4 Conclusion

This chapter provided an account of the Arabic language situation and English as a global language, with a particular focus on Saudi Arabia that represents the focus of this study. It is believed that the present study will further enhance the body of research on language variation in the digital communication setting by explaining the users’ linguistic practices on Twitter. Nevertheless, in the light of the discussion in this chapter, it is important to ascertain how digital communication affects the language variation and change in Saudi Arabia. This can be revealed by considering the different social media platforms, including Twitter, and will offer insight into the different language practices of the users. In addition, the global

spread of technology and the media have significantly contributed to the expansion of English in Saudi Arabia as a pertinent language, specifically in the context of digital communication and marketing.

The following chapter reviews the literature relevant to this study, providing an essential theoretical background to scaffold the analysis of the collected data and highlight new avenues for exploring the potential applications of the study.

Chapter 3:
Literature Review

3.1 Introduction

This chapter provides the reader with the broader context of the study and is divided into five main sections. First, section 3.2 explores digital communication, the development of the technology in this field and the current research trends, in addition to a focus on Twitter, the main platform used for this research. After that, section 3.3 defines multilingualism in general, and then discusses multilingual digital communication, including the creative use of orthography, language choice, code-switching online, the functions of code-switching, and the use of paralinguistic cues. Section 3.4 then considers self-presentation and online identities, and how language is employed to negotiate self and multiple identities online. Next, section 3.5 focuses on branding and discusses several definitions within this domain, alongside the related concept of self-branding, in addition to corporate digital communication and an overview of social media influencers. Finally, the chapter ends with a summary in section 3.6.

3.2 Digital communication

As well as defining digital communication, this section discusses the development of technology, which is an important factor that has contributed to the research in this area. It also explores current trends and perspectives in digital communication, and concludes with an in-depth consideration of Twitter.

3.2.1 Defining digital communication

Since the 1990s, the study of language and electronic communication has piqued the attention of many researchers and provoked discussions on how the Internet is rapidly changing the manner in which language is used (Crystal, 2006). This area of research is referred to as either ‘computer-mediated communication’ (CMC), which became widely recognised when it appeared in the title of the influential *Journal of Computer-Mediated Communication* in 1995, or Computer-Mediated Discourse Analysis (Herring, 2007). In addition, the emergence of devices such as smartphones, and the various applications that are utilised for communication have extended the notion of computer communication as users do not feel that they are holding computers when they communicate using their portable devices. Hence, other inclusive terms such as ‘online communication’ (Lamy &

Hampel, 2007) and ‘digital communication’ are also used (Georgakopoulou & Spilioti, 2015). The current study employs the definition for digital communication provided by Herring and Androutsopoulos (2015, p. 127), that is, “the communication produced when human beings interact with one another by transmitting messages via networked or mobile computers, where ‘computers’ are defined broadly to include any digital communication device”. In support of this, Tagg (2015) argues that digital communication encompasses all kinds of devices that involve the Internet such as computers, laptops, and smartphones, as well as platforms such as Twitter. Therefore, this study will use the term ‘digital communication’, while CMC will also be employed in some instances for quoting references that used it as this was the preferred term in some of the previous literature.

3.2.2 Development of technology and digital communication research trends

The development of digital technologies has reshaped academic research on digital communication over time. The various technological possibilities and constraints that are characteristic of the different digital communication platforms have also changed since the 1990s, when this area of research first came to linguists’ attention. The literature categorises studies on digital communication into three waves. The first and second wave are discussed in this section, while the third wave is explored in section 3.2.3. The first wave of linguistic digital communication studies, as referred to by Androutsopoulos (2006a), primarily focused on the linguistic features of the language employed in emails and e-chat, which led to claims of the existence of a special Internet language. For example, Crystal (2004) was the first to coin the term ‘netspeak’ to describe this specific type of language displaying features unique to the Internet. In the first wave of research, studies focused on areas such as the combination of written and spoken features, the use of emoticons and unconventional spelling, and the differences between modes of communication, namely, synchronous (i.e., real-time CMC such as IRC and instant messaging) and asynchronous (i.e., time-delayed interaction such as emails and forums) (cf. Werry, 1996; Yates, 1996; Witmer & Katzman, 1997; Paolillo, 2001; Androutsopoulos & Georgakopoulou, 2003). However, with the advancement of technology and social media, distinctions such as synchronous and asynchronous are rarely used nowadays. It is also worth noting that some more recent studies reflect digital communication as being characterised by the same structured variation and linguistic change currently underway in contemporary varieties of English. For instance, Tagliamonte and Denis (2008) investigated a 1.5-million-word instant messaging discourse among 71 teenagers in Toronto, and concluded that instant messaging was both a medium at the forefront of change, and a unique new hybrid register exhibiting a fusion of the full range of variants from the speech community—formal and

informal. This represents one of the focus areas in the current study, which considers the formal and informal varieties of the Arabic language, and how different users utilise them for communication on Twitter.

The second wave of digital communication research shifted the focus from medium-related to user-related patterns of language use. Androutsopoulos (2006a) mentions that the unique characteristics of digital communication language, as identified in the first wave of digital communication research, were understood as resources for the users, with the analysis in the second wave of research moving towards contextualised approaches. The majority of the recent digital communication research is placed in the second wave, including the current thesis, where different topics are discussed such as language choice and code-switching, as discussed in sections 3.3.4.2 and 3.3.4.3, respectively. The different research included in Georgakopoulou and Spilioti's (2015) study also explored the trends and topics within the area of digital communication, such as digital communication research methods and perspectives, language resources and discourses, digital literacies, digital communication in public, digital selves and online–offline lives, networks and relationships, in addition to further directions for study.

There is no doubt that new technological uses of digital communication have resulted in more devices and platforms that extend beyond what is perceived as typical e-chat style. Despite the main focus of the initial digital communication research placed on email and chatrooms, digital communication modes now include different social media platforms (e.g., Twitter, Facebook, Instagram, and Snapchat), Wikipedia, corporate websites and blogs, online games and videos (cf. Marwick & boyd, 2011; Kashian et al., 2017; Nowak & Fox, 2018; Waterloo et al., 2018). These novel platforms and the development of digital communication approaches have inspired researchers to explore new domains, as discussed in the following sections. Another important technological advancement that has provided a rich area for investigation is the ability to use different scripts, that is, other than Romanised and non-Romanised alphabets, in typing, as is further explained in section 3.3.4.1. Nowadays, other languages are easily accessible on the Internet, and the use of the different languages online could imply different purposes. The many languages utilised on the various platforms, in addition to their contextual use, represent fertile territory for linguists to investigate and analyse, as explored in section 3.3.

The current research trends in digital communication include different areas such as multimodal analysis to accommodate the multimodal nature of digital communication (e.g., Jones & Hafner, 2012; Jewitt, 2013) and interaction between the online and offline linguistic landscapes (e.g., Lee, 2015). Furthermore, Georgakopoulou and Spilioti (2015) present a

glimpse of an emerging third wave that focuses on two main areas: critical and ethical agendas. The critical approaches focus on discourses and ideologies of digital communication, that is, relations of power and issues of control and surveillance. In terms of ethics, Georgakopoulou and Spilioti (2015) stress the importance of a transition towards new ethical procedures and practices in digital communication.

Moreover, Lee (2016) argues that different topics remain to be explored within the digital communication domain such as translingual practices and the interplay between language choice and identity performance, which represent one of the main foci of this research. Among the potential research areas are the use of different codes among multilingual users, including code-switching; cross-cultural communication between users in different global locations; the use of lexical items or orthographic features among different social groups; the representation of spoken dialectical features in this casual sphere of text; and how the characteristics of the communication medium are employed differently across groups (e.g., corporations, social media influencers, and ordinary users). The public availability of Twitter's data indeed makes this platform an ideal candidate to explore different research opportunities.

3.2.3 Twitter: It's what's happening

Twitter is a microblogging website that was launched in October 2006 and allows users to send messages, referred to as tweets.[1] Mischaud (2007) reported that the majority of the 5,767 tweets analysed in his study included postings related to family, friends, personal information, information-sharing, technology, casual conversation, the workplace, and reporting on activities. However, in addition to communication between friends and family, Twitter is also used to share news and diverse opinions. Tweets can be posted and read on desktop computers or smartphones and other mobile devices through the Twitter application, which allows the user to engage in the instant posting of updates, images, videos, replies to other users, and to participate in hashtags (i.e., a word or phrase that appears after the '#' symbol to identify all tweets related to a particular topic). Users of Twitter can select whether they wish their tweets to be public and appear on their public timeline, in addition to the user's Twitter page; or private, in which case only the user's followers can see the tweets. Twitter posts can be categorised into the following main types:

1) A timeline update created by the user

[1] Tweets were originally limited to 140 characters, but the character count increased to 280 in 2017.

2) A retweet, namely, the reposting of a tweet created by another user
3) A reply addressed to a particular user

Twitter has grown rapidly since its launch, and it is popular worldwide. According to stastista.com, Twitter had 330 million global users as of the fourth quarter of 2017 when the current research started. Figure 3.1 shows the leading countries based on the number of Twitter users as of April 2018, the period when the data were collected.

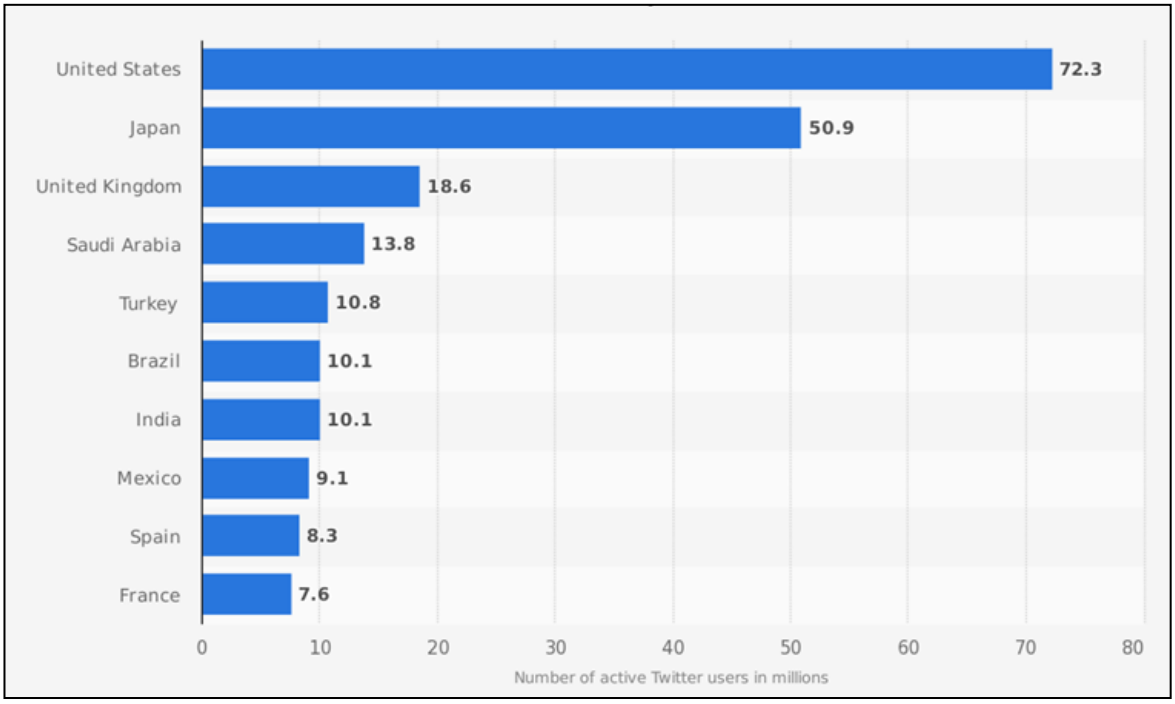


Figure 3.1: Leading countries based on the number of Twitter users (millions) in 2018 (Statista, n.d.)

The statistics presented in Figure 3.1 reveal that the United States is ranked first with 72.3 million active users, and that Saudi Arabia, the main focus of the current study, is ranked fourth globally with 13.8 million active users. These active users comprise 22% and 41% of the countries' 2018 populations, respectively, which highlights that almost double the number of the Saudi Arabian population are active Twitter users compared to the users in the United States.

The massive number of users and interactions on Twitter has attracted linguistic researchers, and there is an substantial body of literature that has investigated different aspects of communication and language usage on Twitter ranging from the functions of tweets, hashtags, and virtual communities, to the use of emoticons and emojis (Gillen & Merchant, 2013; Zappavigna, 2015; Murthy, 2018; Seargeant, 2019). Twitter is described

by Squires (2015, p. 239) as “a popular space for online interaction which allows users to re-use these interactions within and outside of its immediate domain”. Squires considers different specific areas for investigation on Twitter, and mentions the diversity of language use as a phenomenon that still requires additional research. Gillen and Merchant (2013) offer analyses of the dialogic social and linguistic functionalities of Twitter and identify certain emerging practices that include citizen journalism, political activism, maintaining a fan base, event back-channels, corporate advertising, service marketing, crowd sourcing, social networking, and ambient sociability. Hashtags are employed primarily to organise the discussions on Twitter based on certain topics and among users who do not know one another, with Zappavigna (2011, p. 789) characterising their use as discourse where the primary function appears to be affiliation via findability. Marwick and boyd (2011) investigated how content producers on Twitter navigate imagined audiences and present a model of the networked audience through which individuals conceptualise an imagined audience evoked through their tweets. Another interesting area of investigation is the co-occurrence of different features in tweets. Schnoebelen (2012) analysed emotion usage on Twitter and concluded that the most frequently employed emoticons were the smiley, wink, and frown. Furthermore, he analysed the different nose variants of emoticons (e.g., :-) versus :)) and found that they pattern differently with regards to the words that they occur with, in addition to the age of the user, since he found that nose ‘omitters’ were younger than nose ‘includers’. Therefore, it has been shown that the analysis of the form of the paralinguistic cues used in tweets can reveal certain characteristics about the authors, such as their age in the case of Schnoebelen’s (2012) investigation, and perhaps their target audience. These studies reveal that the use of certain linguistic features on Twitter may be linked to particular identities, goals in using the medium, or both (Squires, 2015). It must be highlighted that the language affordances of this platform play an important role in shaping communication on Twitter as shall be discussed in section 3.3.4.2. Hence, language variation on Twitter by diverse groups (e.g., corporations, influencers, and ordinary users) is of great interest to researchers who wish to understand digital communication. The following sections move into a more detailed discussion of multilingualism online.

3.3 Multilingualism and online communication

The current study focuses on multilingual practices in an online setting, with the research questions exploring the different languages and dialects utilised on Twitter, the functionality of the different codes, and the language choice and patterns of communication in the tweets. Therefore, in order to engage with the literature in this field, we must first clarify what is meant by multilingualism, and provide an overview of the key communication and

multilingualism issues in the online domain. Hence, this section considers the definition of multilingualism and explores multilingualism in digital communication. It must be noted that although multilingualism is a broad field in socio-linguistics, the focus in the coming sections is only on multilingualism in digital communication.

3.3.1 Defining multilingualism

Multilingualism has become a major topic of research, attracting the attention of many scholars (e.g., Vildomec, 1963; Weinreich, 1963; Edwards, 1994; Hoffmann, 2000; Wei, 2012; Stavans & Hoffmann, 2015). Furthermore, the research on multilingualism is multidisciplinary, and therefore researchers approach different aspects of multilingualism, which may affect its definition as well as the research method and design applied. As an example, Aronin and Hufeisen (2009) mention that the main strands in multilingualism are situated in the framework of the following research domains: socio-linguistics (cf. Dewaele, 2004; Hoffmann & Ytsma, 2004), psycholinguistics (cf. Jessner, 2006), neurolinguistics (cf. Franceschini et al., 2003), applied linguistics (cf. Hufeisen & Marx, 2007), and teaching and learning (cf. Cenoz, 2009). In the case of the current study, it is situated within the socio-linguistics domain of digital communication or CMC.

Franceschini (2009, p. 34) defines multilingualism as “the product of fundamental human ability to communicate in a number of languages. Operational distinctions may then be drawn between social, institutional, discursive and individual multilingualism”. It can be found either at the individual or the societal level. Therefore, multilingualism also refers to the coexistence, contact, and interaction of different languages at the societal or the individual level (Wei, 2012).

A broad definition of multilingualism is the use of three or more languages in one setting. As an example of an early study, Arsenian (1945) analysed different issues related to multilingualism such as language development, school achievement, personal and social adjustment, and political state specifically in the post-war period in countries like Russia, Switzerland, South Africa, and Belgium. Weinreich (1963) was an early pioneer in the research of bilingualism and multilingualism, and paved the way for further studies in this area. He discussed the phenomenon of bilingual/multilingual practices as one of the results of languages in contact, and primarily French, German, and Romansch in Switzerland. It is also important to mention that in many academic discussions, bilingualism and multilingualism are employed interchangeably, with Wei (2000, p. 131) describing this phenomenon as follows:

> The word ‘bilingual’ primarily describes someone with the procession of two languages. It can, however, also be taken to include the many people in the world who have varying degrees of proficiency in and interchangeably use three, four, or even more languages.

After reviewing the notion of multilingualism, section 3.3.2 looks at the development of the research within the area of multilingualism and its practices.

3.3.2 Multilingualism: recent developments

Generally, earlier definitions of multilingualism as the alternate use of more than one language tended to be restricted to the equal mastery of languages (Edwards, 1994). On the other hand, more recent research on multilingualism allows for greater variation in competence. For instance, Aronin and Singleton (2012) argue that diverse theoretical perspectives such as proficiency level and frequency of use could make the definition complex. Meanwhile, Kemp (2009) debates that although researchers define multilingualism as the use of three or more languages, defining what a language actually is can be complex, and thus she explores the definitions originating from the research purposes and contexts of investigating multilingualism, such as defining what a language is and how languages can be counted with regards to individual proficiency, functional capability, and identity. García (2009) and Canagarajah (2015) discuss the development of the concept of multilingualism over time by reviewing three main models: (i) the subtractive bilingualism model, which implies that learning a second language makes the first language weaker, although this model is criticised because it indicates a conflict between the languages (Canagarajah, 2015); (ii) the additive model, which considers that a bilingual individual has a balanced linguistic competence in both languages (Canagarajah, 2015); and (iii) the recursive model, which indicates that the languages are not added collectively at one time, but rather through practice at different times and in different situations, with those languages supporting one another, and therefore suggesting that multilingualism is linear with each language treated separately (García, 2009). After discussing these three models, García (2009) proposes the model of dynamic bilingualism, which indicates that multilingualism is a dynamic cycle whereby users employ their available linguistic resources depending on the context and the situation (García, 2009). Furthermore, Garcia (2009) stresses that the complexity of the current (twenty-first century) era forces people to communicate in dynamic ways that challenge traditional categories such as a first and second language. Wei (2000) adds that multilingual speakers employ different languages for different purposes, and typically do not have the same level of proficiency in each language. The current study adopts this perception of multilingualism and accounts for the different languages regardless of the proficiency level. For instance, in Saudi Arabia, where

this study takes place, people use different varieties of Arabic in addition to English with different levels of proficiency. Colloquial Saudi dialects are the native languages that Saudis learn at home and use primarily for daily conversations. MSA is taught at school and is used for writing and formal purposes, with Saudis having varying degrees of proficiency based on their education and training. Additionally, Classical Arabic, which is the language utilised for the Quran and prayers, is more limited in terms of its highly professional speakers. Finally, the knowledge of English varies among Saudis, since it is taught as a foreign language in schools and employed widely for business and education purposes. A detailed discussion of these Arabic dialects is provided in Chapter 2.

It also must be asserted that a number of researchers have questioned the traditional concept that languages are objects with clear boundaries (e.g., Jørgensen et al., 2015). Reagan (2004) claims that the concept of the existence of languages with static boundaries such as English or Arabic is problematic, because all languages are constantly evolving over time, while the language also changes from one speaker to another, and in different circumstances. In line with this argument, Makoni and Pennycook (2005) argue that understanding languages as entities with fixed boundaries is the result of colonialism. Furthermore, Canagarajah (2013) emphasises that individuals utilise all the available linguistic resources in their communication without separating between names, languages, or varieties. Therefore, according to Canagarajah (2013), languages are connected and influence each other. These views are aligned with Jørgensen et al. (2015) in terms of drawing clear boundaries between languages or deciding what separates different dialects of the same language, such as Arabic. Otheguy et al. (2015) argue that there are two ways to understand the notion of language: (i) in terms of names that are socially and politically constructed, maintained, and regulated such as English, Arabic, and Chinese; and (ii) as entities without names, but rather comprising of sets of lexical and structural features that make up the linguistic repertoire and are used to enable communication. According to this perception, the speaker engages in the process of ‘languaging’, which indicates that “language users employ whatever linguistic features are at their disposal with the intention of achieving their communicative aims” (Jørgensen, 2008, p. 169).

Consequently, recent advances in multilingualism research have provided alternative terms that describe the use of a plethora of languages for communication where there is greater flexibility in using the different linguistic repertoires. Notably, these include ‘translanguaging’ (García, 2009; García & Li, 2014), ‘metrolingualism’ (Otsuji & Pennycook, 2010), ‘polylingual languaging’ (Jørgensen, 2008), ‘code-meshing’ (Canagarajah, 2011), and ‘superdiversity’ (Blommaert & Rampton, 2012). These new interrelated concepts,

regardless of their intended distinctions, shift the focus from the speakers' linguistic competence in multilingualism to what they do with their linguistic resources in everyday social interactions. In this way, it is related to the current research as it investigates the multilingual practices of corporations and social media influencers who have commercial and popularity objectives for their social media interactions. However, the main idea of these terms is that language is a socially discursive practice without clear boundaries (Creese & Blackledge, 2015). The notion of translanguaging has received particular attention (Bhatt & Bolonyai, 2019), while it has also been applied to digital communication research (e.g., Schreiber, 2015; Kulavuz-Onal & Vásquez, 2018). Furthermore, other digital communication research draws upon translanguaging to introduce new concepts such as trans-scripting, as will be further explained in section 3.3.4.1 (cf. Spilioti, 2019; Androutsopoulos, 2020). Therefore, translanguaging is a pertinent phenomenon worthy of additional focus.

Wei (2011, p. 1221) defines translanguaging as:

> both going between different linguistic structures and systems and going beyond them. It includes the full range of linguistic performances of multilingual language users for purposes that transcend the combination of structures, the alternation between systems, the transmission of information and the representation of values, identities and relationships.

The main idea of translanguaging is that the various linguistic resources utilised by the speaker cannot be easily assigned to one language or another (García & Li, 2014). Translanguaging differs from code-switching, in that it refers to the entire linguistic repertoire of an individual, and the individual's freedom to choose items from their repertoire as they see fit, in contrast to the code-switching view that a speaker shifts or shuttles between two languages to be best understood (García & Li, 2014). Within the framework of translanguaging, bilingualism is seen as dynamic, not simply additive, whereby one keeps adding more languages to one's repertoire (García & Li, 2014). Bilinguals are perceived to have one language system comprised of different languages or varieties, rather than two or more separate language systems. For example, if we wish to apply the process of translanguaging to Arabic, the linguistic repertoire of Arabic speakers is regarded as a single language repertoire, rather than two (e.g., one of MSA, and another for CA). This may explain the practices such as code-switching between the different varieties of Arabic and foreign languages, or writing in different forms such Arabic with Roman script (i.e., Arabizi). Translanguaging implies that speakers are utilising the full scope of their repertoire including MSA, CA, and foreign languages, and constantly choosing the appropriate forms within their repertoire.

Translanguaging, however, has been criticised by a number of researchers, who argue that the notion of separate languages cannot be denied, as evidenced by the existence of the body of research on switching from one language to another (e.g., Jaspers, 2018; Bhatt & Bolonyai, 2019; Jaspers & Madsen, 2019). Furthermore, Bhatt and Bolonyai (2019) assert that within the concept of translanguaging, the linguistic repertoires are still recognised as specific named languages. Seargeant and Tagg (2011) claim that this may lead to a paradox, since it is necessary to analyse the linguistic repertoires through identifying different languages and varieties, despite this problematising the notion of languages as discrete entities. Hence, Bhatt and Bolonyai (2019) argue that translanguaging does not offer a new understanding of multilingual language use that is not covered by the traditional notion of multilingualism and code-switching. Nevertheless, Wei (2018) explains that code-switching manifests a shift between one language and another, which occurs in different patterns. On the other hand, translanguaging assumes that the individual employs his or her linguistic repertoires in such a dynamic manner that it may be difficult to establish patterns for switching between languages or varieties.

According to the discussion above, adopting the concept of translanguaging may not be appropriate for this study as it is important to note that not all online linguistic practices are translingual, with Twitter users appearing to make conscious decisions on the use of the language or variety for communication with their virtual audience. Hence, the current study will use the concept of multilingualism in its broadest sense, namely, what Lee (2016, p. 10) considers “the co-existence of two or more languages, or codes in any communicative context, including various representations of language”. In the present study, for example, the different varieties of Arabic discussed in Chapter 2 are treated as different codes.

Furthermore, as this study focuses on multilingualism online, the concept of ‘networked multilingualism’ as introduced by Androutsopoulos (2015) is a relevant theory for this research. Androutsopoulos (2015) extended the contemporary theorising of multilingualism and coined the new term, which can be defined as an umbrella reference for multilingual practices that are shaped by two interrelated processes: being ‘networked’ (i.e., digitally connected to other individuals and groups), and being ‘in the network’ (i.e., embedded in the global digital mediascape of the Internet). According to Androutsopoulos (2015, p. 88), networked multilingualism “encompasses everything language users do with the entire range of linguistic resources within three sets of constraints: mediation of written language by digital technologies, access to network resources, and orientation to networked audiences”. In the same article, he presents findings from a case study to explore the implications for the theorising of networked multilingualism following an online ethnography

approach, concluding that the students' networked multilingual practices are individualised, genre-shaped, and based on broad and stratified repertoires. The network approach proposed by Androutsopoulos contributes to the understanding of multilingual practices by highlighting the extended opportunities for the creative use of others' signs and voices that the digital culture has to offer, which echoes the research questions of the current study.

After considering the recent developments in multilingualism research and how the theories of multilingualism are applied to digital communication, section 3.3.3 reviews the research carried out in the multilingual digital communication domain.

3.3.3 Moving from English to multilingual digital communication

As noted in section 3.2, the research on digital communication has developed over time, with one of the major technology-related improvements being the ability to use different languages on the Internet. Early research on digital communication in the 1990s focused primarily on the linguistic features of English. For example, Yates (1996) explored a corpus-based comparison between spoken, written, and digital communication discourse, while Collot and Belmore (1996) investigated the electronic language as a new variety of English. In fact, research on online multilingualism in the 1990s indicated that English would dominate the Internet. For example, Gupta (1997, p. 2) discussed English usage on the Internet at that time and claimed, "the dominance of English in the internet needs no arguing for. Computers are in any case English-oriented". The supremacy of English is unsurprising, since the Internet and the majority of the related technologies were invented and first became popular in the United States, where English is spoken as the de facto national language. In support of this, Fishman (1998) reported that 80% of Internet content was written in English and 90% of the world's servers were based in countries where English is employed as the primary language. More recently, however, the importance of English has decreased. Table 3.1 shows the number of native speakers and Internet users for the top five languages (Ethnologue, 2018; InternetWorldStats, 2018).

Table 3.1: Top five languages globally in terms of native speakers and Internet users in 2018 (Ethnologue, 2018; InternetWorldStats, 2018)

Rank	Language	Native speakers (million)	Language	Internet users (million)
1	Chinese	1,299	English	1,052
2	Spanish	442	Chinese	804
3	English	378	Spanish	337
4	Arabic	315	Arabic	219
5	Hindi	260	Portuguese	169

One of the interesting points to be noticed in Table 3.1 is the difference between the number of native English speakers (i.e., 378 million) versus the number of people using English on the Internet (i.e., 1,052 million). While the dominance of English was the initial case, it is evident that the situation of English in digital communication has changed, with other languages now on the rise online. Table 3.1 shows a strong online presence for other languages including Arabic, which is one of the primary languages studied in this research. It should be stressed that the development of keyboard typing in non-Roman scripts, which was discussed in section 3.2.2 above, has supported the increased presence of languages such as Arabic and Chinese online. The Internet has thus provided opportunities for minority languages to become more visible and representative globally (Cunliffe & Herring, 2005), with Lee (2017) highlighting the importance of the presence of 'lesser-written languages', such as the different colloquial varieties of Arabic that represent one of the scopes of this research.

Although the statistics in Table 3.1 offer an indication of the different languages used in addition to English online, it should be borne in mind that these data do not provide a sense of how internet users might use more than one language. Nevertheless, such surveys indicate that language usage other than English is increasing in digital communication, and thus imply the importance of investigating digital discourse beyond the English context (Lee, 2016). Furthermore, besides the aforementioned studies that considered English as a lingua franca or global resource for users, there are researchers who focused their lenses on the localised forms of English. For example, Seargeant et al. (2012), Androutsopoulos (2014), and Spilioti (2019) illustrate that English-related forms are manipulated as local resources. Spilioti (2019) found that the local English–Greek orientation was shown in the types of voices and social persona represented, whereby the respellings of English-related forms do not orient to global cultures or identities, but rather they index local spoken uses of English and identities.

Researchers have also studied the presence of different languages on Twitter, which is the primary platform explored in this research. Honeycutt and Herring (2009) analysed a corpus of naturally-occurring public tweets, finding that English tweets were the most common among other languages such as Japanese, Spanish, Portuguese, and Tagalog. Furthermore, Hong et al. (2011) studied 62 million tweets and discovered that half of these (51%) were in English, while other popular languages including Japanese, Portuguese, Indonesian, and Spanish accounted for 39% of the tweets. The current study is also in line with this research as it explores the different languages used on Twitter in Saudi Arabia, while also comparing the use of the different languages between different groups of users. Hence, the increasing presence of different languages on the Internet has attracted many linguists to investigate multilingualism in digital communication, as explored in the following section.

3.3.4 Multilingual computer-mediated communication

Clearly, digital communication offers a major domain for multilingual communication. In accordance with the discussion of multilingualism in section 3.3.3, Leppänen and Peuronen (2012) mention that multilingualism on the Internet refers to two interrelated phenomena: (i) enabling the choice and diversity of languages as a means of communication, and the analyses of their visibility, accessibility, and status online; and (ii) referring to Internet users' strategies for selecting, drawing on, and employing more than one language in particular modes and environments of digital communication. The first book that tackled the issue of multilingualism in digital communication, and which is extensively cited in the literature, is *The Multilingual Internet*, co-edited by Susan Herring and Brenda Danet in 2007. The volume is a collection of different case studies that investigated the use of English and other world languages in digital communication, which were published in an early special edition of the *Journal of Computer-Mediated Communication* in 2003. In order to highlight the variety of languages and topics first addressed in the area of multilingualism in digital communication, Table 3.2 is adapted from Danet and Herring (2007) and summarises the case studies presented in the book by digital communication mode, population, language, focus, and chapter author. This highlights the diverse range of research topics and foci possible, as well as the numerous languages and varieties employed online even in 2007, thus underscoring the depth of multilingualism that occurs in the digital communication domain.

Table 3.2: Summary of Danet and Herring's (2007) case studies

Mode	Population	Language	Focus	Author
Personal email	High school, university students	Cantonese, English	Code-mixing, representations of Cantonese, Romanisation	Lee
Personal email	Young professionals (24–36 years of age)	Classical, Colloquial Egyptian Arabic, English	Language choice	Warschauer et al.
Bulletin board systems	University students, young people	Taiwanese, English	Writing systems, dialects, language play	Su
Discussion list	Medical students	English, French, German, Italian	Language choice over time	Durham
Discussion list	University faculty, staff	Portuguese	Politeness, gender	Oliveira
Discussion list	EU citizens, all ages	Multiple, English	Language choice, dominance	Wodak & Wright
Bulletin board systems on fan websites	Young people	Japanese	Orthography and typography	Nishimura
Local use-net newsgroups	Computer science students	Catalan, Spanish	Netspeak and Spanish interference in Catalan, issues for machine translation	Climent et al.
Discussion lists, news group	All ages	Greek, Greeklish	Romanisation; uses and features of Greeklish, relations between Greeklish and English borrowing	Tseliga
Web-based, discussion forums	Young adults, migrants, children of migrants to Germany	German, Persian, Hindi, Punjabi, other Indian languages, Greek(lish)	Language, code-switching	Androutsopoulos
SMS	University students, young people	French	Orthography and typography	Anis

(short message service)				
Instant messaging	Female students	Arabic, ASCII-ised Gulf Arabic	Orthography and typography	Palfreyman & Al Khalil
ICQ chat	High school, university students	Cantonese, English	Code-mixing, representations of Cantonese, Romanisation	Lee
Webchat	Young people (11–25 years of age)	Thai	Turn taking, gender	Panyamethee kul & Herring
Chatroom	Housewives	Japanese	Kaomoji (Japanese emotions), gender	Katsuno & Yano
Graphical chat	Unspecified	English, miscellaneous European languages	Language choice	Axelsson et al.

As can be noticed from the case studies included in Table 3.2, a diverse range of topics are investigated within multilingualism in CMC. Lee (2016) states that since the publication of *The Multilingual Internet*, a broad range of platforms, languages, and geographical locations have been researched. Various topics have been discussed within the research domain of multilingualism in digital communication such as language choice, code-switching, language ideology, minority languages, and online translation. The following sections provide an overview of four main topics that are relevant to this study: the creative use of different linguistic features, language choice, and code-switching, and the paralinguistic functions of emoji.

3.3.4.1 Creative use of orthography in digital communication and trans-scripting

Earlier research on multilingualism online described participants' creative use of different features as a solution to the technical restrictions of the American Standard Code for Information Interchange (ASCII), which is based on 128 codes that rely on the Roman alphabet and the sounds of the English language (Danet & Herring, 2007). As mentioned in section 3.3.3, English was the dominant language of the Internet and non-English communities had limited digital-encoding possibilities, which hindered their online communication. The ASCII protocol urged the users of non-Roman languages such as Chinese, Japanese, Arabic, and Greek to use a Romanised transliteration of their

languages. These transliteration schemes were based on mappings between the letters of the standard writing system for a particular language, such as Greek and Arabic, and forms considered unconventional or non-standard in the given context such as Roman letters or numerals (Spilioti, 2019).

Lee (2002) examined the linguistic features of text-based CMC in Hong Kong based on a 70,000 word corpus of electronic instant messaging texts, with her results recognising the novelty and the linguistic specificity of digital communication texts including creative orthography. Table 3.3 presents examples of Romanised Cantonese expressions from her data.

Table 3.3: Examples of Romanised Cantonese (Lee, 2002)

Romanised expression	Target Cantonese expression	Example from data
bai lin	*baai3 nin4* (Relatives visiting, a Chinese custom during Chinese New Year)	we just 'bye lin' and went home! how abt u?
bei sum gei	*bei2 sam1 gei1* 'Work hard!' (an encouragement)	then "bei sum gei": la cu next time!

Nishimura (2003) explored the linguistic properties of informal bulletin board system messages in Japanese and identified the innovative use of scripts and punctuation, in addition to the incorporation of informal spoken features. Moreover, Romanisation has been documented in Greek by different researchers such as Tseliga (2007), who labelled this variety as 'Greeklish' and described the linguistic and socio-cultural perspectives on Roman-alphabetised Greek in CMC. Furthermore, Spilioti (2009) investigated the use of Greek and Roman alphabets in SMS messages, concluding that such marked choices are employed as a means of indexing the participants' affiliation with global popular cultures and enhancing expressivity in a medium of reduced paralinguistic cues.

In terms of Arabic, several researchers described the use of Roman script to write Arabic words, which is called Romanised Arabic or Arabizi. For example, Warschauer et al. (2002) found a Romanised version of Egyptian Arabic being utilised extensively in informal email and online IRC channels. For example, one of the interesting features of the adaptation of

Romanised Egyptian Arabic is the widespread use of the numbers 2, 3, and 7 to represent the phonemes /ʔ/ (voiceless glottal stop), /ʕ/ (voiced pharyngeal fricative), and /ħ/ (voiceless pharyngeal fricative), respectively, which are not easily rendered in the Roman alphabet. The use of these numbers arose among Arabic Internet users, spread spontaneously, and is now widely recognised. In Table 3.4, Warschauer et al. (2002) provide an example of the use the numbers 2 and 3 from their data.

Table 3.4: Examples of Romanised Arabic (Warschauer et al., 2002)

Example from data	Translation
Hello Dalia, *7amdellah 3ala el-salama ya Gameel. we alf mabrouk 3alal el-shahada el-kebeera.* *Keep in touch. I really hope to see you all Soooooooooooooon (Maybe in Ramadan).* *Kol Sana Wentom Tayyebeen.* *Waiting to hear from you…* *Laila*	*Hello Dalia,* *Thank God for the safe return, my sweet. Congratulations for the big certificate [sarcastic].* *Keep in touch…I really hope to see you all Soooooooooooooon (Maybe in Ramadan)* *Happy Ramadan.* *Waiting to hear from you….* *Laila*

Similar findings were reported by Palfreyman and Khalil (2003), who presented a small-scale, exploratory study of how female Arab university students in the United Arab Emirates used the Latin alphabet to write vernacular Arabic in online communication, specifically in instant messaging (using applications such as MSN Messenger, Yahoo Messenger, or ICQ). They found that the variety of Arabic employed online highlighted influences from computer character sets, different varieties of spoken Arabic, Arabic script, English orthography, and other Latinised forms of Arabic employed in contexts that pre-date digital communication. Palfreyman and Khalil (2003) concluded that the users had developed creative solutions to the technical constraints prevailing at that time. It is apparent from the studies mentioned above that the creation of Romanised scripts, including Romanised Arabic, resulted due to the lack of the letters of some languages on the computer keyboard. Interestingly, in the current era, these scripts are still being utilised despite technological advancement allowing the users to type in their own language. Therefore, the use of these scrips is now a language choice for the users that is expected to serve different purposes, as shall be explored in this study.

The earlier studies focused on the use of Roman characters for writing languages that associated with other writing systems such as Greek and Arabic. However, recent studies have explored reversed Romanisation, which involves the use of non-Roman script for writing English forms.

Some studies have documented the existence of such forms in multilingual studies. For example, Jaworska (2014) explored the linguistic practice of an online discussion forum, employed by a community of English-speaking Germans living in the United Kingdom. The article examines how the users deployed linguistic and other semiotic resources on a forum to co-construct humorous code-plays, and documents English-related forms in German. Moreover, Spilioti (2014) investigated the uses of English in social media practices among Greek Internet users, and described the Greek-alphabet English. Spilioti (2019) analysed more than one thousand tokens of Greek-alphabet English ('Engreek') collected from different types of online sources. The research analysed how the respelt forms are created, for what purpose, as well as for whom. The study revealed that the phonetic respellings of English forms suggest a link between language creativity and orality in digital communication.

The initial reason for the creative use of orthography was the technical restrictions of the Roman script. However, when other scripts, such as Arabic and Greek, were enabled for the users, the use of Arabizi or Greeklish became a choice rather than necessity. For many users, the use of these types of transliteration is not triggered by software constraints, but rather is produced through the transient manipulations of linguistic resources that may extend beyond the speakers' assumed national or standard language (Spilioti, 2019).

As discussed above, translanguaging is a theoretical approach to practices in which multilingual people integrate the semiotic resources associated with two or more linguistic systems (García & Li, 2014). In line with this analogy, Androutsopoulos (2015) describes the process of the representation of a language in another spelling or script, and in his research on networked multilingualism coined the term 'trans-scripting' to describe literary digital practices whereby the features of one of the available languages are represented in the spelling or script of another language. Furthermore, Spilioti (2019) argues that previous research on Romanised scripts focused on the languages that use such scripts in terms of how, when, and to what end. However, considering the translanguaging shift in the field, there is a pivot of focus from the language norms and regularities associated with certain groups, to the very act of respelling as a creative or transformative act. Hence, Spilioti (2019, p. 5) defines trans-scripting as "a process or respelling that creatively manipulates available resources associated with multiple languages, scripts and other modes for multiple

networked audiences", while Androutsopoulos (2020) discusses trans-scripting in the context of examining practices of writing and used 'Hellenised English' (i.e., English in Greek script) as a case study. Androutsopoulos (2020, p. 2) refers to trans-scripting as "the practice of selecting non-canonical script or orthography to graphically represent a language in ways that are neither socially expected nor technologically determined". The research analysed the Hellenised Greek that appeared in the subtitles of YouTube videos for a former Greek Prime Minister speaking English, with Androutsopoulos (2020) arguing that trans-scripting is a semiotic strategy that is mobilised for language-ideological purposes by evoking cultural stereotypes rather than a neutral representation of linguistic content.

Studies also explore the attitudes of users towards using Romanised varieties. Hamdan (2017) investigated the attitudes of university students towards the use of Romanised Arabic in CMC, finding the student's major and the language of instruction to be an important variable in determining the use or non-use of Romanised Arabic. For example, the students of English and students of Medicine who received their instruction in English were more in favour of Romanised Arabic than the students of Arabic and Islamic Sharia who received their instruction in Standard Arabic. Moreover, the majority of both the users and non-users of Romanisation agreed that it might endanger the Arabic language if its use continued.

3.3.4.2 Language choice online

Language choice in digital communication is concerned with the resources available to the online users and how they negotiate their language preferences when communicating with others (Lee, 2016). The development of the Internet technology provided opportunities for users to draw upon a broad range of resources that can be referred to as technological affordances (Herring & Androutsopoulos, 2015). Lee (2017) identifies the resources available to online users as representational resources, human resources, ideological resources, and technological resources. The list below shows an example of the range of meaning-making resources drawn on by a multilingual Internet user presented by Lee (2017, p. 24):

1. Representational resources
    a. Languages (e.g., regional and social varieties)
    b. Genres
    c. Scripts (e.g., Romanisation)
    d. Modes (e.g., writing, images, animations, and gestures)

2. Human resources
    a. Interactants
    b. Mediators
    c. Others who contribute directly or indirectly to text
3. Ideological resources
    a. People's perceptions (e.g., values, thoughts, and feelings)
    b. People's everyday experiences in using texts
4. Technological resources
    a. Software (e.g., chat applications)
    b. Hardware (e.g., computers and smartphones)
    c. Access to the Internet
    d. Network resources (e.g., copy/paste and online translations)

Lee created this list following the analyses of resources utilised in multilingual online contexts in Hong Kong (cf. Lee, 2007), where she identified five major linguistic resources available to the study participants: English, standard written Chinese, Cantonese in characters, Romanised Cantonese, and morpheme-by-morpheme literal translation. Similar to Arabic and Greek, her study shows how these different varieties of Chinese are employed for different purposes by different users. Furthermore, the study reveals the participants' consideration of certain factors, namely, the expressiveness of the language, the perceived functions of the communication medium, the user's familiarity with the language, the user's identification with the language, the technical constraints of the input methods, speed, and the perceived practicality of the writing system. This same classification of languages and factors can be applied to many multilingual contexts in the world, including the Arab context that has a similar situation in terms of the non-Romanised script.

It is broadly accepted that English is one of the most important linguistic resources available in multilingual online communication. Wright (2004) compared language use online by educated speakers in several countries (i.e., France, Italy, Japan, Macedonia, Oman, Poland, Tanzania, Ukraine, and Indonesia), finding that the participants' choice of language varied according to the online content and communication mode. The study also suggested that to the extent that Internet resources become available in the users' own language, English language usage would decrease. On the other hand, other research suggests that English is a common language used for communication in multilingual online settings. Durham (2007) examined language choices on a mailing list for members of a pan-Swiss medical student organisation, and found that English was the dominant language of interaction among a group of French-, Italian- and German-speaking students. According

to Durham's findings, English was a natural choice for the participants since it was an accepted lingua franca in Switzerland. Additionally, Lee and Barton (2011) studied the content of 100 Flickr sites (providing image and video hosting, and an online community), in addition to the users' texts, and found that 51% of the sites contained English only and over 70% of the user profiles were in English. In the Arabic context, Warschauer et al. (2002) found that Egyptian professionals used only English in their emails, particularly in formal communications. Surprisingly, the common form of writing in Egypt, as well as all other Arab countries (i.e., MSA) was almost absent in their data. On the other hand, a Romanised version of Egyptian CA, which does not have a standardised writing system, was found to be favoured in personal communication. The current study supports the argument that the mode of communication and formality are important factors behind users' preferences for linguistic resources online.

Several studies also examined how different local dialects are written online. For example, Siebenhaar (2006) found that Swiss–German dialects were prominent in IRC chatrooms among German-speaking Swiss communities, despite not having a standardised written form. Androutsopoulos (2013b) explored the representation of German dialects on YouTube and examined the multimodal performance of dialect in videos and the negotiation of these performances in the audience comments, finding many of the YouTube videos to be tagged with keywords from different German dialects, which indicates the increasing visibility of different local dialects in social media.

3.3.4.3 Code-switching online

This section turns to code-switching and the role it plays in digital communication. It must be highlighted that the majority of the studies on code-switching have focused on speech, since it was assumed that written language tends to have fixed rules of orthography (Sebba et al., 2012). However, Sebba (2012) argues that there is a large volume of ancient written data that includes texts written in more than one language. Moreover, the development in technology and new digital platforms with all the language affordances on offer can now produce written data that resemble spoken language, thus representing a rich area for investigation, as explored here.

Code-switching is a common linguistic phenomenon observed in many multilingual contexts. It has been observed that Internet users do not always employ one single language for communication, and that different patterns and discourse functions for code-switching exist in digital communication (cf. Androutsopoulos, 2013a; Zhang, 2015). The most classic definition for code-switching is provided by Gumperz (1982, p. 59) as: "the

juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems". As Romaine (1995) explains, the term 'code' refers not only to different languages, but also to varieties of the same language, as well as styles within a language (e.g., formal register). Hence, in the current study, the term 'code' will be utilised as a form of linguistic resource that is available for Internet users.

Following Gumperz's conceptualisation of code-switching, different researchers focused their attention on different aspects of the phenomenon. Myers-Scotton (1993) presented the Matrix Language Frame model, which was employed for the analysis of code-switching in speech and shows that one language functions as the foundation to form the grammatical basis for the mixed-language sentence. The theoretical models on code-switching presented by Myers-Scotton (1993) attempt to identify the base (matrix) language of code-switching according to syntactic and morphological criteria. Myers-Scotten (1993) argues that one of the languages involved in code-switching plays a dominant role and is labelled the 'Matrix Language', and that its grammar sets the morphosyntactic frame. In other words, the morphosyntactic frame determines the dominant language of the tweet. On the other hand, the language that plays a less dominant role is referred to as the (embedded) code. This will be further explained in the Methodology chapter as the distinction between the matrix code and the embedded code was important for the coding of the tweets. Furthermore, another popular approach applied in code-switching research is the conversational approach presented by Auer (1999) that provides a distinction between two kinds of switching: (i) insertional switching, in which one language serves as the base language, into which words from another language are inserted; and (ii) alternational switching, which involves switching between languages across sentences.

Different terms have been introduced in the literature based on the definition of code-switching, including the term 'code-mixing'. The distinction between the terms code-switching and code-mixing is controversial. Some scholars use both terms interchangeably (cf. Poplack, 1980), while others differentiate between the two (cf. Hoffmann, 1991). According to those who differentiate between the two terms, code-switches occur across phrase or sentence boundaries, while code-mixes occur within sentences and usually involve single lexical items. Hoffmann (1991, p. 104) discusses the difference between code-switching and code-mixing in terms of the lexical, or the phrasal and sentence level:

> Switches occurring at the lexical level within a sentence (intra-sentential switches) are referred to as 'code-mixes' and 'code-mixing'. On the other hand, changes over phrases or sentences (inter-sentential), including tags and exclamations at either end of the sentence, are called 'code-switches' and 'code-switching'.

This research will not differentiate between code-switching and code-mixing, and the term code-switching will be used to describe the coexistence of multiple languages, varieties, or resources in an online interaction. However, the current research differentiates between intra-tweet code-switching and inter-tweet code-switching, to distinguish between the code-switching instances that manifest within the boundaries of a tweet, and in separate tweets but on the same timeline of the selected accounts, as explained further in the Methodology chapter.

Another common phenomenon in multilingual situations that is usually discussed within code-switching discourse is 'borrowing', which involves the "adaptation of lexical material to the patterns of the recipient language" (Poplack & Meechan, 1995, p. 200). This adaptation takes the form of morphological and phonological integration of different degrees into the speech of both bilingual and monolingual individuals. These borrowed words are even used by monolinguals, who may or may not be aware of their foreign origins (Romaine, 1995, p. 55). Furthermore, Grosjean (1982) describes language borrowings as lexical items which, although originally from language A, have come to be integrated into language B such that language B speakers use them without any awareness of their foreign origin. The language items that have undergone the process of borrowing are referred to as 'borrowings' or 'loans'. For the sake of consistency, the term 'borrowing' will be utilised throughout this study. It will be further explained in the Methodology chapter how borrowings are identified in the current research. Most languages borrow from others. English, for example, has borrowed many words from different languages over the centuries such as *jungle* (Hindi) and *petite* (French). Conversely, English has loaned many words to other languages such as *computer*, *television*, and *telephone*. It is important to remain mindful that the process of adapting a foreign word into the local or native language takes time. Myers-Scotton (1993) hypothesises that high frequency borrowed forms probably enter the language initially as code-switches, and then gradually become integrated into the language as borrowings since they are employed more frequently alongside the native forms they often duplicate. Weinreich (1963) provides three important reasons for the occurrence of borrowing: (i) universality, as all languages need to name new things, persons, places, and concepts that do not have any linguistic equivalent in the native language; (ii) the low frequency of certain words in the language; and (iii) to resolve the clash of homonyms. As in any other language, borrowing in Arabic represents a strategy to expand its vocabulary in order to cope with the new terminologies introduced in the global domain.

It can be inferred from the above that in code-switching situations, the speaker is expected to have a certain proficiency level in the language, while in borrowing situations the

expressions are employed as part of the speaker's native language. However, since the distinction between code-switching and borrowing is not straightforward, scholars have attempted to set different criteria to mark the difference between the two phenomena. The hypothesis that code-switching entails two grammars, whereas borrowing only entails one (Poplack et al., 1988) is an essential assumption of differentiating between borrowing and code-switching. Moreover, the frequency hypothesis criterion, which is the degree of occurrence of a lexical item in the host language usage, has also been considered in distinguishing between code-switching and borrowing. Myers-Scotton (1988, p. 70) asserts that borrowed forms "should be distinguishable from the embedded language which speakers know in some abstract sense, borrowings are available to many (or all) speakers, while embedded forms in switching are not". In other words, borrowings find their way to become part of the natural lexicon of speakers, while code-switched items are still considered foreign words. Myers-Scotton (1992) compares code-switching to borrowing and again argues that borrowed words are usually more recurrent than code-switched items. For example, the English word *Internet* that appears recurrently in many languages is a case of borrowing. The distinction between borrowing and code-switching is essential for the coding process of the data, as shall be discussed in the Methodology chapter.

In addition to the term code switching, the term diglossic switching is used in the current study to describe the code-switching between MSA and CA. As discussed in section 2.2.1, diglossia refers to the idea that a high variety (H) and a low variety (L) can be spoken simultaneously (Ferguson 1959). Following this definition, Heath (1989) uses the term diglossic switching when referring to the switch that occurs between Moroccan Colloquial Arabic and Classical Arabic. Furthermore, the term was also used by Bassiouney (2020) to describe the same phenomenon in Egypatian context. As such, the term diglossic switching is used throughout the analysis chapters.

In online communication, Lee (2017) argues that instances of code-switching contribute to the overall interactivity of an online exchange. For example, if someone is using one language for a caption and responds to a comment in another language, then this may not be a case of code-switching as the two may not be 'dialogically interrelated' (Androutsopoulos, 2013a, p. 673). Gumperz's, Myers-Scotton's, and Auer's approaches, which were developed for spoken interaction and have been discussed above, were adopted by several digital communication studies (eg. Androutsopoulos, 2006b; Siebenhaar, 2006; Leppänen, 2007). The reason for this is that one of the early assumptions about language use in digital communication is its resemblance to speech. However, Sebba (2012) argues that these models were developed for spoken interaction

and may not dovetail into written texts such as digital communication. For instance, one important feature of written and online code-switching not covered in the approaches developed for spoken interaction is 'script-switching', which involves the alternation between orthographic resources (Lee, 2017). If the users are switching the script, then it is clearly a change to another language. However, in cases of borrowing, the change to a different language may not be realised. It is worth mentioning that script-switching is very common in Arabic online communication, with different examples illustrated by Palfreyman and Khalil (2003), as discussed in section 3.3.4.1.

There has been an increase in the interest of exploring code-switching in digital communication, with Androutsopoulos (2013a) and Lee (2017) offering a summary of a selection of studies on code-switching in digital communication in the 1996–2015 period. Table 3.5 lists the works by author, platform, involved languages, and participants to highlight the concepts and themes that were explored.

Table 3.5: Selected research on code-switching in digital communication (adapted from Androutsopoulos, 2013a, p. 674; Lee, 2017, p. 42)

Author	Platform	Code-switching languages	Participants	Concepts and Themes
McClure (2001)	Mailing lists. email	English, Assyrian, Greek, English	Ethnic minorities, friends	Language maintenance, identity, self-presentation
Georgakopoulou (1997, 2004, 2011)	Email	Greek, English	Friends	Identity, self-presentation
Lam (2004)	Chatroom	Romanised Cantonese, English	Immigrants, English as second language students in the United States	Literacy
Hinrichs (2006)	Email	Jamaican Creole, English	University students, Jamaican diaspora	Functions, identity
Siebenhaar (2006)	IRC	Swiss German, Standard German	Youth	Code choice, language variation

Androutsopoulos (2006b, 2007)	Forums	German/Greek, Persian, Hindi, Arabic	Ethnic minorities	Identity
Chen (2007)	Bulletin board systems	Chinese (Mandarin), English, Taiwanese	Adolescents	Discourse functions
Lee (2007)	Email, ICQ	Chinese (Cantonese), English	University students	Effects of synchronicity
Su (2007)	Bulletin board systems	Chinese (Mandarin), Taiwanese, English	College students	Indexicality, playfulness
Deumert and Masinyana (2008)	SMS	Xhosa, English	Young adults	Homogenisation
Warschauer et al. (2007)	Email	English, Egyptian Arabic	Young professionals	Language choice
Huang (2009)	Bulletin board systems, email	Chinese (Mandarin), English, Taiwanese	Chinese–English bilingual university students	Script-switching
Leppänen et al. (2009, 2011)	Blogs, forums, fan fiction	Finnish, English	Young people	Translocality, identity
Spilioti (2009)	SMS	Standard Greek, Greeklish, English	Youth	Code-switching versus borrowing
Vandekerckhove and Nobels (2010)	MSN Messenger	West Flemish, Dutch	Teenagers	Code-eclecticism
Lexander (2012)	SMS	Wolof/Pulaar, French	Students in Senegal	Literacy practices

Seargeant et al. (2012)	Facebook	Thai, English	Young adults, Thai native speakers	Addressivity/ Audience design
Androutsopoulos (2015)	Facebook	Greek, German, English	German–Greek secondary school students	Networked multilingualism
Kytölä (2013)	Forum	Finnish, English	Football fans	Metapragmatics
Bali et al. (2014)	Facebook (pages)	English, Hindi	English–Hindi bilinguals	Natural language processing
Halim and Maros (2014)	Facebook (status updates)	Malay, English	Bilingual adults, English teachers	Discourse functions
Jaworska (2014)	Forums	German, English	English-speaking German expatriates in Great Britain	Language play, networked multilingualism, translanguaging, poly-/metro-lingualism
Themistocleous (2015)	IRC	Cypriot, Standard Greek	Greek-Cypriots	Identity
Thorne and Ivković (2015)	YouTube	Multiple	YouTube commenters	Linguistic landscape, pluralingualism
Zhang (2012, 2015)	Douban, Weibo, Youku	Chinese (Mandarin), English	Commenters on government microblog	Identity, language play

As seen in Table 3.5, researchers examined a range of platforms depending on their popularity at the time the study was conducted. Code-switching research on social media platforms such as Facebook and Twitter became popular in 2012 (e.g., Seargeant et al., 2012), and with these platforms still representing an important medium for online

communication, the area requires further research and analysis, as noted by Androutsopoulos (2013a). It is also clear from Table 3.5 that research concerning young people dominates this field (e.g., Hinrichs, 2006; Siebenhaar, 2006; Su, 2007; Spilioti, 2009). Another common focus of users involves migrants and minorities (e.g., McClure, 2001; Lam, 2004; Androutsopoulos, 2007), while other researchers selected users based on their profession (e.g., Warschauer et al., 2002; Zhang, 2012). However, no study in Table 3.5 compared the practices of different groups of social media users, which represents one of the important scopes of the current research.

Following the discussion on code-switching online, it is important to consider certain websites or social media platforms, where the language of posts may differ to the languages or varieties displayed in the comments. Furthermore, the same user or different online contacts may display wall posts in different languages. These are thus multilingual discourse spaces, but they do not automatically constitute instances of code-switching (Androutsopoulos, 2011). This multilingual discourse space cannot be conceived of as part of one 'episode', but what ties the elements together is their spatial coexistence in product and reception, as opposed to their dialogic orientation to each other. Furthermore, while these units, such as posts and tweets, comprise multilingual web surfaces, they are often monolingual in themselves despite nothing preventing some of these units from containing code-switching (Androutsopoulos, 2013a). Hence, Androutsopoulos (2013a) distinguishes code-switching from other patterns of multilingualism on four counts:

1) The multilingual Internet
2) The coexistence of different languages on a web page or comment thread
3) Language choices for emblems
4) Sequential language choices lacking a dialogical interrelation

Following this distinction between code-switching and digital multilingualism, it is useful for the analysis of Twitter data to consider that multilingualism presents itself in the two forms of inter-tweet and intra-tweet code-switching, as shall be discussed in the Methodology chapter.

3.3.4.4 Functions of code-switching online

In addition to identifying the digital multilingualism on the various CMC platforms, it is vital to comprehend the functions and motivations of the code-switching instances in their context. This section first reviews the functions of code-switching online and then moves on

to presenting hashtags as an example of code-switching on Twitter. Finally, it highlights a sample of the studies on code-switching online in the Arab context.

As mentioned in section 1.1, the use of different language varieties on Twitter is not the result of spontaneous language production, as typing the tweet involves awareness and self-consciousness before posting it publicly. This makes online code switching distinct from face-to-face spoken code-switching. However, considering the assumption that some features of digital communication are similar to speech, it is expected that in online discourse code-switching serves similar functions to that in speech. Scholars have investigated the functions of code-switching (e.g., Grosjean, 1982; Gumperz, 1982; Auer, 1995). For example, Grosjean (1982, p. 152) presented the following reasons for code-switching in a conversational setting:

- Fulfil a linguistic need for a lexical item, set phrase, discourse marker, or sentence filler
- Continue the last language utilised in the conversation
- Quote someone
- Specify the addressee
- Qualify the message through amplification or emphasis
- Specify the speaker's involvement (personalisation)
- Mark and emphasise the group identity (solidarity)
- Convey confidentiality, anger or annoyance
- Exclude someone from the conversation
- Change the speaker's role: raise the status, add authority or demonstrate expertise

While some functions of spoken code-switching resemble those in digital communication, there are new functions that have been identified online. Androutsopoulos (2013a, p. 681) identifies the following functions of code-switching in this online context:

- Formulaic purposes such as greetings, farewells, and best wishes
- Performing culturally specific genres such as poetry and jokes
- Conveying reported speech
- Repetition of an utterance for emphatic purposes

- Switching to index one particular addressee, or to respond to language choices through proceeding contributions
- Contextualising a shift of topic or perspective, to distinguish between facts and opinion, or information and effect
- Switching to mark an utterance as jocular or serious, and to mitigate potential face-threatening actions
- Switching to or from the interlocutor's code to index consent or dissent, agreement or conflict, alignment or distancing, and so forth

Despite the functions provided by Androutsopoulos (2013a) accounting for digital communication in general, other studies look specifically at the functions in the discourse of social media. For example, Halim and Maros (2014) analysed the functions of code-switching on Facebook written by Malay–English bilingual users, where many of the functions echo those provided above by Grosjean (1982) and Androutsopoulos (2013a) such as code-switching for quotation, addressee specification, reiteration, message qualification, clarification, and emphasis. In addition, Halim and Maros (2014, pp. 131–132) identified further new functions:

- Switching for checking, employed to seek approval and confirmation from the speaker. For example, switching to a Malay tag '*kan*' to seek agreement for a statement made: "It's funny when I know what you know… Kan?" (Right?).
- Switching for availability, used when a word or expression is only available in a particular language. For example, in the post "Celebrating my personal '*hijrah*'" (starting to live in a more positive or more righteous manner), where the word *hijrah* does not have an equivalent in English.
- Switching for principles of economy, whereby the user tends to choose the shortest and easiest words with which to communicate.
- Free-switching, referring to instances of code-switching that do not serve any specific pragmatic or discourse-related function.

Another unique feature of social media is the availability of different forms for communication in every platform. For example, users on Twitter can type a post in one language and then use a hashtag in another language.

Research on hashtags also represents a rich area for research in the context of code-switching on social media platforms that has attracted a number of researchers (e.g., Page, 2012; Zappavigna, 2015). According to Lee (2017), it is debatable whether the coexistence of the different languages in the post and the hashtag can be considered as one coherent

text or as separate text, especially since hashtags serve their own specific communicative functions. For the current study, hashtags will be considered as part of the text, as in many instances they are blended with the same text, as shown in Figure 3.2.

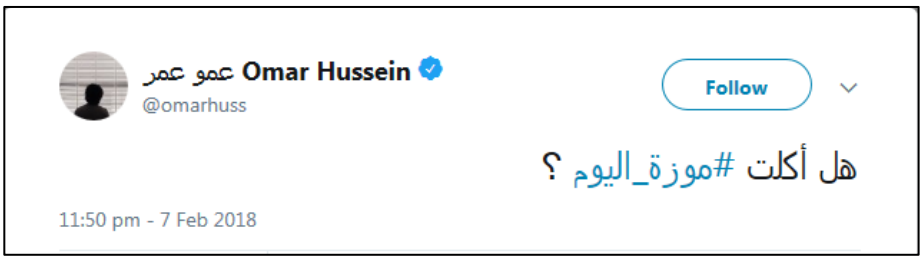


Figure 3.2: A tweet with a hashtag in the text

Moving into code-switching online in the Arab context, different studies have investigated the functions of code-switching on digital platforms. For example, Warschauer et al. (2002) found that participants switched to Egyptian Arabic to express personal emotions specifically, when those emotions could not be clearly expressed in English. Furthermore, Al-Khatib and Sabbah (2008) added more functions for code-switching between Arabic and English in their research on mobile text messages among Jordanian university students, finding that Arabic was utilised for greeting, quotations, cultural, and religious purposes. On the other hand, English was employed for showing prestige, mentioning academic terms, and for discussing taboo or offensive topics. In terms of code-switching between the different varieties of Arabic, Albirini (2016) studied the reasons for switching between MSA and CA on Facebook, and reported that the shift from CA to MSA was utilised to underscore an important point in an utterance, request emphasis and attention, introduce quotations, switch to a serious tone, produce rhyming stretches of discourse, adopt a pedantic tone, and introduce a pan-Arab and pan-Muslim identity. In contrast, switching from MSA to CA was used to adopt a comic tone, simplify ideas, insult others, present everyday sayings, devalue a specific section of a conversation, and introduce indirect quotations.

Code-switching in digital communication represents a complex topic, considering the different formats and resources available for the multilingual users. However, based on the discussion above, the current research will consider the functions recorded in all these mentioned studies when analysing the tweets, with the functions grouped into six categories. This will be further explored with examples from the data in the Methodology chapter, along with examples from the data. Furthermore, it is essential for the analyses of the functions of code-switching on digital communication to be supported by the users' perspectives in order to interpret the meanings of these instances, which will be achieved

in this research by conducting interviews with selected users. This section now moves on to provide a discussion on the functions of the paralinguistic cues, focusing mainly on emojis.

3.3.4.5 The paralinguistic function of emojis

An important element that constantly appears across digital communication platforms is the emoji, since communication on social media differs from face-to-face conversations in that facial expressions and body language are excluded. According to Tantawi and Rosson (2019), this non-verbal communication plays a 'paralinguistic' role in communication.

The word 'emoji' is drawn from Japanese and it is composed of the kanji for 'picture' (e-) and 'character' (-moji) (Seargeant, 2019). Since emojis can be defined as an "abstract representation of facial expressions and body language" (Tantawi and Rosson (2019, p. 275), they are utilised along with text in digital communication to enhance the emotion in the content. The current research examines the occurrence of emojis in tweets and compares how the different groups (i.e., corporations, influencers, and ordinary users) use these paralinguistic cues.

Emojis emerged from an earlier form of keyboard-based emotional expression named 'emoticons', which are keystrokes arranged to visually convey a facial expression (Tantawi & Rosson, 2019). Emojis became more widely employed as they are visually richer, more expressive, and more complex than emoticons (Herring & Dainas, 2017). The popularity of emoji use resulted in the decrease of emoticons employed on social media (Pavalanathan & Eisenstein, 2016), which indicates that emojis now play a more significant role than other paralinguistic cues such as emoticons. Nevertheless, this research also explores other paralinguistic cues employed by users and compares them with emojis. Tantawi and Rosson (2019) argue that when emojis are used in conjunction with words, they are more precisely interpreted with regards to the emotion(s) they convey than when they are utilised independently without any words. However, emojis are still employed alone on different social media platforms, such as Twitter, to convey a message. In terms of the functions of emojis, there are six main uses: (i) to express emotion (Alismail & Zhang, 2018), (ii) for the modification of tone (Herring & Dainas, 2017), (iii) as a substitution for non-verbal behaviour (Pavalanathan & Eisenstein, 2016), (iv) for opening and closing conversations (Danesi, 2017), (v) when the user has nothing to say (Danesi, 2017), and (iv) to react to a response from another user, as well as describing a physical action (Herring and Dainas, 2017). Although the quantitative analysis of the current study does not include the functions of the emojis in the tweets, it will be useful to keep these functions in mind during the discussion.

After considering the literature on multilingualism and online communication, the next section discusses another important area that is relevant to the current research, which is self-presentation and identity in digital communication.

3.4 Self-presentation and identity in digital communication

Digital communication allows users to form relationships and communities with others that they do not necessarily know offline. Therefore, this interaction and the identity negotiated with the virtual audience is an area worthy of investigation. Furthermore, negotiating language choice and alternating linguistic codes serve as an important resource for self-presentation and identity construction online (Lee, 2017). This section commences with a brief review of the linguistic identity research, before progressing to a more focused discussion on identity in digital communication.

3.4.1 Self-presentation theories

This section explores the theory of self-presentation, which offers theoretical perspectives on how individuals present themselves and facets of their identity to the virtual audience on social media in general, and on Twitter in particular, by employing certain languages and varieties. This is in line with the current study's focus on investigating the languages and varieties used by corporations, influencers, and ordinary users within the frame of self-presentation theory.

Erving Goffman pioneered identity research in *The Presentation of Self in Everyday Life* (1959), which discusses how individual and team performances are constructed and maintained. Although Goffman's theory was developed to study individual performance in physical environments, it could be applied to online settings (Vaast, 2007). Goffman (1959) argues that when an individual communicates with other members of the community, he/she begins to manage the impression on the audience through adjustments in appearance, situation, and behaviour. At the same time, the community or the audience will decide whether they want to continue interacting with this individual. The theory of self-presentation has been utilised in different disciplines, including digital communication, with Rui and Stefanone (2013) emphasising that the Internet provides new opportunities for self-presentation, particularly on social networking sites and applications that allow users to strategically create their profile pages and reveal information about themselves. Therefore, Twitter, among other social media platforms, is considered a tool that helps users to present themselves online by expressing their opinions, and therefore the language or variety they use is an important component of this self-presentation.

Researchers have debated the management of self-presentation in online versus offline settings. For example, Vaast (2007) argues that online settings may limit the ability of individuals to control their self-presentation to the audience, since many factors that individuals rely upon in physical environments are absent in online settings (e.g., clothing and body language). In contrast, boyd (2007) and Krämer and Winter (2008) suggest that online self-presentation is more manageable than its offline counterpart as the individual can more carefully select the information, and the language, presented to the virtual audience. However, online self-presentation may become complicated as it is sometimes performed for an anonymous and varied audience (boyd, 2007), such as the case of corporations and social media influencers. In many cases, Twitter users interact with an unknown audience, and adjust their communication based on the audience's responses and comments, thus influencing the user's language usage. In this way, the imagined audience becomes known to the users (Marwick & boyd, 2011).

Graham (2015) presents historical perspectives on linguistic identity research and discusses three major research models in this area: (i) social constructivist theory, which is built on Goffman's work (cf. de Fina et al., 2006); (ii) positioning theory (cf. Van Langenhove & Harré, 1999); and (iii) categorisation membership theory (cf. Antaki & Widdicombe, 1998). Graham (2015, p. 307) argues that all these theories co-exist in any interaction, while observing that "we index different identity *categories* as one strategy to *position* ourselves in relation to others, and as our interactions with others unfold, our identities shift and *emerge* according to our conversational and interactional needs and desires".

After considering the theories that form the basis for self-presentation, the following section discusses self-presentation and identities within the context of digital communication.

3.4.2 Multiple identities online

New technologies and social media are important factors influencing how identity is presented. Therefore, identity or self-presentation online does not merely concern who we are, but also who we want to be perceived as by others, as well as how others see us (Barton & Lee, 2013). Furthermore, individuals present themselves differently online and they may, for example, create one persona on a separate account for their parents and family, and another account for friends (Zhao et al., 2008)

A Slice of Life in my Virtual Community (Rheingold, 1993) is one of the most commonly discussed articles in the online identity domain, where a pragmatic case is formed for the emergence of communal relations online, with the emphasis on people using words on

screens to engage in the full range of social activities. Rheingold proposes that computers have introduced a new form of social life known as 'virtual communities', which consist of groups of people linked by their participation in computer networks. Those embedded in such virtual communities share many of the characteristics of people in ordinary communities, except that they have no face-to-face contact, are not bound by the constraints of time or place, and use computers to communicate with one another. Warschauer et al. (2002), as discussed above, also reveal how identities are negotiated via language, through their conclusion that Egyptian professionals reserved English for formal email, while a Romanised form of Arabic was more prominent in informal email and chat. Furthermore, Georgakopoulou (1997) showed that switching between English and Greek in email by Greek teachers was carried out to enhance solidarity, as well as to indicate professional in-group membership. Meanwhile, in Senegal, Lexander (2012) found French, rather than African languages, to be the dominant language of romantic text messages among lovers.

In examining online relations and identity that influence the strategies employed by online users, Graham (2015) notes three important factors, namely, the perceptions/expectations of the audience, the limitations/capabilities of different media, and the interactional goals. All these factors are crucial in examining identities, especially when different languages are involved (e.g., Georgakopoulou, 2011; Lee, 2016). Previous studies in the area show how online users employ different languages, varieties, or typographies to embed themselves as part of a subculture. Others may use language or variety atypical to daily conversation in order to reflect their local identity. Hence, online participants may index their identities through multiple language switching (Barton & Lee, 2013). It is also worth noting that the conscious positioning of oneself using a certain language or variety is helpful in discovering the attitude towards these languages and varieties, and represents an important area to be explored in the interviews of the current study.

Identity performance on a public platform such as Twitter differs from that on private platforms such as email or Facebook, especially since a different audience is involved. Multilingual practices are salient in public platforms. Barton and Lee (2013) discuss 'glocalisation' and how 'glocal' identities are constructed in public online spaces by Flickr users, underscored by the observation that several of the study informants reported that although they did not need to use English in other areas of life, they interacted with other online members in English. On the other hand, in some situations users want to be part of the global world without conceding their existing identities, and therefore they use their local language. The negotiation between global and local identities, and language online and

writing multilingually, is one of the essential literacy practices for the projection of glocal identities online (Barton & Lee, 2013). For example, the use of Romanised texts in Cantonese and Greek was found to signal technological competence, cosmopolitan outlook, and a glocal identity (Spilioti, 2009; Lee, 2017). Since the language choice employed indicates the identity users aim to project, it can be utilised as an important strategy to include or exclude certain types of virtual audience or followers (Tagg, 2015).

3.5 Branding and self-branding

In digital communication, brand-building has emerged as a new challenge for corporations and public figures, with novel technologies offering the audience the ability to opt-out of advertisements, unlike television channels, for example. Therefore, many corporations hire creative agencies and armies of technologists to insert their brands throughout the digital universe (Holt, 2016). The general strategy is to build direct relationships with the target audience, and connect with them in real time. In this section, the definitions of branding and self-branding are first discussed, along with theories relevant to the current research. Then, corporate digital communication is explored, before the final section provides an overview of social media influencers and celebrities.

3.5.1 Definitions and relevant theories

Wilson (2020, p. 1179) defines branding as:

> a strategic and concise means of expressing, amplifying, and controlling: who you are, what you do, who you are associated with, and what you own – for competitive gains, using a variety of communication methods and media, that are both collaborative and measurable.

This implies that branding represents the strategy of forming the identity and the communication strategy shared with an intended audience.

Branding is always associated with marketing and public relations, as the three concepts are vital components in the communication strategy of corporations. Hence, it is useful to consider the definitions of marketing and public relations, as well as the theories relevant to branding.

The American Marketing Association (2022) defines marketing on their webpage as “the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large”. At the heart of the definition is communication and exchange, which implies that such communication is now bi-directional. Kotler et al. (2019) claim that the character of

marketing has shifted from product-centric (Marketing 1.0) to customer-focused marketing (Marketing 2.0), and is now founded on the target customer and referred to as 'value-driven marketing' (Marketing 3.0). Furthermore, Kotler et al. link the changes in technology and the evolution of social media to a new wave of marketing, which is relevant to this study.

In terms of public relations, the practice has also shifted from uni-directional to bi-directional communication, as highlighted in the following definition provided by the Public Relations Society of America (2022) on their webpage: "Public relations is a strategic communication process that builds mutually beneficial relationships between organizations and their publics".

In summary, these definitions indicate that successful branding occurs when the brand becomes and remains known to the audience through the use of marketing and public-relations strategies. It is important to consider what the brand means for the audience, as well as the aforementioned importance of two-way communication and community interaction, which are pertinent to the notion of 'crowdculture'. Holt (2016) argues that companies should target crowdculture on social media to achieve effective branding. Social network platforms, including Twitter, allows fans to create a huge community around known entertainment personalities such as performers and athletes, who have considerable popularity on social media. For instance, the numbers of Twitter followers of sport stars and teams such as Cristiano Ronaldo (99.5 million), Neymar (56.6 million), FC Barcelona (41.2 million), and Real Madrid (13.9 million) are far higher than the sports brands Nike (9.1 million) and Adidas (4.1 million). These online celebrities have a community of followers ('crowdculture') that interact with them through likes and tweets. Therefore, companies now employ brand ambassadors who reach out to and engage with the followers in real time to advertise the brand. The current study thus considers the language employed by corporations and social media influencers, as they sometimes work in unison to deliver the communication strategy of a brand.

Khamis et al. (2017) assert that self-branding involves individuals developing a distinctive public image for commercial gain and/or cultural capital. This public identity is responsive to the needs and interests of the target audience. Marwick (2010) claims that self-branding primarily involves a series of marketing strategies applied to individuals, while in his article titled *The Brand Called You*, Peters (1997) argues that everyone has the potential to stand out, and that individuals must identify with their brand identity to be distinctive and project a memorable self-image. Peters adds that applying the concept of branding to individuals is not only possible, but also imperative and inevitable. However, the concept of self-branding may raise practical and ethical concerns, especially since it constructs an implicit

assumption that everyone is expected to self-brand to reach his or her potential (Khamis et al., 2017).

Despite these issues, the popularity of self-branding has prevailed, with many individuals, whether famous offline or not, seeking fame on social media and thus building a large number of followers. The various social media platforms, including Twitter, provide an opportunity for everyone to become famous, even if they lack a strong public identity or the resources to self-promote on the scale of established celebrities (Khamis et al., 2017). For example, Labrecque et al. (2011) emphasise that experts in website development are no longer necessary since anyone can easily upload text, images, and videos from their own personal devices. Moreover, Page (2012) mentions that self-branding operates on a spectrum that includes corporations who personalise their identity, the use of branded products to signify status and identity, and the production of public personae. As the technological barriers are crumbling, online platforms are ideal for personal branding. In addition, social media provides the potential for all to share their daily lives with virtual friends and audiences. This is exemplified in the story of Faris Al Turki, a Saudi entrepreneur who started the Arabic hashtag #Farisbreakfast in 2011, and asked Twitter users to share images of their breakfasts using the hashtag. He later became an influencer on Twitter, and currently has over 260,000 followers. Moreover, he developed his idea into a restaurant that serves all-day breakfast dishes in seven branches in three major cities of Saudi Arabia: Riyadh, Jeddah, and Makkah. This highlights the potential of self-branding to transform ordinary users' lives through sharing their identities and interests with large audiences.

3.5.2 Corporate digital communication

One of the new areas of investigation within the domain of digital communication research is corporate communication, with the current study aiming to investigate the different language varieties employed within digital communication. The online presence of corporations, institutions, and different individuals through websites and social media plays a significant role in how they communicate. Kaplan and Haenlein (2010) stress the potential of social media for redefining corporate communication and provide guidelines for companies seeking to integrate social media into their communication strategy. The growing number of Internet users has attracted companies and institutions to develop creative social media strategies, with companies increasingly realising the brand impact of tweeted news, publicised 'likes', and shared and uploaded content, and thus encourage their followers to share their opinions, images, and videos related to the brand (Puschmann & Hagelmoser, 2015). This online presence is now an important requirement for any business, especially in terms of branding and public relations (Kaplan & Haenlein, 2010). In fact, as discussed

in section 3.5.1, the presence on digital platforms is even more important than other conventional media platforms that lack the interactional component of interpersonal communication (Puschmann & Hagelmoser, 2015). In social media, this interactional element is crucial for the success of a company's social media communication strategy.

There is, of course, a difference between the communication of corporations and the personal communication of individuals, as discussed in section 3.5.1. Companies and institutions strive to develop corporate social media strategies characterised by a strong emphasis on strategic aims, which are pursued by professionals (e.g., marketing and public-relations experts) and are generally stable and consistent over a period of time (Horton, 1995). On the other hand, private personal communication revolves around the negotiation of social relationships and identity representation through the expression of thoughts or emotions (Walther, 2007). There are a range of reasons why corporations and institutions should pay close attention to their digital communication strategies. Horton (1995) mentions that marketing is the most typical form of external corporate communication that seeks to provide customers with product-related information to persuade them to make a purchase. Since the Internet and social media provide access to millions of potential customers, it is crucial for corporations to reach out to this massive audience. In addition, other aims for corporate communication involve disseminating information regarding the company or institution to create, build, and maintain a positive perception in the individuals of the society. Another essential aim is monitoring and responding to criticism or negative situations to protect the reputation of the individual brand or company.

Due to the value of considering the different perspectives of corporate communication, researchers have examined corporate social media from a range of perspectives employing different methodologies. Early research on corporate digital communication began with examining the role of workplace emails (e.g., Sproull & Kiesler, 1986). Later, Auger (2014) studied the two-way symmetrical/asymmetrical communication of non-profit organisations on Twitter, while Romenti et al. (2016) discussed the quality of dialogic conversations among companies and their audiences on social media. Then, Park and Kang (2020), in addition to Troise and Camilleri (2021), explored the role of corporate digital communication in the context of corporate social responsibility, while Wang and Yang (2020) examined the corporate digital communication on the Facebook and Twitter profiles of non-profit and for-profit organisations. Furthermore, Okazaki et al. (2020) studied the potential of strategic corporate social responsibility communications through Twitter. (Note that corporate social responsibility is included among the functions of the tweets analysed in this study.) Moreover, Carnevale et al. (2017) proposed a framework of brand linguistics that is defined

as the study of language effects on consumers in brand-related settings, although brand linguistics focuses on the consumer as the unit of analysis, as opposed to the company/corporation.

The theories and research in this area imply the importance of corporations regularly evaluating their digital communication and implementing the necessary changes to their communication strategies (Gregory, 1996). This study provides an overview of the current practices of corporations in digital communication in terms of language use, and could also contribute to the evaluation of the corporations' performance.

3.5.3 Social media influencers

Influencer marketing is on the rise, with many marketers planning to either start using influencers, or to increase such engagement in their media plans (Campbell & Farrell, 2020). Raven (2008) defines social influence as a change (i.e., the target of the influence) in an individual's thoughts, feelings, attitudes, or behaviours as a result of another person's (i.e., the influencer's) action. Another recent definition for social media influencers is provided by Kay et al. (2020), which describes social media influencers as individuals with large online followings that attract significant engagement (e.g., likes), and who are able to harness this popularity for marketing efforts in a specific industry. Influencers differ from traditional celebrities through the manner in which they gain fame specifically to be considered a personal brand or celebrity. Whereas traditional celebrities become famous through their pursuits such as acting, music, sports, and politics, and then gain a following through their work, interviews, and media relations (Campbell & Farrell, 2020).

Nevertheless, there are varying classifications of social media influencers. For instance, Porteous (2018) suggests three levels—micro, macro, and celebrity—whereas others suggest two levels: micro and macro (Dhanik, 2016). Campbell and Farrell (2020) draw from a variety of sources to develop five distinct influencer categories: celebrity influencer, mega-influencer, macro-influencer, micro-influencer, and nano-influencer. These categories are based not only on the follower counts, but also the perceived authenticity, accessibility, expertise, and cultural capital. It must be mentioned that although the current study does not differentiate between the investigated influencers' accounts, the influencers included in this study either belong to the category of micro-celebrity or celebrity influencer, and hence these are defined below.

Senft (2008, p.25) defines the micro-celebrity as "a new style of online performance that involves people amping up their popularity over the Web using technologies like video,

blogs and social networking sites". Unlike celebrities, who can become public icons with large-scale followings, the micro-celebrity "is a state of being famous to a niche group of people" (Marwick, 2013, p. 114), and involves the curation of a persona that feels authentic to followers. On the other hand, a celebrity influencer is any individual who enjoys public recognition outside of social media, and is leveraged by brands for their large follower base. For example, celebrity influencers such as Ed Sheeran and Marcus Rashford experienced musical and footballing fame, respectively, prior to or independent from the evolution of social media, despite using their social media presences to support their careers and propagate brand partnerships (Campbell & Farrell, 2020).

3.6 Conclusion

This chapter discussed the development of digital communication research while focusing on multilingualism in the digital world. Section 3.2 presented research discussing the nature of digital communication and other considerations such as technological advancement and its impact on digital communication research, and the current trends in research, in addition to focusing on Twitter specifically. Then, in section 3.3, an overview of multilingualism was provided that addressed previous studies on multilingualism in general, and multilingualism in digital communication specifically, covering key concepts in linguistics (i.e., creative orthography, language choice, code-switching, and paralinguistic cues) that are pertinent to the current research. Finally, an overview of other important concepts related to this research was presented in sections 3.4 and 3.5, namely, self-presentation, online identities, branding, corporate digital communication, and social media influencers. The concepts and research discussed in this chapter reveal that the relationship between language, discourse, self-presentation, and identity comprises a rich area of investigation. In the next chapter, research methodologies are explored, as well as justification for the approach employed in this study.

Chapter 4:
Methodology

4.1 Introduction

This study examines multilingual practices on Twitter, as employed by Arabic speakers in Saudi Arabia, in order to identify the usage patterns and functions of different language varieties. In addition, it explores how the language choices of social media users are used for the purposes of both branding and self-branding. To address the research questions listed in Chapter 1 (see section 1.2), the data collection for this study included (i) the observation of online activities, (ii) the collection and linguistic analysis of Twitter data, and (iii) interviews with users. The research design combines quantitative analysis with qualitative observations. The following sections discuss the overall methodology for the data collection and analysis. Section 4.2 explores the approach adopted for the current research, in addition to quantitative and qualitative research in general. After that, sections 4.3 and 4.4 examine the data collected from Twitter, and the coding process for the languages, patterns, and functions appearing in the tweets, respectively. Next, section 4.5 outlines the interviews conducted following the collection of the data from Twitter, while section 4.6 provides an account of the ethical guidelines followed for collecting and presenting the data, and section 4.7 presents the pilot study, including discussion of its benefits for shaping the research design. Finally, section 4.8 draws the chapter to a close with a conclusion.

4.2 Research approach

The current research is located within the field of linguistics, resulting in the research methods being discussed through the lens of linguistic research. Litosseliti (2018) notes that linguistics is a multidisciplinary and interdisciplinary field of study, investigating a wide range of language phenomena, and being characterised by a variety of theoretical, epistemological, and methodological approaches. In addition, linguistic research involves the study of language, both within and across different branches (e.g., socio-linguistics and psycholinguistics), or related fields in the humanities, the social sciences, and the natural sciences (e.g., psychology, education, and sociology). Thus, linguistics forms a continuous process of bridge-building across its branches, in collaboration with other fields, in order to draw on various opportunities and challenges. This current research explores a specific linguistic phenomenon (i.e., the use of languages and varieties on Saudi Twitter), and builds

a bridge with the marketing field in order to identify the use of various languages and varieties for the purposes of branding.

Researchers can follow either a deductive or inductive approach, depending on the relationship between theory and research (Bryman, 2016). The deductive research approach depends on the development and testing of a particular theory, with researchers commencing by proposing a hypothesis based on an established theory in a known field of study, followed by the testing of this theory on the proposed hypothesis. The results of deductive research either confirm the accuracy of the theory, or modify it in accordance with the findings (Creswell, 2014). Meanwhile, the inductive research approach begins by collecting the data, followed by an analysis and the development of theory, which is then compared to previous relevant literature (Bryman, 2016). Researchers employing the inductive approach tend to focus on the context in which a certain phenomenon takes place, and therefore use a smaller sample than that of the deductive approach (Bryman, 2016). The current research employs the inductive approach as the data were collected and then analysed in order to answer the research questions, whereby the findings were subsequently compared with the relevant literature.

The research methods and techniques adopted in a research project depend upon both the research questions and the focus of the researcher. Quantitative research tests theories by examining the relationships between variables, which are analysed using statistical procedures (Creswell, 2014) and with the aim of investigating numerical outputs and achieving generalised meaning from these numbers. On the other hand, qualitative research involves an interpretive approach, analysing the data in their natural setting and focusing on making sense of the phenomenon in terms of the meanings accorded by the associated individuals (Denzin & Lincoln, 2011). The mixed-methods research method is defined as a procedure for collecting, analysing, and combining quantitative and qualitative data during the research process within a single study, in order to better understand the research problem (Creswell, 2014). This current research employs a mixed-methods approach, due to this being deemed the most suitable for answering the research questions listed in Chapter 1. Therefore, the tweets were analysed quantitatively, using percentage values for the results, alongside some of the items emerging from the interviews. Furthermore, the answers provided by the participants in the interviews were analysed qualitatively using thematic analysis, as shall be explained in section 4.5.4. The rationale for using a mixed-methods approach for the current study is to have a deeper understanding of the Twitter's users linguistic practices. The quantitative analysis provides an overview of the proportion of the languages and the functions while the qualitative analysis allowed for

further understanding of the attitudes and motivations that shape the users' linguistic choices on Twitter. A similar approach is used by Androutsopoulos (2015) for the data collection, which combines observation of online activities, collection and linguistic analysis of screen data, and data elicited through direct contact with users. Focusing on the participants' linguistic choices for self-presentation, and the performance of multilingual talk online.

As discussed previously, this study aims to achieve a holistic perspective of the languages and varieties employed on Twitter in Saudi Arabia, in addition to gaining an in-depth understanding of the motivations and reasons for using these languages and varieties. Twitter provided an ideal opportunity to obtain a general view of the languages and varieties employed in Saudi Arabia on digital platforms, as well as the more detailed descriptions provided by the users. For this reason, the research was designed to incorporate two forms of data collection: the collected tweets, and the interviews. Therefore, a more in-depth understanding of the findings was obtained by the analysis of the tweets in relation to the views expressed by the users in the interviews. Sections 4.3 and 4.4 discuss in detail the methods utilised to collect and code the data, respectively.

4.3 Collection of the tweets

The current research considers the collected tweets as the main data source for the analysis, supplemented by the data gathered via the interviews. The following sub-sections examine the strategies followed for obtaining the tweets for this study, including the sampling strategy, details about the data contained within the tweets and how they were collected, and the participants.

4.3.1 Sampling strategy

In order to answer the research questions established in this study, tweets were collected from the timelines of 100 public profiles. As discussed in Chapter 1 and Chapter 3, three separate groups were selected to ensure representation from a broad range of Twitter users: corporations, influencers, and ordinary users. One of the main objectives of this study is to investigate the relationship between the use of different language varieties in digital communication and branding. Therefore, half of the accounts came from corporations (50%), and the other half were comprised of well-known social media influencers (30%) and ordinary users (20%). The reason for the slightly higher percentage of influencers compared to ordinary users is that the former have more branding activities. The corporate accounts represent a range of interests, including government, industry, entertainment, products, and

restaurants. Although it may be argued that government tends to communicate in a separate manner to corporations, some government accounts were included as they also primarily use Twitter to disseminate information about their services and to promote their effectiveness to the Twitter community. Similarly, social media influencers are known for their involvement in a variety of fields such as media, technology, law, medicine, and sports. Section 4.3.2 discusses the criteria for identifying the participants, while section 4.3.3 describes details about the tweets such as the quantity collected for each account, and the data-collection period.

4.3.2 Research participants' criteria

The profiles chosen for the collection of tweets were selected based on the following inclusion criteria:

Corporations:

1) The selected accounts were official accounts authenticated by Twitter, that is, they all had a blue Verified badge.

2) The selected accounts were active, showing at least 80 Twitter interactions during the period of the data collection.

3) The Twitter accounts had at least 10,000 followers at the time of the data collection.

Influencers:

1) The selected accounts were official and authenticated by Twitter, that is, they all had a blue Verified badge.

2) The selected accounts were active, with at least 60 Twitter interactions during the data-collection period.

3) The sample contained both genders.

4) The Twitter accounts had at least 10,000 followers at the time of data collection.

Ordinary users:

1) The selected accounts were active accounts, having a minimum of 50 Twitter interactions during the period of the data collection.

2) The sample contained both genders.

3) The Twitter accounts had a range of between 300 and 10,000 followers at the time the data were collected.

As noted above, the sample included a number of corporations, while the influencers represent various interests and are famous for a variety of reasons. In order to fulfil the aims of the study, the accounts included needed to show different language variation. In addition, both during the pilot study, and when selecting the accounts, several patterns were noticed in the use of language varieties on Twitter in all three groups. For example, some combined different Arabic varieties, while others used both Arabic and English. Therefore, in order to achieve diversity, the variation in the languages and varieties seen on the different Twitter accounts was considered during the data collection, as per the aforementioned inclusion criteria.

It is important to highlight that the social media influencers selected for this study not only consisted of celebrities in the offline world since self-branding, as discussed in Chapter 3, is practised both by celebrities expected to have a strong public image (i.e., sportspeople and performers), and also by ordinary citizens through exploiting the novel technologies of social media to build an audience (Khamis et al., 2017). Thus, social media platforms such as Twitter and Instagram have accelerated the process of building a personal brand across several platforms for those not previously known to online audiences.

Finally, it is important to point out that there were a number of difficulties when it came to collecting tweets from the ordinary users, due to the challenges of locating accounts that were active during the period of data collection. The initial plan had been to focus on popular hashtags, but reviewing the hashtags proved insufficient for the purposes of this study, resulting in a need to locate active accounts by tracking interactions with corporations and influencers.

4.3.3 Information in the tweets

Twitter's format means tweets are limited to 280 characters, while posting is undertaken through the Twitter website or applications on smartphones and other devices. In addition, users can choose to share their tweets publicly or to keep their accounts private and thus only accessible to selected followers. For the current study, a total of 13,426 tweets were collected from 100 public accounts. In order to ensure the users' geographic location, the only profiles utilised for the analysis were those indicating that they were based in Saudi

Arabia and showing related content, thus the participants primarily consisted of users from major cities in Saudi Arabia: Riyadh, Jeddah, and Al Khobar.

The data were collected between November 2017 and May 2018, depending on the activity and the number of tweets collected from each account. Some of the accounts posted frequently (e.g., 1,800 published tweets in one month), while others posted infrequently (i.e., 20 or fewer tweets in one month). Due to this discrepancy, the data-collection period was extended for the accounts with less frequent posting, in order to ensure that the minimum number of tweets could be gathered, that is, 80 tweets by the corporations and 60 each for the influencers and ordinary users. In addition, the data collection period for the highly active accounts was limited to two weeks, in order to balance the contribution from the accounts in the data.

The tweets collected from the corporations, influencers, and ordinary users' Twitter accounts were categorised into updates, retweets, and replies, with only updates and retweets being analysed. It was decided to include retweets because (even in cases where another's words were posted without further comment), the original tweet is still shown on the retweeter's timeline and employed as an aspect of self-presentation. Replies were excluded as the original tweeter's language generally influences the language in the reply. Table 4.1 presents a summary of the number of tweets collected from each of the three groups of Twitter users.

Table 4.1: Number of collected tweets for the corporations, influencers, and ordinary users

Group	Corporations (N=50)	Social media influencers (N=30)	Ordinary users (N=20)
Number of tweets	9,370	2,593	1,٤٦٣
Mean per account	171.3	97.9	99.6
Percentage of total collected	70%	19%	11%

Table 4.1 presents the range of tweets per account for each of the three groups. It must be noted that most of the corporations' accounts tend to draw up plans for their presence on Twitter, including a schedule for their tweets and interactions, as shall be further explained in 5.3.5. This means that most corporations are active on a daily basis and therefore will have the highest number of tweets. Similarly, social media influencers are expected to have a frequent presence, and therefore their number of tweets is higher than the ordinary users, who have more flexibility and less obligation in terms of their presence on Twitter.

The tweets for this study were mainly collected though TAGS, one of the Search API tools created by Martin Hawksey (2010), where the main benefit is the provision of a spreadsheet that contains a range of information, including the author, the actual message text, hashtags, the number of user followers, the number of user friends, and the user location. The tweets were then coded for the language varieties used in the tweets, including their patterns and functions. Then, the proportion of multilingual tweets were calculated. However, as TAGS only provides live tweets with specific time limitations, it was impossible to retrieve all the data to cover the required minimum number of tweets for each user. Hence, tweets were also obtained from Podargos (www.podargos.com), which provides publicly available data from online websites and applications such as Twitter and Facebook, to support the data collection and complete the data-set for the users.

4.4 Coding of the tweets

The current study investigates multilingual interactions and their functions in digital communication, primarily in relation to Twitter. Hence, the study utilises a quantitative approach to analyse the proportions of the different varieties and their functions, as employed by the corporations, influencers, and ordinary users. It should also be noted that the study commenced with a qualitative descriptive analysis during the pilot study, in order to identify the codes for the quantitative analysis. This approach proved helpful in understanding the use of the different languages and varieties, considering the multimodal aspect of digital communication.

The coding of the tweets followed the methodological framework of Computer Mediated Discourse Analysis (CMDA) presented by Herring (2004). CMDA adds to the traditional assumption of discourse analysis that it is being shaped by the technological features of digital communication. According to Herring (2004), CMDA is used to analyse different parts of language such as characters, words, utterances, messages, exchanges, threads, and archives. CMDA has been applied by several researchers. For example, Yates (1996) used CMDA framework to investigate the similarities and differences between a corpus of digital, spoken, and written data. Therefore, the present study adopted the CMDA approach for observing, coding, and interpreting data, supplemented by qualitative analysis. The primary focus of the research design is the exploration of a phenomenon, with the use of qualitative interpretations intended to assist the quantitative analysis. The results of these two approaches are then integrated during the interpretation of the findings. The CMDA approach adopted in the current study is manifested in identifying the codes that appeared in the data and counted in terms of their relative patterns.

The tweets in this study were gathered along with related information, including the author, the actual message text, the date and time of the tweet, and a link to the tweet itself. In addition, the following categories were coded: (i) the main language utilised, (ii) other languages, and (iii) the function of the tweet. The following are the main common language varieties found in the data: (i) Classical Arabic, (ii) MSA, (iii) CA, (iv) Arabic with Latin script (Arabizi), (v) English, and (vi) English using Arabic script. Section 4.4.1 discusses the important considerations that were followed in the analysis of the tweets, with section 4.4.2 providing an explanation for the coding of each variety and language, and providing examples from the data. Then, section 4.4.3 presents the coding process relating to the functions of the tweets. Finally, section 4.4.4 provides a brief overview of the statistical analysis to compare the results from the three groups.

4.4.1 Analytical approach to multilingualism on Twitter

This section provides an overview of the analytical approach from three important perspectives. First, it considers categorising the code-switching levels in the multilingual discourse on Twitter, then it discusses identifying the base (matrix) language of code-switching according to the criteria presented by Myers-Scotton (1993), and finally it considers how instances of borrowings and code-switching were distinguished.

4.4.1.1 Inter-tweet vs intra-tweet multilingualism

As discussed in the Literature Review chapter, code-switching is generally defined as a "juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems" (Gumperz, 1982, p. 59). When we examine different digital platforms, such as Twitter as the platform employed for the current study, we find different languages and varieties used on the timeline of the users. According to Androutsopoulos (2013), these are multilingual or heteroglossic discourse spaces, but they do not automatically constitute instances of code-switching. The posts differ in terms of authorship and production process, but what binds them together is their spatial coexistence in product and reception, as opposed to their dialogic orientation to each other. Therefore, although these tweets can be monolingual, they may include several languages or varieties in the same timeline, which constitutes a multilingual space. Furthermore, within this same Twitter timeline space, there are tweets that contain intra-sentential code-switching. In the current research, multilingualism presents itself in two ways:

1) Inter-tweet: using different languages in different tweets within a single timeline.

2) Intra-tweet: using different languages and varieties within a single tweet.

The presentation of the results in the analysis chapters will focus on these two aspects by first presenting the multilingualism in the monolingual tweets, and then secondly, presenting the results of the multilingualism in the mixed tweets.

4.4.1.2 Matrix language vs embedded language

As discussed in section 3.3.4.3 on code-switching online, the theoretical model of the matrix code and embedded code by Myers-Scotton (1993) is utilised to determine the dominant language of the tweet. This criterion applies to many forms of intra-tweet code-switching, such as the example displayed in Figure 4.1.

Figure 4.1: Matrix code vs embedded code: example 1

There is also another form of code-switching that shows a separation between the matrix and embedded code. For example, the tweet could start with a language (matrix code) and then switch to another language (embedded code), such as the following example in Figure 4.2.

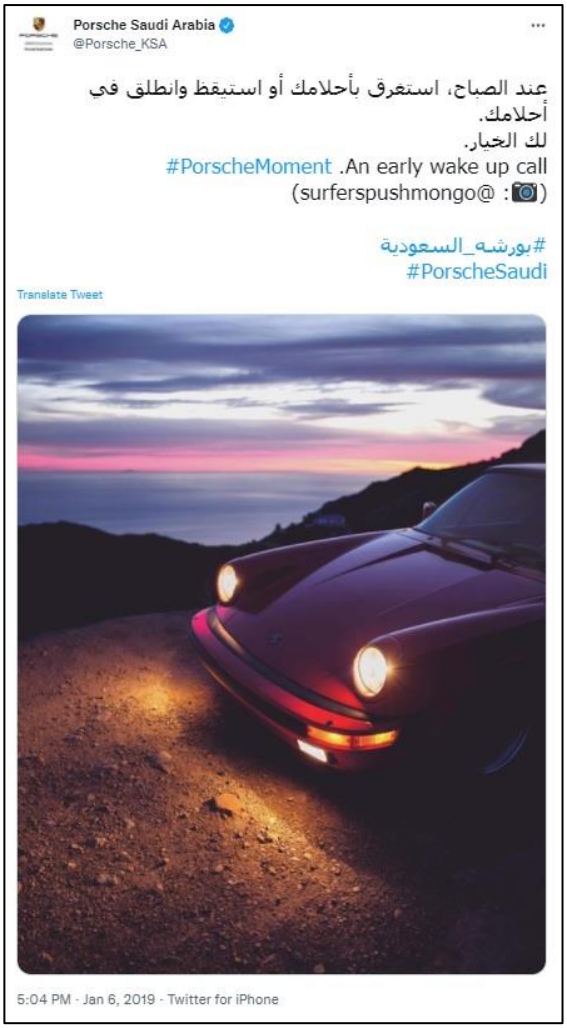


Figure 4.2: Matrix code vs embedded code: example 2

Having discussed how the matrix and embedded codes are determined in the tweets, we now turn to how instances of code-switching are differentiated from borrowings.

4.4.1.3 Code-switching vs borrowing

The third perspective of multilingualism is distinguishing code-switching and borrowing. According to the discussion in the Literature Review chapter, the following criteria by Poplack and Sankoff (1984) in addition to Muysken (1990) are followed to determine the borrowing instances in the data:

1) Lexical borrowed items are inserted in the lexicon of the native speaker
2) Borrowings show phonological, morphological, and syntactic adaptation
3) Borrowings substitute a language's own words
4) Borrowings are frequently employed in everyday conversations

For the current research, an item that fulfils all these criteria is considered a borrowing rather than code-switching. For example, the word 'laptop' (لابتوب) appeared in the data several times, and it fits with all the criteria mentioned above as it is well known to native speakers; it shows phonological, morphological, and syntactic adaptation as it appears taking the plural form of Arabic *-at* as in *labtob-at* (لابتوبات); and it uses the *b* letter instead of *p*, which is not part of the Arabic system of letters and sounds. Furthermore, it substitutes an Arabic equivalent of the word, and it also appears in everyday conversations. On the other hand, the word 'tutorial' (توتريال) is considered a code-switching instance since although it is written with Arabic script, it is not commonly used in everyday conversations and cannot be described as inserted in native speakers' lexicon, since not all Arabic speakers understand its meaning.

4.4.2 Languages and varieties

As discussed in Chapter 2, there were different varieties of Arabic presented in the data, with six languages and varieties coded in the tweets:

1) Classical Arabic
2) MSA
3) CA
4) Arabizi
5) English
6) English with Arabic script

In addition, the data revealed that it was common for some of the corporations to tweet the same content twice, using two languages (i.e., Arabic and English), as illustrated in Chapter 2. During the coding, when tweets with the same content but in different languages directly one after the other occurred within an account, the individual tweets were coded for the language used and Pair. This means that two tweets with the same content in MSA and English would be coded as: MSA Pair and English Pair. All the pairs were with English with MSA or CA as the other language.

It must be highlighted that in the analysis the tweets in different languages are counted separately. Hence, a pair does not count as one instance.

Furthermore, there were other world languages in the data such as French, Spanish, and Urdu, whereby each language had its own category in the coding process. However, they are sometimes grouped together in the analysis chapters due to their infrequent occurrence.

The following section considers the communicative functions and how they were coded in the tweets.

4.4.3 Functions of tweets

When coding the tweets for this research, it was also important to explore their functions to establish whether this influences the language choices. Therefore, as discussed in section 3.3.4.4 on the functions of code-switching online, 19 different functions were identified and analysed in order to create a typology of the different functions of the tweets in the present study. After that, the functions identified in the coding process were compared to the functions described in previous work of a similar nature (cf. Androutsopoulos, 2013; Halim & Maros, 2014), with these being categorised into six main functions: (i) formulaic purposes, (ii) culturally specific genres, (iii) reported speech, (iv) dialogical interrelation, (v) marketing, and (vi) self-expression. Hence, the data were coded using the sub-functions with the six main functions discussed in the subsequent analysis chapters for the corporations, influencers, and ordinary users. Figure 4.3 shows the 19 sub-functions and how they were grouped into the main six functions.

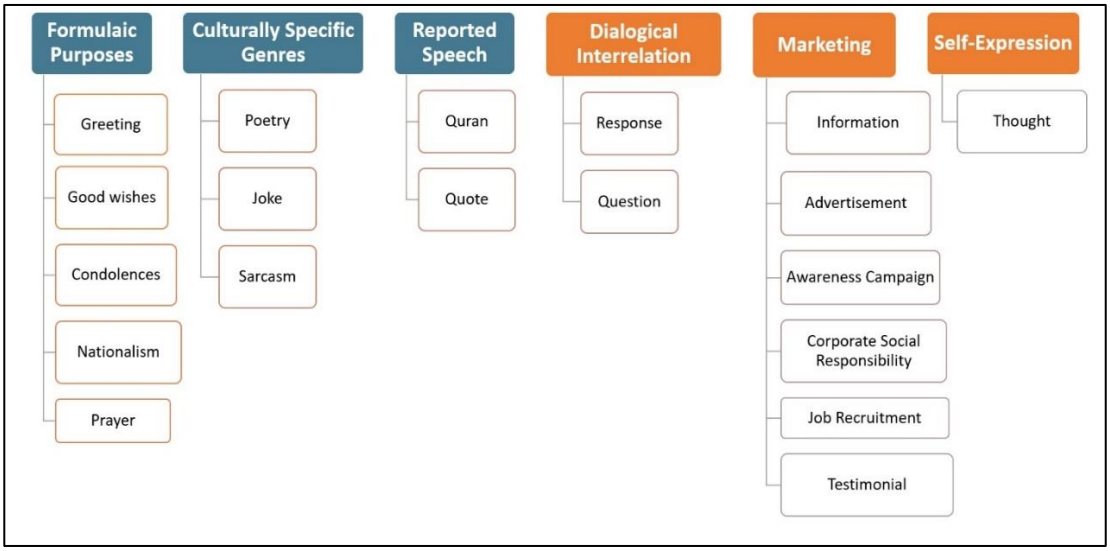


Figure 4.3: Functions and sub-functions of the tweets in the data

It must be noted that different types of accounts will use more or fewer functions. For example, the marketing function is utilised more by the corporations' accounts and by some of the influencers, while the self-expression function is only employed by the influencers and ordinary users. Furthermore, some of the functions may overlap in tweets, and some posts may show more than one function. For example, a tweet can provide general information and ask followers if they have any impressions or previous experiences. In

these cases, the function deemed more visible to the audience in terms of the proportion of the tweet was selected as the function of that tweet.

The following sections present examples from the data for each of the main functions coded in the tweets.

4.4.3.1 Formulaic purposes

Figure 4.3 shows that the sub-functions for formulaic purposes are greetings, good wishes, condolences, nationalism, and prayer. Formulaic purposes were found to provide opportunities for the users from the three groups in terms of engagement with their audience, including being competitive in the most effective manner of presenting themselves to their virtual audience. For example, tweets were employed to share greetings and good wishes with followers, particularly in different seasons and in the morning. Figure 4.4 is an example of a formulaic purpose of offering good wishes from the account of the King Faisal Foundation for the holy month of Ramadan.

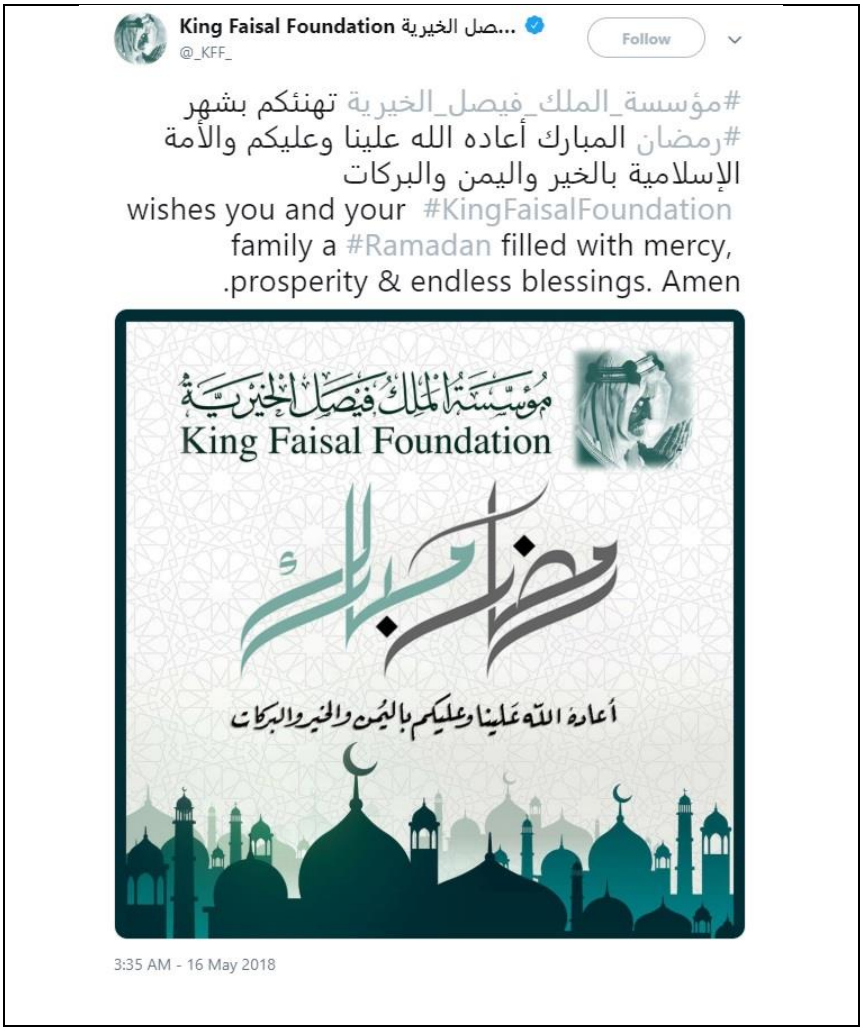


Figure 4.4: A tweet offering good wishes

Also, some tweets showed that one of the most desired attributes the users tend to reflect in their account is their sense of belonging and love for the country. Users therefore share posts to demonstrate their nationalism during different events, including the National Day and football games, as well as for new government announcements or appointments. The following example in Figure 4.5 is taken from the account of the National Commercial Bank, posted on the anniversary of the fourth annual allegiance of the King.

[We pray to Allah the highest and the capable to support them and help them to serve Islam, then the nation and the people #fourth_allegiance_anniversary]

Figure 4.5: A tweet that serves the function of nationalism

Furthermore, users of Twitter also express their condolences, showing sympathy for others when loved ones are lost for example, as seen in Figure 4.6.

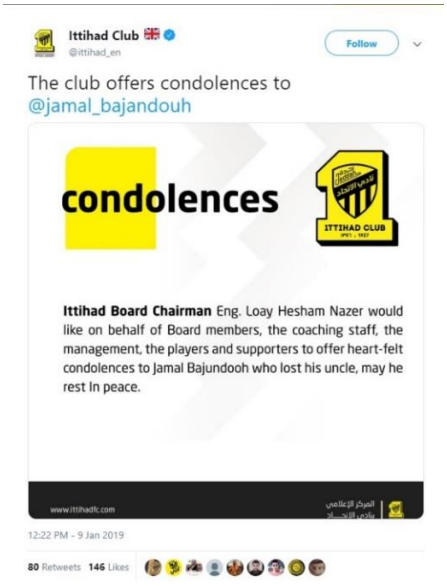


Figure 4.6: A tweet offering condolences

4.4.3.2 Culturally specific genres

The culturally specific genres include poetry, jokes, and sarcasm tweets, since Twitter posts are not always serious and professional. Some users post humorous content, such as anecdotes, to appeal to their audience and increase the opportunity to initiate engagement, as seen in the following example in Figure 4.7 of a joke from an ordinary Twitter user.

[There is an ant studying on the roof. Why? Higher education student. Ha-ha]

Figure 4.7: A tweet featuring a joke

On the other hand, sarcasm enables users to discuss issues in an indirect manner, including making humorous statements regarding different current affairs in the society. In the following example in Figure 4.8, a social media influencer criticises the way that food delivery applications (despite requesting the address of the customer before an order can be placed), always contact customers by means of WhatsApp to find their location.

Figure 4.8: A tweet with sarcasm

4.4.3.3 Reported speech

Another widely used function for the tweets concerns the use of quotations from the Quran, and famous political or historical figures, in addition to the globally famous, such as in the following example in Figure 4.9 where one of the social media influencers posts an inspirational quote.

Figure 4.9: A tweet with a quote

4.4.3.4 Dialogical interrelation

One of the techniques utilised on social media platforms to encourage engagement between the user and their followers is to pose a question, or to initiate a discussion in the tweet's thread of replies. These questions can be posted on the timeline, or through a Twitter Poll, as seen in the following example in Figure 4.10 from a social media influencer on the topic of unused medicine in the home.

[How many different brands of excess medication do you have at home right now?
#retweet #health
34% Less than 10 brands
19% Between 11 and 20 brands
22% We can be a branch for Al Nahdi (a famous pharmacy in Saudi Arabia)
25% I just want to see the results]

Figure 4.10: A tweet posing a question with a poll

Furthermore, it was found in the data that when a large number of clients, customers, or followers ask the same question, instead of replying separately to each enquiry, users either reply to each tweet or create a post with a general response. The following tweet presented in Figure 4.11 from the National Commercial Bank shows an example of such a blanket informative response to all customers.

Figure 4.11: A tweet featuring a general response

4.4.3.5 Marketing

The marketing of services and products could be perceived as one of the most important aims for the users, and in particular the corporations. Marketing can be viewed as the overarching aim of digital communication, but the classification of marketing in this study focuses on direct marketing messages provided by users, such as the following example in Figure 4.12 from Effat University offering discounted scholarships to potential students.

Figure 4.12: A tweet serving the marketing function

Social media, including Twitter, provides an effective platform for promoting awareness of a wide range of causes such as diabetes, cancer, and general health campaigns. The study observed that users generally show enthusiasm for participating in these campaigns, including by creating their own posts or sharing those from other users, as seen in the following example in Figure 4.13 from Bupa Arabia on World Diabetes Day.

[Are you diabetic? Make it a habit!

#World_Diabetes_Day

#Tebtom

Picture: Keep having breakfast]

Figure 4.13: A tweet for an awareness campaign

Twitter users in general, and particularly corporations, tend to take advantage of the wide audience on the platform to advertise job vacancies. The following example in Figure 4.14 from Hyundai shows a job advertisement posted solely in English, thus indicating that, although not overtly stated, English proficiency is prerequisite.

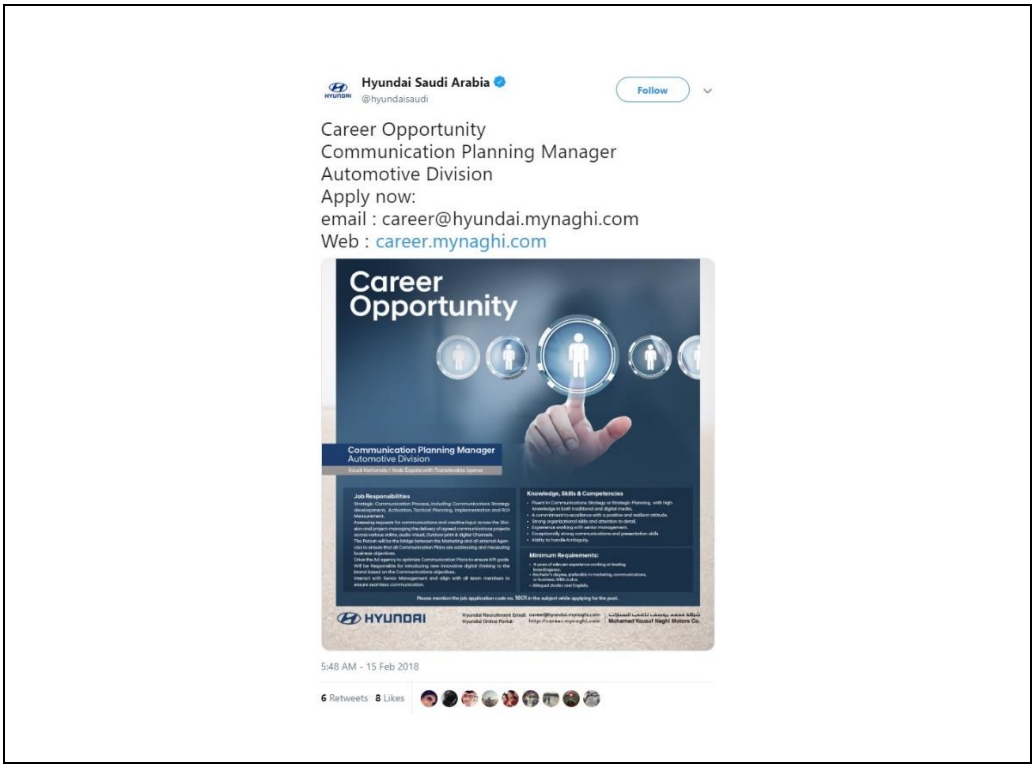


Figure 4.14: A job recruitment tweet

It is currently imperative for businesses, institutions, and celebrities to show the impact they have on society, and Twitter is therefore employed to publicise corporate social responsibility activities, with the example tweet in Figure 4.15 from Al Ittihad football club illustrating such an event.

Figure 4.15: A tweet showing corporate social responsibility

4.4.3.6 Self-expression

It was found in the data that social media influencers and ordinary users tweet about their thoughts regarding current affairs either occurring globally or locally in their daily lives. Figure 4.16 is an example by one of the social media influencers expressing a general thought on aspirations.

Figure 4.16: A tweet conveying an aspiration as self-expression

4.4.4 Statistical analysis

Following the data collection from Twitter and the coding of the tweets, the study then quantitively analyses the linguistic features that arise from the analysis of the corporations, influencers, and ordinary users. The data are then presented in numerical and percentage form for each of the three groups in a separate chapter, and thoroughly discussed to answer the research questions. After that, the final analysis chapter reports on the statistical analysis conducted using chi-square tests to explore the association between the target variables in the research, that is, the languages and functions, and the three groups under study.

With sections 4.3 and 4.4 discussing the data collection and coding methods for the data collected from Twitter, section 4.5 now considers the second data source for the current research: the interviews.

4.5 Interviews

Interviews are one of the most commonly used methods for data collection. As noted previously in section 1.2, this research aims to understand the motivation for using different language varieties, as well as how Twitter users in Saudi Arabia employ these to brand themselves to their large virtual audience. In order to understand important insights that cannot be observed from analysing the tweets alone, a second set of data was collected through interviews, as outlined in the following sections.

4.5.1 Research participants

The total of fifteen interviews were conducted for the current research, with a sample drawn from each of the three groups, as presented in Table 4.2.

Table 4.2: Interview categories and the number of participants

No.	Category	Number of interviews
1	Corporations	7
2	Influencers	5
3	Ordinary users	3
	Total	**15**

The sample size for qualitative research methods such as interviews is smaller than for quantitative research. Following the frameworks of Ritchie et al. (2013) and Creswell (2014) for interviews, the current research conducted 15 interviews, whereby similar themes and

insights were appearing that led the study to reach the saturation point. Furthermore, a larger sample might start to affect the quality of the data, as it could become unmanageable for the researcher to provide in-depth analysis.

4.5.2 Process of interviews

Interviews are considered appropriate for this study, as they are widely utilised in qualitative and digital communication research (cf. Androutsopoulos, 2015). This research employed semi-structured interviews to maintain the focus on the topic being discussed, which were conducted on a one-to-one basis in April and May 2019. The interview questions were specifically designed based on the initial analysis of the data collected from Twitter. The findings offer important insights not accessible through analysing the first set of data, including (i) the preferred language choice, (ii) the motivations and reasons for using each language variety, (iii) attitudes towards the use of the different language varieties, and (iv) differences between the users' online language choice on various digital platforms.

Each of the interviewees were contacted by official email, as per the example included in Appendix 1. The interviewes were audio recorded. Each interview lasted approximately 30 minutes, and in line with the semi-structured approach, the interviewees were asked pre-planned questions listed in Table 4.3 below, with minor adjustments made to suit each of the three groups. Furthermore, in a section at the end of each interview the interviewees were shown actual Twitter data from their accounts, allowing them to discuss their practices. At the end of the interview, each participant was given a copy of the Debriefing Document (see Appendix 2).

As the data were collected in Saudi Arabia, the interviews were primarily held in Arabic, with excerpts subsequently transcribed and translated into English by the researcher, and then validated by colleagues who are native Arabic speakers and fluent English speakers.

4.5.3 Interview question guide

The interviews consisted of pre-determined questions, but were also flexible to allow space for any additonal questions that might arise. Furthermore, the questions were derived from the research questions and guided by the literature. A sample of the interview questions is provided in Table 4.3.

Table 4.3: Sample of interview questions

1. General information
A. Date of interview: B. Name and position of interviewee: C. Company name: D. Company age: E. Company size (1= multinational; 5= self-employed business person) F. Place of interview:
2. Aims and goals on social media
1. What are the company's main aims to achieve in social media? 2. Are there any specific aims for Twitter? If yes, what are they? 3. Do you focus more on increasing the number of followers, or more engagement with the followers? Please explain. 4. How many posts do you plan every day or every week? Are there any specific aims for the posts (engaging discussion/share quote/say good morning)? 5. What device do you usually use for posting? Does this affect the language employed? 6. What is the main language that you use for social media communication?
3. Intended audience
7. Who is your target audience on Twitter? 8. Do you consider different posts for every group?
4. Multilingual practices
9. State whether you use the following language varieties and what the purposes are for using each variety: a. Modern Standard Arabic b. Colloquial Arabic c. Classical Arabic d. Arabizi e. English f. Any other language
5. Semiotic resources
10. Do you use emojis or punctuation? 11. Do you think it is essential to use an emoji or punctuation with every tweet? 12. Do you think the audience will understand your message more if you use an emoji or punctuation?
6. Opinion on the different language varieties

13. What do you think is the ideal language variety to be used on Twitter?
14. Do you think all tweets should ideally be in MSA? Why? Is it easy for you to use?
15. Why do you think people use Colloquial Arabic online? Do you think it affects Arabic? Do you find it difficult to read and understand? Do you think it may replace MSA in online communication?
16. Is it easier for social media users to communicate certain ideas in English? Do you think English posts should be translated into Arabic? Do you retweet in English? Do you translate what you tweet into English?
17. Do you use Arabizi? Do you think it is useful? Who do you think is using it now?
18. Does your language choice differ on different social media platforms?

7. Multilingualism and self-branding

19. What do you think is the expected perception of your audience on Twitter when using:
    a. Modern Standard Arabic
    b. Colloquial Arabic
    c. Classical Arabic
    d. Arabizi
    e. English
    f. Any other language
20. Does the audience interact more when you use a certain variety?

As discussed earlier, the use of different language varieties on Twitter, which is a widely used public platform, is not the result of spontaneous language production, as typing the tweet involves awareness and self-consciousness before posting it publicly. Hence, the different insights on how the accounts are managed linguistically by the interviewees proved highly beneficial in complementing the analysis of the tweets.

4.5.4 Interview analysis

The analysis of the interviews followed thematic analysis in order to identify the themes related to the research questions. Thematic analysis involves searching for frequent patterns of meaning across the data (Braun & Clarke, 2012). The first step in analysing the data from the interviews was listening carefully to the recordings and reviewing the accompanying interview notes to become familiar with the data. After that, one interview from each of the three groups was selected to be fully transcribed (CORP.1, INF.3, and ORD.1). Then, partial transcripts and a summary for the remaining interviews were produced following the selective transcription process proposed by Gillham (2005). Furthermore, as the majority of the interviews were primarily conducted in Arabic, the

transcripts and the summary were translated into English. After that, the data were coded and the relevant interview section for each code highlighted. The major themes emerging from the interviews were identified as follows: languages and varieties employed for communication, opinion on the languages and varieties used on Twitter, using paralinguistic cues, aims and goals on Twitter, frequency of posts, and target audience. The responses were then quantified, and relevant extracts utilised for the analysis, as shall be presented in chapters 5–7. The interviews offered an opportunity to gain insight into users' own perspectives and attitudes about their language use online. As will be discussed in the analysis, their opinions do not always match their actual use, which provides insight unconscious language ideologies as well.

4.6 Research ethics

The various social media platforms provide a rich space for the collection of data for research in disciplines including the social sciences, politics, marketing, language and communication, medicine, and psychology. However, as per research in general, it is vital to consider the ethical standards to ensure that the data are obtained, utilised, and published in a safe and acceptable manner.

The current research was granted ethical clearance by the Research Ethics Committee at Cardiff University's School of English, Communication, and Philosophy. To fulfil the requirements of the Research Ethics Committee, the following key ethical areas were considered:

1) Public Twitter accounts: Corporations, social media influencers, and ordinary users with public accounts on Twitter expect the public audience to view their accounts and profiles, and therefore the data are not private. Furthermore, the data are not sensitive, as most posts are general in nature, containing information regarding the companies or individuals, along with quotations and greetings. Finally, and most importantly, Twitter's terms and conditions state that the users' data may be redistributed or employed for other purposes. Therefore, it is deemed unnecessary to anonymise the data from Twitter when stored or included as examples in a thesis. It should also be noted that no potentially sensitive examples are included in the study. However, as ordinary users are different in publicity from corporations and social media influencers, the collected tweets from the ordinary users only are encrypted.
2) Informed consent: The interviewees were asked to sign consent forms prior to the interviews. As part of their consent, they permitted the researcher to include

extracts from their interview in the thesis and academic publications. Appendix 3 shows a copy of the Consent Form document signed by the interviewees.

3) Anonymity of the interview participants: To protect the identity of the participants, the selected transcripts of the interviews are anonymised in all presentations and publications.
4) Recordings: The recorded interviews will not be shared with anyone outside of the supervisory team.
5) Disposal of data: The data from the interviews will be retained for two years, and then destroyed.

Following the approval of the Research Ethics Committee, the data were collected, and the research conducted as outlined in this chapter.

4.7 Pilot study

A pilot study was conducted in order to validate the tools and methodology for collecting and coding the tweets. This proved beneficial for identifying areas of ambiguity, such as the expected number of tweets for the research data, the different language varieties utilised on Twitter, and the proportion of these language varieties. The pilot study also enabled the researcher to estimate the time needed to collect and analyse the data. It should be noted that the data employed for the pilot study were also included in the research data. However, the pilot study included all of the Twitter interactions from the accounts, while as discussed above, the research data excluded the replies and only include updates and retweets.

The data for the pilot study were collected from seven Saudi Twitter public accounts, from the three groups examined by this study. The data contained 2,370 tweets, collected between December 2017 and March 2018. The analysis of the pilot study focused primarily on the language varieties used on Twitter by the Saudi users. Tweets were collected through the TAGS Search API tool. The results showed that the majority of the tweets (94%) were in a single language, although they also revealed that several languages and language varieties were utilised by the three groups to create a multilingual timeline. This aspect formed the focus of the main research.

In addition, the clarity and coherence of the questions employed for the interviews were tested prior to conducting the research interviews with volunteer participants who were bilingual English and Arabic speakers, and who were asked to imagine that they were a marketing executive, a social media influencer, or an ordinary Twitter user for the purposes of the pilot study.

This pilot study of the tweets and interviews thus proved beneficial in shaping the research design and structuring the data analysis.

4.8 Conclusion

This chapter outlined the overall methodological approach for the data collection and the analysis of the current study. It discussed the research design, covering three important elements: (i) how the data from Twitter were collected, including the combination of quantitative and qualitative methods, as well as an in-depth examination of the coding of the different language varieties and functions of tweets; (ii) the process of the interviews; and (iii) the ethical considerations followed in the study.

This chapter therefore paves the way for the analysis and presentation of the findings of the research in terms of the corporations, social media influencers, and the ordinary users, as undertaken in chapters 5, 6, and 7, respectively.

Chapter 5:
Corporations' Data Analysis and Results

5.1 Introduction

The primary objective of this research is to examine multilingual practices on Twitter by Arabic speakers in Saudi Arabia in order to identify the patterns and functions of using different languages and varieties on the social media platform.

The focus of this chapter is to present the results of the data collected from the corporations. As explained in Chapter 4, the total number of tweets (i.e., updates and retweets) utilised for the analysis of the coporations is 9,370. The previous chapter presented a qualitative outline of the Twitter data to explain how the categorisation of the tweets was achieved, which formed the basis for the analysis presented in this chapter.

Section 5.2 presents the quantitative results extracted from the Twitter accounts of the 50 corporations selected for analysis in terms of the different languages and varieties that appear in the data, in addition to the functions and how the languages and varieties are distributed in the various functions. As explained in Chapter 4, in the current research, multilingualism manifests in two ways:

1) Inter-tweet: using different languages and varieties in different tweets within a single timeline.

2) Intra-tweet: using different languages and varieties within a single tweet.

The sub-sections in section 5.2 focus on these two aspects when presenting the results. First, multilingualism across the tweets is presented, followed by the results for multilingualism manifesting within single tweets.

As discussed in the Chapter 4, this research employs two data sources: the analysis of tweets and interviews with a sample of the users. Therefore, section 5.3 reports the findings from the seven interviews conducted with marketing executives from the corporations included in the sample.

5.2 Analysis of the corporations' tweets

This section reports on the results of the corporations' tweets. The analysis of the data in this section commences in section 5.2.1 with a general overview of the tweets to determine the proportion of monolingual, bilingual, and multilingual tweets. Then, section 5.2.2 and 5.2.3 explore the distribution of languages and varieties in the corporations' monolingual and multilingual tweets, respectively, by following Myers-Scotton's (1997) Matrix Language Frame model of code-switching, as presented in Chapter 3, while reflecting on the matrix languages and embedded languages. In section 5.2.4, the different general patterns of languages that appear in the corporations' data are considered to reveal how corporations typically have patterns that they employ in their communication strategy on Twitter. After that, section 5.2.5 explores how other paralinguistic features (i.e., punctuation, letters, and emojis) are employed with or without text in tweets to either support the communcation process or to be used solely as the communication medium. Then, sections 5.2.6 and 5.2.7 consider the various functions extracted from the tweets, and relate these to the languages and varieties that appeared in the tweets. Finally, section 5.2.8 presents a summary of the analysis of the corporations' tweets.

5.2.1 Tweets' overview

The data were categorised according to whether the tweet was written in a single language or variety, or more than one. Figure 5.1 shows the distribution of tweets featuring one, two, or three codes that appeared in the corporations' data (N=9,370 tweets).

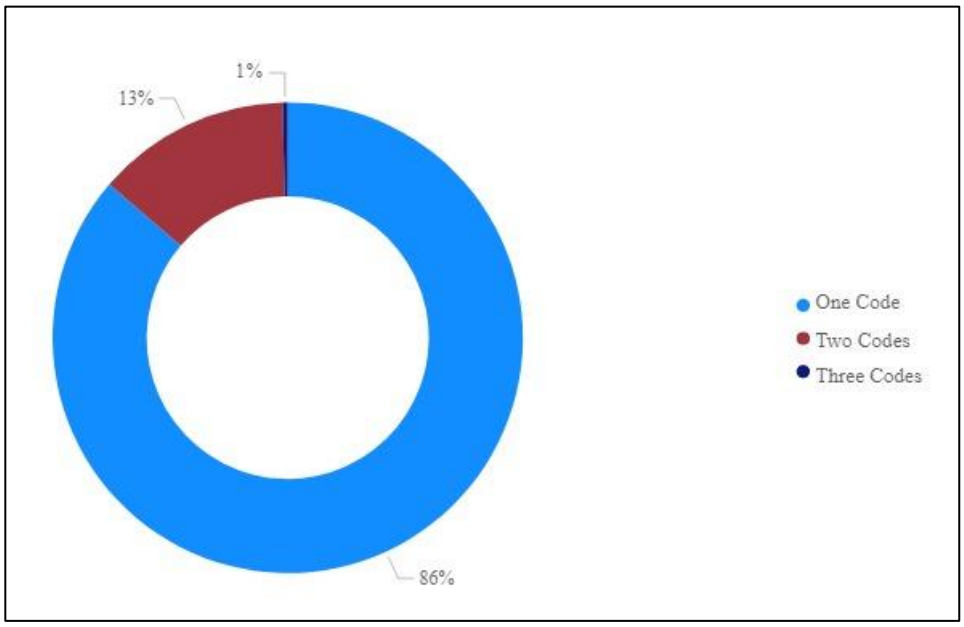


Figure 5.1: Distribution of languages and varieties in the corporations' tweets

It can be seen from Figure 5.1 that the overwhelming majority of tweets (86%) are in a single language or variety. Therefore, despite the common perception that there are instances of code-mixing in single tweets on Twitter, this result suggests that the majority of the code-switching for the corporations may happen at the inter-tweet level rather than at the intra-tweet level. However, it is worth noting that the corporations' tweets in the data commonly tweet the same content twice, using two languages (i.e., Arabic and English), as illustrated in Chapter 2.

Despite Figure 5.1 showing that 86% of the analysed tweets are in a single language or variety, it is worth noting that 1,312 (14%) of these are in fact parallel tweets. This highlights the extent to which code-switching or translation is found across tweets, whereby the same content is tweeted again, for example, once in Arabic, again in English, and very occasionally again in French. Therefore, besides the 86% of monolingual tweets, a further 14% are bilingual or multilingual tweets. It is also worth noting that both tweets are totalled in the 14%, and therefore a pair does not count as one instance.

Additionally, the analysis revealed that some of the monolingual tweets are paired with the multilingual ones. For example, there are cases where the first tweet is posted in English, and then a parallel tweet is written in MSA that also contains English. It was also found that 2% of the parallel texts use more than one language or variety.

After analysing the overall use of monolingual, bilingual, and multilingual tweets, it is essential to examine the distribution of the different languages and varieties that coexist either inter-tweet or intra-tweet.

5.2.2 The languages and varieties in the corporations' monolingual tweets

As discussed in the previous section, code-switching on Twitter occurs more frequently across the tweets rather than within single tweets. This section will first consider the monolingual tweets, before the next section moves on to examine the multilingual tweets. Figure 5.2 presents the distribution of codes in the monolingual tweets (N=8,085).

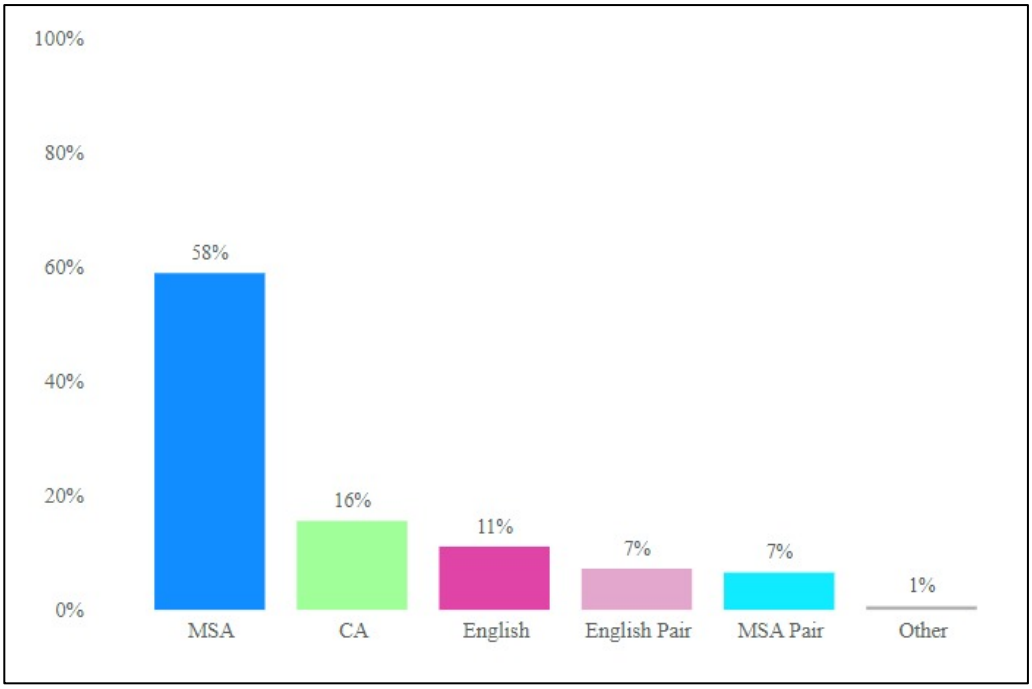


Figure 5.2: Distribution of languages and varieties in the monolingual corporation tweets

With monolingual tweets in different languages and varieties constituting 86% of the data collected from the corporations' Twitter accounts, as stated above in section 5.2.1, Figure 5.2 shows that the majority of the monolingual tweets are written in MSA (58%), followed by CA (16%) and then English (11%). It is not surprising that MSA is the most commonly used code in the corporations' data, considering that it is the official language in Saudi Arabia. This also highlights that regardless of all the language affordances available to the users on Twitter, MSA still dominates as the preferred variety for official communication through the Saudi corporations' Twitter accounts. There is also a fair amount of English used in the corporations' tweets (11% in monolingual tweets, in addition to 7% paired with Arabic tweets), suggesting a prominent role for English in these corporations' tweets and that their followers are exposed to a fair amount of English tweets on a daily basis. It should be noted that the 'Other' category in Figure 5.2 contains the other languages and varieties that appeared in the data, namely CA Pair, Classical Arabic, Arabic with Latin script, English with Arabic script, and Urdu.

5.2.3 The languages and varieties in the corporations' multilingual tweets

Regarding the multilingual tweets in the corporations' data, Myers-Scotton's (1997) Matrix Language Frame model of code-switching was again employed as the basis for the analysis. When two or more languages or varieties are used within a tweet, the source of

the grammatical frame of the constitutent or tweet is the matrix language, while the other language or variety that contributes with limited material is the embedded language. The analysis first focuses on the marix language, and then considers what other languages and varieties are used. Figure 5.3 illustrates the distribution of the matrix codes in the bilingual and mutlilingual tweets (N=1,285).

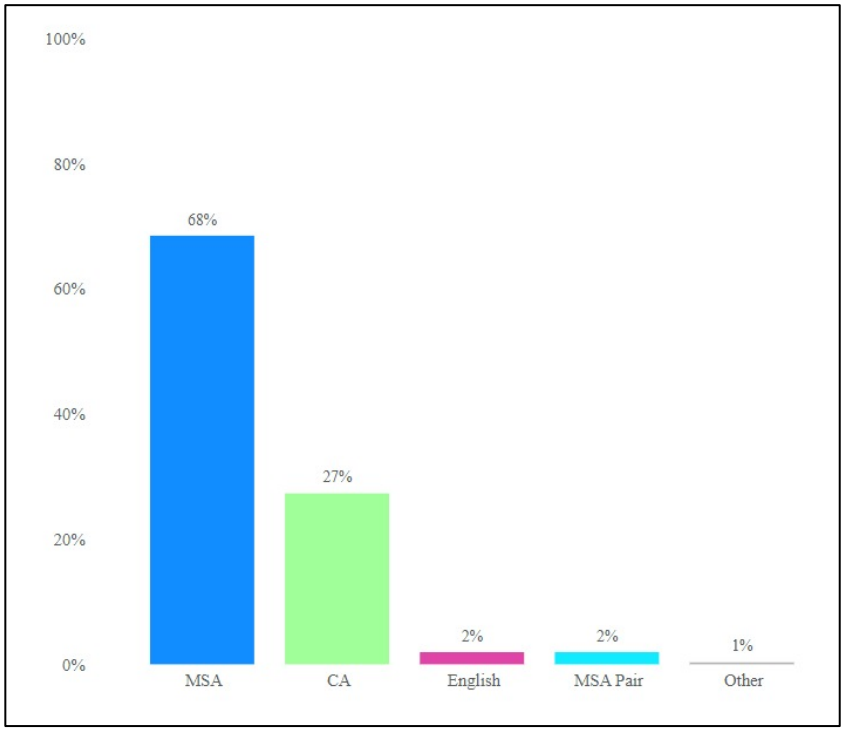


Figure 5.3: Distribution of matrix codes in the multilingual corporation tweets

Similar to the case of the monolingual tweets, MSA still dominates the matrix codes in the corporations' tweets, with 68% of the tweets using the grammatical frame of MSA, followed by CA (27%). The proportion of English matrix tweets is lower in the multilingual tweets (2%) than in the monolingual tweets (11%). Figure 5.4 demonstrates that this is because English is frequently the embedded language. As discussed above in section 5.2.1, some of the parallel text tweets are not always bilingual only as the parallel text, particularly when in Arabic, may include content from other languages and varieties. This is reflected in Figure 5.3, which shows that 2% of the tweets use an MSA Pair, the parallel text of an English equivalent, as the matrix code included content from other languages or varieties. It should be mentioned that Classical Arabic and English with Arabic script can be found among the other codes used as the matrix language. Furthermore, there are some instances of CA when used as a parallel tweet.

Figure 5.4 shows the embedded codes in the corporations' multilingual tweets (N=1,285).

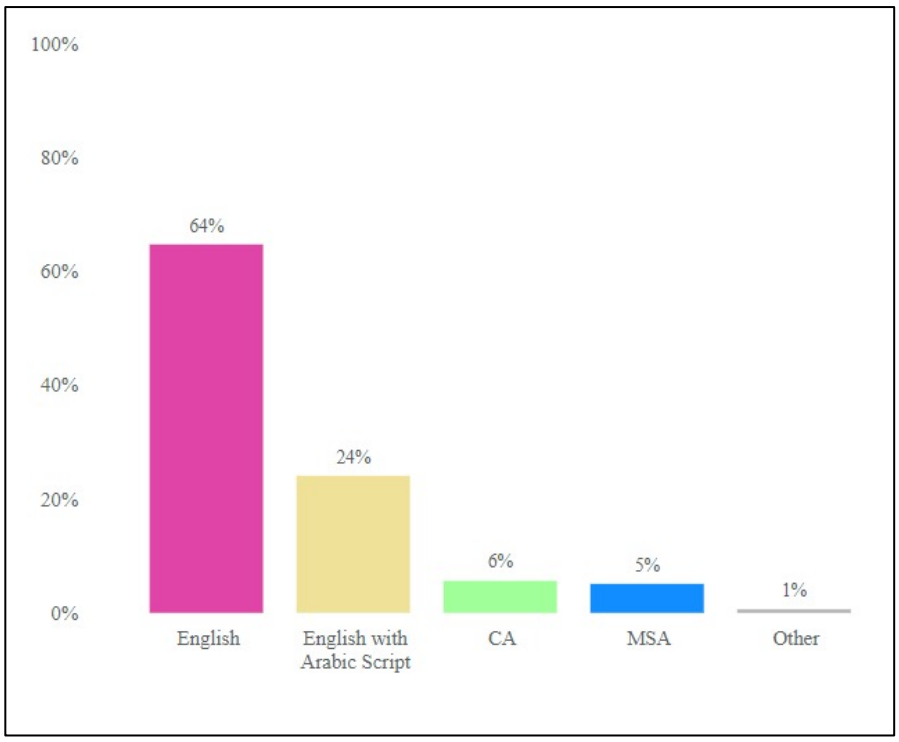


Figure 5.4: Distribution of embedded languages and varieties in the multilingual corporation tweets

The results in Figure 5.4 reveal that English dominates this category with 64%, followed by English with Arabic script at 24%. There are different rationales for why English is primarily used as an embedded language. First, if the tweet is in Arabic, then an exact English translation may follow the Arabic content, or the writer will include a brief English summary. Furthermore, although the Academy of the Arabic Language attempts to identify and establish Arabic equivalent terminology to cope with the rapid technological and scientific development globally (Al-Shbiel, 2017), English words sometimes still appear in the data such as ‘server’ and ‘VR’ (Virtual Reality). Words were also noted that have a well-established Arabic equivalent such as ‘offer’. Furthermore, there are several examples of English with Arabic script words like ‘tutorial’, ‘weekend’, ‘rainbow’, and ‘mention’.

The Arabic varieties CA and MSA are also utilised as embedded codes, although they only constitute 6% and 5%, respectively. This is mostly the case when the tweet starts with English and then continues with the parallel content in Arabic, as either a full translation or a summary in a few words. The other codes that appear in this category are Arabic with Latin script, French, and Spanish with Arabic script. This means that even if the tweet is in English, the writer or editor of the tweet may include a translation in Arabic to expand the number of potential readers.

In addition to comparing the matrix codes and embedded codes in the tweets, the data also offer insights into how the different languages and varieties combine. Table 5.1 presents a list of the embedded codes that appear with each matrix code.

Table 5.1: Distribution of embedded languages to matrix languages in the corporations' tweets

Matrix language	Embedded language	Total number	Percentage for each matrix language	Percentage of the overall multilingual tweets (N=1,285)
MSA	English	698	79%	54%
	English with Arabic script	111	13%	9%
	CA	70	8%	5%
	Arabic with Latin script	1	-	-
	Total	***880***		
CA	English with Arabic script	168	48%	13%
	English	140	40%	11%
	MSA	42	12%	3%
	Spanish with Arabic Script	1	-	-
	Total	***351***		
English	MSA	20	80%	2%
	Arabic with Latin script	3	12%	-
	CA	1	4%	-
	English with Arabic script	1	4%	-
	Total	***25***		
MSA Pair	English with Arabic script	13	52%	1%
	English	12	48%	1%
	Total	***25***		
CA Pair	English with Arabic script	2	100%	-
Classical Arabic	MSA	1	100%	-
English with Arabic Script	English	1	100%	-
	Total	***4***		
Total number of multilingual tweets		***1,285***		

Table 5.1 highlights how the different languages and varieties are used together in mixed tweets, whereby the majority of the tweets with MSA as the matrix language have English

as the embedded language (54%), followed by English with Arabic script (9%). Furthermore, in terms of CA, English with Arabic Script is employed in 13% of the mixed tweets, and English in 11%. On the other hand, when English is the matrix code, 80% of the tweets have MSA as the embedded code and only one tweet has CA as the embedded code. Therefore, it can be seen that when blending English and Arabic, English inclusion is an important requirement in many Arabic tweets, and vice versa. This is again reflected in the parallel text tweets (i.e., MSA Pair and CA Pair), which despite being translations of English content, still embed English or English with Arabic script. Furthermore, the corporations' utilisation of primarily MSA and CA in mixed tweets shows that when they use bilingual text, either parallel or not, the trend in the majority of the tweets is for Arabic followed by English. This is not surprising, as Saudi Arabia is continuously expanding its economic relationships with other countries, with new ventures regularly being undertaken between Saudi corporations and foreign investors and companies. Therefore, English is vital for this global communication.

Table 5.2 details the proportions of the different languages and varieties that appear in the corporations' data.

Table 5.2: Proportion of languages and varieties in the corporations' tweets

No.	Language or variety	Appearance in monolingual tweets	Appearance as a matrix code	Appearance as an embedded code	Instances of variety in tweets	Percentage of corporation tweets (N=9,370)
1	MSA	4,765	880	64	5,709	61%
2	CA	1,259	351	77	1,687	18%
3	English	895	25	835	1,755	19%
4	MSA Pair	529	25	-	554	6%
	English Pair	584	-	-	584	6%
	CA Pair	28	2	-	30	<1%
5	Classical Arabic	15	1	-	16	<1%
6	Arabic with Latin Script	2	-	-	2	<1%
7	English with Arabic Script	1	1	306	308	3%
8	Emoji	6	-	-	6	<1%

9	Urdu	1	-	-	1	<1%
	Total	**8,085**	**1,285**		**9,370**	

Table 5.2 considers the proportions of the different languages and varieties in all the corporations' tweets, whether monolingual, bilingual, or multilingual. With 14% of the corporations' tweets either bilingual or multilingual, as discussed above in section 5.2.1, the total number displayed in Table 5.2 (9,370) counts the monolingual tweets and the matrix language; however, the number of embedded languages is given but not totalled, to ensure consistency in the total number of monolingual tweets. Furthermore, the proportions are calculated based on the number of the corporation multilingual tweets. As Table 5.2 shows, the variety most frequently employed is MSA, which appears in 67% of the tweets (MSA + MSA Pair). CA emerged as the second most utilised variety in the corporations' monolingual tweets, and the third most used variety as the matrix code in multilingual tweets, as shown in Figure 5.3 and Figure 5.4, respectively. However, regarding the proportion of all languages and varieties in tweets, it was found that English is the second most employed variety in the corporations' tweets, appearing in 25% of the tweets (English + English Pair). Table 5.2 also shows the other languages and varieties that appear in the corporations' data, and it is important to highlight that English with Arabic script appears in 3% of the tweets, primarily in lexical items. The single tweet solely in English with Arabic script contained the word 'cute' to describe a picture included in the tweet.

The first research question of this study aims to explore the different proportions of the languages and varieties utilised on Twitter by corporations. The analyses above reveal that MSA is the most used variety, even though previous studies on digital communication and especially social media showed that users tend to use less formal and creative forms of the language that imitate face-to-face communication (cf. Warschauer et al., 2002; Danet & Herring, 2007; Dorleijn, 2016). Furthermore, Arabic with Latin script, which has been reported to be very popular on social media (cf. Yaghan, 2008; Hamdan, 2016), only appears in one tweet. However, there are still creative means of using language on Twitter, as seen in the use of CA that does not have a formal writing system, and the use of English with Arabic script. After considering the proportion of the different languages and varieties that appear in the corporations' data from Twitter, it is worth concluding this section with a summary of the key findings from the analysis:

- Although there is little code-mixing within the tweets, many varieties are visible across the tweets.

- MSA is the main language used, but English still appears in approximately 20% of the tweets (as the matrix/only language, or as an embedded language).
- CA is considered an important resource for the users, utilised solely in the tweets or alongside MSA and English.
- Arabizi does not have the same popularity compared to previous years and studies (cf. Warschauer et al., 2002; Palfreyman and Khalil, 2003). On the other hand, English with Arabic script now exists and is employed by the users in Saudi Arabia.

The following section breaks down the different languages and varieties by patterns to illustrate how the languages and varieties are employed in the timeline of the corporations' official accounts.

5.2.4 The patterns of languages and varieties in the corporations' Twitter data

The previous sections presented an overview of the different languages and varieties utilised by the corporations on Twitter. However, different accounts employ differing patterns for using these languages and varieties, and therefore it is important to identify these patterns. This section discusses how the different codes are displayed on the Twitter timelines of the corporations, and the different patterns that emerged from the data. First, this section shows the overall usage of the languages and varieties by the corporations. Then, the main patterns of languages and varieties that appear in the data will be presented. Next, these patterns will be discussed using samples from different corporations' accounts. After that, examples from the most common pattern found in the corporations' tweets will be considered to respond to the third research question related to the patterns of language use by corporations on Twitter.

Table 5.3 displays the frequency of the languages and varieties that appeared in the corporations' tweets. Furthermore, it displays the number of accounts and tweets that used each language and variety.

Table 5.3: Frequency of languages and varieties in the corporations' data

Main language or variety	Total number of corporation accounts	Percentage of accounts	Total number of tweets	Percentage of tweets
MSA	50	100%	4,765	51%
MSA Pair	18	36%	529	6%
English	31	62%	896	10%
English Pair	18	36%	584	6%

CA	24	48%	1,259	13%
CA Pair	3	6%	28	<1%
MIX	39	78%	1,285	14%
Classical Arabic	7	14%	15	<1%
Emoji	2	4%	6	<1%
Arabic with Latin Script	1	2%	1	<1%
English with Arabic Script	1	2%	1	<1%
Urdu	1	2%	1	<1%
Total			**9,370**	**100%**

Table 5.3 demonstrates that all the corporations' accounts utilise MSA in their tweets. After that, English follows with 62% of the accounts. It is also shown from the table that 36% of the corporations accounts utilise parallel text in English and MSA. However, the table reveals that the CA pair is not as commonly used as MSA since only 6% of the corporations' accounts use it parallel text bilingualism. This indicates that this strategy is mostly used for formal tweets as MSA is the variety that is associated with formality, as discussed in section 2.2 in Chapter 2. As for tweets with mixed languages and varieties, Table 5.3 shows that 78% of the accounts utilise more than one code in their tweets. Finally, the limited use of Arabic with Latin script and English with Arabic script by one account for each of these two varieties must be highlighted.

As for the patterns of multilingualism in the corporations' tweets, the languages and varieties that appear in the timeline of every account were recorded, with the language/variety included in the pattern of languages used in the account if it appeared in at least three tweets. Table 5.4 presents the main patterns of languages utilised, together along with the proposed term for each pattern, in addition to the number of corporation accounts that tweet in this pattern of languages and varieties, with the proportions of these accounts from the overall corporations' accounts (N=50) also provided.

Table 5.4: Patterns of multilingualism in the corporations' data

	Pattern	Number of corporation accounts	Percentage of corporation accounts
1	**Standard Parallel Text Bilingualism** MSA/English/MSA Pair/English Pair/MIX	15	30%
2	**Standard/Colloquial Parallel Text Bilingualism** MSA/English/MSA Pair/English Pair/CA/CA Pair/MIX	3	6%
3	**Bilingualism and Diglossic Switching** MSA/CA/English/MIX	9	18%
4	**Diglossic Switching** MSA/CA/MIX	9	18%
5	**MSA Only**	7	14%
6	**English Only**	3	6%
7	**Other**	4	8%
	Total	**50**	**100%**

The term 'parallel-text bilingualism' was used by Coupland (2012) in his research on framing Welsh and English in Welsh public spaces, and it is used here to describe the same phenomenon but in an online setting. The pattern of using parallel Arabic and English tweets for all content is widely used, particularly by the corporations and institutions. It should be noted that even images and videos attached to a tweet are usually posted in Arabic and English. There are three main ways of using the two languages, that is, English and Arabic: (i) a translation in the same tweet, (ii) a separate tweet for each language, and (iii) separate Twitter accounts for each language.

Since Arabic is the official language in Saudi Arabia, it is therefore to be expected that MSA is a variety that is used in public corporations' accounts, except for those accounts that only post English tweets. It should be noted that the English-only accounts have parallel Arabic-only accounts, such as the official Twitter accounts for the Al Ittihad football club (i.e., @ittihad and @ittihad_en). The account employed Arabic and English in the same account until 2015, when the club's executives decided to post all future English tweets from a new English-only account. The target audience for this English account is current and potential players, the team's global audience, and the international media since the club uses this as an official account that disseminates various information and updates about the team. In the bilingual Twitter accounts, the content is typically the same but translated into English, or vice versa. However, the content of two of the three corporations that have separate accounts for English and Arabic is not always identical, and it can be inferred from this

differing content that there are different editorial teams for the accounts. It is worth adding that the Arabic-only accounts for these corporations also include some English tweets, such as the following example presented in Figure 5.5.

Figure 5.5: An English tweet displayed on an Arabic-only account

Figure 5.5 shows an example of an English tweet in the middle of an account that generally only posts in Arabic, as stated in the description of the account. The use of English in this tweet was purely for advertising purposes and the target was to recruit local followers. It was expected to be familiar to some of the followers as it plays on a theme from the HBO show Game of Thrones (*Winter is coming*), which was popular at the time of this tweet.

Figure 5.6 presents another example from the same account that includes parallel text bilingualism for the expression ‘clean sheet’, which is popular among football fans. The picture in the tweet thus includes the expression in Arabic, as well as a translation in English.

Figure 5.6: A tweet with the same expression in English, Arabic, and English with Arabic script

In addition, there are accounts that only tweet in Arabic, either in MSA only (pattern 5 in Table 5.4) or with diglossic switching through MSA and CA (pattern 4 in Table 5.4). In total, 14% of the accounts use MSA only, which is not limited to a specific category of the corporations as the approach is utilised in official government, bank and magazine accounts. However, it cannot be generalised that only MSA is used in government accounts, as there are government accounts that use other patterns that include different languages (mainly English). For example, the official account of the General Entertainment Authority (@GEA_SA) only uses MSA in their tweets, although it is expected that there will be non-Arabic speakers in their audience. On the other hand, the official account of the Saudi Arabian Monetary Authority (@SAMA_GOV) uses pattern 1 (i.e., Standard Parallel Text Bilingualism) in their timeline.

The other patterns included in Table 5.4 utilise the different codes in differing proportions. Figure 5.7 presents the proportions of the different languages and varieties employed by a sample from the corporations' accounts.

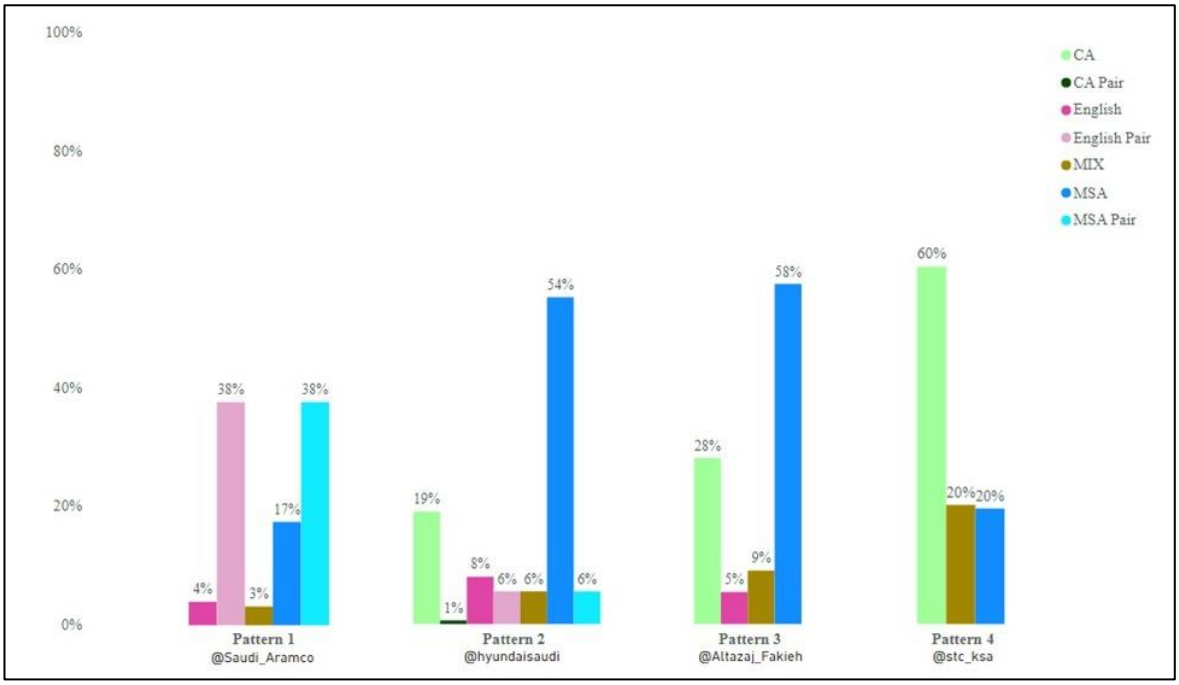


Figure 5.7: Sample of the different languages/varieties and percentages utilised by the corporations

Figure 5.7 shows that in terms of pattern 1, the Aramco corporation has an equal distribution (38%) of parallel texts in separate tweets posted in MSA (MSA Pair) and English (English Pair), while 3% of the tweets use parallel texts in the same tweet. The reason for this may be related to Twitter's word limit, whereby if the content in both languages fits in the same post, it will be posted in a single tweet, but if the content is longer or another element is utilised in the tweet (e.g., a picture or a video that is bilingual), then the content is separated into two different posts. However, this may not always be the case, and there are occasionally separate tweets for Arabic and English, even if the content is compact and could fit in one post. On the other hand, there is some content that is only communicated in MSA or in English, which can be related to the target audience of the tweet. Or, if the post concerns the live coverage of an event, for example, which appears repeatedly in different accounts, it might be faster and more practical for the editors to post it in one language only. Figure 5.8 presents an example from the MISK Foundation, where they announced an event only in one language, English.

Figure 5.8: Example from a corporation tweet about an event posted in English only

The other patterns in Figure 5.7 generally show less consistency in terms of language use. Pattern 2 was found in the Hyundai account, and pattern 3 for Al Tazaj, a local fast-food restaurant, with MSA dominating the posts at 55% and 57%, respectively, while parallel text in CA is the least employed in pattern 2 (Standard/Colloquial Parallel Text Bilingualism) at only 1%. The other languages (i.e., English, parallel texts in English, and MSA) and posts with mixed languages and varieties are used at varying levels. However, in terms of pattern 4 (diglossic switching) used in the Saudi Telecom Company (STC) account, CA is the most utilised variety (60%) followed by an equal distribution of posts using MSA (20%) and posts with mixed codes (20%). Many of the corporations tend to personify the identity of the brand, which is reflected in how they communicate with their target audience on different media platforms including social media. Therefore, CA could be the most suitable option if the intended persona of the brand is to be congenial in order to appeal to the intended audience, which is why it may be prominent in some corporations' accounts. The same brands, such as the Saudi Telecom Company in Figure 5.7, still use MSA for formal posts, while they mix different languages and varieties depending on the function of the post, as will be explained later in sections 5.2.6 and 5.2.7 in this chapter.

Despite a variety of patterns employed by the different corporations, the most common pattern that appears in the corporations' data is Standard Parallel Text Bilingualism (see Table 5.3). In this pattern, which is used by 30% of the corporations in their tweets, the parallel text in MSA and English is displayed either in separate tweets or in the same tweet. Figure 5.9 shows a sample from the corporations that employ this pattern in order to highlight the distribution of each of the languages and varieties utilised by individual accounts.

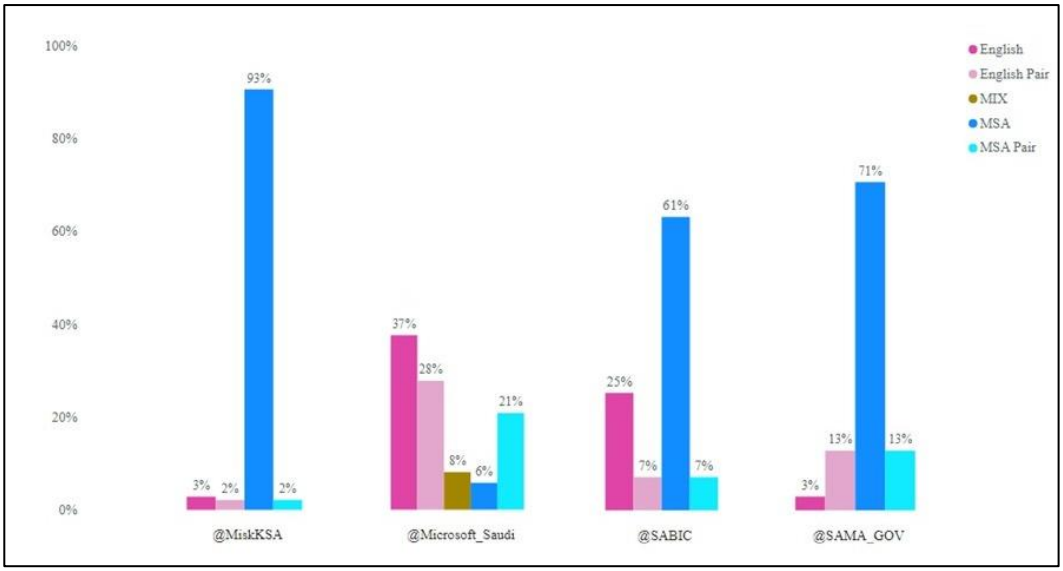


Figure 5.9: Distributions of languages and varieties employed in the standard parallel text bilingualism pattern

Figure 5.9 shows that MSA in some accounts is the most used language and variety, as seen in the first example from MiskKSA (93%), while being the least utilised in the case of Microsoft_Saudi (6%). Moreover, the Microsoft_Saudi example illustrates how the type of business could influence the language or variety employed in the communication strategy, with English used more in monolingual and parallel bilingual tweets in this case. This can also be explained in light of the target audience of the company, as highlighted in Figure 5.9, since the majority of Microsoft_Saudi's posts would concern innovative technology products that may not have equivalent terms in Arabic. Therefore, in parallel posts with the same content, the English pair constitute 28% of the tweets, while the MSA pair is only 21%, even though they would be expected to have similar distributions. Therefore, even when the post is translated into Arabic, there is still embedded content in either English or English with Arabic script within the Arabic matrix, as explained in section 2.3.3.

The third example from SABIC again supports the argument that the choice of language or variety depends on the content and the target audience, as there is a considerable difference between MSA only (61%) and English only (25%), while the parallel bilingual posts constitute 14% in total. The final example of SAMA_GOV exemplifies how government accounts also tend to use parallel bilingual posts (collectively 26%), MSA only (71%), or English only (3%).

This section explored the most common patterns for languages and varieties utilised by the corporations, which is one of the research questions of the current study. The sample from

the different corporations' accounts that use the same pattern in Figure 5.7 and Figure 5.9 helped to visually highlight the different proportions of languages and varieties in use.

5.2.5 The use of paralinguistic cues in the corporations' tweets

The coding of the data also considered the use of paralinguistic cues employed in the tweets, which were primarily emojis used to express feelings and emotions to imitate face-to-face interactions. Figure 5.10 shows the distribution of the different paralinguistic cues found within the corporations' data.

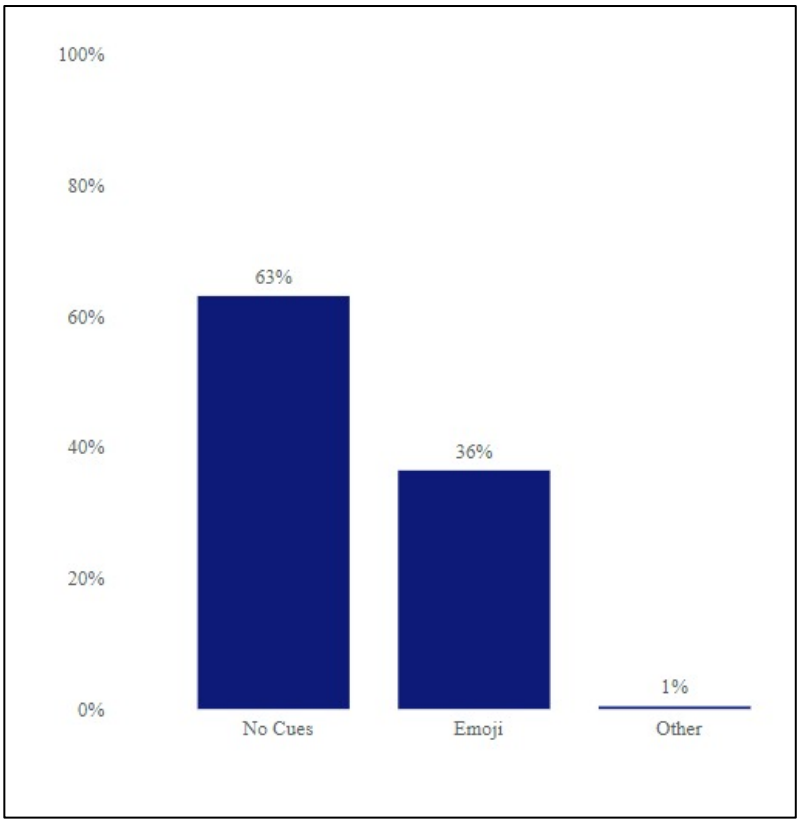


Figure 5.10: Distribution of paralinguistic cues in the corporations' tweets

Figure 5.10 shows that 63% of the total corporations' tweets (N=9,370) do not include any paralinguistic cues besides the text in the tweet. Since the data in this category involve corporate communication, it is perhaps unsurprising that the tweets do not always include paralinguistic features since the main reason for using paralinguistic cues (e.g., emojis and emoticons) is expressing emotions. However, 37% of the corporations' tweets do use paralinguistic cues besides the text, with the emoji emerging as an important element in the Twitter data since 36% of the corporations' tweets include one or more of these icons. As mentioned in section 5.2.4, it could be argued that emojis are utilised when corporations

attempt to personify the brand and to communicate in a congenial tone with the target audience. Furthermore, emojis are employed to support the text in conveying a message.

The other category in Figure 5.10 consists of emoticons (e.g., :)) or using multiple punctuation marks (e.g., !!!!), and letters such as *Nooooo.* It should be noted that the usage rates of emojis and emoticons are expected to be higher with the influencers and ordinary users.

It is also essential to explore the languages and varieties used with paralinguistic cues to determine if the general trend remains the same, or if there is a difference. As emojis are the most utilised paralinguistic cue with text in tweets, they were selected to investigate the associated languages and varieties used. Table 5.5 presents the proportion of tweets with emojis from each language or variety, in order to match the rates of each language and variety that appear in the corporations' data.

Table 5.5: Proportion of tweets with emojis in the corporations' data for each language/variety

No.	Language or variety	Number of tweets	Number of tweets with emoji	Percentage
1	MSA	4,765	1,612	34%
2	MSA Pair	529	19	4%
3	CA	1,259	1,028	82%
4	CA Pair	28	3	11%
5	English	896	279	31%
6	English Pair	584	33	6%
7	MIX	1,285	433	34%
8	English with Arabic script	1	1	100%
9	Emoji only	6	-	-

Table 5.5 explores the frequency of emoji usage with each language and variety. Although the table shows that the emoji is primarily used with MSA, it is also noticed that it only appears in 34% of the tweets written in MSA. On the other hand, 82% of the tweets written in CA employ an emoji. This indicates that CA and the emoji are frequently utilised together as a mechanism to personify the tweet and make it more congenial to the audience. On the other hand, since 34% and 31% of the tweets in MSA and English, respectively, use an emoji, this shows that MSA and English may be considered more formal. The other varieties and languages include MSA Pair, CA Pair, Classical Arabic, and English with Arabic script,

which means that the emoji appears with all the languages and varieties in the data. Furthermore, only six posts by the corporations utilise the emoji without any text. Figure 5.11 presents the emoji usage data from Table 5.5 in visual format in terms of each language and variety, to allow easy comparison between the proportions of usage.

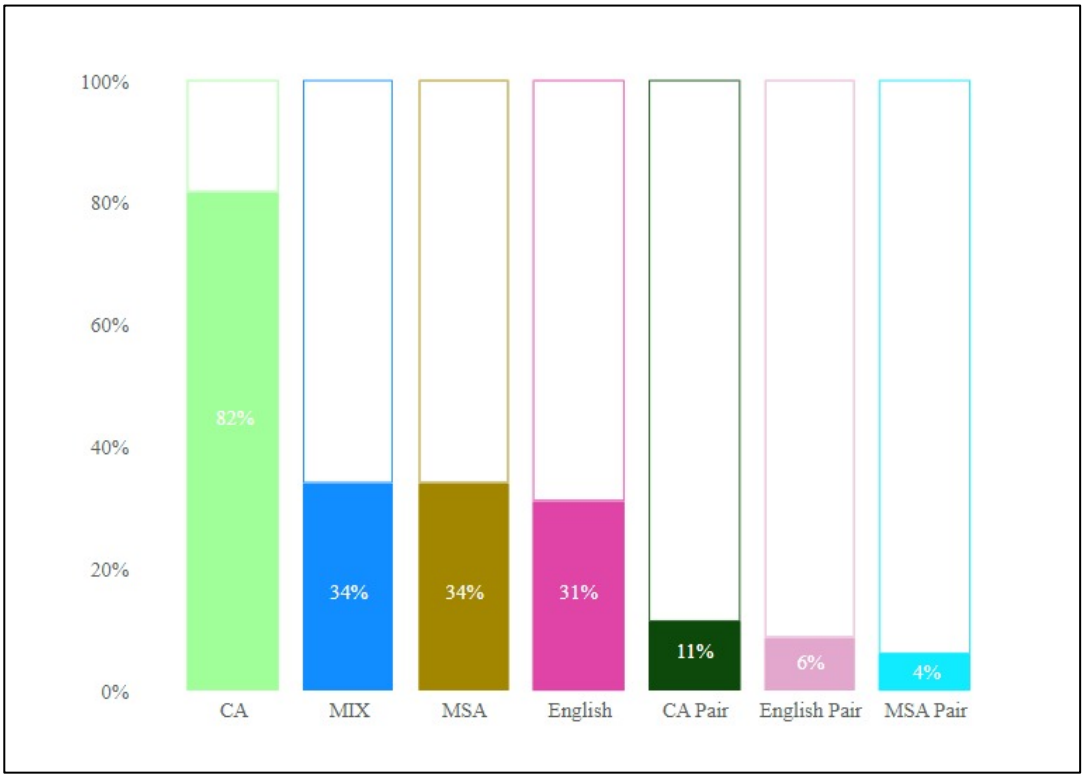


Figure 5.11: Distribution of emojis employed with each language and variety in the corporations' Twitter data

After considering the different languages and varieties used in the corporations' tweets, section 5.2.6 explores the functions of the tweets extracted from the same data, and links the functions to the language choice in the corporations' posts.

5.2.6 The functions in the corporations' tweets

The aim of this section is to answer the second research question, which concerns the functions of the posts made on Twitter. As explained in Chapter 4, the functions that emerged from the data are categorised into six main groups:

- Formulaic Purposes
- Culturally Specific Genres
- Reported Speech

- Dialogical Interrelation
- Marketing
- Self-expression

In the corporations' data, only the first five functions were found, with the self-expression category emerging in Chapter 6 and Chapter 7 in relation to the influencers' and ordinary users' tweets, respectively, regarding their personal thoughts. Table 5.6 presents the number and proportion of tweets for each of the main functions.

Table 5.6: Number and proportion of tweets for each main function

No.	Function	Number of tweets	Percentage
1	Formulaic purposes	397	4%
2	Culturally specific genres	164	2%
3	Reported speech	161	2%
4	Dialogical interrelation	438	5%
5	Marketing	8,210	87%
	Total	**9,370**	**100%**

Since corporations' primary purpose for maintaining an active Twitter account is marketing, as discussed in previous chapters, it is unsurprising that the dominant function found in the corporation's tweets is directly marketing content (87%). It should be noted that the corporations do not necessarily promote all their products to their followers, but they all promote their activities to different stakeholders and the global audience, which represents a slightly less conventional form of marketing. For example, while Aramco, the official oil company of Saudi Arabia, does not sell any products to individuals, the company still has a very strong presence online, as measured by the number of followers and the degree of engagement on Twitter and social media in general.

The second most frequent function is dialogical interrelation (5%), which corporations employ as a strategy to interact with their social media audiences; for example, through asking questions or introducing a competition. Then, formulaic purposes (4%) are utilised in different instances such as national celebrations. Finally, tweets that include reported speech and culturally specific genres are the least frequently used at 2% for each function.

As the majority of the corporations' tweets involve direct marketing, it is important to explore the specific categorisation of the tweets in this category. Table 5.7 presents the distribution of the functions in the direct marketing category, along with the number of tweets and the proportion.

Table 5.7: Number and proportion of the subfunctions in the corporations' Twitter data for marketing

No.	Marketing function	Number of tweets	Percentage
1	Information	3,997	49%
2	Advertisement	3,933	47%
3	Corporate social responsibility	139	2%
4	Awareness campaign	122	2%
5	Recruitment	17	
6	Testimonial	2	
	Total	**8,210**	**100%**

The subcategory of information within the marketing category, which can be linked to the representative speech act, is the most frequently employed subfunction (49%), with corporations tending to share considerable information about their activities in order to promote their brands. Although the information they mention does not include a call-to-action for the customers or consumers, the intention is to persuade the audience to have a positive perception of the brand. The next most frequent function for the tweets is direct advertising (47%), whereby corporations promote their brands or offer benefits when their audience use them, which is therefore classified as a commissive speech act. The next subfunction is corporate social responsibility, which tends to involve the initiatives carried out by corporations to support the community, and is only rarely utilised (2%). Finally, the least frequently used functions (collectively 2%) are awareness campaigns on different occasions, personnel recruitment, and testimonials by different users of the brands. The categorisations of the functions will now be utilised in section 5.2.7 to explore how the different languages and varieties are used to express them.

5.2.7 The languages and varieties employed with functions in the corporations' tweets

This section explores the response to the third research question, which concerns the use of different languages and varieties by the corporations in relation to the functions.

Figure 5.12 provides an overview of the proportions of languages and varieties utilised in each function, while Table 5.8 details the number of tweets for the languages and varieties employed with each main function.

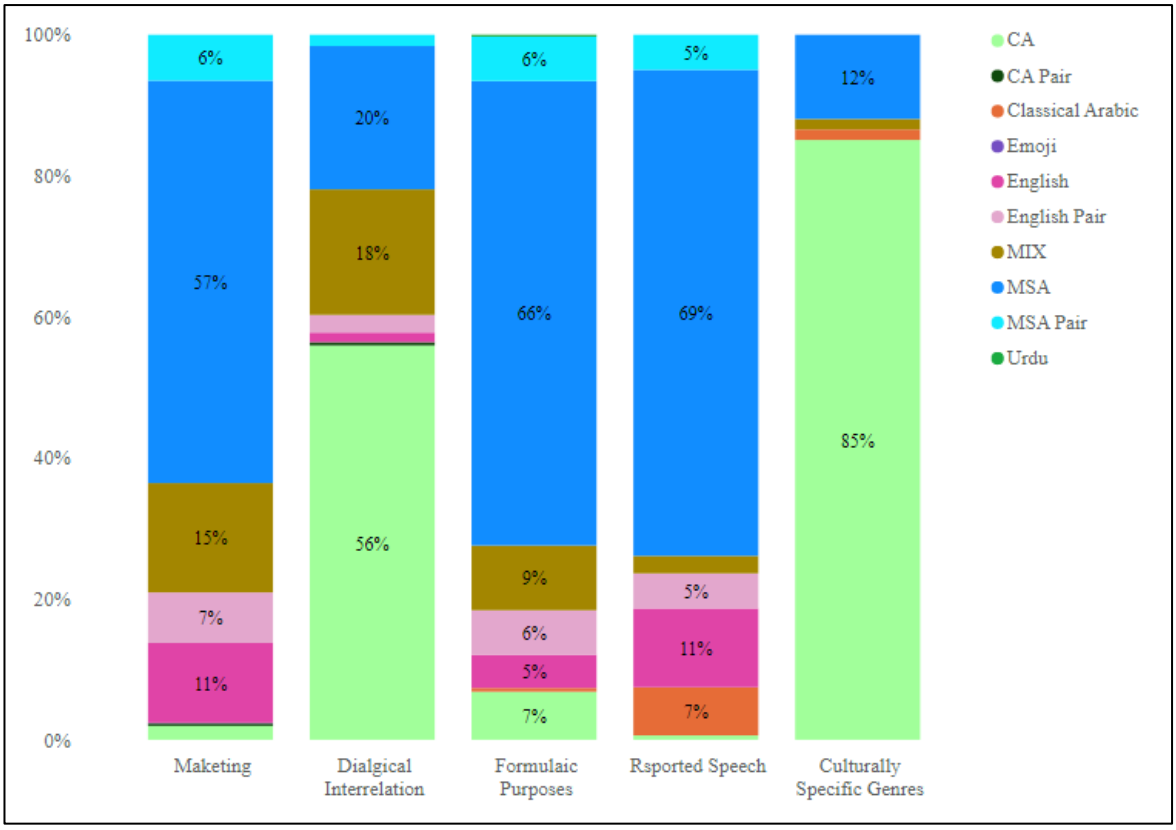


Figure 5.12: Proportion of the functions and languages/varieties in the corporations' data

Table 5.8: Number of functions with the languages and varieties in the corporations' tweets

Main function	Main language	Number of tweets
Formulaic purposes	CA	42
	Classical Arabic	2
	English	18
	English Pair	24
	MIX	35
	MSA	251
	MSA Pair	24
	Urdu	1
Total		397 (4%)
Culturally specific genres	CA	57
	Classical Arabic	1
	MIX	1
	MSA	8
Total		67 (1%)

Reported speech	CA	1
	Classical Arabic	11
	English	18
	English Pair	8
	MIX	4
	MSA	111
	MSA Pair	8
Total		161 (2%)
Dialogical interrelation	CA	245
	CA Pair	2
	English	6
	English Pair	11
	MIX	78
	MSA	89
	MSA Pair	7
Total		438 (5%)
Marketing	CA	914
	CA Pair	26
	Classical Arabic	1
	English	854
	English Pair	541
	MIX	1,169
	MSA	4,306
	MSA Pair	490
	Emoji	6
Total		8307 (88%)
Total number of tweets		**9,370**

As Figure 5.12 and Table 5.8 show, there is variation in the different languages and varieties employed by corporations on Twitter, depending on the function. Since MSA is the variety most utilised by corporations (see section 5.2.1), Figure 5.12 highlights that MSA is primarily used in reported speech (69%), formulaic purposes (66%) and marketing (57%). Moreover, CA is the variety most frequently employed in culturally specific genres (85%) and dialogical interrelation (56%). On the other hand, English is the language least utilised for dialogical interrelation (6%), and is never used for culturally specific genres. As CA is the variety employed for daily conversation, it can be explained that these tweets are written in CA to personify the brand and attempt to communicate naturally with the audience. For example, it may be expected that a daily conversational enquiry such as *What are you having for breakfast today?* may receive more responses if it is written in CA rather than MSA. English,

monolingually in one tweet or bilingually in parallel texts, is mostly utilised in marketing (18%), reported speech (16%), and formulaic purposes (11%). This shows the importance of English in the formal communication of the corporations, and can also be interpreted as an attempt to attract a global as opposed to merely a local audience. Mixed tweets, as Figure 5.12 shows, are more common in dialogical interrelation and marketing. Considering the word limit of posts on Twitter, a simple greeting in two languages such as Arabic and English can be accommodated within the word limit in one tweet, which is the main reason why the majority of mixed tweets are in the mixed category. The code-mixing can also use the different varieties of Arabic, namely MSA and CA. Sometimes, the reason or occasion can be in one variety (e.g., MSA), and the greeting in the other (e.g., CA). It also should be noted that several of the tweets use only emojis without any text (N=6), with one tweet in Urdu.

5.2.8 Summary of the analysis of the corporations' tweets

To summarise the findings, the analysis of the corporations' tweets reveals the following:

- Although there is some code-switching within the tweets, the main evidence of code-switching is between the tweets displayed on the timeline of the accounts. The data also reveal that some of the corporations employ a strategy whereby they send two paired tweets, generally one in MSA and one in English.
- The main language employed across the tweets is MSA (61%), followed by English (19%) and CA (18%). There are different patterns for the languages and varieties utilised by the corporations; some use MSA almost exclusively, while others have a balance of MSA, English, and CA.
- The emoji is used more frequently with CA than other varieties, and is employed to make the tweet appear less formal.
- The functions of the tweets affect the variety chosen, in some cases. For example, CA is primarily utilised in tweets for formulaic purposes (85%).

Following the analysis of the corporations' tweets, the next section presents the analysis of the interviews with the marketing executives to examine their perspectives regarding the languages and varieties used on Twitter in Saudi Arabia.

5.3 Analysis of the corporations' interviews

This section reports the findings of the interviews conducted with marketing executives from seven corporations, comprising of 3 males and 4 females. While the interviewees had

different positions, they were all key members of the team that handles social media; some were the actual individuals handling the account, while others were involved in the decision-making process. For ease of reference, they are all referred to as 'marketing executives' in this section. The introductory part of the interviews covered general topics, which enabled the participants to discuss their background and the corporations' goals, target audience, and frequency of posts, before focusing on the research area, namely, the languages and varieties utilised on Twitter, in addition to the paralinguistic cues and the translation of tweets. The interview results are presented as follows. First, sections 5.3.1–5.3.3 present the themes and opinions related to the languages and varieties, in addition to the paralinguistic cues. Then, section 5.3.4 discusses the aims and goals of the corporations on social media, and relates these to the functions of the tweets discussed in section 5.2.6. After that, sections 5.3.5 and 5.3.6 explore the frequency of posts and the corporations' target audience on Twitter. Finally, a summary of the findings is presented.

The following sections present summaries of the results in tables, in addition to interview extracts related to each section. It must be noted that some of the marketing executives provided several responses. However, the proportions displayed in the tables are counted based on the number of respondents who provided each answer.

5.3.1 Languages and varieties employed for communication on Twitter

This section aims to help respond to the first and fourth research questions that explore the languages and varieties utilised on Twitter, and the patterns for using these languages and varieties. The interviews thus investigate the research aim, which is exploring the use of the different languages and varieties on Twitter. Table 5.9 presents the languages and varieties primarily used in the corporations' accounts, as stated by the interviewed marketing executives.

Table 5.9: Corporations' languages and varieties used on Twitter

No.	Theme	Number of interviewees	Percentage of interviewees
1	MSA only	3	43%
2	MSA/CA (white dialect)	3	43%
3	English	2	29%
4	CA only	1	14%

Before discussing the different languages and varieties mentioned by the interviewees, it is important to highlight the 'white' or 'plain' dialect referred to by the marketing executives

and the social media influencers. This variety is a hybrid between MSA and CA (see section 2.2.2.3) that may adopt the syntax and morphology of CA, and the lexical items from MSA, which means that there is no regional marking for the white dialect.

The general purpose for using this variety was stated as achieving a formal yet congenial tone. Furthermore, since it adopts lexical items from MSA, local colloquial terms are not employed so the origin of the user will potentially be concealed, while ensuring that the post is more accessible for all Arabic speakers. As was also apparent from the interviews, the users are aware of this variety, and they consciously edit their posts to represent it. They referred to it by different terms such as the 'white dialect' and 'plain dialect', and it was described by many of the interviewees as "*neither Fusha [MSA] nor A'amia [CA]*".

As discussed in Chapter 4, the coding of the tweets did not include the white dialect as a separate variety, since it cannot be identified based on a clear criterion and it could either be categorised as CA or MSA based on the syntax. Moreover, a post may have the features of MSA, but could present as CA, depending on how it is read by the audience.

In support of this, interviewee CORP.1 mentioned that using the white dialect could be interpreted by the audience as a language mistake, rather than the intentional use of a congenial variety of Arabic:

Extract 5.1

> *We used to write in an official language [MSA] but then we looked at studies that recommend using a language that is close to the audience and not rigid but also keeps the same language foundation…the audience is diverse and if you use a colloquial term they may attack you for it. However, I feel that this variety can be used by individuals but not corporations as they should only use the standard form. If you use colloquial language and somebody criticises you, they are right. But if you write standard, nobody will criticise you. Also, people may think that we have made a mistake and not that we are using the colloquial terms on purpose.*

The extract above also shows that market research encourages some corporations to use the white dialect. However, the respondent stated that their corporation still prefers MSA. The analysis revealed that the majority of the corporations' tweets are in MSA (61%) and MSA Pair (6%), but this was not reflected in the interviews as only three out of the seven marketing executives (43%) reported that they use only MSA in the tweets. However, another three interviewees mentioned that they mostly use the plain white dialect that may have features of MSA, as discussed above. One of the interviewees discussed one of the

limitations of using MSA, which "*is used rarely for new company announcements*". However, the data from the same account reveals that MSA is employed in 73% of the tweets. Others expressed different purposes for using MSA and the white dialect. For example, one marketing executive reported that "*MSA is used for invitations and plain Arabic for storytelling in events*". The respective account for this corporation was found to utilise MSA in 17% of the tweets. Furthermore, the white dialect can be used to encourage the audience to engage with the posts, as expressed by interviewee CORP.5: "*In formal posts, we use only Standard Arabic, but we have some posts in the white dialect to encourage people to give us their opinions and feedback. We don't use slang, only white dialect*". This interviewee perceived the white dialect to be a better option for communication than CA, which is referred to as 'slang', even though in Arabic the word *ameiia*, which is CA, is commonly employed, while slang is utilised to describe a less formal variety than CA. This highlights that for branding, some marketing executives find using CA in corporate communication to be similar to using slang in ordinary communication, although there are some exceptions.

In terms of code-mixing in the tweets, interviewee CORP.2 reported that this may be determined by the length of the post: "*If there is space, we post in two languages, but if the post is too wordy, I stick to one language with some mixing*".

Furthermore, different languages and varieties, specifically English in Arabic script as mentioned in the extract below from interviewee CORP.2, could be used based on the editing team's expectations regarding the audience's ability to understand them:

Extract 5.2

> *We sometimes use English words in Arabic script for words that could be understood by our audience on Twitter such as 'tutorial' and 'masterclass' because they are known. Honestly, nobody until now asked for the meaning of the words and the words are more clear than the Arabic equivalent 'Jalsa Esteshariia'.*

On the other hand, it was shown in the analysis of the tweet data for the corporations and the interviews with the marketing executives that Arabizi is not a common option for communication on Twitter, with interviewee CORP.7 stating, "*we don't use Arabizi as we prefer to write in English or Arabic*". Similar unfavourable attitudes towards Arabizi were expressed by other interviewees, including CORP.1: "*Arabizi is very unprofessional. Also, I don't even continue to read the post if it is in Arabizi. It is a disaster if companies use it*".

One of the relevant themes that emerged in the interviews was whether the device used to post the tweet affects their choice of language, especially if the tweet is not pre-planned based on a schedule. The majority of the interviewees stated that they usually employ mobile phones for posting on Twitter, with one of the marketing executives expressing this device as a preference since using a mobile phone to post is easier than computers, especially during events, because the pictures are taken and ready on the device, with only the typing of the accompanying message required.

Interviewee CORP.2 expressed the difference noticed when posting via a computer versus a mobile phone: "*We usually post using mobile phones and I noticed that when I post on a computer, my language becomes more formal, and I don't use Internet language and characters. On mobile phones, my language is more friendly*". This also may correlate with the function of the tweets. For example, planned tweets might be posted from a computer, while tweets covering events would be posted from a phone for practical reasons.

5.3.2 The interviewees' perceptions regarding the languages used on Twitter

The marketing executives' opinions were sought regarding the ideal language and varieties employed for the Twitter audience in Saudi Arabia, with the following sections presenting the findings about the language that should be utilised on Twitter, the language with the most audience interaction, and whether tweets in English should be translated into Arabic, and vice versa.

5.3.2.1 The ideal language or variety for use on Twitter

The marketing executives discussed the expectations of the languages and varieties that should be utilised in Saudi Arabia on Twitter, with Table 5.10 summarising their responses when they were asked about their opinion on the language.

Table 5.10: Corporations' opinions regarding the optimum language or variety for use on Twitter

No.	Theme	Number	Percentage
1	MSA	3	43%
2	Depends on the content	3	43%
3	MSA/CA (white dialect)	2	29%
4	English	1	14%

In line with the results of the analysis of the corporations' tweets, it can be seen from Table 5.10 that MSA was deemed the optimum tweeting language or variety for three of the seven interviewees. The following extract by CORP.3 shows the commitment of one marketing executive's corporation to MSA:

Extract 5.3

> *As an institution, I think we are responsible for preserving the standard form of the language and making it part of our mission. Even if some people may be against this, in the future we will be remembered as an institution that is protecting the identity [national/Arab/Saudi]. I think every institution must keep at least 80% of its content in Standard Arabic.*

On the other hand, 43% of the interviewees asserted that the language or variety employed depends on the content of the tweet. Some of these interviewees said that no one specific language or variety should be utilised on Twitter. One of the important factors affecting the choice of language is the persona of the brand, as expressed by CORP.5:

Extract 5.4

> *The persona of the brand, the positioning, and the communication guidelines dictate the language that should be used on the different social media platforms. There should not be a standard for the language being used on social media, but it depends on the content and how the character of the brand should communicate it to the target audience.*

While the persona of the brand remained an important factor in choosing the language, other interviewees provided different views on the ideal language utilised on Twitter, such as the view from CORP.3:

Extract 5.5

> *If the post is targeting average locals in Saudi Arabia and it uses English for the tweet, then it is considered wasted content because most followers may not understand it. But sometimes the post is in English for a certain reason, and it targets people who are expected to be bilingual.*

Extract 5.5 highlights that using English in different posts will limit the audience size by targeting only those users that can understand both languages, who may be from a certain socio-economic background. Furthermore, English, in addition to the white dialect, was again favoured by interviewee CORP.2 due to its international appeal, although the interviewee cautioned not to become too colloquial or use influencers who might employ offensive language:

Extract 5.6

> *If you ask me, I think it should be English because it is international, but also white dialect is important. However, corporations should not go beyond the white dialect. Using a 'very colloquial' dialect is not good, especially if you want an audience from different regions to follow you and be familiar with your comments. This also affects our choice of influencers we work with, as if they use vulgar language we will not be interested to work with them and associate ourselves with them.*

5.3.2.2 The ideal language or variety for audience interaction

Although the analysis of the tweets did not consider the interactions of the audience (i.e., the number of likes, retweets, or replies), it was possible through the interviews to ask the marketing executives' opinions regarding the most appealing language or variety for audience interaction, with the responses summarised in Table 5.11.

Table 5.11: Ideal language or variety for audience interaction

No.	Theme	Total	Percentage
1	MSA/CA (white dialect)	4	57%
2	MSA	3	43%
3	CA	1	14%

Table 5.11 shows that the most ideal interactions are with the different varieties of Arabic, whereby none of the interviewees mentioned English. Interviewee CORP.1 indicated that "*there is no interaction at all with the English posts and most interactions are with the white dialect and rarely with Standard Arabic*", which reflects exactly the results presented in Table 5.11.

As noted earlier, some corporations allocate different Twitter accounts for different languages (e.g., one Twitter account for Arabic tweets, and another for English), while one of the corporations also has Twitter accounts in French and Spanish, although the interaction is not the same across all accounts. The following is an extract from CORP.3:

Extract 5.7

> *Our official account is in Arabic, but we have a separate account in English and other accounts for Spanish and French. This is a trending phenomenon in our sector, and we had to create accounts for the different languages. Honestly, there is not much engagement on these pages, but they are important for our presence.*

5.3.2.3 The need for English posts to be translated

While the results above indicate that the use of a certain language or variety may be intended to serve purposes related to the brand persona or the specific function of the tweet, 14% of the corporations' tweets are translated into either English or Arabic. The interviewees were asked if all tweets in English should be translated, considering that they involve corporations in Saudi Arabia, with their responses summarised in Table 5.12.

Table 5.12: Interviewees' opinions on the need to translate posts

No.	Theme	Number	Percentage
1	Yes	3	42%
2	No	2	29%
3	Depends on the user and the audience	2	29%
	Total	**7**	**100%**

Table 5.12 shows that the majority of the interviewees (42%) believed that posts should be translated, 29% disagreed and another 29% felt that the translation of tweets depends on the user and the audience. This indicates that the strategy of parallel tweets followed by some of the corporations as discussed in section 5.2.1 is favoured by the majority of the interviewed marketing executives.

5.3.3 Using paralinguistic cues

The pilot study and the initial analysis of the tweets revealed that emojis are the paralinguistic cue most employed in the tweets in comparison to the other paralinguistic cues that appear in the data (i.e., punctuation and letters). Therefore, the interviews also focused on the use of emojis, with Table 5.13 showing the frequency of emoji usage with the tweets, as reported by the marketing executives.

Table 5.13: Corporations' use of the emoji on Twitter

No.	Theme	Number	Percentage
1	Sometimes	3	43%
2	Never	3	43%
3	Always	1	14%
	Total	**7**	**100%**

The analysis of the corporations' tweets discussed in section 5.2.4 revealed that 36% of the tweets contained an emoji. In line with these results, Table 5.13 indicates that the use of the emoji was not preferred by the majority of the interviewed marketing executives, as 86% reported never or only sometimes using this paralinguistic cue. It was expressed in the interviews that using emojis may not match the brand guidelines of the corporation, especially if the brand has a 'neat' and 'classic' persona. It was also expressed that corporations attempt to maintain their brand identity rather than experimenting with languages that may or may not appeal to a wider audience on Twitter.

For other corporations, perhaps with different brand guidelines or greater flexibility regarding the language being used on Twitter, it would be acceptable to use emojis in certain situations such as when responding and in certain congenial posts, as expressed in the interviews. It appears that the corporations did not necessarily view the content in individual tweet responses as part of their branding, so were more flexible in their use of emojis. Individual responses are generally closer to having a conversation and emojis could allow the discussion to be more casual.

It also appears that some corporations categorise emojis as formal and informal. Apparently, official emojis are those utilised to indicate the time or location, such as the calendar or building emoji. On the other hand, informal emojis include smileys and people. It was mentioned by interviewee CORP.3 that "*the formal emojis [time & location] are often used because they help with the word limit and make the tweet more attractive*". It can therefore be assumed that formal tweets written in MSA for the announcement of events by this corporation, for example, could include the time and location icons, but not facial emojis.

5.3.4 Aims and goals in social media

This section discusses the findings related to the second research question:

- What are the functions of the codes used on Twitter?

This research question aims to investigate the different functions of the languages and varieties employed in the tweets. Therefore, it is important to understand the context of the tweets and the main goals of the users when constructing them. The interviews explored the social media goals of the corporations, which can be linked to the functions of the tweets discussed in the analysis in section 5.2.6. The analysis of the tweets' functions presented five main functions—formulaic purposes, culturally specific genres, reported speech, dialogical interrelation, and marketing—with Table 5.14 presenting the themes and

associated functions identified from the interviewees' responses when asked about their aims for social media in general, and specifically on Twitter.

Table 5.14: Corporations' aims and goals on social media

No.	Theme	Total	Percentage
1	**Engagement** *(formulaic purposes, culturally specific genres, and dialogical interrelation)*	6	86%
2	**Increasing the number of followers** *(formulaic purposes, culturally specific genres, reported speech, dialogical interrelation, and marketing)*	5	71%
3	**Awareness** *(formulaic purposes and marketing)*	4	57%
4	**Presence** *(formulaic purposes and marketing)*	3	43%
5	**Other**	4	57%

It can be seen from Table 5.14 that the two most common goals for corporations on social media are engagement and increasing the number of followers. In many instances, the two themes may overlap as some of the interviewees considered that increasing the number of followers is a result of engagement. In one of the interviews this was emphasised, with the interviewee reporting that they always noticed an increase in the number of followers during engagement campaigns.

For some of the marketing executives, the number of followers displayed on the account was very important, while others were more focused on the quality of the followers rather than merely the quantity, as expressed by interviewee CORP.5:

Extract 5.8

> *It matters to us who is following the account from officials and media and news accounts. There are many techniques for increasing the number of followers, but what we really care about is the quality of the followers we have rather than the number of the followers.*

The corporations' consideration of the groups of followers is one of the factors that determines the choice of language and variety. For example, MSA would be the chosen variety for posts targeting the media, as reporters can use the same content as a press release to publish in conventional media platforms such as newspapers and magazines.

On the other hand, Extract 5.9 by interviewee CORP.1 shows that some corporations prefer engagement in marketing strategies over increasing the number of followers:

Extract 5.9

> *Our target on social media started by increasing the number of followers. Then, after some years we figured out that it is not just the number that matters… Is it better to have one million who don't have any interaction or 200,000 who engage and respond to our campaign? Since last year, we don't consider the increase in the number of followers as we care about the campaigns and promoting certain posts to get more engagement.*

The same interviewee continued by explaining how they measure the engagement rate of their corporation's account through engagement rate criteria, which consider the number of comments, likes, and the people who view the post. The engagement factor depends on how motivating the post is for the audience to interact, and the language plays an important role. As highlighted in previous sections, some of the interviewees noticed that the audience engage more with certain languages and varieties, such as the white/plain dialect or CA.

The next most common goal on social media for the interviewed marketing executives was spreading awareness about the brand and the services it provides. It can be argued that the overarching goal for corporations on social media is marketing their services or products, such as providing general information about what the corporations offer in order to increase awareness, as explained by some of the marketing executives in the interviews. As discussed in section 5.2.6 in the tweets' analysis, the marketing of products or services is one of the most frequent functions that uses different languages and varieties, with MSA the preferred variation involved in 57% of the analysed tweets.

The next theme discussed by the interviewees was the importance of the corporations' presence on social media in general, and specifically on Twitter. Interviewee CORP.5 mentioned that their corporation "*spends heavily...heavily on putting strategies in place and designing campaigns for social media, and the focus on social media is Twitter*".

There are also other specific goals for the marketing executives on social media such as responding to questions from the consumers or customers, with interviewee CORP.1 highlighting that social media platforms must be linked with the call centre to answer customer queries, with this facility now being carried out by some corporations on Twitter. Furthermore, another goal discussed in the interviews was drawing traffic and directing people to stores to increase visits and sales, with event coverage being of great importance.

5.3.5 Frequency of posts

The frequency of posts was one of the challenges in collecting the data of this study, as explained in the Methodology chapter. Therefore, the second interview question aims to understand the intentions of the marketing executives in terms of the frequency of the tweets displayed on the timeline of the corporation. Furthermore, discussing the frequency of the posts provided important insight into how corporations plan and structure the tweets. Knowing the frequency of the posts provides an indication about how the tweets are produced, and whether there is prior planning and consideration before the tweets are posted, or whether it is ad hoc. This planning may also have implications for the language being used. Table 5.15 presents a summary of the responses noted in the interviews when the respondents were asked about the frequency of posting.

Table 5.15: Corporations' frequency of posts on Twitter

No.	Theme	Total	Percentage
1	No plan	4	57%
2	3–5 times per week	2	29%
3	Daily	1	14%
	Total	**7**	**100%**

Section 5.3.4 revealed that for some of the corporations, their presence on social media is an important component of their marketing and public awareness strategy. While for some of the marketing executives, a daily presence was important, others disagreed with this strategy and noticed that posting extensively on a daily basis decreases the number of followers and reduces their focus on the content.

Therefore, for some of the interviewees engagement was more important than the frequency of the posts, with one asserting the importance of audience engagement rather than the frequency of tweeting, and that some corporations have recently started reaching out to specific social media accounts to retweet the posts, even if this involves payments.

In the analysis of the multilingual tweets in section 5.2.3, it was highlighted that some of the corporations' tweets involve the live coverage of events. This was discussed with the marketing executives in order to understand how these tweets are treated, and whether there are specific protocols and considerations before they are posted. Some of the answers highlighted that there are considerable preparations for the tweets from these events. For instance, in one of the interviews, it was mentioned that there are some ad hoc

visits that show how the corporation is presented to the outer and international community. Therefore, the marketing team prepare high quality posts with pictures and post them immediately, after passing through a cycle of approval stages.

The approval for the tweets was also discussed in the interviews, where it was emphasised that checking the image of the brand and the appropriateness of the post prior to posting is an important stage. Some marketing executives stated that they usually create and approve a plan with a professional agency that handles the posting of the tweets. However, even in the case of events, the posts require final approval from the corporations' marketing team before posting. This process reveals that before posting, the different elements of the tweets, including the language, are revised and edited until approval in order to match the communication strategy.

The marketing executive CORP.3 discussed the challenges of the posting process due to the impact of language or content issues:

Extract 5.10

> *There should be only one person responsible for posting on the account. It is not an easy job...I get scared whenever I need to post something because the account is followed by millions and I will be accountable for any mistake. Any tweet must be approved by the language editor and myself before it is posted.*

Some corporations' accounts can be particularly active during certain periods, after which the activity becomes less frequent. This was confirmed by interviewee CORP.6 who reported that the number of posts depends on the activities at that time; sometimes they post 20 tweets in a day, and other times none.

Finally, it is worth noting that while the majority of the interviewees did not specify a categorisation of tweets, one of them stated that they usually consider two kinds of tweets: general tweets that concern daily news, and specific tweets that involve managerial policies and announcements. This emphasises the planning for tweets, as explained in this and previous sections.

5.3.6 Target audience on social media

The expected virtual audience of Twitter users plays an important role in corporations' choice of communication strategies and the language used. It was also frequently raised in the interviews that the content depends on the target audience. Hence, this section

considers the target audience, with Table 5.16 quantifying the most common responses related to the target audience of the corporations, specifically on Twitter.

Table 5.16: Corporations' target audience on Twitter

No.	Theme	Total	Percentage
1	Consumers and the general public	6	86%
2	Businesses	5	71%
3	Media	3	43%
4	Government	2	29%

As highlighted in different parts of the interviews, the target audience of the corporations affects the choice of language employed by the marketing team. For example, interviewee CORP.5 stated, "*some posts that target the international community are written in English and they constitute 15% to 20% of the account's content, and the rest is in Arabic*".

5.3.7 Summary of the corporations' interviews

The analysis of the interviews revealed the following findings:

- The majority of the marketing executives agreed that MSA is the most utilised variety on Twitter.
- The white/plain dialect was discussed by the majority of the interviewees, and was described as a variety between MSA and CA. The rationale for using this variety was to be congenial, as per CA, while excluding local expressions, as per MSA.
- A negative attitude towards Arabizi (i.e., Arabic with Latin script) was expressed in different parts of the interviews, with this variety being regarded as unprofessional. This is in line with the results of the tweets' analysis in section 5.2.3, since the variety only appeared in one tweet.
- The discussion on the functions, frequency of posts, and target audience revealed that the corporations plan and edit the posts to echo the intended communication strategy of the brand, in terms of the language or variety employed in the post.

5.4 Conclusion

This chapter outlined the findings from the two analysis phases exploring how corporations employ different languages and varieties on Twitter. The first phase involved the quantitative analysis of the tweets, while the second phase involved the qualitative analysis of the interviews with seven marketing executives from the selected corporations.

The findings responded to the first, second, third, and fourth research questions regarding the corporations' group through analysing the different languages and varieties used on Twitter. Furthermore, the chapter presented a discussion on the functions and aims of the tweets, which is an important aspect that affects the choice of language and variety employed. The aim of the interviews was to elicit the motivations and personal opinions of the marketing executives regarding the languages and varieties utilised on Twitter, aside from the languages and varieties used on the corporate social media accounts in general.

The next chapter will present the findings for the same analysis applied to the influencers' data, in order to understand how they utilise the languages and varieties for their self-branding on Twitter.

Chapter 6:
Influencers' Data Analysis and Results

6.1 Introduction

This chapter reports the results of the analysis of the influencers' tweets and interviews. The categorisation of the languages and varieties discussed in Chapter 4 forms the basis for the analysis of the findings presented in this chapter.

Section 6.2 presents the quantitative results that emerged from analysing the tweets posted by the 30 social media influencers' Twitter accounts selected for the research. As with the analysis of the corporations carried out in the previous chapter, the results are presented based on inter-tweet (i.e., using different languages and varieties in different tweets within a single timeline) and intra-tweet (i.e., using different languages and varieties within a single tweet) code-mixing. After that, section 6.3 reports the findings of the interviews conducted with five social media influencers.

6.2 Analysis of the influencers' tweets

As mentioned in the chapter 4, the total number of tweets (i.e., updates and retweets) used for the analysis of the social media influencers is 2,593. The analysis in this section commences with a general appraisal of the tweets in section 6.2.1 to measure their monolingual, bilingual, and multilingual distribution. The distribution of the languages and varieties in the monolingual and multilingual tweets is then explored in sections 6.2.2 and 6.2.3, respectively, following the Matrix Language Frame model of code-switching. Then, section 6.2.4 discusses the various general patterns of languages that appeared in the influencers' data. Following that, section 6.2.5 considers how other paralinguistic features (i.e., punctuation, letters, and emojis) are utilised either with or without text in the influencers' tweets to support the communication process or as the sole communication medium. After that, section 6.2.6 investigates the various functions extracted from the tweets, while section 6.2.7 relates the functions to the languages and varieties that appeared in the tweets. Finally, section 6.2.8 summarises the findings that emerged from the analysis of the influencers' tweets.

6.2.1 Tweets' overview

The data were classified according to the number of languages or varieties used in the tweets. Figure 6.1 shows the distribution of the influencers' tweets (N=2,593) based on one, two, or three codes.

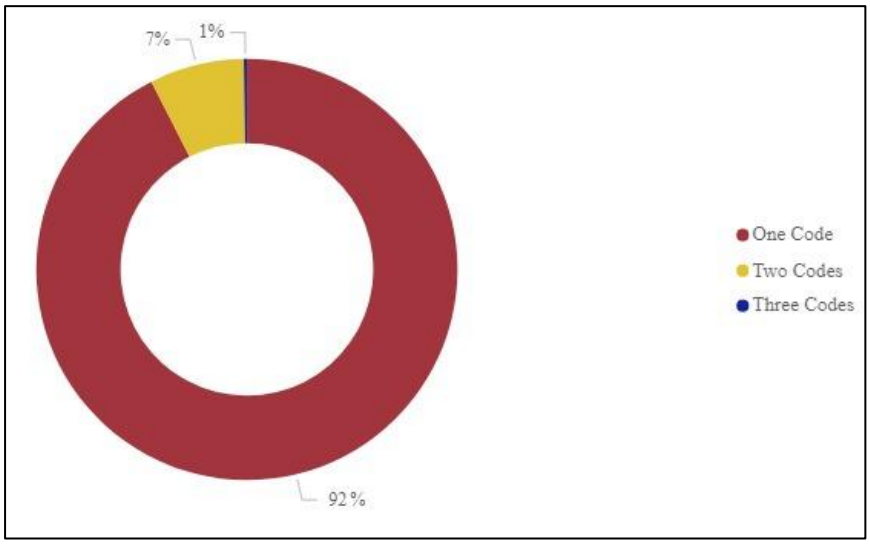


Figure 6.1: Distribution of the number of languages/varieties in the influencers' tweets

It can be seen from Figure 6.1 that the majority of the tweets (92%) are in one language or variety, which is similar to the results for the corporations' tweets (i.e., 86% featured one code). However, while it was found that the corporations' tweets used a strategy of parallel tweets in Arabic and English 14% of the time, this strategy only represents 8% of the influencers' tweets. Furthermore, in the parallel text tweets in the influencers' data there were no mixed tweets in any of the paired tweets, which was not the case in the corporations' tweets as 2% of the parallel texts used more than one language or variety.

The overall view of the influencers' data reveals that the majority of the tweets are in one language or variety, which means that multilingualism mainly manifests inter-tweet. The following section explores the different languages and varieties that appeared in the timelines of the selected influencers.

6.2.2 The languages and varieties in the influencers' monolingual tweets

With the majority of the influencers' tweets found to be monolingual, the various languages and varieties employed by the social media influencers in their different posts create a multilingual timeline and a discourse space. This section discusses the monolingual tweets,

while the following section considers the multilingual tweets. Figure 6.2 presents the languages employed in the monolingual tweets (N=2,397) gathered from the social media influencers.

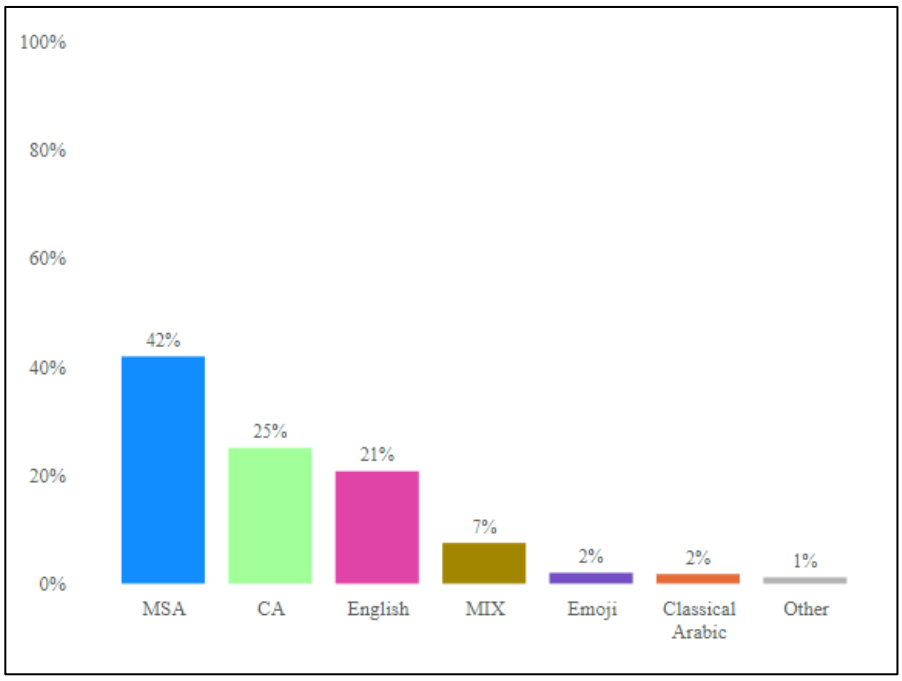


Figure 6.2: Distribution of languages in the monolingual influencer tweets

The majority of the influencers' monolingual tweets are written in MSA (42%), followed by CA (25%) and then English (21%), which echoes the trend of the corporations' monolingual tweets at 59% in MSA, 16% in CA, and 11% in English. Here, again MSA is the main language used in tweets. The results indicate, however, that the influencers use more CA (25% vs 16%) and English (21% vs 11%) than the corporations. The reason for more English being used in the influencers' tweets in comparison with the corporations could be because influencers do not use paired tweets in English and Arabic, as per the corporations, but they still need to connect with their international followers. Figure 6.2 also reveals that the influencers use intra-tweet bilingualism/multilingualism, with 7% of the tweets featuring a mixed language or variety, which is also seen in Figure 6.1 above. Furthermore, 2% of the tweets use emojis only without any text, such as the example presented in Figure 6.3.

Figure 6.3: Example of an influencer's tweet using only emojis

The other languages and varieties utilised in the influencers' tweets include English with Arabic script and French. The following section considers the multilingual tweets that appeared in the influencers' data.

6.2.3 The languages and varieties in the influencers' multilingual tweets

Figure 6.4 demonstrates the matrix codes used in the influencers' bilingual and multilingual tweets (N=194), while Figure 6.5 presents the other languages and varieties utilised in the same tweets. It must be noted that the results first show the matrix language, and then discuss the other languages and varieties that are utilised.

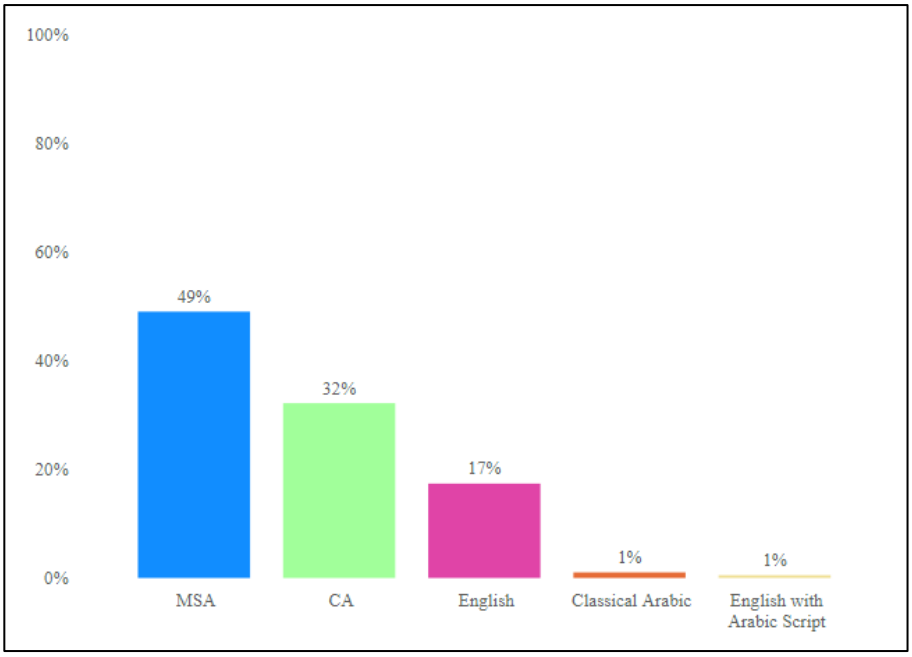


Figure 6.4: Distribution of matrix codes in the multilingual influencer tweets

There are three principal varieties and languages employed by the influencers as the matrix language in the multilingual tweets: MSA, CA, and English. In similarity to the corporations' multilingual tweets, MSA (49%) dominates the grammatical frame in the influencers' tweets, followed by CA (32%). Interestingly, the influencers are more likely to use English as the matrix language than the corporations, where it only represented 2% of the multilingual tweets.

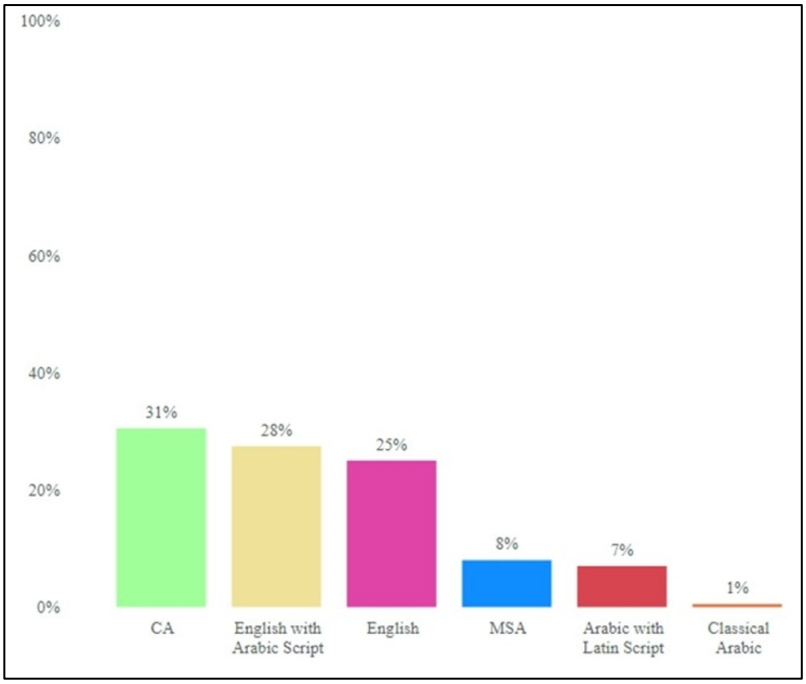


Figure 6.5: Distribution of embedded codes in the multilingual influencer tweets

Figure 6.5 presents the embedded codes in the influencers' multilingual tweets, where it is confirmed that CA is the most utilised embedded code (31%), followed by English with Arabic script (28%) and then English (25%). Surprisingly, the results indicate a difference from the corporations' results, which revealed the dominance of English (65%) and English with Arabic script (24%) in this category. There are also instances where the influencers use MSA, Arabic with Latin script, and Classical Arabic.

Table 6.1 explains how the different languages and varieties that appeared in the social media influencers' data relate, while presenting the embedded codes that appeared with each matrix code.

Table 6.1: Distribution of embedded codes to matrix codes in the influencers' data

Matrix code	Embedded code	Total number	Percentage for each matrix language	Percentage of overall tweets
MSA	CA	47	49%	24%
	English	31	32%	16%
	English with Arabic script	17	18%	9%
	Arabic with Latin Script	1	1%	1%
	Total	***96***		

CA	English with Arabic script	39	64%	20%
	English	17	27%	8%
	MSA	4	7%	2%
	Classical Arabic	1	2%	1%
	Total	***61***		
English	Arabic with Latin script	13	38%	6%
	CA	11	32%	5%
	MSA	10	30%	5%
	Total	***34***		
Classical Arabic	English with Arabic script	1	50%	1%
	CA	1	50%	1%
	Total	***2***		
English with Arabic Script	CA	1	100%	1%
	Total	***1***		
	Total	***194***		***100%***

Table 6.1 reveals that the MSA and CA varieties are the most employed combinations in the influencers' tweets (24%). Furthermore, it can be seen that CA is used as an embedded code with all other matrix codes (i.e., MSA, English, Classical Arabic, and English with Arabic script). This highlights the strong presence of CA in the tweets, and could imply that the influencers use CA phrases in tweets to convey congeniality or casual expressions.

The next most common combination in the influencers' tweets is CA with English with Arabic script. This again indicates the popularity of CA among the influencers, and how English with Arabic script (20%) in addition to English (8%) is utilised to support the communication in the tweet. Furthermore, Table 6.1 shows that the combination of MSA with English is often employed by the influencers (16%), although with less frequency than the same combination in the corporations' tweets (54%). This also indicates that when the influencers use MSA, they are less likely to use English in the same tweet.

Table 6.2 presents the distributions of the different languages and varieties that appeared in the influencers' tweets (N=2,593).

Table 6.2: Proportion of languages and varieties in the influencers' tweets

No.	Language or variety	Appearance in monolingual tweets	Appearance as a matrix code	Appearance as an embedded code	Instances of variety in tweets	Percentage of all influencers' tweets
1	MSA	1,096	96	14	1,206	46%
2	CA	655	61	60	778	30%
3	English	542	34	48	624	24%
4	French	2	-	-	2	<1%
5	Classical Arabic	46	2	1	49	2%
6	Arabic with Latin Script	-	-	14	14	<1%
7	English with Arabic Script	3	1	57	61	3%
8	Emoji	52	-	-	52	2%
9	Number	1	-	-	1	<1%
	Total	**2,397**	**194**		**2,593**	

Table 6.2 shows the proportion of the different languages and varieties in all the influencers' tweets, whether monolingual, bilingual, or multilingual. With 8% of the influencers' tweets being either bilingual or multilingual, as discussed in section 6.1.1, the total number of instances displayed in the table (2,593) includes the monolingual tweets and the matrix language. However, the embedded languages and varieties in the influencers' tweets are not counted in the total to avoid affecting the total number of tweets. Table 6.2 reveals that the majority of the influencers' tweets employ MSA (46%), followed by CA (30%) and then English (24%). In comparison, the corporations' tweets employ a higher degree of MSA (61%), with less usage of CA (18%) and English (19%). The other languages and varieties in Table 6.2 include English with Arabic script, which was employed in 3% of the influencers' data and is the exact same percentage of this variety used in the corporations' tweets (see Table 5.2). It was also noticed that there are more instances of Arabic with Latin script (N=14) in the influencers' tweets than in the corporations' (N=1). There was also a number used in a tweet, along with emojis, as shown in Figure 6.6.

Figure 6.6: A tweet featuring a number and emojis

The tweet shown in Figure 6.6 is interesting as it shows how numbers and emojis can be used to create a sentence. The author is engaging with his followers by mentioning a football player who wears the number 13 shirt and who moved to Al Ahli football club (represented by the green heart to reflect the new team's main brand colour) from Al Wehda football club (represented by the red broken heart to reflect the old team's main brand colour), with the hand representing a wave to say goodbye to the old team and hello to the new one.

After considering the different languages and varieties that appeared in the influencers' tweets, the following is a summary of the key findings that appeared from the analysis:

- The influencers' tweets show only 8% of inter-tweet code-mixing, but employ a range of languages and varieties across tweets.
- Similar to the corporations' tweets, MSA dominates the influencers' tweets (46%), with English appearing in 24% of their tweets versus 20% of the corporations'.
- CA appeared more frequently in the influencers' tweets (30%) in comparison to the corporations' (18%).
- Arabic in Latin script appeared in only 1% of the influencers' tweets, while English in Arabic script appeared in 3% of the tweets.

It is essential to keep in mind that the corporations' accounts reflect the branding goals of their marketing strategies, while the influencers have more flexibility in their accounts. Therefore, there are differences in the languages and varieties that are used by the corporations and the social media influencers. The next step in the analysis is to investigate the different languages and varieties utilised in the influencers' tweets by using patterns to investigate how the languages and varieties are employed in the influencers' official accounts.

6.2.4 The patterns of languages and varieties in the influencers' tweets

Having considering the different languages and varieties used by the influencers on Twitter, this section explores the various patterns that emerged from the influencers' data regarding how the different languages and varieties are utilised inter-tweet on the timeline. First, the common patterns that appeared in the data are listed. After that, the section investigates the different samples that utilise the languages and varieties in the patterns identified in the influencers' tweets.

It is first important to explore the overall usage of the languages and varieties by the social media influencers. Table 6.3 displays the frequency of the languages and varieties that appeared in the tweets, while also showing the number of accounts and tweets that used each language or variety.

Table 6.3: Frequency of languages and varieties in the influencers' data

Main language or variety	Total number of influencers' accounts	Percentage of accounts	Total number of tweets	Percentage of tweets
MSA	30	100%	1,096	42%
CA	29	97%	655	25%
MIX	27	90%	196	6%
English	24	80%	542	22%
Emoji	17	57%	52	2%
Classical Arabic	13	43%	46	2%
English with Arabic Script	2	7%	3	<1%
French	2	7%	2	<1%
Number	1	3%	1	<1%
Total			**2,593**	**100%**

It can be seen from Table 6.3 that all the influencers use MSA on their timeline on Twitter, which shows the importance of this Arabic variety to all the social media influencers. CA follows, with 97% of the social media influencers using it to engage with the virtual audience on Twitter. Then, a mix of languages and varieties in the same tweet is used by 90% of the social media influencers' accounts, which means that 10% of the accounts (N=3) did not have any intra-tweet code-mixing. The table also shows that although English appears in 22% of the tweets, it is used by 80% of the social media influencers under study. In terms of using emojis only as a tool for communication, the data shows that 57% of the social media influencers use emojis only on their timeline. Table 6.3 also reveals that two of the

social media influencers use English with Arabic script, and another two influencers use French.

After considering the overall frequency of the languages and varieties used by the influencers, the following is a discussion on how these languages and varieties appeared together on the timelines of the analysed social media influencers' accounts. Similar to the analysis of the other groups, the languages and varieties that appeared in at least three tweets in the timeline of the account were included in the pattern. Table 6.4 presents the main patterns of the languages employed, together with the proposed term for each pattern. Furthermore, the table shows the number and percentage of the influencers' accounts (N=30) that tweet in this specific pattern.

Table 6.4: Patterns of multilingualism in the influencers' data

Pattern	Number of influencers' accounts	Percentage of influencers' accounts
1 Bilingualism	**3**	**10%**
MSA/English	3	10%
2 Bilingualism and Diglossic Switching	**18**	**60%**
CA/MSA/English/MIX	11	36%
CA/MSA/English/MIX/Emoji	5	17%
CA/MSA/English/Classical Arabic/MIX/Emoji	2	7%
3 Diglossic Switching	**6**	**20%**
CA/MSA/MIX	6	20%
4 Other	**3**	**10%**
Total	**30**	**100%**

Section 5.2.4 found that MSA was the common variety displayed in all of the patterns of the corporations' tweets. This trend is also seen in the influencers' data, with all of the patterns including MSA among the codes used. It was also discussed that there are corporation accounts that only tweet in MSA or in English. However, this is not the case with the influencers, where it can be seen from Table 6.3 that multilingualism was demonstrated in all the accounts (N=30). However, it must be highlighted that three of the influencers utilise only MSA and English (Pattern 1 in the table), which is different to the parallel text

bilingualism or paired tweets in the corporations' data, as the tweets in the influencers' accounts do not have the same content. Furthermore, the table shows that the most common pattern is bilingualism and diglossic switching, utilised by 60% of the influencers in this study. This combination of languages and varieties could indicate the use of formal versus friendly posts, as will be explored later in the functions' sections 6.2.6 and 6.2.7 below.

Figure 6.7 presents a sample of the different languages and varieties employed from the influencers' accounts for each of the patterns listen in Table 6.4.

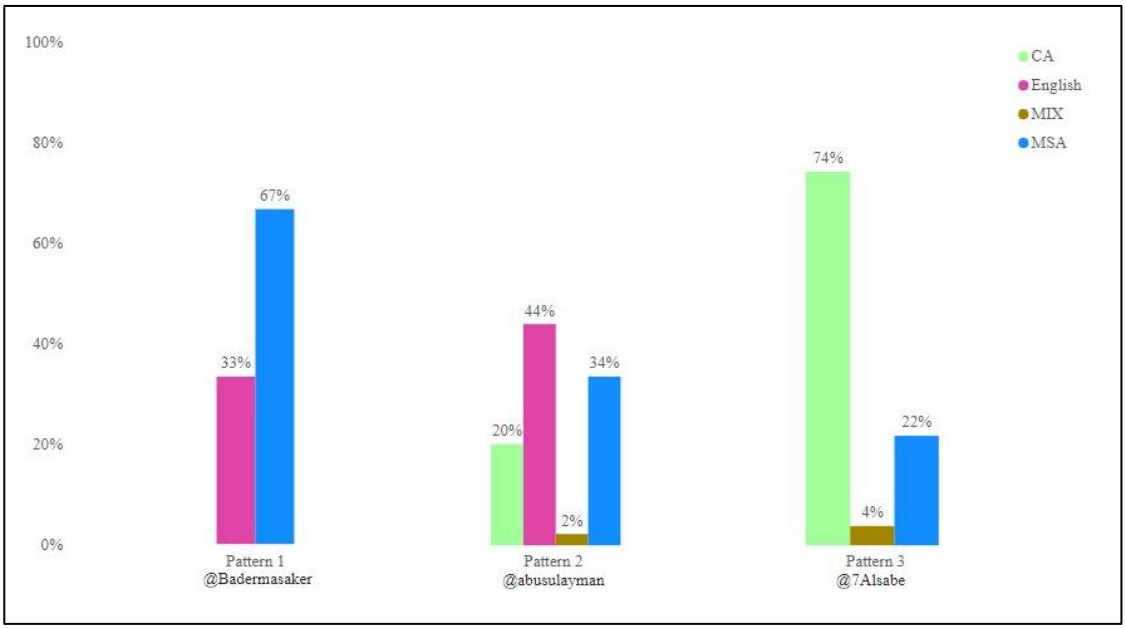


Figure 6.7: Sample of the different languages and varieties used by the influencers

Figure 6.7 shows an example for each pattern listed in Table 6.4. The different patterns display variation in the use of different languages and varieties by the influencers on Twitter. The first example demonstrates one of the bilingual influencers' accounts, where the Twitter account of Bader Al Asaker (@Badermasaker) resembles some of the corporations' communication strategies by using only MSA and English in his communication, with more tweets in MSA. Pattern 2, which is exemplified by the account of Mona AbuSulayman (@abusulayman) shows that she uses English in most of her tweets, which could be due to her diverse international audience. Interestingly, in February 2020 (i.e., after the period when the data for the current research were collected) Mona AbuSulayman started a new Twitter account in English only. However, because this account was only recently established, it does not have a comparable number of followers to her original account, and

it has not yet been authenticated by Twitter. Furthermore, she still uses her original account to post English tweets. It must also be noted that Mona AbuSulayman is probably among the first influencers to follow the corporations' approach of having separate Twitter accounts (one for English and another for Arabic). The third pattern, exemplified by Abdullah Al Sabe (@7alsabe), shows the use of different Arabic varieties (i.e., CA and MSA), in addition to mixed tweets using English or English with Arabic script such as *save login info* and ستوري (story). Therefore, even though the account seems to alternate between MSA and CA, within the tweets there is some English and English in Arabic script. Using a wide range of languages and varieties to communicate with the audience on Twitter could be linked to the functions of the tweets, as discussed later in this section.

As most of the accounts appear to have primarily bilingual and diglossic content, that is, 60% of the influencers' account in this study, it is worth examining a sample from the accounts that use this pattern. This should illustrate how the languages and varieties have different proportions in these accounts, even if they follow the same pattern in general.

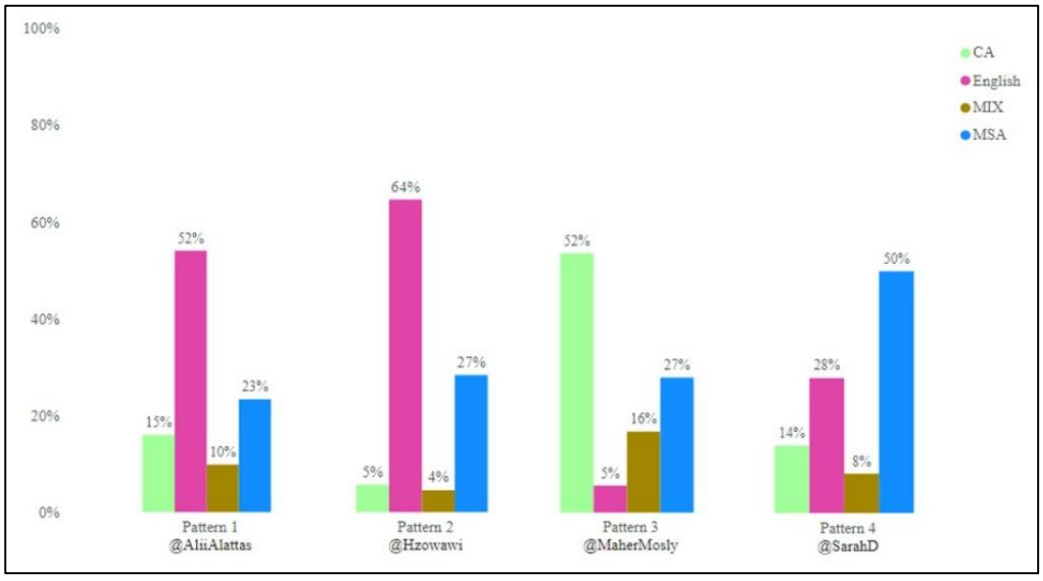


Figure 6.8: Proportion of languages and varieties used in the bilingual and diglossic switching pattern by a sample of influencers

Figure 6.8 reveals additional variation in the use of different languages and varieties by the influencers who post bilingual and diglossic content on Twitter. For example, English is the most employed language in Pattern 1 from Ali Al Attas (@AliiAlattas) at 52%, and in Pattern 2 from Hosam Zawawi (@HZowawi) at 64%. Conversely, English is the least utilised language in Pattern 3 from Maher Mosly (@MaherMosly) at 5%. Furthermore, MSA is the most used variety in Pattern 4 from Sara Dundarawy (@SarahD) at 50%, and the second

most frequently used variety in the other patterns. This again indicates the preference for MSA by the influencers on Twitter. In terms of CA, it is the most utilised variety in Pattern 3 from @MaherMosly at 52%, but it is the least used variety in the Pattern 2 from @Hzowawi at 5% of the tweets, and only 1% more than the use of mixed tweets on his timeline. The same discrepancy of using the languages and varieties also appeared in the corporations' data, as illustrated in Figure 5.9. It was noticed that although MSA has the greatest incidence in some of the corporations' accounts, it was also the least employed variety by other corporations' accounts.

After considering the different languages and varieties utilised on the timelines of the influencers, the next section explores the use of paralinguistic features in the influencers' tweets.

6.2.5 The use of paralinguistic cues in the influencers' tweets

Several paralinguistic tools such as emojis and punctuation marks are utilised by the influencers to convey social and emotional aspects to their audience on Twitter. Figure 6.9 illustrates the distribution of the different paralinguistic cues used by the influencers in the data.

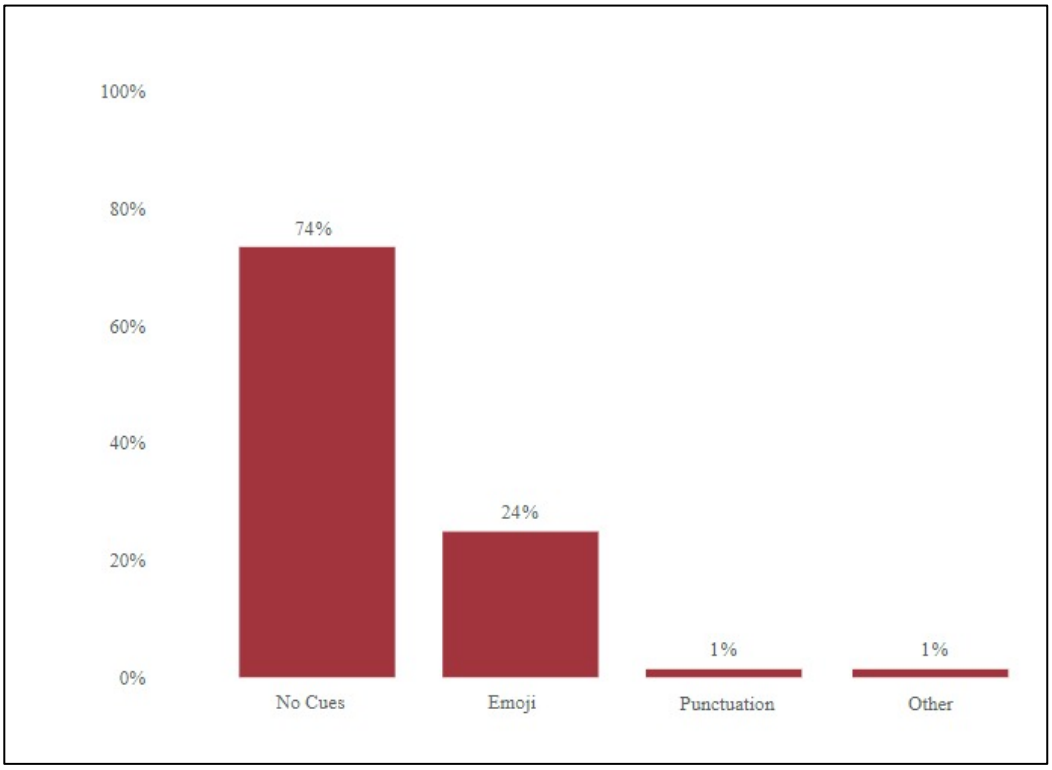


Figure 6.9: Distribution of paralinguistic cues in the influencers' tweets

Figure 6.9 indicates that 74% of the influencers' tweets do not utilise any paralinguistic cues besides the textual content. However, 2% of all the influencers' tweets in the data (i.e., 52 tweets out of the 2,593 sample) contain only emojis without any text. Additionally, 24% of the tweets include emojis as well as text, while 1% of the influencers' tweets include punctuation or emoticons such as :). These paralinguistic tools are employed to replace the facial expressions and body language used in face-to-face communication, and therefore they perform an important role in the interactions with virtual audiences on Twitter.

The Other category in Figure 6.9 concerns multiple punctuation marks (e.g., !!!!) and letters (e.g., Noooo), which can also be utilised in Arabic script. It is surprising that the percentage of emoji and emoticon use in the influencers' tweets (26%) is almost one-third less than that seen in the corporations' tweets (37%). This could be because some of the corporations' accounts use many emojis in an attempt to personify the brand when communicating with the Twitter audience. Furthermore, corporations may utilise the emoji as a tool to minimise the word count and make the tweet more appealing, such as using the clock emoji to indicate the time of an event.

The analysis of the data also considered the languages and varieties used with the paralinguistic cues, in order to compare the general results. Emojis were selected to be explored alongside the associated languages and varieties because they represent the most frequently employed paralinguistic cue in the tweets. Table 6.5 presents the proportion of tweets featuring emojis from each language or variety, to match the frequency that appeared in the influencers' data.

Table 6.5: Proportion of tweets with emojis in the influencers' data for each language/variety

No.	Language or variety	Total number of tweets	Number of tweets with emojis	Percentage
1	CA	655	232	35%
2	MSA	1,096	220	20%
3	English	542	74	14%
4	MIX	194	63	32%
5	Emoji only	52	-	-
6	English with Arabic script	3	1	33%
7	Number	1	1	100%

Table 6.5 presents the frequency of emoji use with each language and variety. It can be seen that 35% of the influencers' tweets in CA use emojis, compared to 20% of the influencers' tweets in MSA, and 14% in English. It may be that the social media influencers consider the English posts more formal and therefore they include fewer emojis. It is also apparent that although MIX tweets come fourth in terms of the emoji being included in the influencers' tweets, it has a relatively high percentage (i.e., 32%) when calculating the number of mixed tweets that use emoji. The table also illustrates that CA and MIX are very similar in terms of the distribution of tweets that utilise text with emojis as a paralinguistic cue at 35% and 32%, respectively. These results differ from the general trend of the influencers that suggests the dominance of MSA in monolingual tweets and the matrix code in mixed tweets.

In comparison with the corporations' results, it can be noticed that emojis in the influencers' tweets appear mostly with CA (35%), as opposed to with MSA in the corporations' tweets (34%). However, it must be noted that CA is used in 30% of the influencers' tweets versus 18% in the corporations' tweets, which could be the reason for this difference. This enhances the notion that influencers use CA to appear congenial to their virtual audience, and that they use the emoji device to support this.

Following the analysis of the different languages and varieties that appeared in the influencers' data, the following section considers the functions of the influencers' tweets and the link between these functions and the language.

6.2.6 The functions of the influencers' tweets

This section considers the functions that emerged from the influencers' tweets in order to link the functions with the languages and varieties employed by the influencers in section 6.2.7. The functions that appeared in the influencers' data are similar to the corporations' functions, with the addition of self-expression. The following are the main identified functions:

1) Formulaic Purposes
2) Culturally Specific Genres
3) Reported Speech
4) Dialogical Interrelation
5) Marketing
6) Self-Expression

Table 6.6 shows the number of tweets and distribution for each of the main functions.

Table 6.6: Number and distribution of tweets for every main function

No.	Function	Number of tweets	Percentage
1	Formulaic purposes	307	12%
2	Culturally specific genres	65	3%
3	Reported speech	192	7%
4	Dialogical interrelation	274	10%
5	Marketing	899	35%
6	Self-expression	852	33%
	Total	**2,589**	**100%**

Table 6.6 illustrates that the majority of the influencers' tweets (35%) are utilised for the marketing of products or services, either by announcements or sharing an advertisement. Figure 6.10 presents an example of a job advertisement posted by one of the social media influencers.

Figure 6.10: Example of a marketing tweet by one of the influencers

The next most used function in the influencers' tweets is self-expression (33%), where the influencers usually share their thoughts about current affairs, such as: "*Usually am optimistic about the future but the moral questions being raised about robot capabilities are a bit frightening .. Will it be normal by 2100 ? And I am not going to be alive to find out*" (@abusulayman 05/01/2018). Or the tweet could be about personal events that occur such as: "*Dark sea salted chocolate* 🤩 *that sinful pleasure* 🔥" (@SafanaSejini 03/01/2018). Furthermore, the self-expression might concern a general observation of life such as "*Take a selfie, fake a life*" (@AliiAlattas 10/02/2017). The next most frequent function is formulaic purposes (12%), which becomes very popular during festive periods such as Ramadan, Eid, and National Day. Then, dialogical interrelation constitutes 10% of the influencers' tweets to interact with the virtual audience on Twitter, followed by reported speech (7%) and culturally specific genres (3%).

Having considered the different functions, the next section explores the languages and varieties employed with each function reported in this section.

6.2.7 The languages and varieties employed with functions in the influencers' tweets

This section relates to the third research question, which concerns using the different languages and varieties in relation to the functions. Figure 6.11 illustrates the functions along with the languages and varieties that were utilised with each function in the influencers' tweets. Furthermore, Table 6.7 details the number of languages and varieties used in each main function.

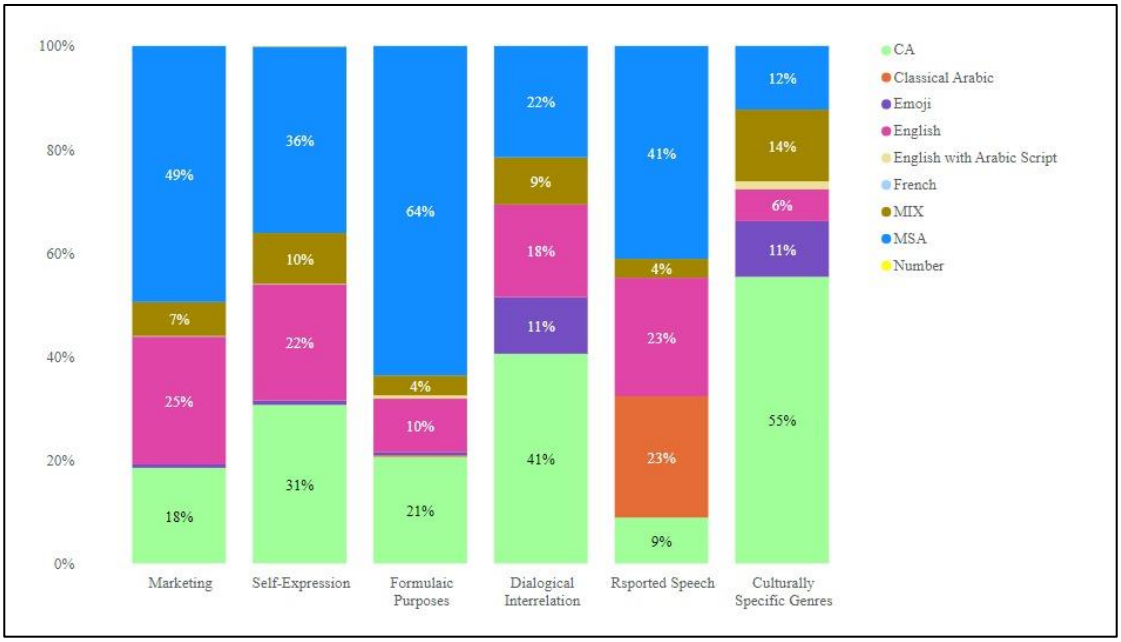


Figure 6.11: Comparison of the functions and languages/varieties in the influencers' tweets

Table 6.7: Number of functions with the languages and varieties in the influencers' tweets

Main function	Main language	Total
Marketing	CA	166
	Emoji	6
	English	222
	French	1
	MIX	59
	MSA	445
Total		899 (35%)
Self-expression	CA	261
	Emoji	7
	English	191
	French	1
	MIX	84
	MSA	307
	Number	1
Total		852 (33%)
Formulaic purposes	CA	64

	Classical Arabic	1
	Emoji	2
	English	32
	English with Arabic Script	2
	MIX	12
	MSA	198
Total		311 (12%)
Dialogical interrelation	CA	111
	Emoji	30
	English	49
	MIX	25
	MSA	59
Total		274 (10%)
Reported speech	CA	17
	Classical Arabic	45
	English	44
	MIX	7
	MSA	79
Total		192 (7%)
Culturally specific genres	CA	36
	Emoji	7
	English	4
	English with Arabic Script	1
	MIX	9
	MSA	8
Total		65 (3%)
Total number of tweets		**2,593**

It can be seen from Figure 6.11 and Table 6.7 that some languages and varieties are more frequently employed with certain communicative functions of the tweets. MSA, which is the most utilised variety by the influencers (see Figure 6.2), is primarily used for formulaic purposes (64%), marketing (49%) and reported speech (41%). On the other hand, CA is mostly frequently employed for culturally specific genres (55%) and dialogical interrelation (41%). Similar to the case in the corporations' results, CA is likely utilised for these functions to imitate daily conversations and to be congenial with the virtual audience. In the influencers' tweets, English is generally used in marketing (25%), reported speech (23%), self-expression (22%), and dialogical interrelation (18%). The use of English by the influencers differs to that of the corporations, for whom English is sometimes an important

language to formally communicate the brands. With the influencers, English is employed significantly in dialogical interrelation and self-expression, indicating that English is not merely utilised to present information to an international audience, but also to interact with the influencers' followers regardless of whether they are Saudi, Arab, or from other parts of the world. As for mixed tweets, they are generally less frequently used by the influencers, and are employed primarily in culturally specific genres. Finally, the data show that Classical Arabic is only significantly utilised in reported speech.

6.2.8 Summary of the analysis of the influencers' tweets

The following are certain highlights from the analysis of the influencers' tweets:

- Similar to the analysis of the corporations' tweets, multilingualism mostly presents itself in the influencers' accounts across the tweets on the timeline, as opposed to within a single tweet.
- The majority of the influencers' tweets are in MSA (46%), followed by CA (30%) and English (24%). While the proportion of MSA is higher in the corporations' tweets (61%), the proportion of CA (18%) and English (19%) in the corporations' tweets is lower.
- There are different patterns for using the languages and varieties, but most of the influencers' accounts tend to employ a mix of MSA, English, and CA in their Twitter accounts.
- The emoji in the influencers' accounts is frequently employed with CA (35%) and in MIX tweets (32%).
- The function of the tweet appears to affect the language chosen by the influencer, with MSA primarily employed in tweets for formulaic purposes (64%) and marketing (49%), while CA is mainly utilised in tweets for the culturally specific genres (55%).

Besides the analysis of the influencers' tweets that included a detailed description of the languages and varieties employed on their timelines, along with the functions, this study also considers the input of the selected influencers that were interviewed. Section 6.3 thus presents the results of the interviews with the social media influencers.

6.3 Analysis of the influencers' interviews

This section presents the results of the interviews conducted with five of the social media influencers: two males and three females. As per the analysis of the corporations' interviews, the introductory part of the interviews covered themes such as the influencers'

goals on social media, the frequency of posts, and the target audience. Then, the discussion focused on the main research aspect, namely, the languages and varieties employed on Twitter, in addition to the paralinguistic cues.

In each section, the interview results are presented in tables and extracts. Sections 6.3.1–6.3.4 feature the themes that emerged in relation to the languages and varieties, in addition to the paralinguistic cues. After that, section 6.3.5 considers the aims of the social media influencers and relates these to the functions that emerged from the tweets. Then, sections 6.3.6 and 6.3.7 discuss the frequency of posts and the influencers' target audience on Twitter, respectively. Finally, section 6.3.8 presents a summary of the findings from the influencers' interviews.

It should be noted that the influencers gave several answers to some of the questions. Therefore, the proportions shown in the tables are based on the number of respondents who provided each answer.

6.3.1 Languages and varieties employed for communication on Twitter

This section discusses the responses to the first and fourth research questions related to the languages and varieties utilised by the social media influencers on Twitter. Table 6.8 presents the languages and varieties employed by the influencers, as discussed in the interviews.

Table 6.8: Influencers' languages and varieties used on Twitter

No.	Theme	Number	Percentage
1	MSA only	3	60%
2	MSA/CA (white dialect)	2	40%
3	CA only	1	20%
4	English	1	20%

It can be seen that the majority of the interviewed influencers (60%) use MSA in their tweets, followed by the white dialect (40%), which as per the marketing executives of the corporations was described and recognised by the influencers. This is in line with the findings of the analysis of the tweets, whereby the majority (46%) are in MSA. The following sections detail the views of the influencers on the different languages and varieties and how they use them solely or together on Twitter.

6.3.2 The use of Arabic and English in tweets

Some of the influencers discussed Arabic and English in general, without differentiating the varieties of Arabic, which indicates that some of the influencers do not focus on selecting the variety and only consider Arabic and English. The following extract from interviewee INF.4 explains how she uses English and Arabic:

Extract 6.1

> *I use both English and Arabic, and I once did a poll to ask my followers about the language they would prefer for me to use and it was tight: 48% to 52%. I have many international followers. My Twitter posts are mostly English, but I use Arabic when I want to support a local business, event or campaign.*

Interviewee INF.5 also emphasised the importance of the followers' analysis to the decision of the optimum language to be employed:

Extract 6.2

> *It depends on the results you receive on your type of followers. If the results show that most of your followers are from within the country then I think it is an insult if you use English in most of your tweets.*

The influencers spoke further in the interviews regarding intra-tweet code-mixing, which constitutes 8% of their tweets, with interviewee INF.3 mentioning how he uses different languages and varieties within a tweet: "*I sometimes use Colloquial Arabic or English words or phrases within a very rigid Standard Arabic post to give it some flavour*".

The influencers also discussed inter-tweet code-switching on the timeline and the functions of each language or variety, as expressed by interviewee INF.2:

Extract 6.3

> *I usually tweet in Standard Arabic, but I could use Colloquial Arabic in informal posts such as comments on football games. Sometimes I tweet quotes in English.*

The extract above supports the results of the social media influencers' tweets, which showed that MSA is used in 46% of the tweets, followed by CA (30%) and English (24%). The extract also explains how the different varieties of Arabic and English can be used for different functions on the timeline of the influencers.

6.3.2.1 The advantages of using certain languages/varieties in tweets

Interviewee INF.3 shared some advice on how to be successful on Twitter and mentioned the use of Arabic as one of the key factors for success: "*The recipe for success on Twitter is to continue more Arabic, more lives, videos and show more face*".

Some of the influencers were proud to be using MSA, as expressed by interviewee INF.3:

Extract 6.4

> *I am proud of my Standard Arabic and I prefer to use it. Also, when I aim for heavy criticism, I've noticed that saying it in Standard Arabic makes it less harmful than colloquial dialect and gives the same effect. It gives me a character and people accept it better.*

Interviewee INF.1 also mentioned feeling pride when using MSA accurately and fluently:

Extract 6.5

> *I think using Colloquial Arabic reflects that the user doesn't know how to write in Standard Arabic. We used to show off that we know English. but I feel now using Standard Arabic correctly makes me proud.*

The same influencer (INF.1) added that using MSA helps to position the tweeter among a certain group of people or professionals: "*When I use MSA, it shows seniors and officials that I am part of their professional circle*".

The analysis of the tweets showed that MSA appears in 46% of the influencers' tweets and CA in 30% of the tweets. This was recognised and explained by some of the influencers in that while using CA is acceptable, they still favour MSA and appreciate its importance to the identity of the Arabic community, as the following observation by interviewee INF.5 suggests:

Extract 6.6

> *It is I think natural that language changes and Arabic is not an exception, but I don't think anything can replace Standard Arabic…even in countries where the colloquial dialects are more dominant such as Egypt. They still use Standard Arabic in the end. If everyone sticks to their colloquial dialects, we will not be one nation.*

Since CA is the daily informal spoken variety, it is preferred by the influencers who aim for greater engagement and interaction. The following extract by interviewee INF.5 explains one of the influencers' point of view, who primarily uses CA on Twitter:

Extract 6.7

> *The Quran was revealed in 7 Ahruf [varieties] of Arabic so I don't think it is harmful if we use different varieties of Arabic in writing today. It is about how we communicate and how to deliver the message. Arabic will never die as it is the language of the Quran and Allah has been protecting it for hundreds of years and He will still protect it...the spread of Colloquial Arabic is not a threat at all to Arabic. The danger is when you downgrade the language and you communicate with the people in a disrespectful way. It is important to practise and work on your Standard Arabic and increase the knowledge of words but in communication it is OK to use the Colloquial Arabic. I think Standard Arabic and Colloquial Arabic are parallel and each one has its own communication functions. The language must serve my communication goals, and the other way around.*

Interviewee INF.2 presented a similar perspective:

Extract 6.8

> *Sometimes I write what I think so I use Colloquial Arabic. I am not showing my excellence in Arabic...what matters to me is that people get the message and interact with it, so I don't focus a lot on the language.*

The white dialect was also discussed by some of the influencers, and how it is useful because it has some elements of the colloquial dialect but without using specific terms that may signal the origin of the tweeter. For instance, INF.5 stated that: " *When I use the white dialect, I don't use local colloquial terms which helped to expand my audience in other regions of the Kingdom*".

6.3.2.2 The disadvantages of using certain languages/varieties in tweets

The influencers also discussed receiving criticism regarding the language that they use, with interviewee INF.1 describing how using MSA may have a negative impact: "*People think if you are using Standard Arabic that you are showing off that you are educated, and this is very dangerous*". This may indicate that the choice of Arabic is not always easy for the influencers, and that they may need to 'simplify' the language used. For that reason, interviewee INF.1 preferred to mix different languages and avoid any negative perceptions because of the language:

Extract 6.10

> *What I noticed is that when I tweet in English, I receive criticism asking why I am using English and when I tweet in Standard Arabic, I also receive criticism that I am showing off. From my experience, it is better to have a language that is in-between.*

The analysis of the patterns used by the influencers, presented in section 6.2.4 above, show that there is variation in how the influencers utilise the languages and varieties on Twitter. Interviewee INF.4 mentioned a different perspective for using the languages on Twitter. The extract below shows that she sometimes consciously avoids using Arabic in an attempt to control the followers of her account.

Extract 6.11

> *I personally avoid tweeting in Arabic because it attracts people that may cause me problems…I hate to say this but knowing more languages means you are more educated and you have a different mindset.*

The extract above also reveals that social media influencers use the different languages as a tool to attract certain groups of followers, and therefore finding variation in the patterns is to be expected, as explained in section 6.2.4.

As per the negative attitudes of the marketing executives, the majority of the social media influencers expressed a similar view when asked about using Arabizi. Some were convinced that using Arabizi could push the audience away from the account, as implied by interviewee INF.5:

Extract 6.12

> *Arabizi…May Allah vanish it. Who writes in Arabizi disgraces two languages. It is not a communication tool…it is a very ignorant way. Arabizi was cool in 2005, but not now. There is a clear Arabic keyboard now and there is no excuse.*

Although most of the responses were not in favour of Arabizi, interviewee INF.4 expressed the need for it in certain situations whether on Twitter or other social media platforms:

Extract 6.13

> *I think using Arabizi or words in other languages depends on what you are using for your keyboard...if you are typing in English and there is a word that must be in Arabic in your sentence, it is easier to use Arabizi instead of switching the keyboard.*

The analysis of the tweets revealed only 14 instances of Arabic with Latin script, that is, Arabizi, and all the instances were as an embedded code. Furthermore, 13 of the instances were with English as the matrix code and only one case with MSA as the matrix code. This supports the view in the previous extract that users may prefer not to switch the keyboard, and hence they use Arabizi.

6.3.3 The interviewees' perceptions regarding the languages utilised on Twitter

The interviews also provided the social media influencers with an opportunity to voice their opinions regarding their general attitude towards the various language affordances available to the users on Twitter. This section presents the responses noted by the influencers regarding the optimum language or variety to be employed on Twitter, the language with the most audience interaction, and whether tweets in English should be translated into Arabic, and vice versa.

6.3.3.1 The optimum language or variety for use on Twitter

During the interviews, the social media influencers discussed their expectations regarding the ideal language or variety that should be utilised on Twitter, with Table 6.9 summarising their responses.

Table 6.9: Influencers' opinions regarding the optimum language or variety for use on Twitter

No.	Theme	Total	Percentage
1	MSA/CA (white dialect)	2	40%
2	Depends on the content	3	60%
	Total	**5**	**100%**

It can be seen from Table 6.9 that a majority of the influencers (60%) felt that there is no single language or variety that should be utilised on Twitter, with the choice of language depending on the content of the tweet. Nevertheless, two of the interviewed influencers (40%) believed the MSA–CA hybrid to be the ideal variety, which they referred to as the white dialect. It was also noticed that, unlike the corporations, the social media influencers focused in the interviews on the Arabic dialects rather than the use of English, which was strategic for the corporations.

6.3.3.2 The ideal language or variety for audience interaction

Similarly to the corporations, the analysis of the tweets did not include the interaction of the audience with the tweets of different languages. Therefore, in the interviews the influencers were asked about the languages and varieties that usually receive greater interaction on Twitter. Table 6.10 shows the languages and varieties with most audience interaction, as expressed by the influencers.

Table 6.10: Influencers' opinions regarding the ideal language/variety and audience interaction

No.	Theme	Total	Percentage
1	MSA/CA (white dialect)	3	60%
2	Depends on the content	2	40%

It can be seen from Table 6.10 that the social media influencers' audience primarily engage and interact with tweets that use the white dialect. Meanwhile, two of the influencers emphasised that the interaction depends on the content of the tweet, rather than just the language, thus suggesting that content and language function in unison to achieve successful interaction on Twitter.

Interviewee INF.5 cited one of the advantages of Twitter as being a more effective platform for measuring interaction and evaluating the performance of the account:

Extract 6.14

> *Snapchat does not give any value because it is a one-way relationship, but on other platforms including Twitter you receive comments and likes from your followers so you can always evaluate what you share with people.*

6.3.3.3 The need for English posts to be translated

The analysis of the influencers' tweets showed that only 1% of the influencers' tweets are translated from English into Arabic, or vice versa. The social media influencers were asked if they thought the tweets should be translated, with Table 6.11 summarising their responses.

Table 6.11: Influencers' opinions on the translation of tweets

No.	Theme	Total	Percentage
1	No	2	40%
2	Depends on the user, the audience, and the content	2	40%
3	Yes	1	20%
	Total	**5**	**100%**

Table 6.11 shows that two of the influencers did not think that the content should be translated to English. Furthermore, the table again supports the influencers' view that the

language usage depends on the content of the tweet. However, the influencers also cited the importance of the type of audience and user. The following extract from interviewee INF.2 shows how the influencer stressed the importance of the audience:

Extract 6.15

> *If I want to feel empathy with the followers, I think it is better to write in Arabic and English and translate the content…even if the content is different, it shows that that you care for the audience.*

Even for the influencers who asserted that tweets should not be translated, they still acknowledged the benefits of other users translating their tweets, as expressed by interviewee INF.4:

Extract 6.16

> *Although I don't think tweets must be translated, I respect the users who translate their tweets [English to Arabic and Arabic to English] as it gives the chance for everyone to understand the content.*

6.3.4 Using paralinguistic cues

The influencers were asked about using paralinguistic cues in the tweets. The analysis of the tweets revealed that the influencers use mostly emojis, as 24% of the tweets include an emoji in which 2% are only an emoji. Table 6.12 presents the frequency of using emojis with tweets, as stated by the social media influencers.

Table 6.12: Influencers' opinions on using emojis in tweets

No.	Theme	Total	Percentage
1	Sometimes	3	60%
2	Always	2	40%
	Total	**5**	**100%**

The analysis of the influencers' tweets revealed that only 24% incorporated an emoji. However, three of the influencers said that they sometimes use emojis in their tweets, while two claimed that they always use an emoji in their tweets.

In section 5.3.3 in the previous chapter, the marketing executives seemed to categorise emojis into formal and non-formal. The same categorisation appeared in the influencers' interviews, with interviewee INF.3 stating that he only uses formal emojis: "*I use formal emojis but not smiley faces, and I prefer any posts without emojis because not everything*

is a joke". Furthermore, interviewee INF.2 reported that the emojis are generally utilised for a specific function, namely, in replies: "*I prefer using emojis in replies…I think people look down at posts with emojis, but this makes it more acceptable*". On the other hand, interviewee INF.1 thought that emojis should be employed only in specific contexts, and not consistently in all content: "*Emojis can kill a tweet if not used properly. For example, if it is a condolences tweet and you use a crying emoji, you have lost the meaning*". Therefore, a tendency can be inferred from the influencers' responses that emojis should be utilised only in specific contexts.

6.3.5 Aims and goals in social media

Similarly to the analysis of the corporations, this section examines the findings related to the second research question regarding the functions of the different codes employed on Twitter. Although the focus of this thesis is the languages and varieties of the users on Twitter, this section provides a number of examples from the interviewees regarding the reasons they use Twitter. All the interviews with the social media influencers began with a discussion on how they started using social media and their general aims and goals, which are linked to the functions of the tweets discussed in sections 6.2.6 and 6.2.7 above. Table 6.13 presents the themes identified from the social media influencers' responses when asked about their aims for social media, along with the associated functions extracted from the Twitter data.

Table 6.13: Influencers' aims and goals on social media

No.	Theme	Total	Percentage
1	**Awareness** *(formulaic purposes, marketing, and self-expression)*	5	100%
2	**Criticism** *(culturally specific genres and self-expression)*	3	60%
3	**Engagement** *(formulaic purposes, culturally specific genres, and dialogical interrelation)*	2	40%
4	**Increasing the number of followers** *(formulaic purposes, culturally specific genres, reported speech, dialogical interrelation, and marketing)*	2	40%
5	**Quality of followers** *(formulaic purposes, reported speech, and self-expression)*	2	40%
6	**Blogging** *(self-expression)*	1	20%

The awareness functions for the social media influencers differ to those mentioned by the corporations. For the corporations, awareness is more focused on disseminating information about the products or services of the brands. On the other hand, awareness for the influencers, specifically on Twitter, means that they want to keep the community up to date with various news or information, whether directly related to them or to the society in general. The following extract from interviewee INF.2 mentions the different topics he likes to discuss on Twitter, and how he started his own business through spreading awareness about food on the platform:

Extract 6.17

> *I started on Twitter by sharing pictures and advice on food with family and friends, and then my audience expanded, and I started to give food information and advice on places and when I gained many followers. I started my own restaurant, although I am not a chef and I don't position myself as a chef. I became famous and many people know me. I started my own business because of social media, which is very nice.*

The same purpose was also expressed by interviewee INF.4, who began by recalling successful stories about employees and job opportunities to spread impact: "*I started tweeting when I worked at Taqat [a government employment programme] and I wanted to share stories of the cases I worked on and what was going on to spread the impact*".

Another theme that emerged from the interviews with the influencers is criticism, whereby the social media influencers mentioned that they use Twitter as a platform to criticise certain behaviours in society, or the services provided by particular companies or corporations. Interviewee INF.3 mentioned criticism and how the language affects the manner in which it is received by others: "*On social media I criticise and mock others, sometimes harshly, but the language I use makes it sound like a joke and acceptable*".

Increasing the number of followers was also an interesting topic for discussion with the social media influencers. Most of the functions that appeared in the analysis of the tweets in sections 6.2.6 and 6.2.7 above (i.e., formulaic purposes, culturally specific genres, reported speech, dialogical interrelation, marketing, and self-expression), could be used to attract more followers. Two of the influencers believed this to be an important factor, as expressed by interviewee INF.3:

Extract 6.18

> The number of followers is important for advertisements...the first impression for marketers to approach you for advertisements is the number of followers. It even helps with my job as potential clients visit my

> account and see how many followers and how many posts I have. The average person may get the impression that I am good and hire me...building my image is important in this way.

On the other hand, two other influencers were more concerned with the technique of adding the followers and the quality of the followers rather than the quantity, and they stressed that they do not agree with paying to add followers as this could simply add to the criticism and negative feedback that they receive.

The extract above shows how some of the influencers give attention to the quality of the followers rather than superficial followers who may be more disruptive. Finally, interviewee INF.4 highlighted that the topic may determine which platform is used and stated that the followers on Twitter differ from those on other platforms such as Instagram, and thus the content needs to match the follower typology: "*At some point I stopped linking my Instagram account from my Twitter account because I didn't want my Instagram community to be with my Twitter community*". This again indicates that the social media influencers consider the content and the audience of the platforms before posting the tweets.

6.3.6 Frequency of posts

While collecting the data, it was noticed that at some periods the users are very active and post frequently, while at other times they tweet less frequently. All the interviewees confirmed that they do not have any specific schedule for the tweets, and only post when they feel that they have something of value to share. It was mentioned by some of the influencers that they prefer to think of quality content rather than high quantity output, as exemplified by interviewee INF.5:

Extract 6.19

> *I don't post if I don't have anything to say…it is not mandatory. I am following someone who I really admire but he is spamming my timeline with so many posts. If I just care about the number, I think I will lose people. It is all about content.*

Furthermore, some influencers mentioned that because of their large number of followers, they feel a responsibility to tweet from time to time. While interviewee INF.1 recognised the value of having a plan, she still does not have one: "*Unfortunately, I don't have a plan and I could have been better if I had someone to help me plan the posts, but I don't as social media is a side thing for me and not my main job*".

As evidenced by interviewee INF.1 above, not all social media influencers have agencies or social media professionals to handle their Twitter accounts and plan their posts, with interviewee INF.5 addressing this issue: "*I handle my account myself and you can easily figure out if an influencer answers by himself/herself or if it is handled by an agency, as the style of replies and posts are different*".

The analysis of the interviews with the marketing executives revealed that there is a strong emphasis on the planning of the tweets in terms of what and when to post. This contrasts with the approach of the social media influencers, who although emphasising that they give attention to the language and content that they post, do not have any formal posting schedule.

6.3.7 Target audience on social media

As the target audience may affect the choice of language used by the social media influencers, they were asked about this during the interviews. Table 6.14 presents the target audience groups mentioned by the influencers. Again, the percentages reflect the number of responses from the overall number of interviewees (N= 5).

Table 6.14: Influencers' target audience on social media

No.	Theme	Total	Percentage
1	The general public	3	60%
2	Professionals	2	40%
3	Family and friends	1	20%
4	Colleagues	1	20%
5	Businesses	1	20%

It can be seen from Table 6.14 that most of the influencers' tweet for the general public. However, interviewee INF.5 acknowledged the different groups of audiences among her followers, and tends to address them at the beginning of the tweet: "*If I am targeting a specific group, I mention this at the beginning of the tweet such as, "Dear educators…*".

6.3.8 Summary of the influencers' interviews

The findings from the interviews with the social media influencers can be summarised as follows:

- The social media influencers agreed that MSA or the plain/white dialect, which is described as an MSA–CA hybrid, are the varieties they utilise most on Twitter, and

that while CA is sometimes acceptable on Twitter, there is more commitment to MSA.

- Discussion of the functions and frequency of posts revealed that the influencers do not tend to plan their posts, but they do consider the content and language of the tweet before posting.
- The discussion with the influencers revealed that they mostly tweet to spread awareness in the society about different news and information related to themselves or the public in general. Furthermore, the influencers expressed that they tweet to criticise certain common behaviours or incidents in the society.
- The negative attitude towards Arabizi was also expressed by the influencers, which echoes the analysis of the tweets that shows a usage of less than 1%.
- The discussion on the languages and varieties employed by the influencers (section 6.3.1) and the aims of the influencers on Twitter (section 6.3.5) provide insights on when and why different languages and varieties are utilised. For instance, English is used to appeal to the international followers, while the use of MSA positions the tweeter among a certain group of professionals.

6.4 Conclusion

This chapter reported the findings from the social media influencers' data, which was collected from Twitter and five semi-structured interviews. The mixed-method approach utilised for the analysis in this study combined both quantitative and qualitative data from the tweets and the interviews, and provides a multidimensional analysis of how the languages and varieties are employed on Twitter in Saudi Arabia. This proved helpful for the analysis of the corporations' data in the previous chapter, and similarly for the influencers' data in the current chapter. The next chapter presents the analysis and the findings of the ordinary users on Twitter.

Chapter 7:
Ordinary Users' Data Analysis and Results

7.1 Introduction

This chapter presents the analysis of the third group, the ordinary users, whereby the number of the tweets used for the analysis is 1,463. As per the analysis of the corporations and social media influencers, the categorisation of the languages and varieties presented in Chapter 4 was utilised for coding the tweets and for the discussion in the interviews.

Section 7.2 discusses the analysis results of 20 ordinary users' tweets, considering the use of the different languages and varieties, in addition to the functions of the tweets. As per the analysis in the previous chapters for the corporations and social media influencers, the findings are presented following the inter-tweet and intra-tweet code-mixing explained in the Chapter 4. After that, section 7.3 provides the analysis of the interviews conducted with five ordinary users on Twitter.

7.2 Analysis of the ordinary users' tweets

This section discusses the quantitative results of the ordinary users' tweets. First, in section 7.2.1, a general overview of the tweets is provided in order to measure the proportions of multilingualism that presented. Then, sections 7.2.2 and 7.2.3 detail the distribution of the different languages and varieties in the monolingual and multilingual tweets, respectively. Next, section 7.2.4 considers the general patterns of the languages and varieties that appeared in the ordinary users' tweets on Twitter. After that, the analysis of the paralinguistic cues employed by the ordinary users is presented in section 7.2.5, along with a discussion of the relation of these paralinguistic cues, mainly the emoji, with the languages and varieties used in the tweets. Then, sections 7.2.6 and 7.2.7 respectively consider the functions of the tweets and their relation to the languages and varieties that appeared in the ordinary users' tweets. Finally, a summary of the results of the ordinary users' tweets is presented in section 7.2.8.

7.2.1 Tweets' overview

The ordinary users' data were categorised according to whether the tweets are written in one language or variety, or more. Figure 7.1 presents the distribution of the ordinary users' tweets based on one, two, or three codes.

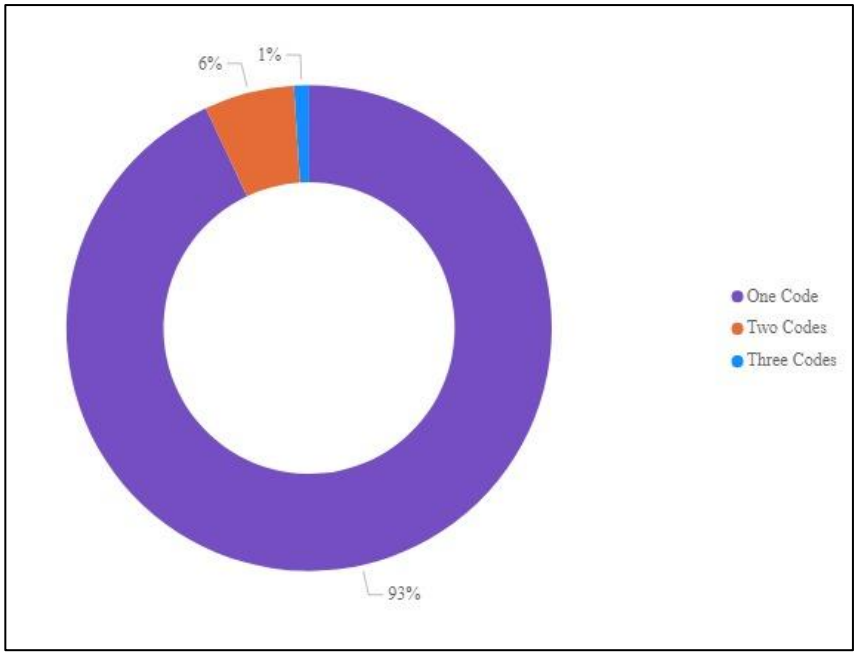


Figure 7.1: Distribution of the number of languages and varieties in the ordinary users' data

As per the previous groups (i.e., the corporations and influencers), the majority of the ordinary users' tweets (93%) are in one language or variety, while 6% utilise two and 1% employ three languages or varieties. The frequent use of parallel texts in separate tweets was highlighted in the corporations' analysis with 14% of those tweets using this means of communication. However, similar to the influencers' data, for the ordinary users parallel texts in separate tweets are not commonly utilised, with only one case found in the data. The parallel bilingual tweets in Figure 7.2 are the exact translation of the text provided in English and MSA without any intra-tweet mixing. It is not clear why the hashtag of knowledge is employed only with the Arabic tweet, but using the hashtag may be the reason why the tweet received two likes as it had more exposure.

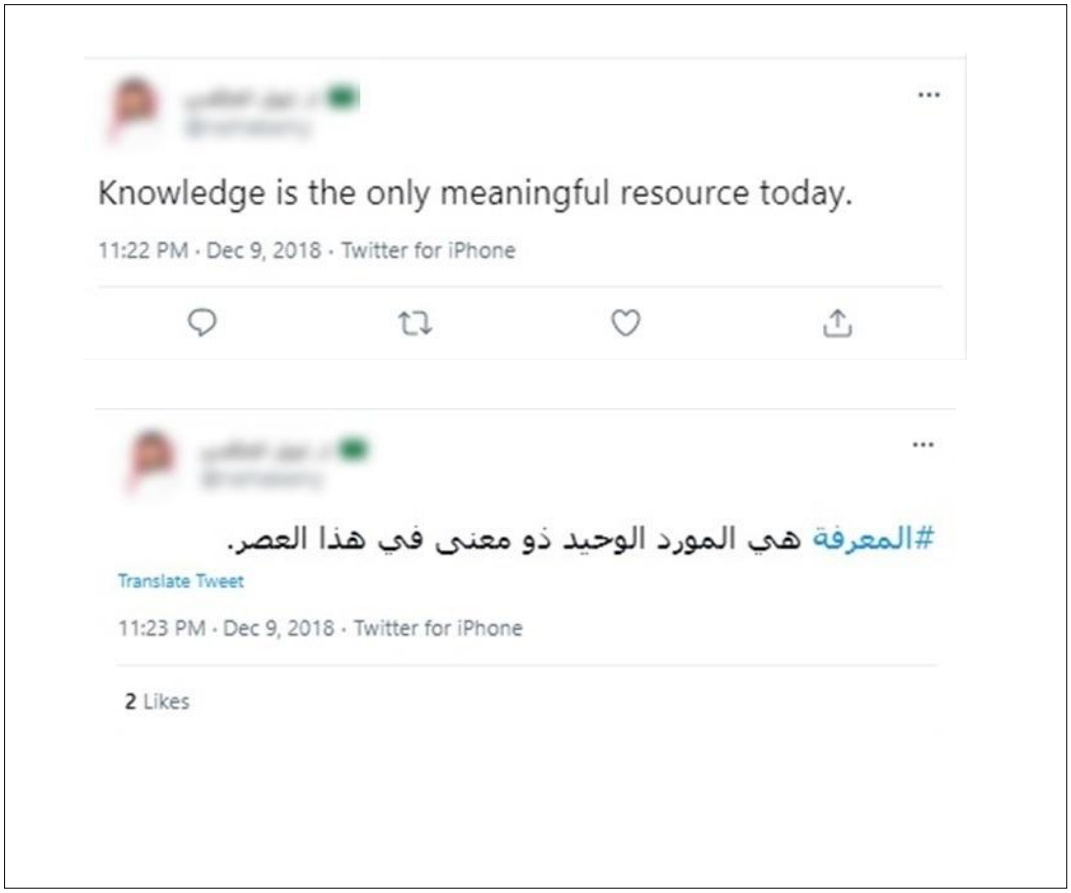


Figure 7.2: Parallel bilingual tweets by an ordinary user

The multilingualism in the case of the ordinary users' tweets thus occurs inter-tweet, with 93% of the tweets being conducted in one single code. The following section discusses the various languages and varieties identified in the ordinary users' data.

7.2.2 The languages and varieties in the ordinary users' monolingual tweets

Despite the high proportion of ordinary users' tweets in one language presented in the previous section, the languages and varieties used in the different tweets serve to create a multilingual discourse. Figure 7.3 shows a breakdown of the languages and varieties of the monolingual tweets (N=1,370) found in the ordinary users' data.

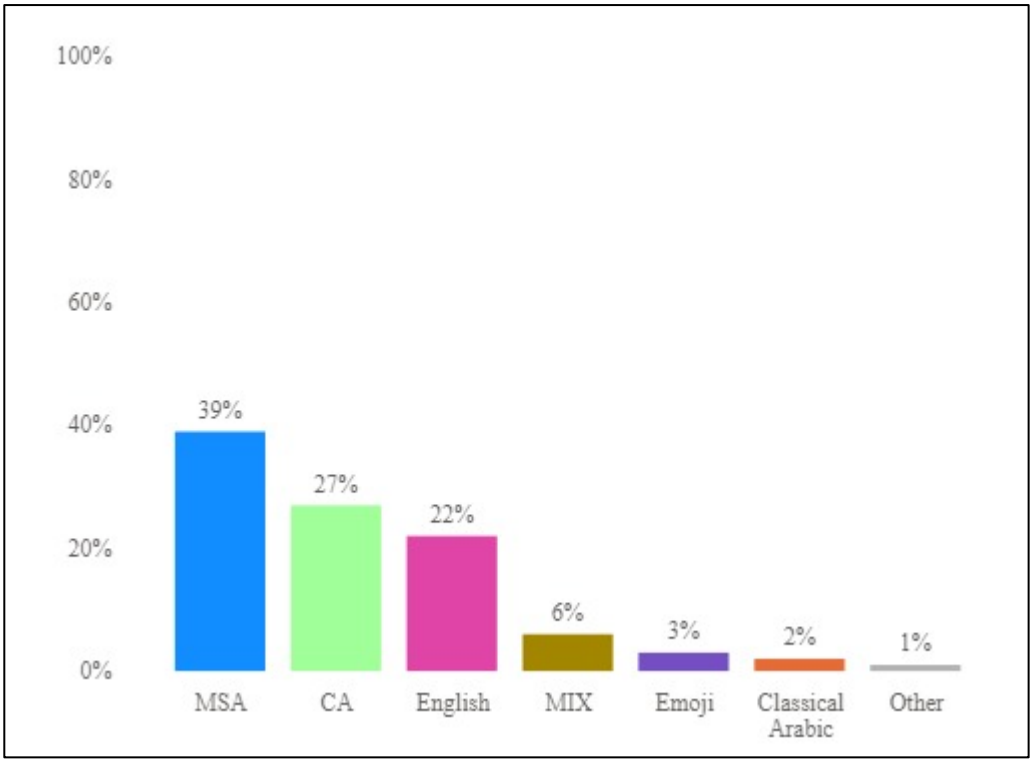


Figure 7.3: Distribution of languages in the monolingual ordinary users' tweets

It can be seen from Figure 7.3 that the ordinary users employ MSA (39%), CA (27%), and English (22%) most often in their monolingual tweets. It is also noteworthy that the ordinary users' hierarchy of languages and varieties echoes the influencers' results shown in Figure 6.2 in Chapter 6. Furthermore, in all three groups (i.e., the corporations, influencers, and ordinary users), the top three are always MSA, CA, and English in this same order. However, the proportions differ, especially when comparing the influencers and ordinary users against the corporations. For example, 27% of the ordinary users' tweets are in CA, while for the corporations' tweets this drops to 16%. In terms of English, 22% of the ordinary users' tweets are in English versus 11% in the corporations' monolingual tweets, which are not paired with an Arabic tweet. Figure 7.3 also reveals 6% mixed tweets in the ordinary users' data, which will be further analysed in the next section.

7.2.3 The languages and varieties in the ordinary users' multilingual tweets

Following the Matrix Language Frame model employed for the analysis of the corporations' and the influencers' tweets in Chapter 5 and Chapter 6, respectively, Figure 7.4 presents the distribution of matrix codes in the multilingual ordinary users' tweets (N=93).

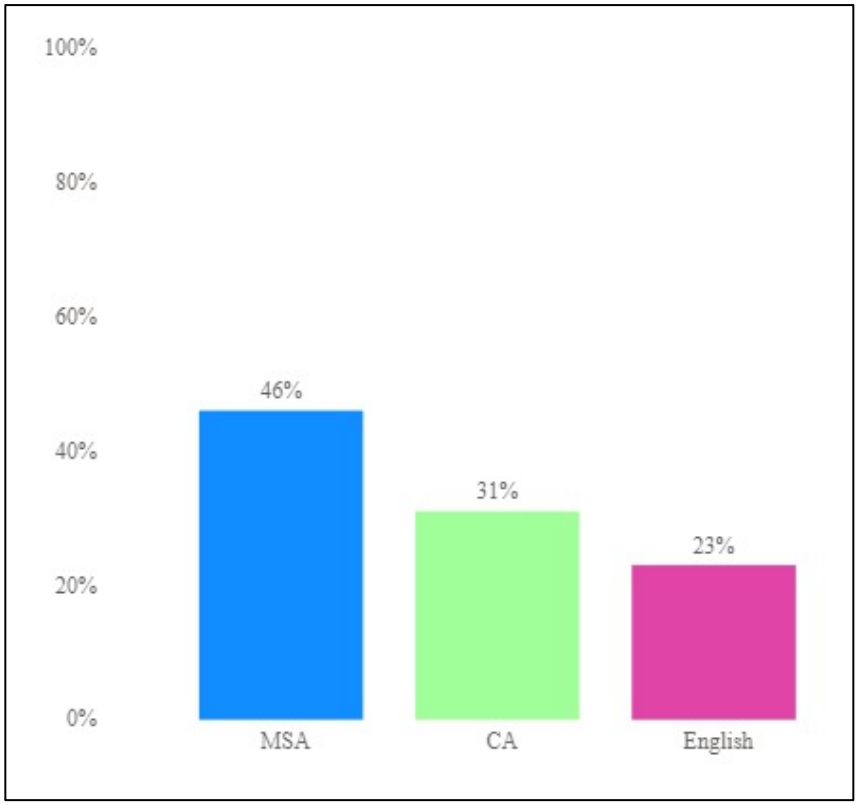


Figure 7.4: Distribution of matrix codes in the multilingual ordinary users' tweets

Figure 7.4 shows the matrix codes utilised in the bilingual and multilingual tweets, revealing that only MSA (46%), CA (31%), and English (23%) are employed as matrix codes in the ordinary users' tweets. Again, the proportions of these languages and varieties are similar to those of the influencers' tweets in ratio and hierarchy; however, the corporations' multilingual tweets feature English much less frequently (2%) than either the ordinary users (23%) or the influencers (17%).

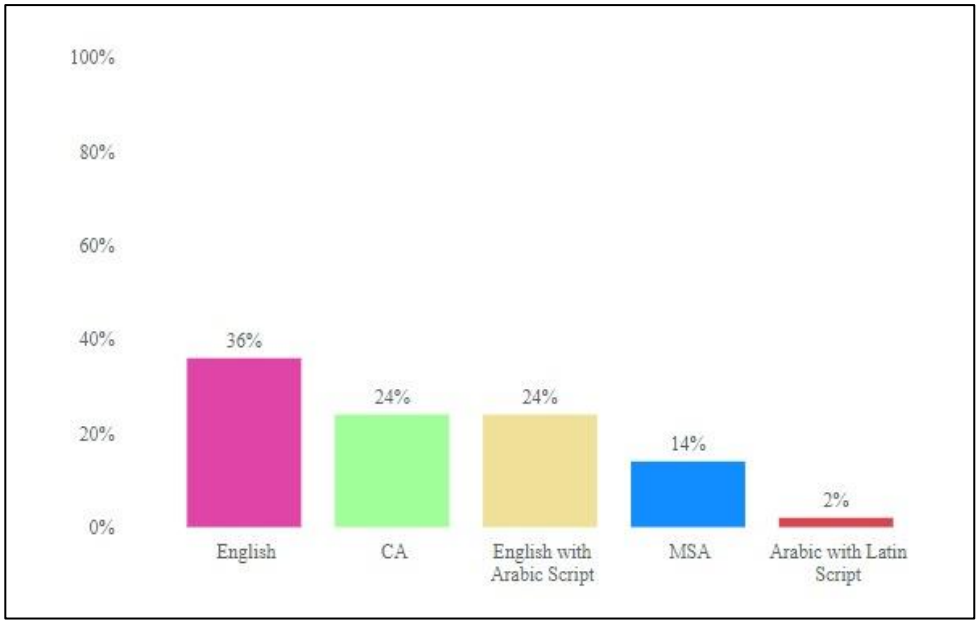


Figure 7.5: Distribution of embedded codes in the multilingual ordinary users' tweets

Figure 7.5 illustrates the languages and varieties embedded in the ordinary users' tweets, revealing that the majority of the mixed tweets utilise English as an embedded code (36%), followed by CA and English with Arabic script at the same distribution (24%), then MSA (14%) and finally Arabic with Latin script (2%). There is variation in the results of the codes embedded in the multilingual tweets employed by the three different groups. For instance, the influencers primarily use CA in 31% and English with Arabic script in 28% of the mixed tweets. On the other hand, the corporations primarily utilise English (65%) and English with Arabic script (24%).

Table 7.1 presents the different languages and varieties employed by the ordinary users combined, and displays the embedded codes that appeared with each matrix code.

Table 7.1: Distribution of embedded codes to matrix codes in the ordinary users' data

Matrix code	Embedded code	Total number	Percentage for each matrix language	Percentage of overall tweets
MSA	English	22	51%	25%
	CA	16	37%	17%
	English with Arabic script	5	12%	5%
	Total	***43***		
CA	English with Arabic script	15	52%	16%
	English	13	45%	14%
	MSA	1	3%	1%
	Total	***29***		
English	MSA	13	62%	14%
	CA	6	29%	6%
	Arabic with Latin script	2	9%	2%
	Total	***21***		
Total number of tweets		***93***		***100%***

It can be seen from Table 7.1 that the most frequently used varieties and languages together are MSA with English (25%), followed by MSA with CA (17%). This differs from the influencers' data, which shows that MSA and CA are the most utilised varieties together (24%). The next most frequent combination in the ordinary users' data is CA in conjunction with English with Arabic script that constitutes 16% of the mixed tweets, followed by CA with English, and English with MSA, where both combinations constitute 14% of the mixed tweets.

Table 7.2 presents the proportion of the different languages and varieties that appeared in the ordinary users' tweets (N=1,463).

Table 7.2: Proportion of languages and varieties in the ordinary users' tweets

No.	Language or variety	Appearance in monolingual tweets	Appearance as a matrix code	Appearance as an embedded code	Number of tweets	Percentage of all ordinary users' tweets
1	MSA	570	43	14	627	43%
2	CA	397	29	22	448	31%
3	English	326	21	35	382	26%
4	French	3	-	-	3	<1%
5	Classical Arabic	25	-	-	25	2%
6	English with Arabic script	5	-	20	25	2%
7	Arabic with Latin script	-	-	2	2	<1%
8	Emoji	44	-	-	44	3%
	Total	**1,370**	**93**	**93**	**1,556**	

Table 7.2 presents all the languages and varieties in the ordinary users' data (i.e., monolingual, bilingual, and multilingual). It must be noted that the total of 1,556 tweets counts the bilingual and multilingual tweets twice, while the percentages are calculated based on the number of the ordinary users' tweets, which is 1,463. According to Table 7.2, the majority of the ordinary users utilise MSA at 43%, followed by CA at 31%, and English at 26%, which corresponds closely to the data from the influencers of MSA at 46%, CA at 30%, and English at 24%. The other languages and varieties that appeared in the ordinary users' data constitute less than 8% and include Classical Arabic, French, English with Arabic script, Arabic with Latin script, and emoji only.

Reflecting on the different languages and varieties employed by the ordinary users reveals the following key findings:

- As per the influencers' findings, the ordinary users engage in more intra-tweet code-mixing, with only 7% of the tweets featuring two or more languages and varieties.
- The majority of the ordinary users' tweets employ MSA (43%), followed by CA (31%) and then English (26%).
- Overall, the distributions of the languages and varieties utilised in the ordinary users' tweets are very close to those of the influencers, thus indicating that individuals,

whether influencers or ordinary users, use the languages or varieties in a similar manner.

The next section discusses the patterns of the different languages and varieties, and how they are employed on the timelines of the ordinary users.

7.2.4 The patterns of languages and varieties in the ordinary users' data

The patterns of the several languages and varieties that appeared on the timelines of the ordinary users are discussed in this section in terms of the most common patterns utilised by the ordinary users. Furthermore, certain samples from these patterns are considered in greater detail.

The first step for analysing the patterns is presenting the frequency of the languages and varieties across all the examined ordinary users' Twitter accounts. Table 7.3 presents the languages and varieties that appeared in the tweets, along with the number of ordinary accounts that displayed these languages and varieties on their Twitter timeline.

Table 7.3: Frequency of languages and varieties in the ordinary users' data

Main language or variety	Total number of users' accounts	Percentage of accounts	Total number of tweets	Percentage of tweets
MSA	20	100%	570	39%
CA	16	80%	397	27%
English	18	90%	326	22%
MIX	18	90%	93	6%
Emoji	10	50%	44	3%
Classical Arabic	9	45%	25	2%
English with Arabic Script	4	20%	5	<1%
French	2	10%	3	<1%
Total			**1,463**	**100%**

In line with the analysis of the ordinary users' tweets in the previous sections, Table 7.3 confirms the dominance of MSA in the ordinary users' accounts since it appeared in all of the examined accounts. This is also similar to the results of the influencers' tweets, in which

all of the accounts utilised MSA. Surprisingly, English is employed in 90% of the ordinary users' accounts, while CA is used in 80% of the accounts. However, English is utilised in 22% of the overall tweets, compared with 27% for CA. Hence, despite English being utilised by more of the ordinary users, it is employed less in the tweets as the main language for communication on Twitter. Table 7.3 also reveals that 90% of the accounts employ a mix of languages and varieties within the same tweet, which is the exact same percentage of the influencers' accounts using mixed tweets.

Following the analysis of the frequency of the languages and varieties in the ordinary users' accounts, this section now considers the languages and varieties that appeared together on the timelines of the ordinary users. Again, if the language or variety was utilised at least three times on the timeline of the ordinary user, it was counted as part of the pattern. Table 7.4 shows the main patterns that emerged from the analysis, along with the proposed term for each pattern and the number and proportion of the ordinary users' accounts (N=20) that use this specific pattern.

Table 7.4: Patterns of multilingualism in the ordinary users' data

Pattern	Number of accounts	Percentage of total accounts
1 Bilingualism	**4**	**20%**
MSA/English	2	10%
MSA/English/MIX	2	10%
2 Bilingualism and Diglossic Switching	**14**	**70%**
CA/MSA/English	5	25%
CA/MSA/English/MIX	3	15%
CA/MSA/English/MIX/Emoji	5	25%
MSA/English/Classical Arabic	1	5%
3 Diglossic Switching	**2**	**10%**
CA/MSA	2	10%
Total	**20**	**100%**

Table 7.4 demonstrates the most common patterns employed by the ordinary users on Twitter, whereby the most common pattern used is bilingualism and diglossic switching (70%). However, it can be noticed that within this pattern there is variation in the use of

languages and varieties. This variation in the ordinary users' tweets is more frequent than the variation that appeared in the influencers' (60%) and the corporations' (18%) patterns. This may be the case because the ordinary users' tweets are generally not planned and structured like those of the corporations' accounts, while ordinary users may be more flexible in terms of language when compared with the influencers who seek visibility to a wider audience. Furthermore, it can be inferred from Table 7.4 that MSA is an essential variety of Arabic that is employed in all patterns by the ordinary users. Moreover, multilingualism is shown in all patterns and all the ordinary users' accounts, except for two accounts that use two varieties of Arabic, namely MSA and CA.

Figure 7.6 shows the proportions of the different languages and varieties utilised by a sample of the ordinary users for each pattern.

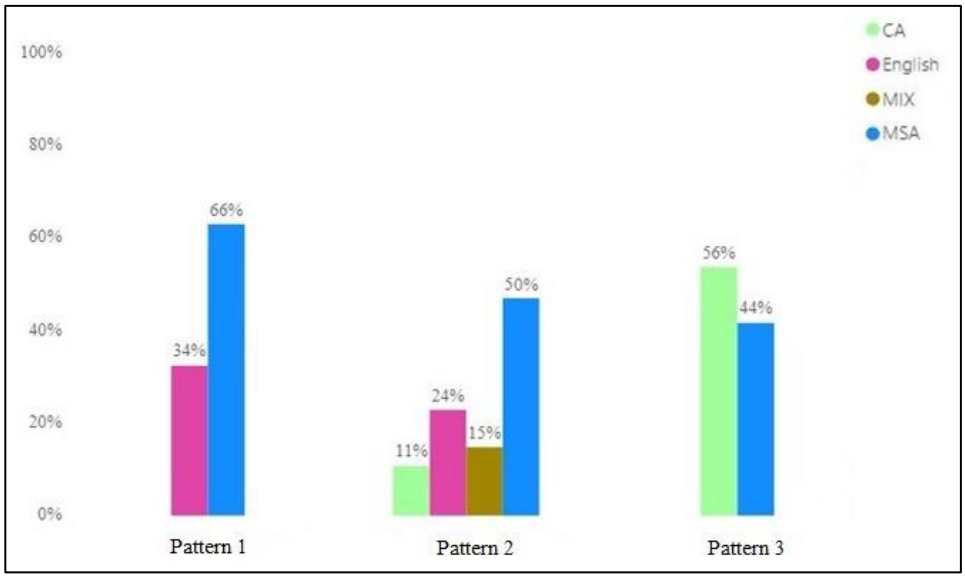


Figure 7.6: Sample of the different patterns employed by the ordinary users

The examples presented in Figure 7.6 illustrate the patterns mentioned in Table 7.4. It is important to note that the names of the users have been concealed for privacy, since these accounts are for ordinary users whose anonymity is required. The data displayed in Table 7.4 reveal some discrepancy in using the languages and varieties by the ordinary users. The first example shows the same pattern, that is, bilingualism. For this user, the tweets are in MSA and English, but with more tweets in MSA. It must be noted that this pattern, used by two of the ordinary users (10%), resembles some of the corporations in using only MSA and English in their communication with their audience on Twitter. The bilingualism/dialecticism pattern exemplified by Pattern 2 in Figure 7.6 shows more variation in the languages and varieties utilised, whereby further examples showing how this pattern

can be employed with different proportions of the languages will be presented later in this section. The third pattern (i.e., diglossic switching) manifests itself in example 3 in Table 7.4, and shows how some ordinary users utilise only Arabic in their tweets but with different varieties, namely CA and MSA. It is also interesting that the ordinary users' **diglossic switching** is only for the different varieties of Arabic with no mix of English, as per the case of the influencers. The inconsistency of using the different languages and varieties can be explained in light of the functions, as will be discussed later in section 7.2.7.

Figure 7.7 demonstrates how the ordinary users employ the languages and varieties in one of the most common patterns identified in Table 7.4, namely bilingualism and diglossic switching (70%). While all of the examples presented in Figure 7.7 use this pattern, they show different proportions of the languages and varieties.

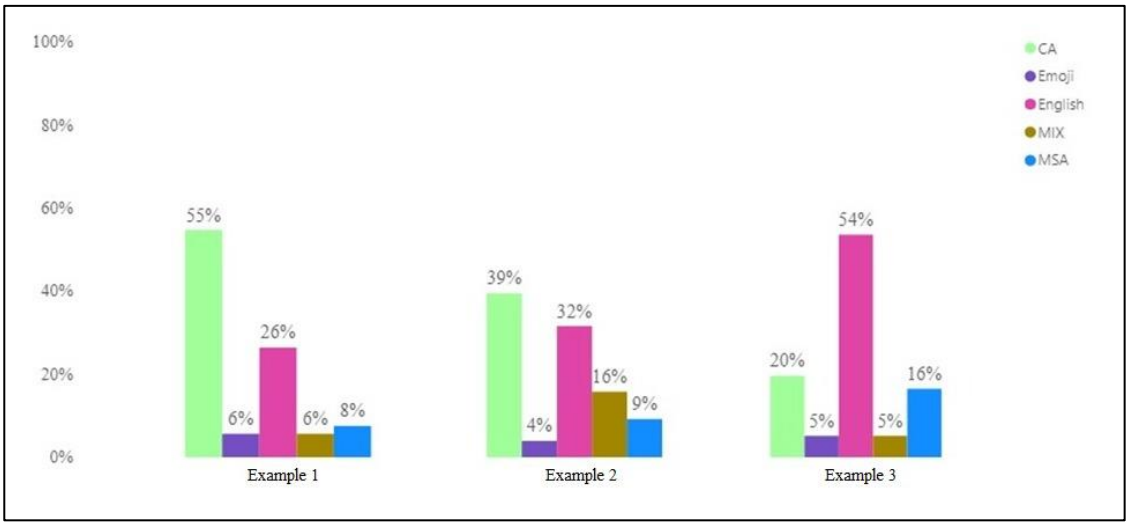


Figure 7.7: Proportion of languages and varieties used in the bilingual and diglossic switching pattern by a sample of ordinary users

Figure 7.7 demonstrates the variation of how the ordinary users utilise the languages and varieties featuring the most common pattern (i.e., bilingualism and diglossic switching) that features in 70% of the ordinary users' accounts, as illustrated in Table 7.4. Overall, it can be seen from the figure that the most utilised variety and language are CA and English. The first and second patterns show that CA dominates the timelines of these users, while the third pattern indicates that English is the language primarily employed on the timeline of this user. Furthermore, in the first two patterns, English is second in frequency after CA, while in the third pattern CA is second in frequency after English. These results indicate the preference for CA and English by the ordinary users. Surprisingly, MSA is ranked third in the first (8%) and third (16%) patterns, and ranked fourth (9%) in the second pattern. In

general, these samples reveal that there is less discrepancy in using languages and varieties in the ordinary users' data when compared to the results of the influencers and corporations presented in the earlier chapters.

After exploring the different languages and varieties used in the ordinary users' tweets, the following section considers the paralinguistic features that appeared in their data.

7.2.5 The use of paralinguistic features in the ordinary users' tweets

There are different paralinguistic cues utilised in the tweets to communicate social and emotional aspects to the virtual audience. Among these cues are emojis, which appeared many times in the ordinary users' data. Figure 7.8 demonstrates the distribution of emojis, as well as other paralinguistic cues employed by the ordinary users on Twitter.

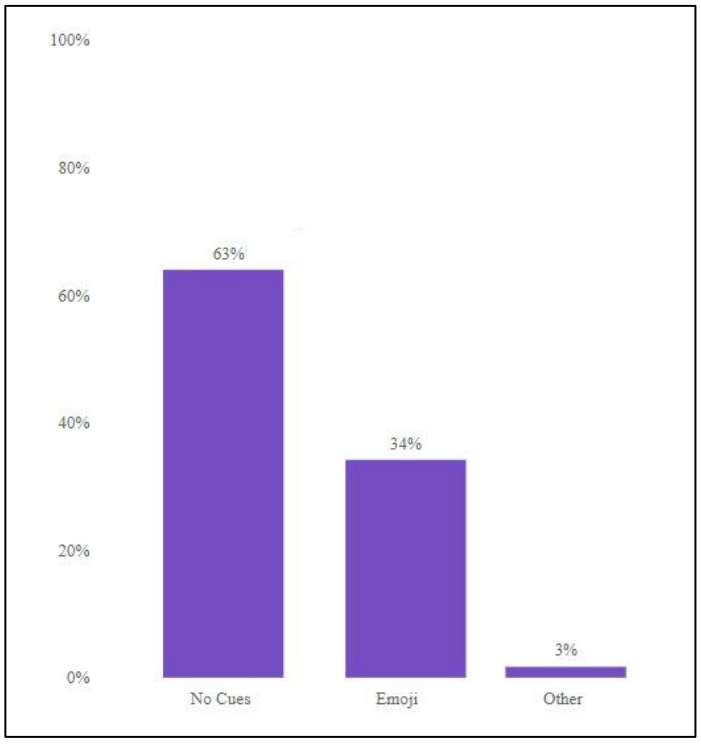


Figure 7.8: Distribution of paralinguistic cues in the ordinary users' tweets

Figure 7.8 shows that 63% of the ordinary users' tweets only use text-based content without mixing any other paralinguistic cues. Meanwhile, 3% (44 tweets out of 1,463 in total) utilise emoticons, punctuation, or letters without any text, and 34% of the tweets contain an emoji as well as the text. As mentioned in the previous chapters, the emoji is employed in the tweets to mimic the facial expressions and body language typical of face-to-face communication, and therefore it is expected to be found in the majority of the tweets of the

influencers and ordinary users, but not the corporations. The 'Other' category in the figure includes punctuation marks (e.g., !!!) and letters (e.g., Nooo), which can be also used with Arabic script.

Table 7.5 displays the number and proportion of tweets that utilised emojis with each language or variety.

Table 7.5: Proportion of tweets with emojis in the ordinary users' data for each language/variety

No.	Language or variety	Total number of tweets	Number of tweets with emojis	Percentage
1	CA	448	226	51%
2	English	382	115	30%
3	MSA	627	109	17%
4	MIX	93	25	27%
5	French	3	3	100%
6	Classical Arabic	25	3	12%
7	English with Arabic script	25	2	8%

Table 7.5 shows the proportion of emoji use with each language and variety. It is observed that the emoji was employed in all three of the French tweets (100%), which were extracted from three different Twitter accounts. Moreover, the proportion of ordinary users employing the emoji with CA is particularly high at 51%, while 30% of the English tweets feature the emoji. The results of using emojis with the languages and varieties in the ordinary users' data differ from the influencers' results, which revealed that 35% of the CA tweets use emojis, followed by 20% of the MSA, and 14% of the English tweets.

Sections 7.2.2, 7.2.3, and 7.2.4 discussed the languages and varieties that appeared in the ordinary users' data, while section 7.2.5 explored the paralinguistic features utilised with the verbal elements of the tweets. In the next section 7.2.6, the functions that appeared in the same data are considered

7.2.6 The functions of the ordinary users' tweets

This section considers the functions of the ordinary users' tweets in order to link the languages and varieties with the functions. It must be noted that the functions extracted from the ordinary users' data are similar to the influencers' functions, which are as follows:

1) Formulaic Purposes
2) Culturally Specific Genres
3) Reported Speech
4) Dialogical Interrelation
5) Marketing
6) Self-Expression

Table 7.6 presents the number of tweets and proportions for each of the main functions.

Table 7.6: Number and distribution of tweets for every main function

No.	Function	Number of tweets	Percentage
1	Formulaic purposes	125	9%
2	Culturally specific genres	40	3%
3	Reported speech	139	9%
4	Dialogical interrelation	180	12%
5	Marketing	7	1%
6	Self-expression	972	66%
	Total	**1,463**	**100%**

As explained in Chapter 4 and Chapter 5, the category of self-expression is only employed by the influencers and ordinary users. Table 7.6 shows that the majority of the ordinary users' tweets concern self-expression (66%) to share their thoughts or anecdotes from their lives with the followers. It is noticed that the ordinary users tweet more about their lives than the influencers, with only 33% of the latter's tweets concerning self-expression. Furthermore, Table 7.6 shows that although these are ordinary users who are not likely to be paid for any advertisements, seven of the tweets involve marketing products or events such as the following example seen in Figure 7.9, which shows an advertisement for a children's camp.

Figure 7.9: Example of a marketing tweet by one of the ordinary users

The next most common function for the ordinary users is dialogical interrelation, which constitutes 12% of the tweets and is thus slightly higher than the percentage of the influencers in the same category at 10%. Furthermore, the reported speech function is also higher in the ordinary users compared to the influencers at 9% and 7%, respectively. Finally, the formulaic purposes function involves 9% of the ordinary users' tweets, while the culturally specific genres involves 3%, compared to 12% and 3%, respectively, for the influencers' tweets.

The next section discusses how these functions are linked with the languages and varieties that appeared in the ordinary users' tweets.

7.2.7 Languages and varieties used with functions in the ordinary users' tweets

This section explores how the different languages and varieties displayed on the timelines of the ordinary users are utilised with the functions extracted from the same data. Figure 7.10 shows the functions along with the languages and varieties employed with each function in the ordinary users' tweets, while Table 7.7 provides a detailed list of the languages and varieties used with each main function.

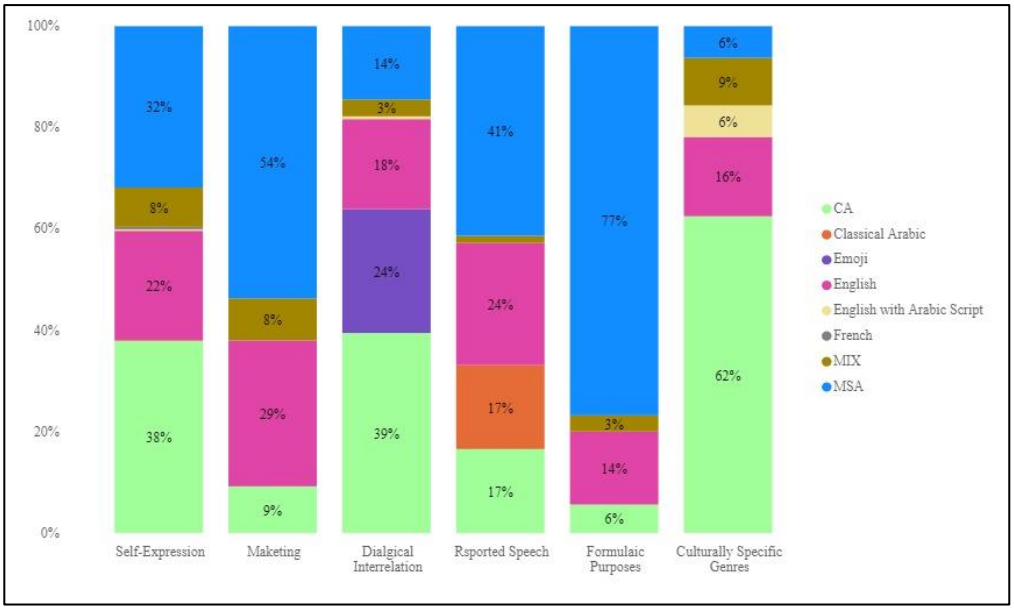


Figure 7.10: Comparison of the functions and languages/varieties in the ordinary users' tweets

Table 7.7: Number of functions with the languages and varieties in the ordinary users' tweets

Main function	Main language	Total
Self-expression	CA	273
	Classical Arabic	1
	English	232
	French	3
	MIX	77
	MSA	384
	English with Arabic script	2
Total		972
Marketing	CA	2
	English	3
	MIX	1
	MSA	1
Total		7
Dialogical interrelation	CA	71
	Emoji	44
	English	32
	MIX	6
	MSA	26

	English with Arabic script	1
Total		180
Reported speech	CA	24
	Classical Arabic	18
	English	35
	MIX	2
	MSA	60
Total		139
Formulaic purposes	CA	7
	English	18
	MIX	4
	MSA	96
Total		125
Culturally specific genres	CA	20
	English	6
	English with Arabic Script	2
	MIX	3
	MSA	3
	Classical Arabic	6
Total		40
Total number of tweets		**1,463**

It can be observed from Figure 7.10 and Table 7.7 that the function of the tweet may determine the language or variety utilised in the post. In general, the results presented above are similar to those of the influencers. For example, MSA is the most employed variety in formulaic purposes (77%), as per the influencers' results (64%). Furthermore, CA is primarily used in culturally specific genres posts (62%), which resembles the case for the influencers (55%). As for English, it is primarily utilised by the ordinary users in terms of marketing (29%), reported speech (24%), and self-expression (22%).

7.2.8 Summary of the analysis of the ordinary users' tweets

The following are certain highlights from the analysis of the ordinary users' tweets:

- As per the findings of the corporations and the influencers, the various languages and varieties are utilised by the ordinary users across the tweet, rather than intra-tweet code-mixing.

- The analysis of the tweets reveals that 43% of the tweets are in MSA, 31% in CA, and 26% in English. These results are very similar to those of the influencers at 46% in MSA, 30% in CA, and 24% in English. However, the corporations use more MSA in their tweets, comprising 61% of the data.
- While the ordinary users have different patterns for employing the languages and varieties, the majority (70%) employ a mix of MSA, CA, and English, among other languages and varieties. The percentage is close to the results of the influencers, where 60% of the accounts use the same pattern.
- The ordinary users employ emojis distinctly, with the results showing that they are employed frequently with CA (51%) and English (30%), while for the influencers, emojis are employed in 35% of the CA tweets and 14% of the English tweets.
- The ordinary users engage in more self-expression tweets than the influencers, at 66% and 33%, respectively. On the other hand, 35% of the influencers' tweets involve marketing, compared to 1% of the ordinary users' tweets. The remaining functions have an approximately similar distribution.
- Similar to the corporations' and social media influencers' finding that the function of the tweet affects the language or variety used, the ordinary users' tweets show the same trend. Furthermore, in line with the influencers' tweets, MSA is primarily utilised in formulaic purposes (77%) and marketing tweets (54%), while CA is primarily utilised in tweets for culturally specific genres (62%).

7.3 Analysis of the ordinary users' interviews

This section explores the results of the three interviews with the ordinary users on Twitter, comprising one male and two females. In similarity to the analysis of the corporations' and social media influencers' interviews, the aims and goals of the ordinary users on social media, their target audience, the languages and varieties, and the paralinguistic cues employed on Twitter are investigated.

The analysis of the themes extracted from the ordinary users' interviews are only counted without adding percentages since there are only three interviewees.

7.3.1 Languages and varieties employed for communication on Twitter

In general, the discussion of the languages and varieties with the ordinary Twitter users was not as detailed as that with the corporations and social media influencers. Furthermore, none of the interviewees described using the white dialect, with the discussion only involving

MSA, English, and CA. Table 7.8 shows the interviewed ordinary users' responses regarding the languages and varieties used.

Table 7.8: Ordinary users' languages and varieties used on Twitter

No.	Theme	Number
1	MSA	3
2	English	1
3	CA	1
	Total	**5**

All the ordinary users reported using MSA in their tweets, which differs from the results in section 7.2 where only 17% of the ordinary users' tweets were found to be in MSA, with 51% in CA and 30% in English. Interviewee ORD.3 exemplified an extreme rigidity toward using MSA:

Extract 7.1

> *I only use Standard Arabic. I don't even think of posting jokes because I don't want to use Colloquial Arabic. As I told you earlier, I have a different personality on social media which is formal, so I don't use Colloquial Arabic. I sometimes even use Twitter to contact companies such as STC [Saudi Telecom Company] to solve issues and I avoid any discussion in Colloquial Arabic. I have zero percent Colloquial Arabic in my posts.*

In terms of preparing the tweets for posting. Interviewee ORD.1 mentioned that he edits the tweets prior to posting to ensure that the language is correct: "*I edit my Arabic language many times and I use diacritics to make sure it is read correctly by others*".

As for CA, Interviewee ORD.1 reported using other Arabic dialects than the Saudi default for different purposes such as humour: "*I sometimes use the Egyptian dialect for some of my funny posts*".

The analysis of the ordinary users' tweets revealed that 30% are in English. Two of the interviewees stated that they only use English occasionally to post quotes, but not their own content, which indicates that one of the functions of using English in the ordinary users' tweets is reported speech.

7.3.2 The interviewees' perceptions regarding the languages utilised on Twitter

During the interviews, the ordinary users provided their opinions on the different languages employed on Twitter. This section presents their responses regarding the language that they felt most appropriate for use on Twitter, the language with the most audience interaction, and whether tweets should be translated into English or Arabic.

7.3.2.1 The optimum language or variety for use on Twitter

The ordinary users in the interviews expressed their opinion regarding the ideal language or variety to be utilised on Twitter, while the analysis of the ordinary users' data revealed the occurrence of MSA in 43% of the tweets. Table 7.9 shows the opinions of the ordinary users on the language or variety that should be employed on Twitter.

Table 7.9: Ordinary users' opinions regarding the optimum language or variety for use on Twitter

No.	Theme	Total
1	MSA	1
2	Depends on the user and the audience	1
3	Depends on the content	1
	Total	**3**

Table 7.9 shows that the interviewees each had a different opinions regarding the optimum language or variety for use on Twitter. Interviewee ORD.2 claimed that tweets should only be posted in MSA or CA:

Extract 7.2

> *I think we need to use Arabic even if it is Colloquial Arabic…we got used to English and this should not be the case. I even criticise the people who write their usernames in English. I understand that some people may have international followers, but Arabic still needs to be the dominant language.*

Although some users may prefer MSA, they can be more flexible about the languages and varieties used, as explained by Interviewee ORD.2:

Extract 7.3

> *I prefer my tweets to be in the same Arabic used in Saudi Arabia, Bahrain, Morocco, Syria, Yemen and every place that speaks and writes in Arabic. Especially if the tweet is general, I prefer that it is understood by all my followers. However, I don't criticise people who use Colloquial Arabic as sometimes it may be necessary to convey the meaning. I think Standard Arabic is important in published tweets, but Colloquial Arabic can be used in replies.*

Interviewee ORD.1 believed that the aims and goals for having a presence on Twitter, in addition to the target audience, play a major role in choosing the language or variety:

Extract 7.4

> *I think this depends on the reason you are on Twitter. If you are there for literacy, it should be Standard Arabic, but if you are on Twitter to give advice and be a role model then the language choice will be different. For example, if you are talking to teens, maybe Colloquial Arabic or Arabizi, so the target audience is also important. You need to know what is your personality style and the people you want to follow you.*

Although the aim of this part of the interview was not to explore how the ordinary users perceived tweets in different languages and varieties posted by social media influencers, Interviewee ORD.2 expressed her view about this in terms of the dilemma of remaining respectful to the local audience while reaching out to international followers:

Extract 7.5

> *If a Saudi influencer posts in English, I don't think it is respectful to the people. I also think at the same time that it is important for some influencers who have an international audience to use English to give them a sense of our society, especially if it is used in a positive way.*

It should be noted here that in the previous chapter, the influencers asserted that English posts can also be important to ensure that the maximum number of followers have the opportunity to understand the posted content.

7.3.2.2 The ideal language or variety for audience interaction

Similar to the analysis of the corporations' and social media influencers' tweets, that of the ordinary users' tweets did not incorporate the number of likes, retweets, and replies. The ordinary user interviewees were asked about the languages and varieties that the audience tends to interact with to a greater extent on Twitter, with their responses presented in Table 7.10.

Table 7.10: Ordinary users' opinions regarding the ideal language/variety for audience interaction

No.	Theme	Total
1	MSA	1
2	CA	1
3	Depends on the content	1
	Total	**3**

Again, the answers of the ordinary users presented in Table 7.10 highlight their different points of views. Interviewee ORD.3, who believed that only MSA should be employed on Twitter, had not considered the interaction of the audience, but still rejected the idea as a rationale for modifying the posting language: "*I haven't thought about the audience interaction depending on the language, and I will still use Standard Arabic even if other languages may have more interaction*".

On the other hand, Interviewee ORD.1 asserted that if the user is aiming for more interaction, then CA should be employed: "*There is more dialogue in the interaction and that's why Colloquial Arabic could be used more*".

7.3.2.3 The need for English posts to be translated

The analysis of the ordinary users' tweets revealed the translation of content in their tweets to be uncommon, as only one case was found in all the ordinary users' tweets. During the interviews, the ordinary users were asked if they thought the tweets should be translated, with Table 7.11 presenting their responses.

Table 7.11: Ordinary users' opinions on the translation of tweets

No.	Theme	Total
1	Yes	1
2	No	1
3	Depends on the user and the audience	1
	Total	**3**

It can be seen from Table 7.11 that the three interviewees had varying responses to this question. To recall, in the previous chapter it was found that although the influencers leaned towards not translating the content, they also stressed the importance of the type of content and the audience.

7.3.3 Using paralinguistic cues

The analysis of the tweets reveal that the ordinary users mostly employ emojis, as 34% of the tweets included an emoji with the text and 3% of the ordinary users' tweets in the data only featured emojis, as shown in Table 7.3. Table 7.12 below presents the frequency of using paralinguistic cues with tweets, as expressed by the interviewees.

Table 7.12: Ordinary users' opinions on using paralinguistic cues in tweets

No.	Theme	Total
1	Always	2
2	Sometimes	1
	Total	**3**

The analysis of the ordinary users' tweets showed that 34% included an emoji. However, the responses presented in Table 7.12 indicate that the ordinary users always or sometimes use a paralinguistic cue in their tweets. Interviewee ORD.1 stressed the importance of using emojis to make the tweets more engaging and accessible:

Extract 7.6

> *Emojis can make the tweet look more attractive, and it also summarises what you want to say. I feel that emojis and moving pictures are more important in responses rather than posts.*

Although the analysis of the tweets showed that only 3% of the tweets included paralinguistic cues other than emojis, interviewee ORD.3 expressed his preference for the use of punctuation as he felt that emojis did not match his formal identity: "*I prefer punctuation such as exclamation marks to emojis. I feel that emojis don't match my formal personality on Twitter*".

7.3.4 Aims and goals on social media

The ordinary users had different aims for social media, with nine different themes identified from the interviews, even though there were only three ordinary user interviewees. Table 7.13 displays the themes that emerged from the ordinary users' responses.

Table 7.13: Ordinary users' aims and goals on social media

No.	Theme	Frequency
1	**Blogging** *(self-expression)*	3
2	**Engagement** *(formulaic purposes, culturally specific genres, and dialogical interrelation)*	2
3	**Awareness** *(formulaic purposes, marketing, and self-expression)*	1
4	**Increasing the number of followers** (*formulaic purposes, culturally specific genres, reported speech, dialogical interrelation, and marketing)*	1
5	**Quality of followers** *(formulaic purposes, reported speech, and self-expression)*	1
6	**Presence** *(formulaic purposes, reported speech, and self-expression)*	1
	Total	**9**

Generally, all the interviewees believed that being on Twitter and social media is important, and that anyone who is not on social media is 'illiterate', as described Interviewee ORD.3.

As shown in Table 7.13, all of the interviewed ordinary users discussed using Twitter for blogging, with Interviewee ORD.3 explaining the similarity between Twitter and blogs:

Extract 7.7

> *Twitter for me is a relief. I used to do blogs under nicknames and Twitter gives me a similar space to blogging. Although Twitter is more limited, it was a nice challenge to write what you want to say in one or two lines. It was interesting for me to do blogging in this way.*

The extract above shows the users' acknowledgement of the challenge of expressing themselves within the word limit of the tweets. It is therefore expected that they also make some adjustments and edit the tweet prior to posting. This editing may also include the choice of language or variety. Therefore, even if the ordinary users do not have pre-set

strategies for their communication on Twitter, developing the tweets still requires writing and editing stages.

It was also interesting that the interviewees discussed reflecting a different personality on Twitter. For example, Interviewee ORD.3 stated: "*I feel that my personality on social media is different than my real personality as it is more formal on Twitter. Some people may have funny personalities on social media and also in real life but for me they are different*".

It was reported in section 7.2.6 that 66% of the ordinary users' tweets were self-expressions of their daily lives, which was echoed by Interviewee ORD.1: "*I like sharing interesting information and parts of my daily life, songs or scenes from movies*".

In terms of increasing the number of followers, the interviewees perceived this as a common goal, and that they may retweet for people who have many followers in the hope that these popular users and their followers may like or retweet one of their posts.

The quality of the followers and who engages with the posts was important for Interviewee ORD.2: "*I don't care a lot about the number, but I care about who follows me and I feel happy when somebody I admire follows me and likes my posts*".

7.3.5 Frequency of posts

The presence of ordinary users differs from the corporations and the social media influencers as the ordinary users do not feel obliged to always attend to the audience to maintain or increase their followers. Therefore, many ordinary users could cease tweeting for a long period of time, which represented one of the challenges when collecting the data for the current study. Table 7.14 presents the ordinary users' responses when asked about their frequency of posting.

Table 7.14: Ordinary users' frequency of posting on Twitter

No.	Theme	Total
1	Daily	1
2	Occasionally	1
3	No plan	1
	Total	**3**

Table 7.14 shows the discrepancy in the planning of posts by the ordinary users, with each of the three interviewees providing a different response. For example, Interviewee ORD.1 reported posting on Twitter on a daily basis: "*I usually try to spend two hours every day on*

posts and also on increasing likes and retweets by sharing content from other social media platforms to Twitter".

On the other hand, some of the ordinary users only use the platform to engage with topics that may be of interest to the community on Twitter, as expressed by Interviewee ORD.3: "*I don't use Twitter daily and I don't show my daily life, but I tweet about what is happening in the country; that's why I only tweet occasionally when there is something [interesting to share]*".

There were also different approaches to preparing tweets by the ordinary users, with Interviewee ORD.3 drafting tweets in batches for later posting without any particular plan:

Extract 7.8

> *I write using a pen and paper. I may write 20 tweets and then I publish them on different days...maybe 5 every day, and I also include in my plan retweeting for others. It also depends on my mood as sometimes I don't have the inspiration to write.*

The lack of planning by some of the ordinary users contrasts with the findings from the corporations, where the interviews revealed organised drafting and posting of tweets, which does not occur with the ordinary users or the influencers.

7.3.6 Target audience on social media

The ordinary user interviewees were asked about their target audience on Twitter, with Table 7.15 presenting their responses.

Table 7.15: Ordinary users' target audience on Twitter

No.	Theme	Total
1	Family and friends	2
2	The general public	1
3	Professionals	1
4	No specification	1
	Total	**5**

The ordinary users' main target audience are their families and friends. However, they also have varying audience, with whom they communicate using specific strategies such as employing different languages for particular groups, as explained by Interviewee ORD.2:

Extract 7.9

> *My target audience is divided into three groups. The first group is my friends in real life and my friends at work, and I like to share with them morning posts and quotes and I interact with their tweets. The second group are the intellectuals and professionals who share with me their interest of literacy in the Arab world, not only in Saudi Arabia. The third group are professionals in English, so my posts to them are in English. Some posts of course may be for two of the groups if, for example, it is a morning post in Arabic, so it will be for the first two groups.*

Interviewee ORD.1 reported having different accounts for the different groups of audience: "*I have different accounts. I have a personal one for my family and friends, in which I share personal pictures and posts, and another account for everybody else*". It must be noted that only the public account of Interviewee ORD.1 was analysed as part of this research.

7.3.7 Summary of the ordinary users' interviews

The findings from the interviews with the ordinary users can be summarised as follows:

- All three of the interviewed ordinary users confirmed that they primarily use MSA in their tweets, with CA and English also mentioned by two of the interviewees.
- Unlike the interviews with the marketing executives and the social media influencers, none of the ordinary users discussed the white/plain dialect or seemed to be aware of its existence. Furthermore, none of the interviewees mentioned Arabizi as a choice for posting on Twitter.
- The ordinary user interviewees acknowledged the challenge caused by the word limit on Twitter. Although they do not plan their posts, they spend time to edit the tweet primarily to fall within the word limit, while this also means that they edit the language usage.
- All of the interviewees reported using Twitter mainly as a blog to express their thoughts, and primarily to their friends and family.

7.4 Conclusion

Similar to the corporations' and social media influencers' data analysis, the analysis of the ordinary users' tweets and interviews combined different methods in order to provide an overall perspective of how the different languages and varieties along with the functions are employed on Twitter.

The following chapter will consider different areas from the analysis of the three different groups in order to identify similarities and differences in terms of how the different groups utilise languages and varieties on Twitter.

Chapter 8:
Comparison of the Corporations', Influencers', and Ordinary Users' Results

8.1 Introduction

This chapter builds on the previous three analysis chapters by comparing the patterns found in the corporations', the social media influencers', and the ordinary users' data to gain a more complex understanding of the intertwining of the various linguistic practices by the three groups on Twitter in Saudi Arabia.

The analysis of the different languages and varieties, along with the functions used on Twitter by the different groups of users, as presented in the three previous chapters, has identified similarities and contrasting means in which the corporations, influencers, and ordinary users utilise linguistic and paralinguistic affordances to interact and attract the attention of their virtual audiences on Twitter. This chapter provides an overview of how the languages and varieties are employed on Twitter to answer the research questions listed in section 1.2.

The chapter commences by presenting the general frequencies of the variables across the three groups in section 8.2.1. Then, the discussion focuses on the overall view of intra-tweet code-mixing in section 8.2.2. After that, the analysis shifts its lens to consider the languages and varieties utilised in the monolingual and multilingual tweets in sections 8.2.3 and 8.2.4, respectively. Next, the paralinguistic cues are discussed in 8.2.5. Then, a comparison of the functions across all groups is presented in 8.2.6. Finally, section 8.3 presents the conclusion of the chapter.

In order to conduct the analysis in this chapter, the data were imported into SPSS to enable the statistical analysis of the data collected from Twitter using chi square tests, and to establish whether there are statistically significant differences between the three groups for the various independent factors examined. Furthermore, additional chi-square tests were performed to compare the groups in the following pairs:

- Corporations versus influencers
- Corporations versus ordinary users
- Influencers versus ordinary users

8.2 Tweets' comparative analysis

As mentioned above, this section reports on the findings of the statistical comparative analysis of the data. Furthermore, the analysis in this section aims to contribute towards answering the first research question regarding the linguistic characteristics of the tweets in Saudi Arabia, in addition to the fourth research question about the patterns of language used for branding and self-branding. The second and third research questions will be answered in section 8.2.6 about the functions of the tweets.

8.2.1 Tweets' overall descriptive analysis

As discussed in Chapter 4, the data for the current research comprise 13,426 tweets collected from the public accounts of 50 corporations, 30 influencers, and 20 ordinary users on Twitter. Table 8.1 displays the frequencies of all the study variables for a general perception of the tweets in terms of the data drawn from the three groups.

Table 8.1: Frequencies of the study variables

Variable		*N*	%
Group	Corporations	9,370	70%
	Influencers	2,593	19%
	Ordinary users	1,463	11%
	Total	**13,426**	**100%**
Intra-tweet mixing	Yes	1,574	12%
	No	11,852	88%
	Total	**13,426**	**100%**
Paralinguistic cues	Yes	4,575	34%
	No	8,851	66%
	Total	**13,426**	**100%**
Type of paralinguistic cue	None	8,851	66%
	Emoji	4,457	33%
	Other	108	1%
	Total	**13,426**	**100%**
Monolingual	CA	2,311	17%
	English	1,763	13%
	MSA	6,431	48%
	Other	1,347	22%
	Total	**11,852**	**100%**
Matrix code	CA	443	28%
	English	80	5%
	MSA	1,019	65%

	Other	32	2%
	Total	**1,574**	**100%**
Embedded code	CA	150	10%
	English	934	59%
	English with Arabic	372	24%
	Other	118	7%
	Total	**1,574**	**100%**
Functions	Formulaic purposes	833	6%
	Culturally specific genres	178	1%
	Reported speech	484	4%
	Dialogical interrelation	892	7%
	Marketing	9,525	71%
	Self-expression	1,514	11%
	Total	**13,426**	**100%**

Table 8.1 provides an overview of the data that shows the majority of the tweets (88%) are written in one language or variety, which means that, as noted in Chapters 5 –7, the code-switching tends to occur across the tweets on the timeline as opposed to within the tweets themselves. Nonetheless, 12% of the tweets incorporate two or more varieties, as will be explored in greater detail in section 8.2.4 below. Table 8.1 also shows that 34% of the tweets incorporate a paralinguistic cue, overwhelmingly in the form of emojis, while MSA dominates the monolingual tweets (48%) and is the main matrix code in 65% of the mixed tweets, but is employed much less frequently as an embedded code. English is utilised monolingually in 13% of the tweets and in 5% as the matrix code, but it is widely used as an embedded code (59%). In terms of the functions, it can be seen from Table 8.1 that 71% of the tweets are for marketing purposes. This is because as the corporations' tweets constitute 50% of the data and the influencers constitute 30%, therefore it could be expected that the marketing tweets would dominate the functions in the data.

After considering the overall view of the data, the following sections 8.2.2–8.2.5 discuss the variables presented above in greater detail to compare the performance of the three groups and test the correlation between the target variables and the three groups using chi-square tests.

8.2.2 Intra-tweet code-mixing

Chapters 5–7 showed that the three groups engage in more tweets with no code-mixing than with code-mixing, with Table 8.2 and Figure 8.1 examining the groups to establish whether their rates are similar.

Table 8.2: Intra-tweet code-mixing by group

Group	Mix/No mix		Total% (N)
	Code-mixing	No code-mixing	
Corporations	14%	86%	100% (9,370)
Influencers	8%	92%	100% (2,593)
Ordinary users	6%	94%	100% (1,463)
Average	**12%**	**88%**	**100% (13,426)**

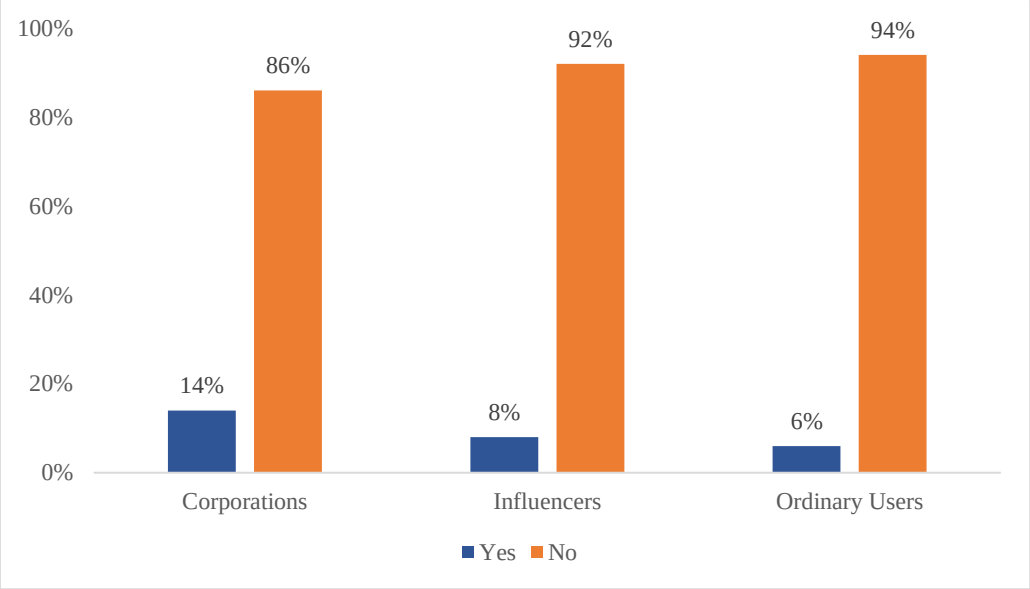


Figure 8.1: Distribution of intra-tweet code-mixing by group

Table 8.3 presents the results of the chi-square tests performed to determine whether there is a statistically significant difference across the groups.

Table 8.3: Intra-tweet code-mixing chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	120.044	$p < 0.001$***
Corporations vs influencers	70.9354	$p < 0.001$***
Corporations vs ordinary users	73.555	$p < 0.001$***
Influencers vs ordinary users	2.0421	$p < 0.153$ (ns)

ns $p > 0.05$ ** $p \leq 0.05$* *** $p \leq 0.01$* **** $p \leq 0.001$*

The results in Table 8.2 and Figure 8.1 indicate that the rates of intra-tweet code-mixing are low for all the groups. The chi-square tests reveal a statistically significant difference across the three groups in terms of their rates of intra-tweet code-mixing ($p < .001$). This difference appears to be driven primarily by the corporations who have a higher rate than the other two groups, as when compared individually with the ordinary users and influencers a statistically significant difference remains, while there is none for the influencers compared with the ordinary users.

Therefore, the following 'intra-tweet code-mixing' pattern is found:

Corporations ≠ Influencers/Ordinary Users

The following section considers in greater detail the 88% of the tweets that use a single language or variety.

8.2.3 Languages and varieties in the monolingual tweets

This section explores the relationships between the groups and the different languages and varieties that appeared in the monolingual tweets, with Table 8.4 and Figure 8.2 presenting the distribution of the monolingual tweets among the three groups, respectively.

Table 8.4: Distribution of monolingual tweets among the three groups

	Monolingual tweets				
Group	**CA**	**Eng**	**MSA**	**Other**	**Total% (N)**
Corporations	13%	10%	51%	26%	100% (8,085)
Influencers	25%	21%	42%	12%	100% (2,397)
Ordinary users	27%	22%	39%	12%	100% (1,370)
Average	**17%**	**13%**	**48%**	**22%**	**100% (11,852)**

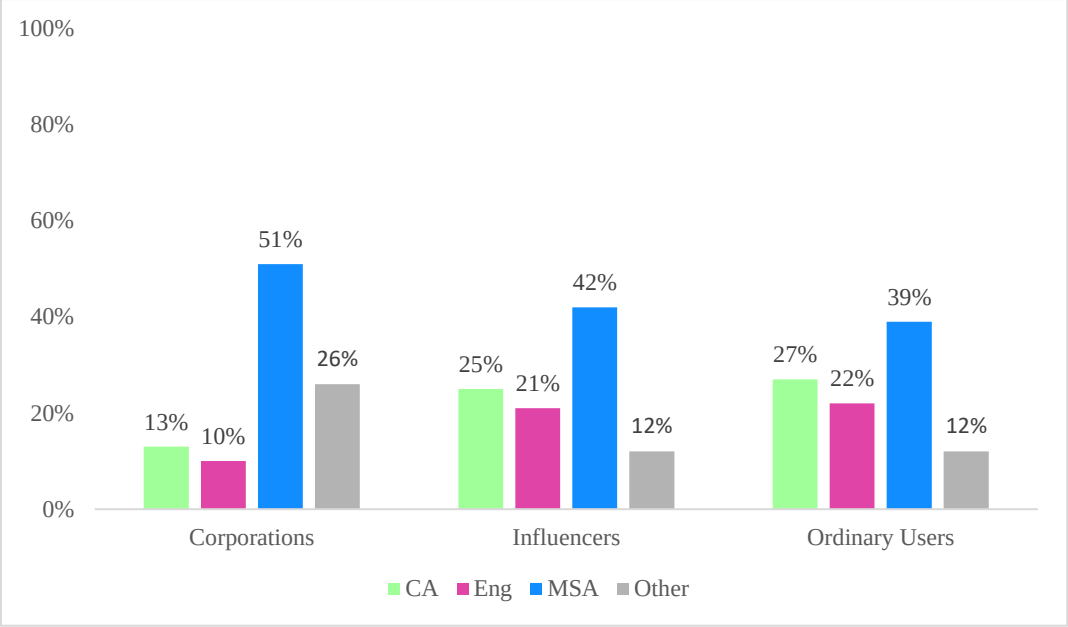


Figure 8.2: Distribution of monolingual tweets by group

Recall that the analysis of the corporations' tweets revealed that some of the accounts use parallel text bilingualism, which was coded as English Pair, MSA Pair, and CA Pair in the data. However, these categories were excluded from this section of the analysis, and thus will not affect the chi-square test results.

Table 8.5 presents the chi-square results, performed to test the relationship between the groups.

Table 8.5: Monolingual tweets' chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	922.843	$p < .001$***
Corporations vs influencers	613.8215	$p < .001$***
Corporations vs ordinary users	484.2159	$p < .001$***
Influencers vs ordinary users	4.5934	$p < 0.2041$ (ns)
ns $p > 0.05$ ** $p \leq 0.05$*	*** $p \leq 0.01$*	**** $p \leq 0.001$*

As Table 8.5 shows, there are statistically significant differences across the three groups, and for the corporations/influencers' and corporations/ordinary users' pairs. As per the case of the intra-tweet code-mixing, the main difference can be found in the corporations' language choices that involve higher rates of MSA (and Other) and lower rates of English and CA than the other two groups. There is no statistically significant difference in the influencers' and ordinary users' use, which suggests that they share the same pattern.

The interviewees from all three groups indicated a preference for MSA, or the white dialect for corporations and social media influencers, as the ideal language or variety to be employed on Twitter. Furthermore, the interview responses revealed that the majority of the respondents did not believe that English and CA should be used on Twitter. For instance, the use of CA was discouraged by only one interviewee from each group, while English usage was discouraged by two marketing executives, one influencer, and another ordinary user. However, the results in Table 8.4 and Figure 8.2 show that the influencers and ordinary users utilise more English and CA, especially in comparison with the corporations. It must also be noted that the majority of the interviewees agreed that the language or variety employed on Twitter depends on the content, while one of the ordinary users surprisingly stated that it depends on the audience.

Therefore, the following 'dominance of MSA in monolingual tweets (less English and CA)' is found: Corporations ≠ Influencers/Ordinary Users.

8.2.4 Languages and varieties in the multilingual tweets

This section reports on the relationships between the groups and the different languages and varieties. First, the distribution of the matrix codes is considered, and then the embedded codes that appear in all the three groups are explored. Table 8.6 and Figure 8.3 present the overall use and distribution of the matrix codes among the three groups, respectively.

Table 8.6: Overall use of matrix codes by the three groups

	Matrix codes				
Group	**CA**	**Eng**	**MSA**	**Other**	**Total% (N)**
Corporations	27%	2%	68%	2%	100% (1,285)
Influencers	32%	17%	49%	2%	100% (196)
Ordinary users	31%	23%	46%	-	100% (93)
Average	**28%**	**5%**	**65%**	**2%**	**100% (1,574)**

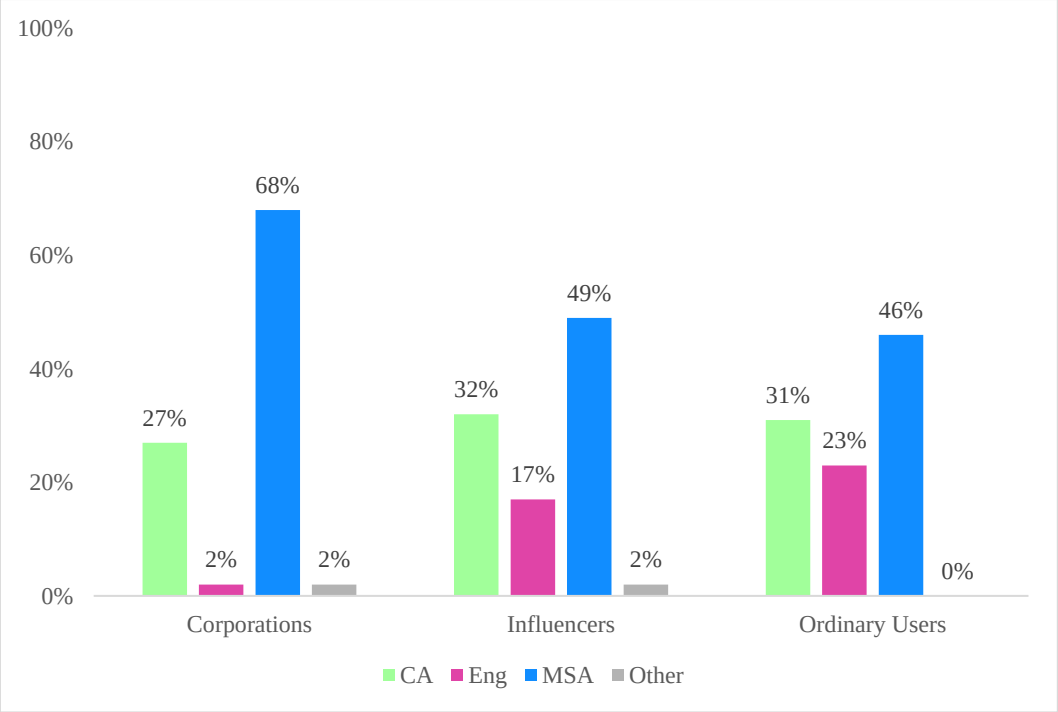


Figure 8.3: Distribution of matrix codes by group

The comparison of all the groups and between each in terms of using the matrix code is shown in Table 8.7.

Table 8.7: Matrix code chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	154.289	$p < 0.001$***
Corporations vs influencers	112.9061	$p < 0.001$***
Corporations vs ordinary users	116.9599	$p < 0.001$***
Influencers vs ordinary users	1.1078	$p < 0.7752$ (ns)
ns p > 0.05	** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*

Table 8.7 reveals the statistically significant differences across the three groups, and also for the corporations/influencers' and corporations/ordinary users' pairs. Again, the main difference can be found in the corporations' language choices, where they have higher rates of MSA (and Other), and lower rates of English than the other two groups.

Therefore, the following 'dominance of MSA as a matrix code in multilingual tweets (less English and CA)' pattern is found: Corporations ≠ Influencers/Ordinary Users. It must be underlined that the difference is mainly in English, as the rates of CA are almost the same.

In terms of the embedded codes in the tweets, Table 8.8 and Figure 8.4 present the overall use of the embedded codes among the three groups, while Table 8.9 shows the chi-square results.

Table 8.8: Overall use of embedded codes by the three groups

	Embedded codes				
Group	**CA**	**Eng**	**EwA**	**Other**	**Total% (N)**
Corporations	5%	66%	23%	6%	100% (1,285)
Influencers	31%	24%	29%	16%	100% (196)
Ordinary users	24%	38%	22%	17%	100% (93)
Average	**10%**	**59%**	**24%**	**7%**	**100% (1,574)**

Note. EwA: English with Arabic script

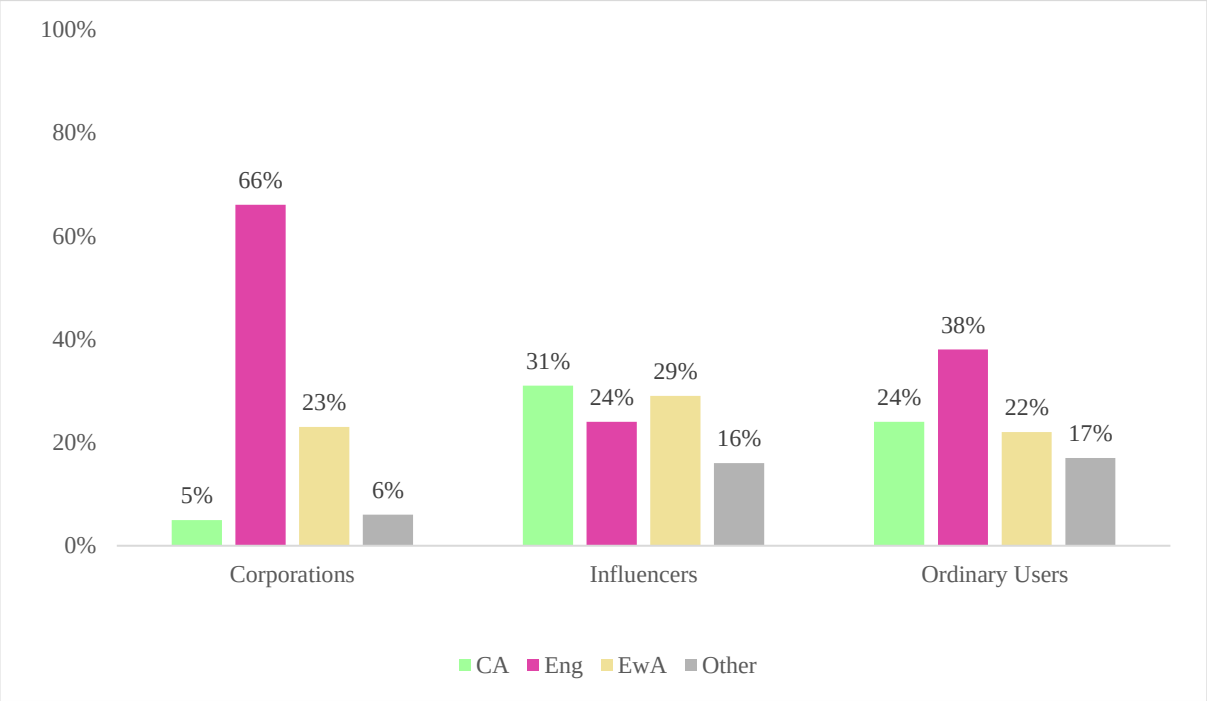


Figure 8.4: Distribution of embedded codes by group

The comparison of all the groups and between each group in terms of using the embedded codes is shown in Table 8.9.

Table 8.9: Embedded codes' chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	232.209	$p < 0.001$***
Corporations vs influencers	203.7851	$p < 0.001$***
Corporations vs ordinary users	74.6219	$p < 0.001$***
Influencers vs ordinary users	6.3037	$p < 0.0977$ (ns)
ns p > 0.05 ** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*

There are statistically significant differences across the three groups, and also for the corporations/influencers' and corporations/ordinary users' pairs. As seen previously, the influencers and ordinary users do not have a statistically significant difference in their rates of use. The difference with the corporations is likely tied to their high use of English as the embedded code. The high percentage of using English and English with Arabic script as an embedded code is not surprising, as the users tend to fill a linguistic gap or to use a term that may be better understood when it is utilised in its original language, English in this case.

Therefore, the following 'dominance of English and English in Arabic script as embedded codes' pattern is found: Corporations ≠ Influencers/Ordinary Users (use more CA and other languages). It must be highlighted that the corporations had less English as the matrix code, which can partly explain why it occurs more as an embedded code.

8.2.5 Paralinguistic cues

This section analyses the use of paralinguistic cues in the tweets, such as emojis and punctuation. Table 8.10 and Figure 8.5 show the three groups' use of a mix of mediums in the tweets, while Table 8.11 presents the results of the various chi-square tests.

Table 8.10: Overall use of paralinguistic cues by the three groups

	Mix medium		
Group	**Yes**	**No**	**Total% (N)**
Corporations	37%	63%	100% (9,370)
Influencers	25%	75%	100% (2,593)
Ordinary users	33%	67%	100% (1,463)
Average	**34%**	**66%**	**100% (13,426)**

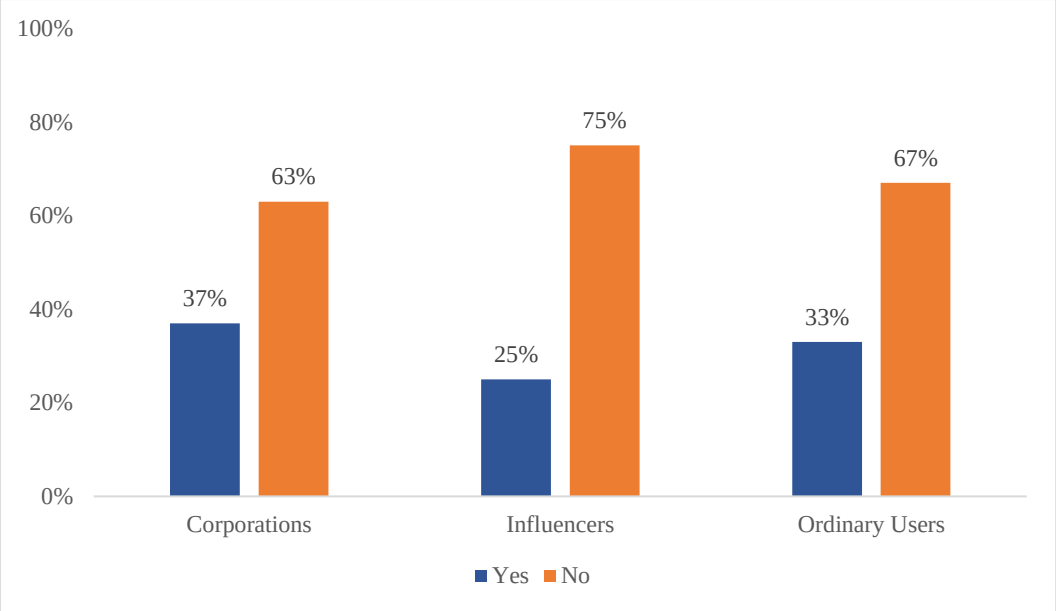


Figure 8.5: Usage of paralinguistic cues by group

Table 8.11: Paralinguistic cues' chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	137.756	$p < 0.001$***
Corporations vs influencers	136.6819	$p < 0.001$***
Corporations vs ordinary users	8.1431	$p < 0.0043$***
Influencers vs ordinary users	33.3965	$p < 0.001$***
ns $p > 0.05$ ** $p \leq 0.05$*	*** $p \leq 0.01$*	**** $p \leq 0.001$*

The chi-square tests reveal statistically significant differences in all cases. Unlike the previous analyses above, the influencers and the ordinary users show a statistically significant difference, which is likely tied to the ordinary users' higher usage of paralinguistic cues.

There may be several reasons for the higher use of the paralinguistic cues by the ordinary users. First, the main audience for ordinary users, as shown in the previous chapter, is family and friends. Therefore, the majority of such communication is personal and informal, and thus using paralinguistic cues supports this through conveying the users' expressions in the tweets. Furthermore, the analysis of the interviews in the previous chapters revealed that the corporations and some of the social media influencers plan their tweets in advance, which may lead to using paralinguistic cues moderately and less frequently than the ordinary users.

In terms of the corporations, Table 8.10 shows that they have the highest rates of paralinguistic cues. The analysis of the interviews with the marketing executives in Chapter

5 indicated that corporations may differentiate between formal (e.g., time and location) and informal paralinguistic cues (e.g., smiling and sad faces). The higher rate of paralinguistic cue usage could be driven by this as one of the main themes in the corporations' tweets is announcing events, which is categorised under tweets for marketing and comprises 89% of the corporations' tweets.

Table 8.12 and Figure 8.6 show the distribution of using emojis among the three groups, which is the most common type of paralinguistic cue. Furthermore, Table 8.13 provides the results of the chi-square tests.

Table 8.12: Overall use of paralinguistic cues by the three groups

		Type of paralinguistic cue		
Group	**None**	**Emoji**	**Other**	**Total% (N)**
Corporations	63%	36%	1%	100% (9,370)
Influencers	75%	23%	2%	100% (2,593)
Ordinary users	67%	31%	2%	100% (1,463)
Average	**66%**	**33%**	**1%**	**100% (13,267)**

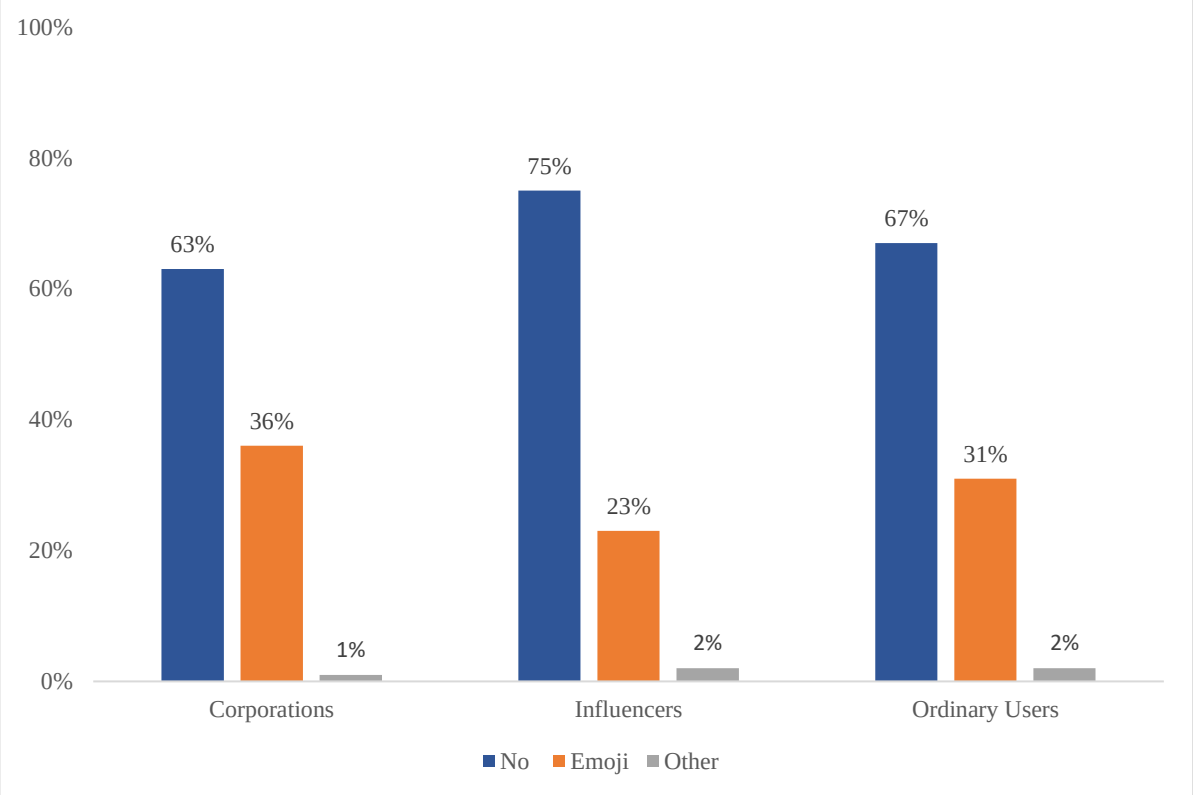


Figure 8.6: Distribution of emojis by group

Table 8.13: Emojis' chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	220.601	$p < 0.001$***
Corporations vs influencers	205.202	$p < 0.001$***
Corporations vs ordinary users	45.712	$p < 0.001$***
Influencers vs ordinary users	61.435	$p < 0.001$***
ns p > 0.05	** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*

The chi-square tests reveal statistically significant differences in all cases. The majority of the interviewees reported utilising emojis sometimes, which echoes the results of the tweets' analysis where none of the groups show a major use of emojis, with 33% of the tweets overall containing an emoji. Furthermore, the marketing executives from the corporations reported that they never use the emoji. This indicates that emoji use is primarily for the personal brands of social media influencers and ordinary users, in addition to corporations that aim to personify their brands and conduct more personalised communication with the audience. On the other hand, other corporations may prefer more formal communication with their audience. Ultimately, this depends on the corporation's marketing strategy.

Investigating the paralinguistic cues is an important aspect for this research as it is part of the diverse affordances available to the users on the various social media platforms, including Twitter. The findings indicate that paralinguistic cues, and especially emojis, are sometimes crucial to communicate the intended meaning to the virtual audience and to imitate face-to-face interactions.

Therefore, the following 'moderate use of paralinguistic cues (including emojis)' pattern is found: Corporations ≠ Ordinary Users ≠ Influencers

8.2.6 Comparison of the tweets' functions

The second and third research questions aim to investigate the communicative functions of the tweets and how the languages and varieties are employed to communicate these functions. Section 8.2.6.1 thus presents a comparison of the functions between the three groups. After that, the subsequent sections compare the different languages and varieties utilised with every main function.

8.2.6.1 Functions' overview

Table 8.14 and Figure 8.7 present an overview of the functions of the tweets that appear in the data: formulaic purposes (FP), culturally specific genres (CSG), reported speech (RS),

dialogical interrelation (DI), marketing (M), and self-expression (SE). Table 8.15 then presents the chi-square test results.

Table 8.14: Overall use of the functions in the tweets

	Main functions						
Group	**FP**	**CSG**	**RS**	**DI**	**M**	**SE**	**Total% (N)**
Corporations	4%	1%	2%	5%	89%	0%	100% (9,370)
Influencers	12%	3%	7%	11%	34%	33%	100% (2,593)
Ordinary users	9%	3%	10%	12%	23%	44%	100% (1,463)
Average	**6%**	**1%**	**4%**	**7%**	**71%**	**11%**	**100% (13,426)**

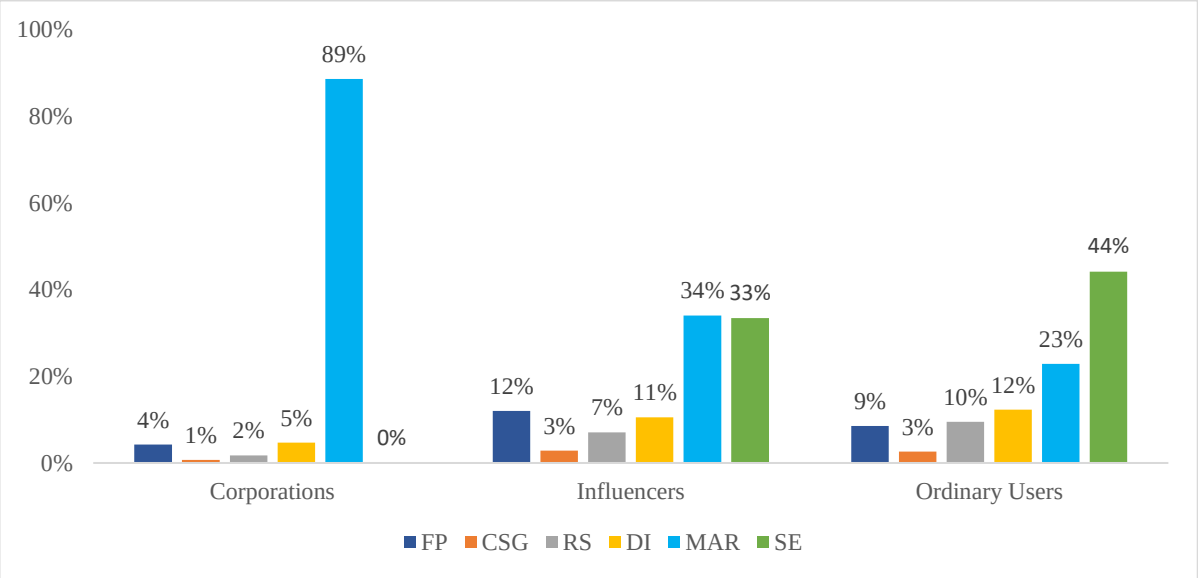


Figure 8.7: Distribution of the main functions by group

Table 8.15 presents the chi-square results for all the functions separately, in addition to the chi-square results with a three way division of marketing, self-expression, and all the other functions combined together (Other).

Table 8.15: Functions' chi-square results of all groups and between groups

Group	Chi-square (All)	P value (All)	Chi-square (M – Other)	P value (M – Other)
All groups	5785.572	$p < 0.001$***	4776.213	$p < 0.001$***
Corporations vs influencers	713.0105	$p < 0.001$***	3400.438	$p < 0.001$***
Corporations vs ordinary users	569.7255	$p < 0.001$***	3391.264	$p < 0.001$***
Influencers vs ordinary users	19.8611	$p < 0.005$ *	55.389	$p < 0.001$***
ns p > 0.05	** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*	

The reason for the additional chi-square test in Table 8.15 is to confirm the validity of the chi-square test for all the groups as formulaic purposes, culturally specific genres, reported speech, and dialogical interrelation are smaller categories than marketing. Furthermore, the self-expression only applies to the social media influencers and the ordinary users. However, as shown in Table 8.15, the significance is almost the same. The results show that there is a statistically significant difference in use across the three groups in the tweets' functions ($p <$ .001). The most common function for the corporations is marketing, while the most common function for the influencers is shared between marketing and self-expression. In terms of the ordinary users, the most common function is self-expression, as seen in Figure 8.7. Aside from marketing and self-expression, there is similar distribution, especially between the influencers and ordinary users, in terms of the functions used in the tweets. The two major differences are in terms of marketing (highest in the corporations) and self-expression (not present in the corporations' tweets). Therefore, the results of the chi-square tests show significant results for the overall as well as the three way distinction chi-square tests. However, it must be noted that the influencers' and the ordinary users' results are more similar than with the corporations. The only difference is that the social media influencers have more marketing tweets than the ordinary users.

Therefore, the following patterns are found:

(i) Dominance of tweets for marketing: Corporations ≠ Influencers/Ordinary Users

(ii) Dominance of tweets for self-expression: Ordinary Users > Influencers (no self-expression for corporations)

8.2.6.2 Formulaic purposes: languages and varieties

The function of formulaic purposes is one of the most employed on Twitter from the data collected for the current study. Table 8.16 and Figure 8.8 show the distribution of the formulaic purposes' tweets among the groups, with the languages and varieties utilised by each group.

Table 8.16: Overall use of languages and varieties in tweets for formulaic purposes

	Formulaic purposes					
Group	**CA**	**Eng**	**Mix**	**MSA**	**Other**	**Total% (N)**
Corporations	11%	5%	9%	63%	13%	100% (397)
Influencers	21%	10%	4%	64%	2%	100% (311)
Ordinary users	6%	14%	3%	77%	-	100% (125)
Average	**14%**	**8%**	**6%**	**65%**	**7%**	**100% (833)**

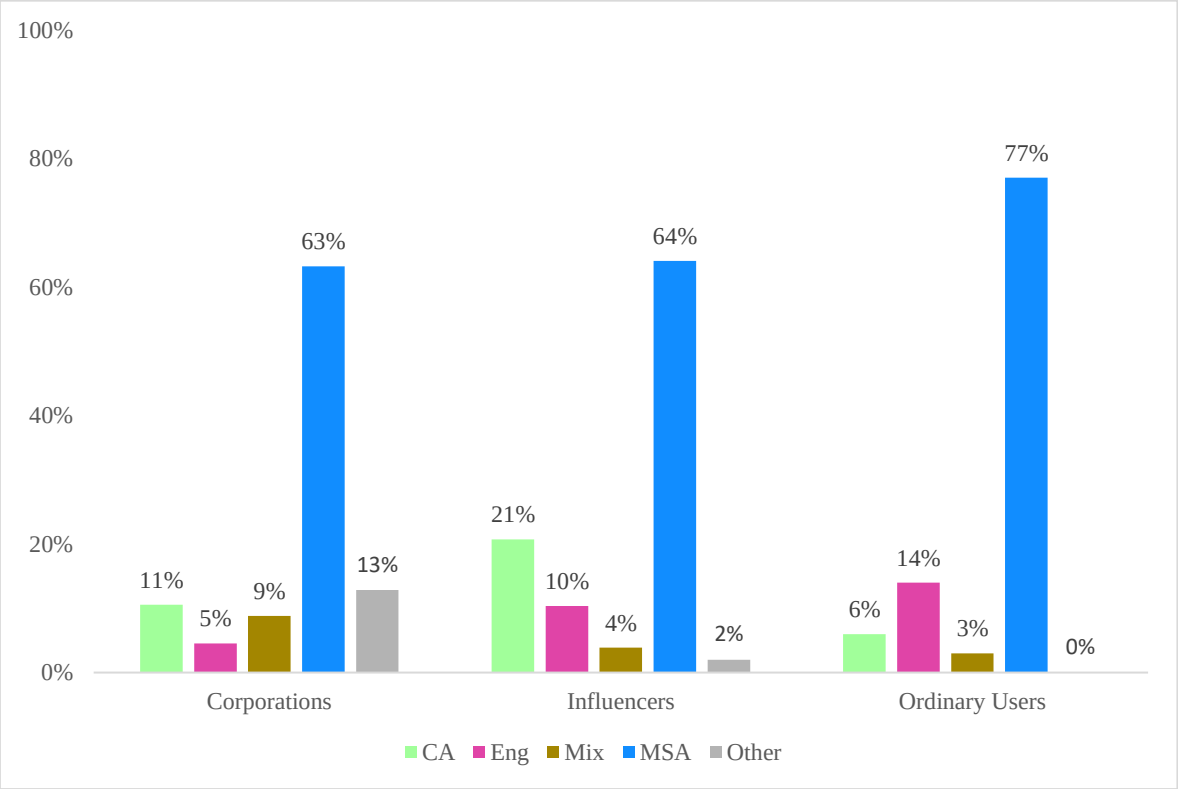


Figure 8.8: Distribution of the languages and varieties used in formulaic purposes tweets by group

The results of the chi-square tests employed to test the relationship between all the groups and between each group are shown in Table 8.17.

Table 8.17: Formulaic purposes' chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	41.771	$p < 0.001$***
Corporations vs influencers	54.1356	$p < 0.001$***
Corporations vs ordinary users	36.27	$p < 0.001$***
Influencers vs ordinary users	16.3465	$p < 0.003$**
ns p > 0.05 ** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*

As can be inferred from Table 8.16 and Figure 8.8, MSA dominates the tweets for formulaic purposes in all groups, followed by CA in the corporations and social media influencers. However, the ordinary users employ more English than CA in the tweets for formulaic purposes. Since formulaic purposes include formal communication, especially for condolences, prayers, and expressing nationalism, it is not surprising that MSA is the most frequently used language. Table 8.17 shows that there is a statistically significant difference across the groups ($p < .001$). Part of this difference may have been driven by the influencers using a higher amount of CA, while the corporations' group use MSA the most frequently (see Figure 8.8).

Therefore, it is found that the three groups use the languages and varieties differently for formulaic purposes, but MSA is the most employed variety for formulaic purposes tweets.

8.2.6.3 Culturally specific genres: languages and varieties

The culturally specific genres' tweets tend to feature least in the data. Table 8.18 and Figure 8.9 present the languages and varieties used by the three groups to communicate this function.

Table 8.18: Overall use of languages and varieties in tweets for culturally specific genres

	Culturally specific genres					
Group	**CA**	**Eng**	**Mix**	**MSA**	**Other**	**Total% (N)**
Corporations	85%	-	1%	12%	1%	100% (67)
Influencers	49%	5%	14%	12%	19%	100% (73)
Ordinary users	53%	13%	8%	5%	21%	100% (38)
Average	**63%**	**5%**	**8%**	**11%**	**13%**	**100% (178)**

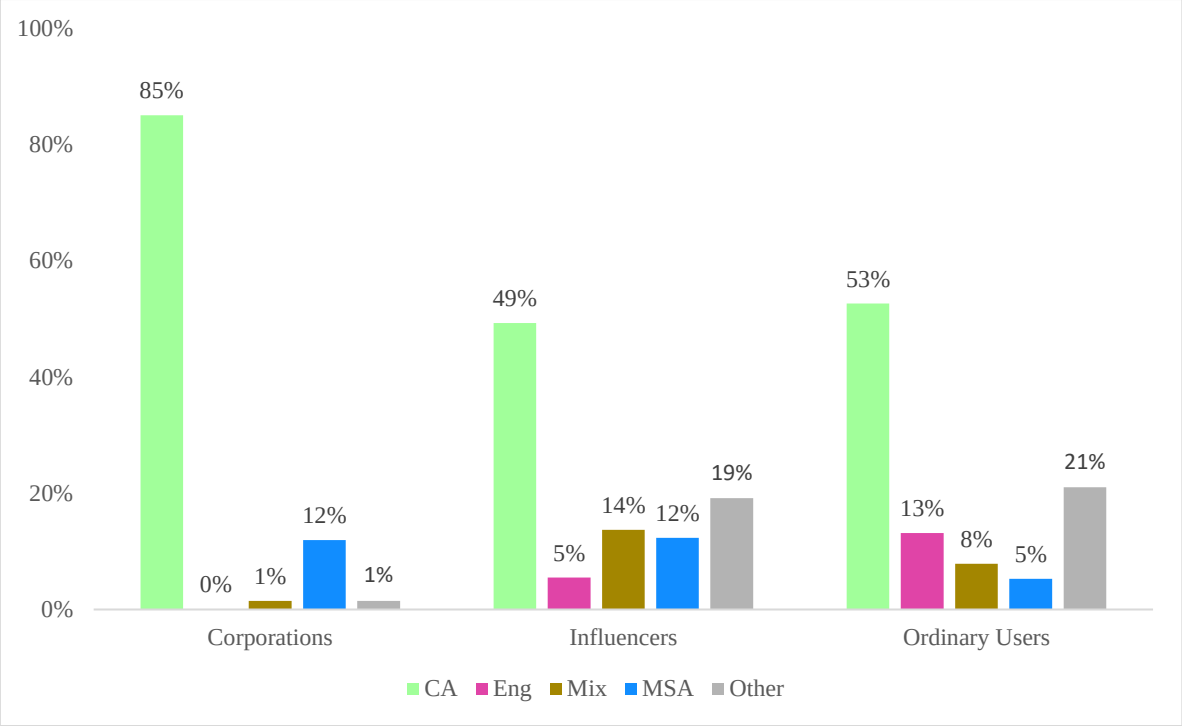


Figure 8.9: Distribution of the languages and varieties employed in culturally specific genres tweets by group

Table 8.19 shows the results of the chi-square tests utilised to test the relationship between all groups and between each group.

Table 8.19: Culturally specific genres' chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	22.419	$p < .001$***
Corporations vs influencers	25.0853	$p < 0.00005$***
Corporations vs ordinary users	23.9154	$p < 0.00008$***
Influencers vs ordinary users	3.8938	$p < 0.4206$ (ns)
ns p > 0.05	** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*

Although Table 8.18 and Figure 8.9 show that all groups mostly use CA, it can be noticed that the social media influencers and the ordinary users have almost similar distributions for CA, MSA, English, Mix, and Other, while the corporations use predominantly CA and MSA without any use of English. It must be noted that the Other category includes mostly Classical Arabic as culturally specific genres also contains tweets with poetry.

Table 8.19 reveals that there is a significant difference between the groups and the language used ($p < .001$). Part of this difference may have been driven by virtually all of the corporations' tweets being in CA, while half of the social media influencers and the ordinary users' language use is distributed between other languages, as discussed. Furthermore, there is a significant difference between the corporations and social media influencers, in addition to between the corporations and ordinary users. Again, the reason for this is that 85% of the corporations' tweets for culturally specific genres are in CA. However, there is no statistically significant difference in the use of the social media influencers and the ordinary users. Although both groups primarily use CA, there are slight differences in the other languages. For example, the ordinary users utilise English (13%) to a greater extent than the influencers (5%), while the influencers have more mixed tweets (14%) in comparison with the ordinary users (8%).

Therefore, it is found that the three groups use the languages and varieties differently for culturally specific genres tweets, although CA is the most employed variety.

8.2.6.4 Reported speech: languages and varieties

The analysis in the previous chapters showed that the corporations, social media influencers, and the ordinary users utilise quotes from different sources in their tweets. Table 8.20 and Figure 8.10 present the proportion of the languages and varieties utilised in the tweets featuring reported speech.

Table 8.20: Use of monolingual tweets for reported speech

	Reported speech					
Group	**CA**	**Eng**	**Mix**	**MSA**	**Other**	**Total% (N)**
Corporations	1%	11%	2%	69%	17%	100% (161)
Influencers	9%	24%	3%	42%	21%	100% (184)
Ordinary users	17%	25%	1%	43%	13%	100% (139)
Average	**9%**	**20%**	**2%**	**51%**	**17%**	**100% (484)**

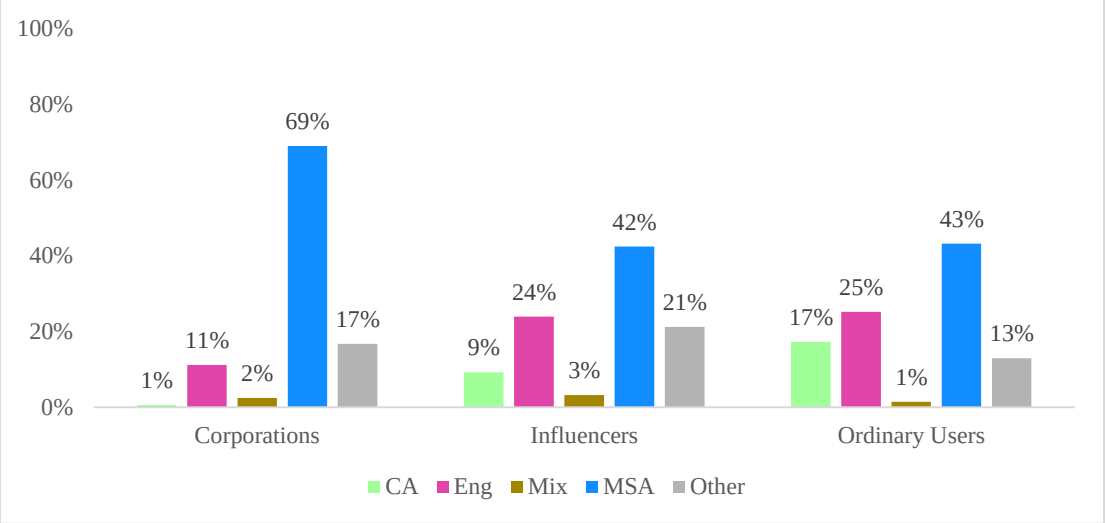


Figure 8.10: Distribution of the languages and varieties used in reported speech tweets by group

Similar to the previous sections, further chi-square tests were run to explore the relationship of all groups and between each group, as shown in Table 8.21. However, the Mix category was grouped with the Other category, so the chi-square result is not affected because all the values for the three groups are less than 5% for the mixed tweets.

Table 8.21: Reported speech chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	51.894	$p < 0.001$***
Corporations vs influencers	32.0785	$p < 0.001$***
Corporations vs ordinary users	42.9071	$p < 0.001$***
Influencers vs ordinary users	8.071	$p < 0.0445$ (ns)
ns $p > 0.05$ ** $p \leq 0.05$*	*** $p \leq 0.01$*	**** $p \leq 0.001$*

For reported speech, Table 8.20 and Figure 8.10 demonstrate that the corporations utilise predominantly MSA, while the influencers and ordinary users use mostly MSA and English. It must be noted that the Other category in reported speech is higher than the other functions because it includes the Quran and Hadith, which are written in Classical Arabic.

Table 8.21 reveals that there is a statistically significant difference between the groups and the language employed ($p < .001$). Furthermore, similar to the previous sections, there is a statistically significant difference between the corporations and the influencers, in addition to the corporations and the ordinary users. Although all of them primarily employ MSA, the corporations utilise less English than the other groups and have a minimal use of CA (1%).

On the other hand, there is no statistically significant difference between the influencers and the ordinary users, as both groups use similar proportions of MSA, CA, English, Mix, and other languages and varieties. It can thus be inferred from the results in this section that the dominant language used for reported speech on Twitter by the different groups is MSA, with many of the reported speech tweets featuring famous quotations that feature in books or the news, and which are chiefly presented in MSA.

Therefore, the three groups have mostly the same distribution of the languages and varieties, with MSA being the most utilised for reported speech tweets.

8.2.6.5 Dialogical interrelation: languages and varieties

One of the main goals for all the three groups on Twitter, as discussed in the previous analysis chapters, is the engagement with the audience. Hence, the dialogical interrelation with the followers is one of the key tools to achieve this goal. Table 8.22 and Figure 8.11 present the languages and varieties used by the three groups for this function.

Table 8.22: Use of monolingual tweets for dialogical interrelation

	Dialogical interrelation					
Group	**CA**	**Eng**	**Mix**	**MSA**	**Other**	**Total% (N)**
Corporations	56%	1%	18%	20%	5%	100% (438)
Influencers	41%	18%	9%	22%	11%	100% (274)
Ordinary users	39%	18%	3%	14%	25%	100% (180)
Average	**48%**	**10%**	**12%**	**20%**	**11%**	**100% (892)**

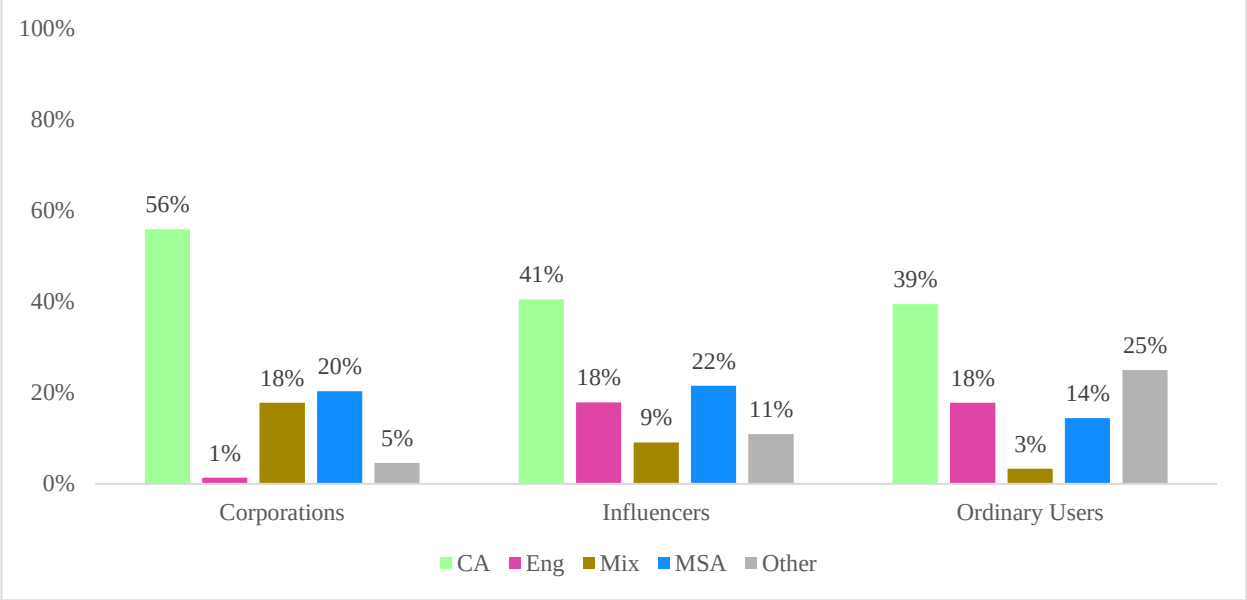


Figure 8.11: Distribution of the languages and varieties employed in dialogical interrelation tweets by group

In addition, Table 8.23 shows the results of the chi-square tests utilised to assess the relationship between all groups and between each group.

Table 8.23: Dialogical interrelation chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	97.587	$p < 0.001$***
Corporations vs influencers	86.2078	$p < 0.001$***
Corporations vs ordinary users	135.3175	$p < 0.001$***
Influencers vs ordinary users	21.2651	$p < 0.0003$ (ns)
ns p > 0.05	** p ≤ 0.05*	*** p ≤ 0.01*	**** p ≤ 0.001*

For dialogical interrelation, Table 8.22 and Figure 8.11 show that the corporations employ a mix of CA, MSA, and Mix tweets; the influencers utilise CA, MSA, and English; and the ordinary users use CA, Other, and English. The dominance of CA in the dialogical interrelation tweets is to be expected, since users tend to imitate face-to-face interactions, and it is prudent to use CA for this communication type with the audience.

Table 8.23 indicates that there is a significant difference between the groups and the language used ($p < .001$). The results of the chi-square tests also show statistically significant

differences among the groups, except for between the influencers and the ordinary users, as they have similar language/variety proportions for the dialogical interrelation tweets.

Therefore, it is inferred that the three groups use the languages and varieties differently for dialogical interrelation tweets, but that CA is the most employed variety, especially in the corporations' tweets.

8.2.6.6 Marketing: languages and varieties

Twitter is an important social media platform, which can also be employed as a search engine for disseminating information and updates about various products, services, or events. Businesses, as well as individuals, engage with the virtual audience on Twitter to increase awareness about their brands and the services they offer. Table 8.24 and Figure 8.12 present the distribution of the languages and varieties among the groups used for the marketing tweets.

Table 8.24: Use of monolingual tweets for marketing

	Marketing					
Group	**CA**	**Eng**	**Mix**	**MSA**	**Other**	**Total% (N)**
Corporations	11%	10%	14%	52%	13%	100% (8,307)
Influencers	18%	25%	7%	49%	1%	100% (883)
Ordinary users	9%	29%	8%	54%	-	100% (335)
Average	**12%**	**12%**	**13%**	**52%**	**11%**	**100% (9,525)**

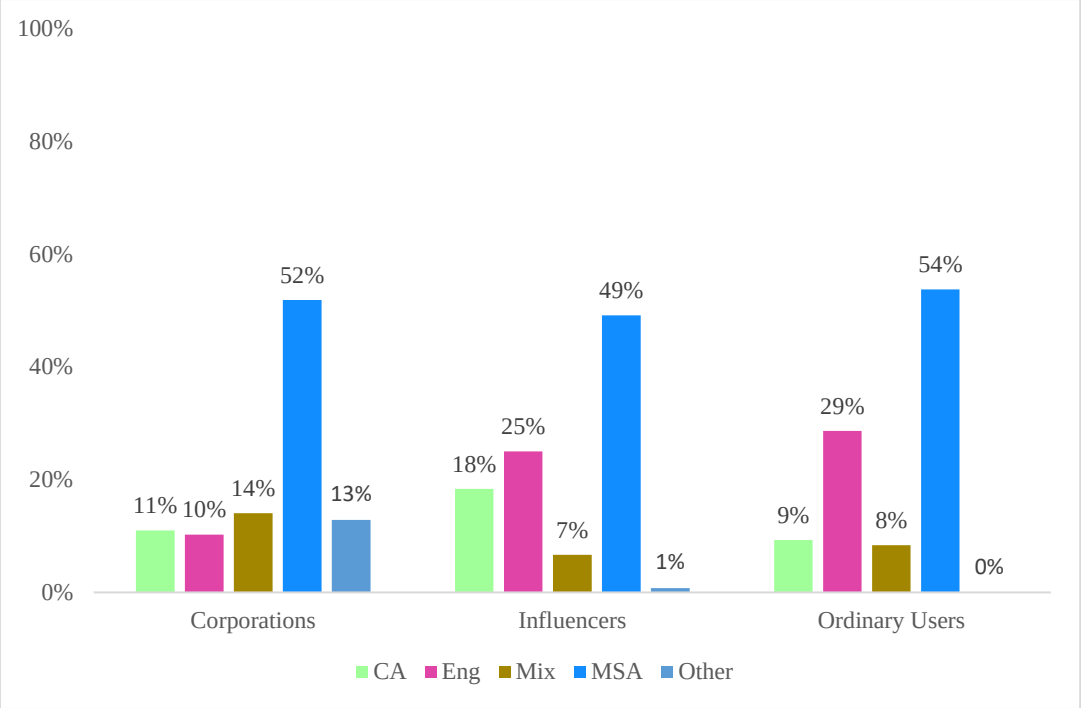


Figure 8.12: Distribution of the languages and varieties utilised in marketing tweets by group

Furthermore, as shown in Table 8.25, chi-square results were obtained to test the relationship between all the groups and between each group.

Table 8.25: Marketing chi-square results of all groups and between groups

Group	Chi-square result	P value
All groups	252.513	$p < 0.001$***
Corporations vs influencers	318.3343	$p < 0.001$***
Corporations vs ordinary users	148.4192	$p < 0.001$***
Influencers vs ordinary users	16.7456	$p < 0.002$**
ns $p > 0.05$ ** $p \leq 0.05$*	*** $p \leq 0.01$*	**** $p \leq 0.001$*

Table 8.24 and Figure 8.12 demonstrate that mainly all the groups utilise MSA for the marketing tweets. However, there are differences in the other languages. For example, 29% of the ordinary users' marketing tweets are in English, compared to 10% for the corporations. However, it must be noted that the number of corporation tweets in marketing (n=8,307) is much higher than that for the ordinary users (n=335). Furthermore, it must be underlined that the number of Mix tweets in the corporations is higher than the social media influencers and the ordinary users, which indicates that English is used as an embedded code within the tweets.

For marketing, there is a statistically significant difference between the groups and the language used ($p < .001$). Although MSA is the dominant language employed by all groups, part of this difference may have been driven by the different rates in which English is utilised between the three groups, as discussed above. The results of the individual chi-square tests show statistically significant differences in all the comparisons, thus highlighting differences in the languages and varieties employed by the groups for marketing. However, MSA again dominates the tweets for marketing by all groups. However, the three groups use the languages and varieties differently for the marketing tweets.

8.2.6.7 Self-expression: languages and varieties

Twitter provides the opportunity for the users to express their thoughts and feelings to the followers and the wider audience. As corporate accounts focus on products and services, this function was only found in the social media influencers' and the ordinary users' tweets. Table 8.26 and Figure 8.13 displays the distribution of the languages and varieties among the groups used for self-expression.

Table 8.26: Use of monolingual tweets for self-expression

	Self-expression					
Group	**CA**	**Eng**	**Mix**	**MSA**	**Other**	**Total% (N)**
Influencers	31%	22%	10%	37%	1%	100% (868)
Ordinary users	38%	22%	8%	32%	1%	100% (646)
Average	**34%**	**22%**	**9%**	**35%**	**1%**	**100% (1,514)**

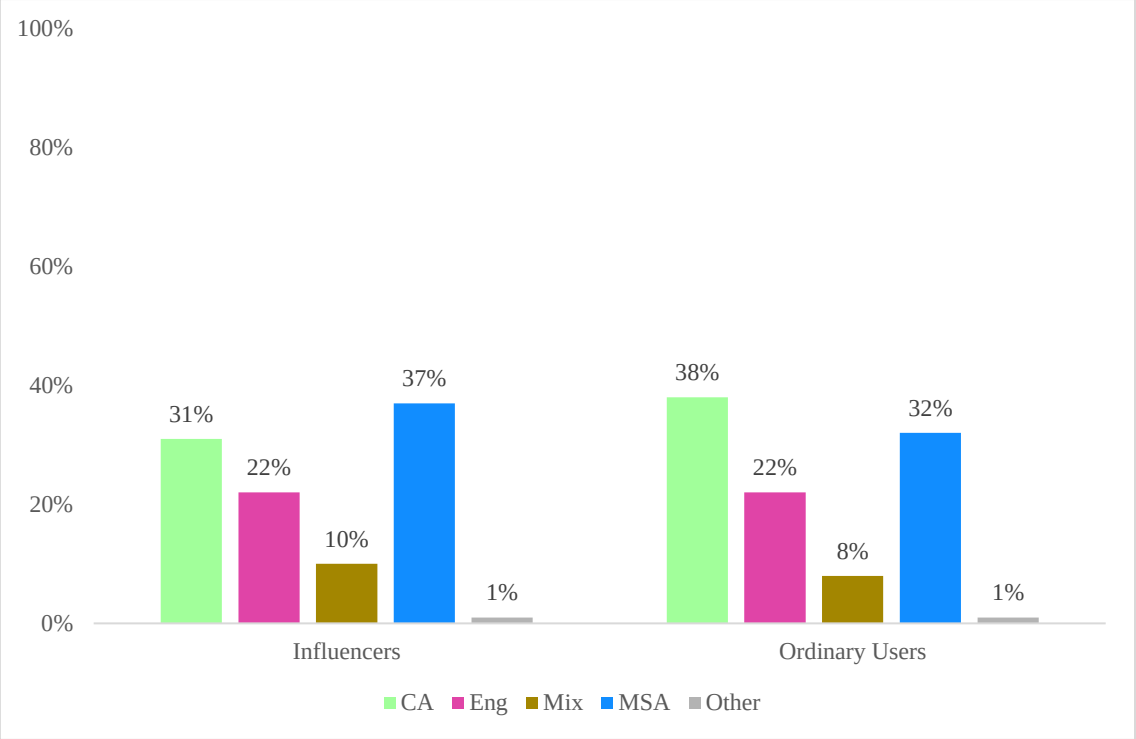


Figure 8.13: Distribution of the languages and varieties utilised in self-expression tweets by group

For self-expression purposes, there are differences that are shown in Table 8.26 and Figure 8.13, such as the influencers being more likely to use MSA than the ordinary users, while the ordinary users are more likely to use CA than the influencers. On the other hand, both groups use English in exactly the same proportion of the tweets, while having similar proportions for the mixed tweets. As for the result of the chi square test, it shows that there is a statistically insignificant difference in the language employed by the social media influencers and the ordinary users ($p < .0432$). Therefore, for the function of self-expression, the social media influencers and the ordinary users have similar distribution of the languages and varieties, and most of the tweets are either in MSA or CA.

8.3 Conclusion

This section summarises the main findings of Chapter 8. The comparison of the individual groups revealed statistically significant differences between the data from the corporations versus the data extracted from the influencers and the ordinary users. On the other hand, the chi-square results revealed statistically insignificant differences and similarities between the social media influencers and the ordinary users, except for the use of paralinguistic cues such as emojis. The following is a summary of the key findings:

- The corporations incorporate more intra-tweet code-mixing than the social media influencers and the ordinary users.
- MSA dominates the monolingual tweets and the matrix code, particularly in the corporations' tweets.
- English and English with Arabic script dominate the embedded codes, and appear in the corporations' tweets more than the influencers' and ordinary users' tweets. However, it must be highlighted that English in monolingual tweets appears more in influencers and ordinary users' tweets.
- There is moderate use of paralinguistic cues, with the corporations and the ordinary users featuring more inclusions than the social media influencers.

The findings confirm that the communication strategy of the corporations differs significantly to that of the influencers and the ordinary users. On the other hand, the influencers and ordinary users tend to share the same language features on Twitter.

Another important finding emerging from this chapter concerns the functions and the language employed with each function, and how the different groups employ the different languages and varieties to communicate these functions. The following is a summary of the findings regarding the functions:

- MSA is the most used variety for formulaic purposes, with the ordinary users employing it to a greater extent than the corporations and the social media influencers.
- CA is the most utilised variety for the culturally specific genres, used by the corporations to a greater extent than the social media influencers and the ordinary users.
- MSA is the most employed variety for reported speech, with the corporations utilising it more than the social media influencers and the ordinary users.
- CA is the most used variety for the dialogical interrelation, with the corporations using it more than the social media influencers and the ordinary users.
- MSA is the most utilised variety for marketing, with the corporations employing it to a greater degree than the social media influencers and the ordinary users.
- MSA and CA are primarily employed for self-expression, with the social media influencers employing slightly more MSA than the ordinary users, and the ordinary users using marginally more CA than the social media influencers.

In terms of the significance of the differences between the groups, the statistical analysis shows a statically significant difference between the corporations and the other two groups, while there is no statistically significant difference between the social media influencers and the ordinary users, except for the language used for the formulaic purposes and marketing tweets, where in both functions the social media influencers utilise more CA, as opposed to the ordinary users who employ MSA to a greater extent. This indicates that the language choice for the users is not random, especially for the corporations, and they consciously think about how they want to be presented to the virtual audience on Twitter, which eventually lead to branding and self-branding.

In sections 8.2.3 and 8.2.5 of this chapter, the results of the qualitative responses from the interviews were highlighted where relevant, which confirm the preference for MSA use by the users, followed by English and CA. Furthermore, the interviews confirmed the moderate use of emojis and other paralinguistic cues, as most of the interviewees said that they use emojis sometimes.

In the next chapter, a concluding discussion is presented to draw this thesis to a close.

Chapter 9:
Conclusion

9.1 Introduction

This thesis explored the linguistic features of three different groups of Twitter users in Saudi Arabia, and how they utilise their communicative repertoires as a tool for branding and self-branding. The analysis examined 13,426 tweets collected from the public Twitter accounts of three distinct groups of users—corporations, social media influencers, and ordinary users—in addition to conducting interviews with a sample drawn from each user group, to investigate their linguistic behaviour on Twitter. These interviews clarified some of the findings resulting from the tweets' analysis to explain the linguistic phenomena observed by Saudi users on Twitter.

The thesis also aimed to better understand the communicative functions of the tweets. This is because it was expected that knowing the purpose of a tweet can help to determine a link between its function and the language selected for its communication. This is particularly important on social media platforms such as Twitter, where users can edit and modify their tweets prior to posting them. It was thus important to explore these functions as it affects the users' choice of language.

This chapter starts with the limitations in section 9.2. Then, section 9.3 presents a summary of the research findings, with the implication discussed in section 9.4. After that, the opportunities for future research are presented in section 9.5. Finally, this chapter ends with closing thoughts in section 9.6.

9.2 Limitations

There are certain limitations to this study that need to be acknowledged. First, the study focused on the linguistic behaviours of the users on Twitter. Hence, the findings are limited to Twitter and may not be applicable to other social media platforms. Furthermore, the study investigated the linguistic variation on Twitter in Saudi Arabia. Therefore, the findings may not be generalisable to other Arab countries. Finally, the results in this study are limited to the time period of the data collection, which was between November 2017 and May 2018 for the tweets, and between April and May 2019 for the interviews. Nevertheless, it must be noted that some of the findings can be extrapolated further and that similar results are likely to be found on other platforms and other time periods.

9.3 The key research findings

In this section, the findings are presented to respond the research questions presented in section 1.2. The discussion below reflects on the theoretical contributions of the research, based on the answers to the research questions.

RQ1) What are the linguistic characteristics of Saudi users' tweets? What languages and varieties are used?

The current research initially expected to find a relatively large number of tweets utilising code-mixing within the tweet, namely, intra-tweet code-mixing. However, the analysis of the tweets revealed that only 12% featured intra-tweet code-mixing. According to Androutsopoulos (2013), there are online spaces that are considered multilingual discourse spaces where different codes coexist. This description applies to the discourse of Twitter, and it was apparent from the data that the majority of the code-mixing occurred across the tweets and formed inter-tweet code-mixing on the users' timelines.

The main finding after exploring the different codes utilised by the Saudi Twitter users is that the users from the three different groups primarily utilise MSA, English, and CA in varying proportions. The analysis also revealed that MSA dominated the Saudi users' tweets, comprising 48% of the monolingual tweets and 65% as the matrix code in the tweets. This was further emphasised in the interviews, with the majority of the interviewees' responses in favour of utilising MSA.

The data analysis also revealed the emergence of new modes of communication in Arabic on Twitter. First, the data showed cases of reversed Romanisation to create English word forms using Arabic characters, which echoes the phenomenon found in Greek online settings reported by Spilioti (2019). In addition, in the interviews the marketing executives and the social media influencers discussed the white dialect and indicated that they primarily used MSA or this dialect, which the interviewees described as an MSA–CA hybrid.

The analysis showed a relatively high use of English, especially as an embedded code. The data revealed that 13% of the monolingual tweets were in English. Additionally, English appeared as a matrix code in 5% of the tweets and in 59% as an embedded code. Studies such as Eldin (2014), Kosoff (2014), and Al Othman (2016) reported the dominance of English in Arab social media. However, the results of this study are not in agreement with other research that does not report any dominance of English among Arab users (cf. Albirini, 2016; Alhejely, 2020). When the interviewees were asked about English usage, they highlighted that the language was employed depending on the content of the tweet and the

intended audience. Furthermore, the analysis of the tweets highlighted that the utilisation of English and Arabic for parallel text bilingualism tweets was a common strategy among the corporations, in which they displayed the same content in English and Arabic either in a single or separate tweets. The corporations used the strategy of parallel text bilingualism in 36% of the accounts, where 6% used CA as the parallel text as opposed to MSA. Furthermore, some corporations had separate accounts for English and Arabic. However, this practice was less common in the tweets by the social media influencers and the ordinary users. The main purpose of this strategy is enabling the users to reach a broader audience nationally and internationally. Moreover, as English is the first linguistic choice of global advertisers and marketers (Bhatia & Richie, 2006), the presence of English on the Twitter page and tweets helps to convey the international qualities of the corporation or the user. The analysis also indicated that the strategy of parallel text bilingualism was growing among the users. However, when the interviewees were asked about the importance of translating the content of the tweets, few of the marketing executives, influencers, or ordinary users agreed on the efficiency of this strategy. Nevertheless, it was found in Chapter 6 that some of the influencers employed parallel text bilingualism, where one even started a separate account for English content only, as per the corporations, but again this was only one case.

RQ2) What are the functions of the languages and varieties utilised on Twitter by Saudi corporations, social media influencers, and ordinary users?

The functions of the linguistic online resources analysed in this study were categorised into six main functions: formulaic purposes, culturally specific genres, reported speech, dialogical interrelation, marketing, and self-expression. These functions resemble those identified in previous studies by Androutsopoulos (2013) and Halim and Maros (2014). However, the functions of marketing and self-expression were added to the current research to reflect what was found in the current study. The overall analysis of the tweets showed that the communicative functions most employed in the tweets were marketing, self-expression and formulaic purposes. Furthermore, the analysis revealed that the majority of the tweets, particularly for corporations, were intended for marketing purposes such as information, advertisement, corporate social responsibility, awareness campaigns, recruitment and testimonials. Furthermore, the social media influencers and ordinary users primarily posted tweets to express their ideas or impressions regarding incidents that they experienced personally, or that unfolded more generally in the society. The third most common function appearing in the data was formulaic purposes, which included tweets for greetings, salutations, condolences, nationalism and prayers. The percentages of the other categories (i.e., culturally specific genres, reported speech and dialogical interrelation) were

fairly distributed among the three groups. This frequency matched the results of the interviews where the marketing, self-expression and formulaic purposes functions were related to the aims of awareness and engagement, the most recurrent themes from the interviews. It must be noted that the aim of awareness for the corporations was primarily concerned with the products or services they offered. On the other hand, awareness for the social media influencers and the ordinary users was primarily associated with societal issues or certain current affairs in national or international settings. In terms of engagement, there were different strategies allowing the users to engage with the audience such as tweeting about different occasions for formulaic purposes, commenting live from events underway, and establishing conversations with the audience such as posing questions to seek opinions or to gather information. The analysis of the functions of the tweets also revealed that some functions are more associated with certain codes. For example, formulaic purposes and reported speech are mostly in MSA while culturally specific genres and dialogical interrelation are mostly in CA. These functions and the languages used with each function are further discussed in the response to the third research question.

RQ3) How are the different groups using the different language choices in relation to the functions?

The third research question helps to shine a light on what occurred to the functions of multilingualism, as identified in this study, in the discourse of digital communication. The answer to this question highlights a general trend regarding the association of the languages and functions, although some cases did not follow the conventional trend, particularly in the varieties of Arabic. For example, there were cases of employing MSA for informal functions and the utilisation of CA for informal tweets. Hence, using and blending different codes of language are employed to serve different functions, which are dependent on the author's intention, as well as the audience's understanding and acceptance of such codes. These issues are related to the branding and self-branding online, which are discussed in the implications section.

RQ4) What are the patterns of communication employed by the different groups for branding and self-branding?

The fourth research question explores the common language patterns appearing in the data. As discussed in the answer to the first research question, parallel text bilingualism was found to be a common pattern among corporations. Furthermore, 18% of the corporations employed MSA/CA/English/MIX in their tweets, while another 18% utilised a mix of MSA and CA. The analysis also showed that neither the influencers nor the ordinary

users employed the parallel text bilingualism strategy. For both groups, the majority of the accounts used bilingualism and diglossic switching (60% for the influencers and 70% for the ordinary users). However, there was variation in the accounts of the social media influencers and the ordinary users in terms of diglossic switching and bilingualism at 20% and 10%, respectively, for the influencers, and at 10% and 20%, respectively, for the ordinary users.

The research also reveals different patterns for the various linguistic codes investigated in this study, namely code-mixing, languages in monolingual tweets, languages in matrix codes, languages as embedded codes, and paralinguistic cues. The statistical results in Chapter 8 indicated that the corporations were more likely to engage in intra-tweet code-mixing to a greater extent than the influencers and the ordinary users. As discussed above, the analysis of the tweets revealed the dominance of MSA in monolingual tweets with less usage of English and CA. The statistical analysis again showed that the corporations used MSA to a greater extent in their monolingual tweets, in addition to matrix codes in the multilingual tweets, than the influencers and ordinary users. On the other hand, the analysis revealed the greater dominance of English and English in Arabic script as embedded codes in the tweets by the corporations than those by the influencers and ordinary users. In terms of the use of paralinguistic cues, moderate usage that included emojis was shown. It was also revealed that the corporations and ordinary users utilised more paralinguistic cues than the social media influencers.

9.4 Implications

Although it is to be expected that MSA would dominate the Twitter accounts of an Arabic-speaking country, the fact that it is still preferred despite all the other language affordances is an interesting finding. In fact, one of the interviewed marketing executives highlighted the commitment of their corporation to publishing grammatically accurate tweets in MSA, which were revised by editors prior to posting. It should also be highlighted that the use of MSA guarantees reaching out to a broader audience, as it is understood by all Arabic speakers, while local varieties may not be easily comprehended by all the audience members with competency in Arabic (Albirini, 2016).

In terms of CA, despite being less frequently employed for written language in offline contexts, it remains an important variety that is employed in digital writing on Twitter. CA is perceived in the literature as the L variety of Arabic (Ferguson, 1959), and it was regarded in the literature as both a corrupt form of Arabic and the language of Arab illiterates (Hoigilt, 2018). In fact, Albirini (2016) argues that the educational level of users on Facebook

determines their use of CA, suggesting that educated users rarely utilise this colloquial form of Arabic. This research, however, finds that CA is an important variety that serves different communicative functions on Twitter, such as assisting users in constructing a cordial identity that might be more challenging to achieve through employing MSA. One of the ordinary user interviewees, for example, reported that she deliberately used Egyptian CA for jokes to establish a humorous identity. In terms of corporate communication, it was also used to convey a friendly persona to the audience. Also, the research has revealed that multilingualism online and the attitudes of the users are rapidly changing. For example, previous studies documented the dominance of Arabizi, while this study reports the fading of this variety.

Digital writing involves writing oneself into being (boyd, 2008). The analysis of the interviews indicated that language choice primarily depends on the target audience, in addition to the content of the tweet. In support of this view, one of the social media influencer interviewees highlighted that she sometimes used English to address a certain group within her audience who were expected to have English competency, while excluding another group without the expected proficiency to decode such tweets. These findings are in line with Seargeant and Tagg (2014), who argue that language choice is a strategy of audience design in social networking sites. The corporations, social media influencers, and ordinary users all had different follower types to whom they needed to present themselves in different ways, thus affecting their language choice. This is what Marwick and boyd (2011) refer to as context collapse. The target audience of the different groups of Twitter users thus represents an important factor that is carefully considered to determine the optimum language or variety to be employed on Twitter in reaching out to them. It must be highlighted that branding and self-branding could be challenging, and especially for anonymous and varied audience (boyd, 2007), as is typically the case of corporations and social media influencers. However, the ordinary users attempted to make generalisations about their virtual audience, and it was agreed by the majority of the interviewees that the audience influences the choice of language in which they decide to tweet. This implies that the users modify their language towards the predicted variety of their audience in order to gain approval (Bassiouney, 2017). Furthermore, one of the important implications of this study stems from the conscious selection of a certain language or variety, since this can facilitate in revealing important aspects about the attitudes towards these languages and varieties.

The current research also confirms the close relationship between language practices online and identity construction which leads to branding and self-branding. Identities in social media do not merely involve who we are, but also entail how we seek to be perceived by others, as well as how others expect us to present (Lee, 2016). Furthermore, when users

post on Twitter they engage in a process of self-promotion and branding (Page, 2013). As discussed above in section 9.3, the analysis of the tweets showed the dominance of MSA across all three groups analysed in this study. The perception that MSA is associated with education and sophistication (Albirini, 2016) implies that the selection of this variety to communicate with the audience on Twitter positions the user as a prestigious individual or brand. Furthermore, the use of CA conveys a sense of congeniality and facilitates informality with the target audience (Al Alaslaa, 2018; Alhejely, 2020). In terms of English, language choice online is a strategy for expressing a person's multilingual identity (Sinatora, 2019), with social media enabling users to assert new forms of local and global identities by employing language (Lee, 2016). As the data collected in the current study have shown, corporations, social media influencers, and ordinary users in Saudi Arabia employ English to project an international identity, with English appearing in all the functions, thus implying that the use of this language is not merely limited to certain contexts. Moreover, English with Arabic script provided the opportunity for the users to present themselves in a new multilingual manner. Finally, the use of the white dialect, which was only recognised by the corporations and the social media influencers, indicates that it is used as a tool for branding such users to a wider audience as it excludes any regional markers.

9.5 Recommendations for future research

The study's focus on Twitter was due to its status as one of the most suitable social media platforms to investigate branding and self-branding, since all accounts are public and not limited to certain followers or groups. However, future research could explore the linguistic behaviours on other social media platforms and compare the results with those reported in this study. Furthermore, additional research could employ the same methodology to report on the linguistic variation in other Arab countries. In addition, the current study relied primarily on the quantitative analysis of the tweets, with the interviews featuring a small sample of 15 participants to supplement the analysis of the tweets with qualitative responses. Therefore, additional research could focus on interviewing a larger sample and making the interviews the primary focus of the study through a purely qualitative study to investigate in more detail the attitudes of the users towards the different languages and varieties used online.

The findings of the thesis highlight issues that merit further investigation; for example, the need to explore parallel text bilingualism in greater depth and to investigate the context of this linguistic phenomenon online. Moreover, an investigation of the use of the white dialect could be explored to establish the general characteristics of this variety, with few mentions of the white dialect in the literature. Furthermore, additional research is recommended to

investigate the phenomenon of trans-scripting and the linguistic features of utilising English with Arabic text. Moreover, the focus of the current study was on the users who write the tweets, and it was highlighted in the implications that the target audience is an important factor that determines language choice. Hence, further research could focus on the audience as the unit of analysis to investigate their interactions with the tweets in different languages and varieties.

Another direction for future research in line with the framework of the current study is collecting tweets from one of the groups and focusing on the analysis of this specific group. Moreover, other groups of users on Twitter could be explored such as academics and government officials, while certain linguistic features from the tweets could also be further investigated such as the use of hashtags along with the languages used for this mechanism, and whether they are bilingual or restricted to a single language. Finally, there is further scope for the context of paralinguistic cues, primarily emojis, to be further studied.

9.6 Closing words

The current research originated from a desire to explore what is happening in terms of language practices on Twitter, mainly for marketing purposes. The increasing number of Internet users means that corporations and social media influencers build creative social media strategies and interactive posts to engage with their followers and encourage them to share the content related to them and their brand. The findings of this study can thus be used as a reference source, which can be consulted for language use to start or maintain a linguistically dynamic and engaging account with the virtual audience on Twitter.

In addition, the study has contributed to the field of digital communication by constructing a general picture of the different multilingual codes and how they are associated with the communicative functions of the tweets as used, and compared, by three different groups of users, that is, corporations, social media influencers, and ordinary users. Generalisations on the different language patterns were formed from the data and analyses, with some of the uses already established in the field and providing further evidence for discussion, while other findings are novel and not supported by the existing literature. The thesis has also served to further demonstrate how the choice of language contributes to the branding of the corporations and the self-branding of the social media influencers and ordinary users. Finally, another important advantage of this study is in documenting the language scene in Saudi Arabia through the lens of social media.

References

Abdulaziz, M. H. (1986). Factors in the development of modern Arabic usage. *International Journal of the Sociology of Language*, *62*, 11-12.

Al-Huri, I. (2015). Arabic language: historic and sociolinguistic characteristics. *English Literature and Language Review*, *1*(4), 28-36.

Al-Khatib, M., & Sabbah, E. H. (2008). Language choice in mobile text messages among Jordanian university students. *SKY Journal of Linguistics*, *21*(1), 37-65.

Al-Rawi, M. (2012). Four grammatical features of Saudi English: Charting the influence of Arabic on the syntax of English in Saudi Arabia. *English Today*, *28*(2), 32-38.

Al Alaslaa, S. (2018). *A sociolinguistic study of code choice among Saudis on Twitter*. PhD thesis. University of Michigan.

Albirini, A. (2016). *Modern Arabic sociolinguistics: diglossia, variation, codeswitching, attitudes and identity*. London; New York, NY: Routledge.

Alfaifi, S. H. (2013). *Code-switching among bilingual Saudis on Facebook*. Southern Illinois University at Carbondale.

Alhejely, G. W. (2020). *Online Identities and Linguistic Practices: A case of Arab Study Abroad Students in the UK on Twitter.* PhD thesis. Manchester Metropolitan University.

Alismail, S., & Zhang, H. (2018). The use of emoji in electronic user experience questionnaire: An exploratory case study. *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hilton Waikoloa Village, Hawaii.

Alothman, E. (2012). *Digital vernaculars: An investigation of Najdi Arabic in multilingual synchronous computer-mediated communication*. The University of Manchester.

American Marketing Association. (2022). *Definition of marketing.* Retrived from: https://www.ama.org/the-definition-of-marketing-what-is-marketing/

Androutsopoulos, J. (2006a). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, *10*(4), 419-438.

Androutsopoulos, J. (2006b). Multilingualism, diaspora, and the Internet: Codes and identities on German-based diaspora websites. *Journal of Sociolinguistics*, *10*(4), 520-547.

Androutsopoulos, J. (2007). Language choice and code-switching in German-based diasporic web forums. *The Multilingual Internet: Language, Culture, and Communication Online*, 340-361.

Androutsopoulos, J. (2011). From variation to heteroglossia in the study of computer-mediated discourse. *Digital Discourse: Language in the New Media*, 277-298.

Androutsopoulos, J. (2013a). Code-switching in computer-mediated communication. *Pragmatics of Computer-Mediated Communication*, 667-694.

Androutsopoulos, J. (2013b). Participatory culture and metalinguistic discourse: Performing and negotiating German dialects on YouTube. *Discourse*, *2*, 47-71.

Androutsopoulos, J. (2014). Moments of sharing: Entextualization and linguistic repertoires in social networking. *Journal of Pragmatics*, *73*, 4-18. https://doi.org/10.1016/j.pragma.2014.07.013

Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, *19*(2), 185-205.

Androutsopoulos, J. (2020). Trans-scripting as a multilingual practice: the case of Hellenised English. *International Journal of Multilingualism*, *17*(3), 286-308.

Androutsopoulos, J. K., & Georgakopoulou, A. (2003). *Discourse constructions of youth identities* (Vol. 110). John Benjamins Publishing.

Antaki, C., & Widdicombe, S. (1998). *Identities in talk*. Sage Publications.

Aronin, L., & Hufeisen, B. (2009). On the genesis and development of L3 research, multilingualism and multiple language acquisition. *The Exploration of Multilingualism*, 1-9.

Aronin, L., & Singleton, D. (2012). Affordances theory in multilingualism studies. *Studies in Second Language Learning and Teaching*, *3*(2), 311-331.

Arsenian, S. (1945). Bilingualism in the post-war world. *Psychological Bulletin*, *42*(2), 65-86. https://doi.org/10.1037/h0061716

Auer, P. (1995). The pragmatics of code-switching: A sequential approach. *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*, 115-135.

Auer, P. (1999). From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism*, *3*(4), 309-332.

Auger, G. A. (2014). Rhetorical framing: Examining the message structure of nonprofit organizations on Twitter. *International Journal of Nonprofit and Voluntary Sector Marketing*, *19*(4), 239-249.

Bali, K., Sharma, J., Choudhury, M., & Vyas, Y. (2014). "I am borrowing ya mixing?" An Analysis of English-Hindi Code Mixing in Facebook. *Proceedings of the First Workshop on Computational Approaches to Code Switching*. Doha, Qatar.

Barton, D., & Lee, C. (2013). *Language Online: Investigating Digital Texts and Practices*. Taylor & Francis Group.

Bassiouney, R. (2020) *Arabic Sociolinguistics*. 2nd ed. Edinburgh University Press.

Bassiouney, R. (2017). *Functions of code switching in Egypt: Evidence from monologues*. Brill.

Bhatt, R. M., & Bolonyai, A. (2019). Code-switching and translanguaging. In: *Handbook of Pragmatics: 22nd Annual Installment* (pp. 59-78). John Benjamins Publishing Company.

Blommaert, J., & Rampton, B. (2012). Language and superdiversity. *MMG Working Paper*, *(12-09)*.

boyd, D. (2007). Social network sites: Public, private, or what. *Knowledge Tree*, *13*(1), 1-7.

Braun, V., & Clarke, V. (2012). *Thematic analysis*. American Psychological Association.

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Campbell, C., & Farrell, J. R. (2020). More than meets the eye: The functional components underlying influencer marketing. *Business Horizons*, *63*(4), 469-479.

Canagarajah, A. S. (2011). Afterword: World Englishes as code-meshing. *Code-meshing as World English: Pedagogy, Policy, Performance*, 273-281.

Canagarajah, A. S. (2013). *Literacy as translingual practice: Between communities and classrooms*. Routledge.

Canagarajah, S. (2015). Clarifying the relationship between translingual practice and L2 writing: Addressing learner identities. *Applied Linguistics Review*, *6*(4), 415-440.

Carnevale, M., Luna, D., & Lerman, D. (2017). Brand linguistics: A theory-driven framework for the study of language in branding. *International Journal of Research in Marketing*, *34*(2), 572-591.

Cenoz, J. (2009). *Towards multilingual education: Basque educational research from an international perspective* (Vol. 72). Multilingual Matters.

Chen, S. Y. R. (2007). *Code switching between English and Mandarin Chinese on postings in the college affiliated bulletin board system in Taiwan: A functional approach*. Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching,

Collot, M., & Belmore, N. (1996). A new variety of English. *Computer-mediated Communication: Linguistic, Social, and Cross-cultural Perspectives*, *39*, 13.

Coupland, N. (2012). Bilingualism on display: The framing of Welsh and English in Welsh public spaces. *Language in Society*, *41*(1), 1-27.

Creese, A., & Blackledge, A. (2015). Translanguaging and identity in educational settings. *Annual Review of Applied Linguistics*, *35*, 20-35.

Creswell, J. (2014). *Research design: qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Crystal, D. (2003). *English as a global language*. Cambridge University Press.

Crystal, D. (2004). *A glossary of netspeak and textspeak*. Edinburgh: Edinburgh University Press.

Crystal, D. (2006). *Language and the Internet* (2nd ed.). Cambridge; New York, NY: Cambridge University Press.

Cunliffe, D., & Herring, S. C. (2005). Introduction to minority languages, multimedia and the web. *New Review of Hypermedia and Multimedia*, *11*(2), 131-137.

Danesi, M. (2017). Emoji in advertising. *International Journal of Semiotics and Visual Rhetoric*, *1*(2), 1-12.

Danet, B., & Herring, S. C. (2007). *The multilingual Internet: language, culture, and communication online*. Oxford; New York, NY: Oxford University Press.

Davis, R. (2014). *Social media branding for small business: The 5–sources model*. Business Expert Press.

de Fina, A. E., Schiffrin, D. E., & Bamberg, M. E. (2006). *Discourse and identity*. Cambridge University Press.

Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage Handbook of Qualitative Research*. Sage Publications.

Deumert, A., & Masinyana, S. O. (2008). Mobile language choices—The use of English and isiXhosa in text messages (SMS): Evidence from a bilingual South African sample. *English World-Wide*, *29*(2), 117-147.

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development*, *25*(2-3), 204-222.

Dhanik, T. (2016). Micro, not macro: Rethinking influencer marketing. *AdAge*. December 12. Retrieved 10 August 2022 from: https://adage.com/article/digitalnext/micro-macro-influencer-marketing-kim-kardashian/307118.

Dorleijn, M. (2016). Introduction: Using multilingual written internet data in code-switching and language contact research. *Journal of Language Contact*, *9*(1), 5-22. https://doi.org/10.1163/19552629-00901001

Durham, M. (2007). Language choice on a Swiss mailing list. In: B. Danet & S. Herring (eds.), *The multilingual Internet: language, culture, and communication online*. Oxford; New York, NY: Oxford University Press.

Durham, M. (2014). *The Acquisition of Sociolinguistic Competence in a Lingua Franca Context*. Multilingual Matters.

Durham, M. (2016). English as a lingua Franca: forms and features in a Swiss context. *Cahiers du Centre de Linguistique et des Sciences du Langage*, (48), 107-118.

Edwards, J. (1994). *Multilingualism*. Routledge.

Eldin, A. A. T. S. (2014). Socio linguistic study of code switching of the Arabic language speakers on social networking. *International Journal of English Linguistics*, *4*(6), 78.

Ethnologue. (2018). *Languages of the World: Summary by language size*. Retrieved 29 June 2018 from https://www.ethnologue.com/statistics/size

Evans, J. R. (2017). A strategic approach to self- branding. *Journal of Global Scholars of Marketing Science*, *27*(4), 270-311. https://doi.org/10.1080/21639159.2017.1360146

Fishman, J. A. (1998). The new linguistic order. *Foreign Policy*, 26-40.

Franceschini, R. (2009). The genesis and development of research in multilingualism. *The Exploration of Multilingualism: Development of Research on L*, *3*, 27-61.

Franceschini, R., Zappatore, D., & Nitsch, C. (2003). Lexicon in the brain: What neurobiology has to say about languages. In: *The multilingual lexicon* (pp. 153-166). Springer.

García, O. (2009). Education, multilingualism and translanguaging in the 21st century. *Social Justice through Multilingual Education*, 140-158.

García, O., & Li, W. (2014). Translanguaging: Language, education, and bilingualism. New York, NY: Palgrave Macmillan.

Georgakopoulou, A. (1997). *Narrative performances: A study of Modern Greek storytelling* (Vol. 46). John Benjamins Publishing.

Georgakopoulou, A. (2004). To tell or not to tell?: Email stories between on- and off-line interactions. *Language@ Internet*, *1*(2004), 0009-0007.

Georgakopoulou, A. (2011). “On for drinkies?”: Email cues of participant alignments. *Language@ Internet*, *8*(4).

Georgakopoulou, A., & Spilioti, T. (2015). *The Routledge Handbook of Language and Digital Communication*. London; New York, NY: Routledge.

Gillen, J., & Merchant, G. (2013). Contact calls: Twitter as a dialogic social and linguistic practice. *Language Sciences*, *35*, 47-58.

Gillham, B. (2005). *Research Interviewing: The range of techniques: A practical guide*. McGraw-Hill Education.

Graham, S. L. (2015). Relationality, friendship, and identity in digital communication. In: *The Routledge Handbook of Language and Digital Communication* (pp. 319-334). Routledge.

Gregory, A. (1996). *Planning and managing a public relations campaign: A step-by-step guide.* London: Kogan Gronstedt.

Gregory, L., Taha Thomure, H., Kazem, A., Boni, A., Elsayed, M. A. A., & Taibah, N. (2021). *Advancing Arabic Language Teaching and Learning*. Washington, DC: World Bank.

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.

Gumperz, J. J. (1982). *Discourse strategies* (Vol. 1). Cambridge University Press.

Gupta, A. F. (1997). The internet and the English language. *First Conference on Postcolonial Theory*.

Halim, N. S., & Maros, M. (2014). The functions of code-switching in Facebook interactions. *Procedia-Social and Behavioral Sciences*, *118*, 126-133.

Hamdan, H. (2017). Attitudes towards Arabic romanization and student's major: Evidence from the University of Jordan. *Arab World English Journal*, *7*(4).

Heath, J. (1989) *From Code Switching to Borrowing: Foreign and Diglossic Mixing in Moroccan Arabic*, London: Kegan Paul International.

Herring, S., & Dainas, A. (2017). "Nice picture comment!" Graphicons in Facebook comment threads. *Proceedings of the 50th Hawaii International Conference on System Sciences*. Hilton Waikoloa Village, Hawaii.

Herring, S. C. (2004). Computer-mediated discourse analysis. *Designing for Virtual Communities in the Service of Learning*, 338-376.

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@ Internet*, *4*(1).

Herring, S. C., & Androutsopoulos, J. (2015). Computer-mediated discourse 2.0. *The Handbook of Discourse Analysis*, *2*, 127-151.

Hinrichs, L. (2006). *Codeswitching on the web: English and Jamaican Creole in e-mail communication* (Vol. 147). John Benjamins Publishing.

Hoffmann, C. (1991). *An introduction to bilingualism*. Harlow: Longman.

Hoffmann, C. (2000). The spread of English and the growth of multilingualism with English in Europe. *English in Europe: The Acquisition of a Third Language*, *121*.

Hoffmann, C., & Ytsma, J. (2004). *Trilingualism in family, school, and community* (Vol. 43). Multilingual Matters.

Holes, C. (2018). *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches* (Vol. 30). Oxford University Press.

Holt, D. (2016). Branding in the age of social media. *Harvard Business Review*, *94*(3), 40-50.

Honeycutt, C., & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. *42nd Hawaii International Conference on System Sciences*. Waikoloa, Hawaii

Hong, L., Convertino, G., & Chi, E. H. (2011). Language Matters In Twitter: A Large Scale Study. *5th International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain.

Horton, J. L. (1995). *Integrating corporate communications: The cost-effective use of message and medium*. Greenwood Publishing Group.

Huang, D. L.-J. (2009). Language use in asynchronous computer-mediated communication in Taiwan. *Australian Review of Applied Linguistics*, *32*(2), 12.11-12.22.

Hufeisen, B., & Marx, N. (2007). Multilingualism? *Receptive multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts*, *6*, 307.

InternetWorldStats. (2018). *Internet World Users by Language*. Retrieved 29 June 2018 from: https://www.internetworldstats.com/stats7.htm

Jaspers, J. (2018). The transformative limits of translanguaging. *Language & Communication*, *58*, 1-10.

Jaspers, J., & Madsen, L. M. (2019). *Critical perspectives on linguistic fixity and fluidity*. Abingdon: Routledge.

Jaworska, S. (2014). Playful language alternation in an online discussion forum: The example of digital code plays. *Journal of Pragmatics*, *71*, 56-68.

Jenkins, J. (2003). *World Englishes: A resource book for students*. Psychology Press.

Jessner, U. (2006). *Linguistic Awareness in Multilinguals: English as a Third Language*. Edinburgh University Press.

Jewitt, C. (2013). Multimodal methods for researching digital technologies. *The SAGE Handbook of Digital Technology Research*, 250-265.

Jones, R., & Hafner, C. (2012). *Understanding digital literacies: A practical introduction*. Routledge.

Jørgensen, J. N. (2008). Polylingual languaging around and among children and adolescents. *International Journal of Multilingualism*, *5*(3), 161-176.

Jørgensen, J. N., Karrebæk, M. S., Madsen, L. M., & Møller, J. S. (2015). Polylanguaging in superdiversity. In: *Language and Superdiversity* (pp. 147-164). Routledge.

Kachru, B. B. (1982). The bilingual's linguistic repertoire. In: *Issues in International Bilingual Education* (pp. 25-52). Springer.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59-68.

Kashian, N., Jang, J.-W., Shin, S. Y., Dai, Y., & Walther, J. B. (2017). Self-disclosure and liking in computer-mediated communication. *Computers in Human Behavior*, *71*, 275-283.

Kay, S., Mulcahy, R., & Parkinson, J. (2020). When less is more: the impact of macro and micro social media influencers' disclosure. *Journal of Marketing Management*, *36*(3-4), 248-278.

Kemp, C. (2009). Defining multilingualism. *The Exploration of Multilingualism: Development of Research on L3, Multilingualism and Multiple Language Acquisition*, *6*, 11.

Khalil, S. (2018). *Fuṣḥá, ʻāmmīyah, or both?: Towards a theoretical framework for written Cairene Arabic*. PhD thesis. University of Leeds.

Khamis, S., Ang, L., & Welling, R. (2017). Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies*, *8*(2), 191-208.

Kotler, P., Kartajaya, H., & Setiawan, I. (2019). Marketing 3.0: From products to customers to the human spirit. In: *Marketing Wisdom* (pp. 139-156). Springer.

Krämer, N. C., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology*, *20*(3), 106-116.

Kulavuz-Onal, D., & Vásquez, C. (2018). "Thanks, Shokran, Gracias": Translingual Practices in a Facebook Group. *Language Learning & Technology*, *22*(1), 240-255.

Kytölä, S. (2013). *Multilingual language use and metapragmatic reflexivity in Finnish internet football forums: a study in the sociolinguistics of globalization*. PhD thesis. University of Jyväskylä.

Labrecque, L. I., Markos, E., & Milne, G. R. (2011). Online Personal Branding: Processes, Challenges, and Implications. *Journal of Interactive Marketing*, *25*(1), 37-50. https://doi.org/10.1016/j.intmar.2010.09.002

Lam, W. S. E. (2004). Second language socialization in a bilingual chat room: Global and local considerations. *Language Learning & Technology*, *8*(3), 44-65.

Lamy, M., & Hampel, R. (2007). *Online communication in language learning and teaching*. Springer.

Lee, C. (2015). Digital discourse@ public space. In: *Discourse and Digital Practices: Doing Discourse Analysis in the Digital Age*. London: Routledge.

Lee, C. (2016). Multilingual resources and practices in digital communication. *The Routledge Handbook of Language and Digital Communication*, 118-132.

Lee, C. (2017). *Multilingualism online*. London; New York, NY: Routledge.

Lee, C. K. (2002). Literacy practices in computer-mediated communication in Hong Kong. *The Reading Matrix*, *2*(2), 1-25.

Lee, C. K. (2007). Linguistic features of email and ICQ instant messaging in Hong Kong. *The Multilingual Internet: Language, Culture and Communication Online*, 184-208.

Lee, C. K., & Barton, D. (2011). Constructing glocal identities through multilingual writing practices on Flickr. com®. *International Multilingual Research Journal*, *5*(1), 39-59.

Leppänen, S. (2007). Youth language in media contexts: Insights into the functions of English in Finland. *World Englishes*, *26*(2), 149-169.

Leppänen, S., & Peuronen, S. (2012). Multilingualism and the Internet. *The Encyclopedia of Applied Linguistics*. Wiley.

Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., Kääntä, L., Räisänen, T., Laitinen, M., & Pahta, P. (2011). *National survey on the English language in Finland: Uses, meanings and attitudes*. Helsinki: Studies in Variation, Contacts and Change in English.

Leppänen, S., Pitkänen-Huhta, A., Piirainen-Marsh, A., Nikula, T., & Peuronen, S. (2009). Young people's translocal new media uses: A multiperspective analysis of language choice and heteroglossia. *Journal of Computer-Mediated Communication*, *14*(4), 1080-1107.

Lexander, K. V. (2012). Analyzing multilingual texting in Senegal–an approach for the study of mixed language SMS. *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 146-169). London: Routledge.

Litosseliti, L. (2018). Introducing Research Methods in Linguistics. *Research Methods in Linguistics*, 1. Bloomsbury Academic.

Makoni, S., & Pennycook, A. (2005). Disinventing and (re)constituting languages. *Critical Inquiry in Language Studies: An International Journal*, *2*(3), 137-156.

Marwick, A., & boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114-133. https://doi.org/10.1177/1461444810365313

Marwick, A. E. (2010). *Status update: Celebrity, publicity and self-branding in Web 2.0*. PhD thesis. New York University.

McArthur, T. (2002). *The Oxford guide to world English*. Citeseer.

McClure, E. (2001). Oral and written Assyrian-English codeswitching. *Trends in Linguistics Studies and Monographs*, *126*, 157-192.

Meiseles, G. (1980). Educated spoken Arabic and the Arabic language continuum. *Archivum Linguisticum*, *11*(2), 117-148.

Mischaud, E. (2007). Twitter: Expressions of the whole self. *An Investigation into User Appropriation of a Web-Based Communications Platform.* MSc dissertation. London School of Economics.

Moshashai, D., Leber, A. M., & Savage, J. D. (2020). Saudi Arabia plans for its economic future: Vision 2030, the National Transformation Plan and Saudi fiscal reform. *British Journal of Middle Eastern Studies*, *47*(3), 381-401.

Murthy, D. (2018). *Twitter*. Cambridge: Polity Press.

Myers-Scotton, C. (1988). *Code-switching and types of multilingual communities*. Washington, DC: Georgetown University Press.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Clarendon Press.

Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual & Multicultural Development*, *13*(1-2), 19-39.

Nishimura, Y. (2003). Linguistic innovations and interactional features of casual online communication in Japanese. *Journal of Computer-Mediated Communication*, *9*(1).

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*, 30-53.

Okazaki, S., Plangger, K., West, D., & Menéndez, H. D. (2020). Exploring digital corporate social responsibility communications on Twitter. *Journal of Business Research*, *117*, 675-682.

Otheguy, R., García, O., & Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review*, *6*(3), 281-307.

Otsuji, E., & Pennycook, A. (2010). Metrolingualism: Fixity, Fluidity and Language in Flux. *International Journal of Multilingualism*, *7*(3), 240-254. https://doi.org/10.1080/14790710903414331

Page, R. (2012). The linguistics of self- branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication*, *6*(2), 181-201. https://doi.org/10.1177/1750481312437441

Palfreyman, D., & Khalil, M. A. (2003). "A Funky Language for Teenzz to Use": Representing Gulf Arabic in Instant Messaging. *Journal of Computer-Mediated Communication*, *9*(1), JCMC917.

Paolillo, J. C. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics*, *5*(2), 180-213.

Park, Y. E., & Kang, M. (2020). When crowdsourcing in CSR leads to dialogic communication: The effects of trust and distrust. *Public Relations Review*, *46*(1), 101867.

Pavalanathan, U., & Eisenstein, J. (2016). More emojis, less:) The competition for paralinguistic function in microblog writing. *First Monday*, *21*(11).

Peters, T. (1997). The brand called you. *Fast Company*, *10*(10), 83-90.

Poplack, S. (1980). Sometimes I'll start a sentence in spanish y termino en espanol: toward a typology of code-switching. *Linguistics*, *18*(7-8), 581-618.

Poplack, S., & Meechan, M. (1995). Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. Cambridge University Press.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. Berlin; New York, NY: Walter de Gruyter.

Public Relations Society of America (PRSSA) (2022). *Learn about public relations*. Retrived from: https://www.prsa.org/prssa/about-prssa/learn-about-pr

Puschmann, C., & Hagelmoser, R. (2015). Corporate blogging and corporate social media. *The Routledge Handbook of Language and Digital Communication*, 226.

Rahman, M. M., & Alhaisoni, E. (2013). Teaching English in Saudi Arabia: prospects and challenges. *Academic Research International*, *4*(1), 112.

Raven, B. H. (2008). The bases of power and the power/interaction model of interpersonal influence. *Analyses of Social Issues and Public Policy*, *8*(1), 1-22.

Reagan, T. (2004). Objectification, positivism and language studies: A reconsideration. *Critical Inquiry in Language Studies: An International Journal*, *1*(1), 41-60.

Rheingold, H. (1993). A slice of life in my virtual community. *Global Networks: Computers and International Communication*, 57-80.

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (2013). *Qualitative research practice: A guide for social science students and researchers*. Sage Publications.

Romaine, S. (1995). *Bilingualism*. Wiley-Blackwell.

Romenti, S., Valentini, C., Murtarelli, G., & Meggiorin, K. (2016). Measuring online dialogic conversations' quality: A scale development. *Journal of Communication Management*, *20*(4), 328-346.

Rui, J., & Stefanone, M. A. (2013). Strategic self-presentation online: A cross-cultural study. *Computers in Human Behavior*, *29*(1), 110-118.

Ryding, K. C. (1991). Proficiency despite diglossia: A new approach for Arabic. *The Modern Language Journal*, *75*(2), 212-218.

Salia, R. (2011). *Between Arabic and French lies the dialect: Moroccan code-weaving on Facebook*. BA dissertation. Columbia University.

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.

Schnoebelen, T. (2012). Do you smile with your nose? Stylistic variation in Twitter emoticons. *University of Pennsylvania Working Papers in Linguistics*, *18*(2), 14.

Schreiber, B. R. (2015). "I am what I am": Multilingual identity and digital translanguaging. *Language Learning & Technology*, *19*(3), 69-87.

Seargeant, P. (2019). *The Emoji Revolution: How technology is shaping the future of communication*. Cambridge University Press.

Seargeant, P., & Tagg, C. (2011). English on the internet and a 'post-varieties' approach to language. *World Englishes*, *30*(4), 496-514.

Seargeant, P., & Tagg, C. (2014). *The language of social media: Identity and community on the internet*. Springer.

Seargeant, P., Tagg, C., & Ngampramuan, W. (2012). Language choice and addressivity strategies in Thai-English social network interactions. *Journal of Sociolinguistics*, *16*(4), 510-531.

Sebba, M. (2012). *Researching and theorising multilingual texts*. Routledge.

Sebba, M., Mahootian, S., & Jonsson, C. (2012). *Language mixing and code-switching in writing: approaches to mixed-language written discourse*. New York, NY: Routledge.

Senft, T. M. (2008). *Camgirls: Celebrity and community in the age of social networks* (Vol. 4). Peter Lang.

Siebenhaar, B. (2006). Code choice and code-switching in Swiss-German Internet Relay Chat rooms. *Journal of Sociolinguistics*, *10*(4), 481-506.

Spilioti, T. (2009). Graphemic representation of text-messaging. *Pragmatics. Quarterly Publication of the International Pragmatics Association*, *19*(3), 393-412.

Spilioti, T. (2014). Greek-Alphabet English: vernacular transliterations of English in social media. *46th Annual Meeting of the British Association for Applied Linguistics*. Edinburgh, Scotland.

Spilioti, T. (2019). From transliteration to trans-scripting: Creativity and multilingual writing on the internet. *Discourse, Context & Media*, *29*, 100294.

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science*, *32*(11), 1492-1512.

Squires, L. (2015). *Design, discourse, and the implications of public text*. New York, NY: Routledge.

Stavans, A., & Hoffmann, C. (2015). *Multilingualism*. Cambridge University Press.

Su, H.-Y. (2007). The multilingual and multi-orthographic Taiwan-based Internet: Creative uses of writing systems on college-affiliated BBSs. *The Multilingual Internet: Language, Culture, and Communication Online*, 64-86.

Tagg, C. (2015). *Exploring digital communication: Language in action*. Routledge.

Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, *83*(1), 3-34.

Tantawi, Y., & Rosson, M. B. (2019). The paralinguistic function of emojis in Twitter communication. *Social Media Corpora for the Humanities*, 68.

Themistocleous, C. (2015). Digital code-switching between Cypriot and Standard Greek: Performance and identity play online. *International Journal of Bilingualism*, *19*(3), 282-297.

Thorne, S. L., & Ivković, D. (2015). Multilingual Eurovision meets plurilingual YouTube. *Dialogue in Multilingual and Multimodal Communities*, *27*, 167.

Troise, C., & Camilleri, M. A. (2021). The use of digital media for marketing, CSR communication and stakeholder engagement. In: *Strategic corporate communication in the digital age*. Emerald Publishing Limited.

Tseliga, T. (2007). It's all Greeklish to me!": Linguistic and sociocultural perspectives on Roman-alphabeted Greek in asynchronous computer-mediated communication. In: B. Danet & S. C. Herring (eds.), *The Multilingual Internet: Language, Culture and Communication Online* (pp.116-141).

Vaast, E. (2007). Playing with masks: Fragmentation and continuity in the presentation of self in an occupational online forum. *Information Technology & People*, 20(4), 334-351.

Van Langenhove, L., & Harré, R. (1999). Introducing positioning theory. *Positioning theory: Moral contexts of intentional action*, 14-31.

Vandekerckhove, R., & Nobels, J. (2010). Code eclecticism: Linguistic variation and code alternation in the chat language of Flemish teenagers. *Journal of Sociolinguistics*, *14*(5), 657-677.

Versteegh, K. (2014). *Arabic Language*. Edinburgh University Press.

Vildomec, V. (1963). *Multilingualism*. AW Sythoff.

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, *23*(5), 2538-2557.

Warschauer, M., Said, G. R. E., & Zohry, A. G. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication*, *7*(4), JCMC744.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, *20*(5), 1813-1831.

Wei, L. (2000). *The bilingualism reader*. Psychology Press.

Wei, L. (2011). Moment analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics*, *43*(5), 1222-1235.

Wei, L. (2012). Conceptual and methodological issues in bilingualism and multilingualism research. *The Handbook of Bilingualism and Multilingualism* (2nd ed). (pp. 26-51).

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, *39*(1), 9-30.

Weinreich, U. (1963). *Languages in contact: findings and problems*. The Hague: Mouton.

Werry, C. C. (1996). Internet relay chat. *Computer-mediated Communication: Linguistic, Social and Cross-cultural Perspectives*, 47-63.

Wilson, J. A. (2020). Understanding branding is demanding…. *Journal of Marketing Management*, *36*(13-14), 1178-1189.

Witmer, D. F., & Katzman, S. L. (1997). On-line smiles: Does gender make a difference in the use of graphic accents? *Journal of Computer-Mediated Communication*, *2*(4), JCMC244.

Wright, S. (2004). MOST and Initiative B@bel funded research: Language on the Internet. In: *UNESCO's International Journal on Multicultural Societies*.

Yates, S. J. (1996). Oral and written linguistic aspects of computer conferencing. *Pragmatics and Beyond New Series*, 29-46.

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, *13*(5), 788-806.

Zappavigna, M. (2015). Searchable talk: the linguistic functions of hashtags. *Social Semiotics*, *25*(3), 274-291.

Zhang, W. (2012). Chinese-English code-mixing among China's netizens: Chinese-English mixed-code communication is gaining popularity on the Internet. *English Today*, *28*(3), 40-52.

Zhang, W. (2015). Multilingual creativity on China's Internet. *World Englishes*, *34*(2), 231-246.

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, *24*(5), 1816-1836.

Zuhur, S. (2011). *Saudi Arabia*. ABC-Clio.

Appendices

Appendix 1:
Sample Email to Interviewees

Dear marketing team at Red Sea Mall,

I hope my email finds you well.

My name is Reem Al Madani and I am a PhD researcher at Cardiff University in UK. The aim of my PhD research is to investigate the use of different language varieties on Twitter and its relation to branding and self-branding by participants in Saudi Arabia. The study investigates three main categories which are corporations, social media influencers, and ordinary users.

I would like to invite the person in charge of social media at Red Sea Mall to participate in an interview to help me understand the motivations and preferences of corporations towards the use of the different language varieties on social media and particularly on Twitter.

It will be highly appreciated if you accept to participate in this academic research and provide me with the contact information of the responsible person to schedule a meeting this week or next week.

Looking forward to hearing from you soon.

Best wishes,

Reem Al Madani
PhD Researcher
Cardiff University
United Kingdom

سعادة مسؤولي التسويق في رد سي مول،

السلام عليكم و رحمة الله و بركاته،

أنا -ريم المدني- باحثة دكتوراة في جامعة كاردف في بريطانيا و يهدف بحثي إلى دراسة اللغات و اللهجات المختلقة على الشبكات الاجتماعية و علاقتها بهوية المؤسسات و الأفراد في المملكة العربية السعودية. تشمل عينة البحث ثلاثة فئات مختلفة و هي المؤسسات والمؤثرين على الشبكات الاجتماعية و مستخدمي تويتر بشكل عادي.

أتقدم لكم بطلب مقابلة مع الشخص المسؤول عن وسائل التواصل الاجتماعي في رد سي مول وتهدف أسئلة المقابلة فهم دوافع و أسباب استخدام اللغات و اللهجات المختلفة على الشبكات الاجتماعية عامةً و تويتر خاصةً.

أتمنى أن يلقى طلبي استحسانكم للمساهمة في هذا البحث العلمي الهادف و تزويدي بمعلومات التواصل مع الشخص المسؤول عن وسائل التواصل الاجتماعي لتنسيق وقت للمقابلة خلال هذا الأسبوع أو الأسبوع القادم.

مع خالص الشكر و التقدير،

ريم المدني
باحثة دكتوراة
جامعة كاردف
المملكة المتحدة

Appendix 2:
Debriefing Document

Research Project

The Role of Networked Multilingualism in Branding and Self-Branding: An Investigation of Practices by Saudi Users on Twitter

Thank you for taking part in this study.

The aim of this research was to investigate the use of different language varieties on Twitter and its relation to branding and self-branding by participants in Saudi Arabia. The study examines the motivations and preferences towards the use of different language varieties by interviewing a selected sample of Saudi Twitter users. The Data of the study combines observation of online interactions in addition to interviews with some of its social actors. The data is collected from Twitter public accounts that represent three main categories which are; corporations, social media influencers, and ordinary Twitter users.

In order to ensure all participants have the same experience during data collection, please do not discuss what you did in this study with anyone who is participating in the study but has not made their contribution.

The data you have provided will be held confidentially. You retain the right to withdraw your data without explanation and retrospectively, by contacting the researcher named below.

If you would like to read more about the area of this research the following resources may be of interest:

Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), 185-205.

Danet, B. and Herring, S. (2007). *The multilingual internet: Language, culture and communication online*. Oxford: Oxford University Press.

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, *13*(1), 114-133.

Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & communication*, *6*(2), 181-201.

Lee, C. (2016). *Multilingualism online*. Routledge: New York.

If you have any questions about this study or your participation in it, please contact:

Researcher	Reem Al Madani – PhD Candidate Cardiff University School of English, Communication and Philosophy	Email: almadanir@cardiff.ac.uk
Supervisor	Dr Mercedes Durham Reader of Sociolinguistics Cardiff University School of English, Communication and Philosophy	Email: durhamm@cardiff.ac.uk

Appendix 3:

Consent Form

Consent Form

Research Project

The Role of Networked Multilingualism in Branding and Self-Branding: An Investigation of Practices by Saudi Users on Twitter

- I understand that my participation in this project will involve an interview about the different language choices online, specifically on Twitter, and that it will require approximately 30 to 45 minutes of my time.
- I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason.
- I understand that I am free to ask any questions at any time. If for any reason I experience discomfort during participation in this project, I am free to withdraw or discuss my concerns with Reem Al Madani, PhD student at Cardiff University.
- I understand that the information provided by me will be held confidentially, such that only the Experimenter can trace this information back to me individually. The information will be retained for up to two years when it will be deleted/destroyed. I understand that I can ask for the information I provide to be deleted/destroyed at any time.
- I understand that information provided by me for this study, including my own words, may be used in the research report, but that all such information and/or quotes will be anonymised.
- I also understand that at the end of the study I will be provided with additional information and feedback.

I, _____
consent to participate in the study conducted by Reem Al Madani, School of English, Communication & Philosophy, Cardiff University under the supervision of Professor Michael Handford and Dr Mercedes Durham.

Signed:

Date:

نموذج موافقة

عنوان البحث

دور تعدد اللغات ضمن الشبكات الاجتماعية في بناء هوية المؤسسات و الأفراد: دراسة ممارسات مستخدمي تويتر في السعودية

- مشاركتي في هذا البحث تتضمن مقابلة عن اللغات المختلفة المستخدمة في الشبكات الاجتماعية عامة وتويتر خاصة وستكون مدة المقابلة 30 إلى 45 دقيقة.
- مشاركتي في هذا البحث تطوعية و أستطيع الانسحاب من البحث في أي وقت من غير توضيح أي أسباب.
- أستطيع أن أسأل أي سؤال في أي وقت وإذا أحسست بعدم الراحة خلال مشاركتي في هذا البحث أستطيع الانسحاب أو التفاهم مع ريم المدني - باحثة الدكتوراة بجامعة كارديف.
- كافة المعلومات التي سأشاركها مع الباحثة ستحفظ بسرية بحيث أن الباحثة فقط تستطيع ربط المعلومات بشخصي. وستحفظ المعلومات لمدة عامين ثم تمحى و يتم إتلافها كما يمكنني أن أطلب أن تتلف المعلومات التي سأشارك الباحثة بها في أي وقت.
- المعلومات التي سأشاركها مع الباحثة و الكلمات و التعبيرات التي استخدمها ستستخدم في تقرير البحث و لكن كل المعلومات و الاقتباسات ستظهر في التقرير مجهولة المصدر.
- سيتم تزويدي بمعلومات إضافية عن البحث بعد المقابلة.

أنا _____
أوافق على المشاركة في الدراسة التي تقوم بها ريم المدني في كلية اللغة الانجليزية و التواصل و الفلسفة بجامعة كارديف بإشراف الدكتورة مرسيدس ديرهام.

التوقيع:

التاريخ: