

From Speech to Screen: The Orthography of Colloquial Arabic  
in Electronically-Mediated Communication

Clara Gordon  
Senior Thesis in Linguistics  
Advisor: Dr. Jason Kandybowicz  
Examiner: Dr. Abbas Benmamoun  
April 15, 2011



From Speech to Screen: The Orthography of Colloquial Arabic  
in Electronically-Mediated Communication  
Clara Gordon

Abstract

What does it mean to communicate textually when one's native language lacks a standardized writing system? This thesis examines the orthographic systems devised by Arabic speaking users of mobile and online text-based media. Such electronic communication has intrigued linguists since the birth of the technologies that facilitate it, leading them to debate whether it can be considered speech or writing. They have observed in the short-form type discourse present in texting, online posting and instant messaging an interesting array of linguistic phenomena, including changes in orthography and spelling, novel phrases and abbreviations. However, Arabic texting and online language remains largely unexplored in the field, and little work has been done on the emergence of new ways of encoding its previously unwritten dialects in electronic media. Electronically-mediated communications (EMC) in various dialects of Arabic world indeed represents a transition from a strictly spoken language to a regularized written code, composed largely in the Roman alphabet. The popularity and rampant censorship of online social media in the Middle East, along with a relative lack of research on languages other than English in text and online communication, makes Arabic dialect in mobile communications a timely topic of study.

After a brief overview of previous literature on the orthography of English and other languages in EMC, I use a small corpus of data collected in the summer of 2010 from students at the University of Aleppo in Aleppo, Syria, as well as from online sources. This data is supplemented by the work of linguists who have investigated EMC in other parts of the Arabic speaking world. This research covers Arabic EMC composed in both the Arabic and Roman alphabets. I focus primarily on Levantine Arabic, but use other varieties such as Algerian and Gulf Arabic for comparison and elaboration. Along with studies conducted in Jordan, Algeria, the United Arab Emirates and Kuwait, I present a portrait of an evolving, user-generated way of transcribing Arabic using the Roman alphabet. In this "Romanized" orthography, users spell out Arabic sounds phonologically in a specialized code of Roman characters and use numeric substitutions

for certain Arabic sounds. This multi-dialectal approach to Romanized Arabic EMC reveals that it has features of both phonetic transcription and transliteration, and follows certain trends in vowel deletion and change. Lastly, I consider the particularities of Arabic language EMC in the “speech versus language” debate that pervades the study of language in mobile communications. *Note: Because of their desire for anonymity, the two University of Aleppo students interviewed for this paper have had their names omitted.*

## 1. Introduction: Arabic in the Digital Age

“I use texting more... I can say what I want more than when I call somebody” (U. Aleppo student, Personal communication, 2010).

The young University of Aleppo student quoted above sat in a empty classroom at the Department of Literature and Languages, reflecting on her communication habits and echoing a sentiment that is strikingly common among teens and young adults worldwide. Text messaging and online communication is increasingly a part of the fabric of everyday social discourse, thanks to a slate of new technologies that allow users to send off packets of written information with a few button clicks or keyboard taps. The volume, brevity and popularity of this mode of communication has serious implications for the orthographic and typographic nature of the written language. Although previously referred to as Computer-mediated communication, current research uses the term electronically-mediated communication (hereafter referred to as EMC) to recognize the increasing array of mobile devices with which this type of language is composed and consumed. EMC encompasses written correspondence carried out using computers, mobile phones and other handheld mobile device, and can occur synchronously (in real-time), or asynchronously (in which messages are left for a recipient who can choose to reply at a later time). This investigation focuses largely on asynchronous communication

in the form of SMS text messages and comments posted on social networking websites.

Within the past ten years, linguists (Baron, 2008; Danet, 2007; Crystal, 2008) have begun to delve into the relatively uncharted territory of EMC, noting the particular ways in which the benefits and constraints of electronic media influence the spelling/orthography, syntax, and discourse structure of language. Such research has uncovered certain processes common to this type of informal messages: novel phrases and abbreviations, deletion of vowels, and typographic representation of extralinguistic features such as facial expressions and prosodic features.

However, little cross-linguistic work has been done, and even less attention has been paid to the particular challenges and resources common to Arabic speaking users of electronic media. The issue of diglossia, the division of different varieties of speech within a single speech community, further complicates Arabic EMC. Every Arabic-speaking community has a spoken, nonliterary vernacular in addition to the standardized register of literary and spoken language, known as *Fusha* or Modern Standard Arabic (MSA). Users who are accustomed to speaking in one register among their friends and writing in another must navigate difficult linguistic terrain in encoding their registers of speech into readily understandable written communication. Users of Arabic EMC in particular have embraced that challenge and devised ingenious methods of Roman character substitution to transcribe the sounds of vernacular Arabic.

Samuel Johnson declared in his *Dictionary of the English Language* that a lexicographer “should be derided, who being able to produce no example of a nation that has preserved their words and phrases from immutability, shall imagine that his dictionary can embalm language” (Baron, 2008, p. ix). Johnson's warning is particularly

relevant in the rapidly evolving field of EMC. What that in mind, this is a exploratory look at a small corpus of data collected by the author in Syria in the summer of 2010, for the purpose of gleaning insight about certain processes of language in electronic media. However, this research proceeds without the illusion that such a language as can ever be truly captured as it is used in electronic media. As Naomi S. Baron notes in her work *Always On*, any effort to fully categorize the particular lexical items of a particular online or texting dialect would be futile, or at least become obsolete within a short time. A sound approach, therefore, is to seek to encompass the “synergy between technology and language,” which happens to be particularly complex in the Arab world (Baron, 2008: x).

The complications of inquiry into EMC has been noted by many linguists. EMC is a popular linguistic phenomenon, prone to fads and the influence of the rapidly changing interfaces that facilitate it. While mobile users were once confined to multiple clicks on a numeric keypad to enter a single letter, they can now use predictive-text (T9) software and full “qwerty” keyboards on their phones, significantly affecting the style and content of their messages. Similarly, Arabic characters are only recently available on mobile phones, and users are no longer required to work around this gap through Romanized spelling, although research indicates that a majority still chooses to. The encoding of Arabic dialect in written form, however informal, has great implications for literacy and its perception in the Arab world; in a covert way, EMC is broadening the possibilities of what Arabic text can be. However, those who write in Arabic dialect must contend with the notion that the vernacular is less legitimate or refined than the more formal MSA. While in Syria collecting data, my friends at the University of Aleppo would scoff or simply act confused when asked if any literature existed in dialect. This

attitude pervades throughout the Arabic speaking world; in her memoir, a former Fulbright researcher in Morocco recalls hearing outraged cries of “You can't write literature in Darija [Moroccan dialect]!” (Melbourne, 2008, p. 50). In general, vernacular Arabic has simply been absent from the literary sphere, but it is becoming increasingly “written” through this ubiquitous medium of communication among the younger generations of Arabic speakers. As Mostari notes in his work on Algerian texting, this development holds possibility for the increased visibility and prestige of spoken dialects: “Thanks to mobile phones, these local varieties may gain a higher status in the Algerian linguistic ecology.” (2009, p. 385).

## 2. EMC Orthography and Its Perception

Despite the tradition in Linguistics of nonjudgmental analysis of language, it is necessary to mention that a certain anxiety about the state of language use lingers in the background of any discussion of electronic shorthand. Complaints and sound bites about new media simplifying and ruining language are rampant in more traditional media outlets, and a slate of sites such as NoSlang.com promise to ease parental worry about the unintelligible net- and text-speak of their children. “The internet is already way too unreadable...” the site's founder states, and “we focus solely on promoting proper English online” (Jones, 2005). In his 2008 work *Txting: The Gr8 Deb8*, David Crystal devotes a whole chapter to the hype surrounding texting, noting that, as with the introduction of just about every new technological development in communication, critics cry out that texting will bring about the end of writing and of literacy itself. For example, he quotes a 2007 Washington Daily article by Eric Uthus dismissing texters as subpar users of language, “...obsessed with taking the vowels out of words and spelling fonetikally”

(Crystal, 2008, p. 8). Baron, meanwhile, ponders whether writing has been reduced in our everyday lives to a quick-paced conversational form, and worries about a “gradually eroding public sense that the quality of our writing matters” (2008, p. 6).

However, other linguists note that texting has brought about the concision and phonetic transparency that has long eluded spelling reformists. Emphasizing that the “notion that a word should always be spelled the same way is a much more recent invention than the language itself,” Ammon Shea investigates the links between the phonetic spelling movement and the organic shift toward such spelling in texting (2010, para. 3). Because “text messaging... comes from the linguistic bottom,” it has a better chance of affecting spelling conventions than top-down measures (para. 9). However, it is important to emphasize that electronic language features are not a simple, uniform condensing of language. The numeric abbreviations and deletions in texting language are largely a response to the inadequacies and inconvenience of the phone interface, and may decrease as more sophisticated technology is developed. Baron adds that “part of the appeal of texting shorthands is their novelty, and that that will fade” (para. 8).

Users of EMC, rather than simplifying “proper” writing, in fact employ a whole host of clever modifications to make up for the shortcomings of the devices with which they compose their writing. Among the most prominent observations of those who study the mechanics of EMC is that it is “characterized by massive and systematic deviations from the orthographic norm” (Yvon, 2009, p. 133). These deviations can be loosely categorized as phonetic transcription (in which *though* becomes *thru*), consonantal orthography or vowel deletion (in which *homework* becomes *hmwrk*), the substitution of characters and numerals for their phonetic value (in which *great* becomes *gr8*), and

abbreviation (in which *tyl* stands in for *talk to you later*). Of course, these processes all exist in tension with the writing conventions of the original language. Such communication is often informal, which licenses certain typographic changes as well, including simplification or omission of punctuation, capitalization, and the omission of grammatical elements. Other work has focused on the metalinguistic function of emoticons, and other typographic approximations of the extra-linguistic qualities of spoken language and non-verbal cues.

The rapid, informal and conversational tone of EMC, as well as the extralinguistic features encoded within them, have led some to consider whether they can be classified as speech, writing, or some hybrid of the two. Some argue that the language goes beyond a mere shorthand, and “mimics or replaces the ability to hear spoken utterances” (Mostari, 2009, pp. 379 – 80). Today, many linguistics favor reframing the “speech versus writing” debate in terms of a spectrum, which accounts for communication that has the tendencies of written or spoken qualities but is not exclusive to either. One researcher sums up these tendencies: “[EMC] is similar to the spoken language in its spontaneity, immediacy, interactivity, associability and lack of formalization, and to the written language in its visual dimension and the possibility for reconsideration and editing” (Borochovsky-Bar-Aba, 2010, p. 3). While there are compelling arguments for a range of positions on this issue, it is fair to say that EMC concerns itself with the replication of informal speech, and does so under the constraints imposed by character limits, hardware and the costs of sending data. The results demonstrate that ingenuity is fostered by such limitations, when combined with the human capacity for linguistic efficiency and the desire for constant communication. The orthographic innovations of



EMC in languages other than English offer a unique perspective on the topic.

## 2.1 Previous Findings on the EMC Customs of Other Languages

Whether or not EMC is composed in English, the language exerts a powerful control over the conventions of the genre. Crystal stresses the predominance of English orthographic rules in his examination of texting in other languages:

Most of the material I have found is for languages which use the Roman alphabet – reflecting the bias toward that alphabet which has been a familiar story in relation to the development of the internet as a whole. Indeed, we might go further and say the bias towards the *English* alphabet...(2008, p. 124)

There is a certain universal nature to English texting abbreviations as well, due to both the dominance of English as a second language worldwide and to its status as the *lingua franca* of electronic media; non-native English speakers in Crystal's multilingual corpus would regularly insert common abbreviations such as *lol* ('laughing out loud') and *brb* ('be right back') in their texts even when those letter combinations were very different from the equivalent phrase in their own language (2008, p. 129).

Crystal also notes extensive use of numerals as syllable replacement in the EMC of languages such as Italian, German, Spanish and French, in which the name of the numeral functions as a syllable in a word. According to the Rebus Principle, this sort of modified pictograph then replaces the characters that spell out that syllable (2008, p. 132). In Spanish, for example, the numeral 2, *dos*, is integrated into the greeting:

- (1)            <sl2>  
                  [saludos]  
                  “greetings” (2008, p. 132)

In fact, English language EMS employs numerals as syllables for the same purpose, using <2> ('to') and <4> ('for') to name a few.

However, languages that use alphabets other than the Roman one rely more heavily on these creative processes, as the use of these alphabets only recently became possible on computers and mobile devices. In early computing, the most common coding system was the American Standard Code for Information Interchange, or ASCII, which was limited to the Roman Alphabet, the numerals 1 to 9 and a limited set of punctuation (Tseliga, 2007, p. 118). Later advances brought the Unicode Worldwide Character Standard, which by summer 2005 included more than 50 different scripts (2007, p. 118). In the meantime, the limitations of ASCII led to orthographic innovations among users who were unable to enter text in their native alphabets. Numeric substitution is used extensively in EMC orthographies developed by speakers of Japanese and Chinese, who have taken the names of numerals and reduced them to syllables represented by a single character. This works, of course, because the language features characters that represent syllables, rather than phonemes. For example, the Japanese greeting *ohayoo*, “good morning” is expressed with a series of numerals whose names roughly resemble the word's characters:

- (2)           0 = *oh*, as in the English numeral  
               8 = *ha*, a shortened form of Japanese *hachi*, “eight”  
               4 = *yo*, a shortened form of Japanese *yon*, “four”  
               0 = *oh* (Crystal 2008, p. 136)

Greek users of EMC have also been found to employ Roman characters and Arabic numerals to represent the original orthography of their language in a hybrid

known as “Greeklish” (Tseliga, 2007). The reasons for these replacements fall under two categories: graphical similarity, in which Roman characters are selected to represent Greek letters with a similar form, and phonetic similarity, in which the Roman character represents the same sound or phoneme as its Greek counterpart. For example, the Greek graph  $\eta$  receives the Roman equivalent  $h$  on a graphical basis, and  $i$  on a phonetic basis (2007, p. 118). Certain numerals are employed because of their visual similarity to Greek characters, including  $\delta$  as a substitute for  $\theta$ . The Greek word  $A\theta\eta\nu\alpha$ , [athina], for example, could be spelled in two different ways:

(3)            orthographic: Athina

                  phonetic: A8hva

Tseliga carried out an investigation of corpus emails composed in Greek and Greeklish, and found that 54 percent of messages employed the Roman alphabet. Furthermore, approximately half of Tseliga's interview professed a preference for orthographic over phonetic transliteration, meaning that they favored a one-to-one correspondence of Greek to Roman characters. “I want the word to look exactly the same [as the Greek one] in terms of orthography” stated one young female student (2007, p. 130). Tseliga hypothesizes that the use of this alphabet motivated more common incorporation of English loan words, particularly those of a technological nature such as “video,” “internet,” and “forum” (2007, p. 125). “If you are technologically literate and you address people who know about computers,” stated one user of the orthography, “you have to write Greeklish...” (p. 134) Users expressed distaste for the aesthetics of phonetic Greeklish, and stressed the clarity of a Romanized orthography that is graphically similar to original spelling (135).

Like Greeklish users, speakers of Arabic insert numerals into a Romanized EMC orthography that employs Roman characters based on their graphical similarity to Arabic characters. Similarly, Arabic speakers employ both an orthographic and phonetic sensibility to translate their language into a foreign alphabet. The processes that occur in Arabic EMC, therefore, not entirely unique; they are part of a global effort to use graphical resources in the most efficient way possible.

## 2.2 A Brief Overview of Arabic Phonology and Orthography

Arabic sounds are represented in two ways: as characters, and as diacritical signs. All consonants have corresponding characters, while vowels can be represented either as characters or as diacritics. There are 28 consonant phonemes, and three vowels, [a], [i], and [u], which may be long or short. This difference between short and long vowels is phonemic, and short vowels are transcribed diacritical marks while long vowels are transcribed as characters. The characters for the long vowels [eː] and [uː] may also serve as the consonants [y] and [w] if they take short vowel diacritical marks. Additionally, a character known as *ta marbuta*, ة, appears word-finally in many feminine nouns and verb and is pronounced as [t] only when there are suffixes following it. This character specifies that the short vowel of the previous syllable is [a].

Characters read from right to left and are connected, so that each letter has an initial, medial, terminal and isolate, or unconnected, form. Short vowel diacritical marks appear above or below the character, and join with it to form a syllable. There are additional diacritics that lengthen a character, signify a glottal stop, and specify that the character has no following vowel. Although present in formal and Qur'anic Arabic, these diacritical marks are absent in most modern print media, and therefore readers infer the

short vowels because of their familiarity with a word. The basic syllable structure for Standard Arabic is CV(C)(C), although dialectal varieties allow for syllable-initial consonant clusters and groups of more than two consonants.

Character	IPA Transcription	Character	IPA Transcription
ا	[aː]	ض	[dˤ]
ب	[b]	ظ	[ðˤ]
ت	[t]	ع	[ʕ]
ث	[θ]	غ	[ɣ]
ج	[d]	ف	[f]
ح	[ħ]	ق	[q]
خ	[x]	ك	[k]
د	[d]	ل	[l]
ذ	[ð]	م	[m]
ر	[r]	ن	[n]
ز	[z]	ه	[h]
س	[s]	و	[w],[uː]
ش	[ʃ]	ي	[y],[iː]
ص	[sˤ]	ء	[ʔ]
ط	[tˤ]		

**Table 1:** Sounds of the Arabic Alphabet.

Even with the absence of transcribed short vowels, the predictable sound-to-character correlation in Arabic explains the relative lack of phonetic shortcuts in Arabic EMC, such as *c* for *see* and *u* for *you*, that occurs in English EMC, in which the relationship between character and sound is considerably more haphazard. Crystal explains: “It is precisely because English has such unpredictable spellings... that [these abbreviations] have a point” (2008, p. 128). Indeed, the abbreviations that preoccupy many of the researchers of English language EMC is markedly absent in Arabic EMC,

with users' creativity far more devoted to the transcription of dialectal sounds than to the abbreviation of words and phrases.

### 2.2.1 Dialectal Differences

In the introduction to one of the first textbooks on Levantine dialect, spoken in Syria and surrounding countries, the author notes the absolute necessity of dialect knowledge to understand the linguistic patchwork of Arab societies. “No linguistic research would be complete,” she adds, “without mastering the colloquial variety of a certain region” (Barakat, 2009, p. 13). There are profound differences between colloquial registers and the language of media, government, and education. Dialectal changes can be classified broadly into two categories: (1) novel features, including distinct morphological and grammatical rules, and lexical items unique to the dialect, and (2) phonological modifications to MSA.

Arabic Character	MSA Pronunciation	Levantine Pronunciation
ث	[θ]	[t]
ذ	[ð]	[d], [s]
ص	[s <sup>h</sup> ]	[z]
ظ	[ð <sup>h</sup> ]	[d <sup>h</sup> ], [z]
ق	[q]	[ʔ]
ء	[ʔ]	(omitted)

**Table 2:** Possible Phonetic Changes in Levantine Dialect  
Adapted from (Barakat, 2009, p. 20).

The first category includes differing verb conjugations, word order, and pluralizing processes for nouns that are distinct from the standard dialect. In the second is phoneme-wide shifts in pronunciation, changes to or deletions of vowels, and changes in the pronunciation certain grammatical features such as prepositions and subject prefixes.

Epenthesis also occurs in some dialects, including Levantine, with word-final consonant clusters. This is to say that while dialects have certain unique features, they also consist largely of MSA pronounced in a different way. Phonemic changes are far from regular, however; they may depend on the particular word in which they appear, or on the phonetic environment. As Table 2 demonstrates, these changes are not always one-to-one; phonemes in MSA may be pronounced as different sounds in different environments in the dialect.

Arabic is a particularly divided language with a wide range of dialects, many of which are mutually unintelligible. Despite these differences, there are certain processes that occur out in a variety of dialects, as well as some common lexical items. For example, the Levantine dialect term [ʃu], “what,” has a counterpart [ʃnu] in certain Moroccan varieties. While we can hardly look at these dialects as a unified class, orthographic systems in EMC feature a correspondingly similar set of processes, including numeric substitution for Arabic characters and vowel deletion. This investigation of Arabic focuses on Levantine dialect, but makes use of examples from fieldwork in other areas of the Arab world. Together, they they present a view of EMC in the Arabic speaking world as a whole, as seen through the lens of one particular dialect.

### 3. Finding on Arabic EMC

*Note: The phonetic transcriptions for the original samples in the following sections are the my estimations based on experience and recollection from the fieldwork area. The pronunciation of the Levantine dialect varies by region, and these transcriptions are intended as a general indication of the Syrian pronunciation. In samples with Arabic orthography, transcriptions line up vertically with the original terms and thus read right to left.*

The original research for this paper was conducted over the course of a month at the University of Aleppo in Aleppo, Syria. Two university students there supplied samples of text and online discourse, and two extended personal interviews were conducted with one of the original subjects and one other student. All three were in their early to mid-twenties and either undergraduates or graduate students. Furthermore, several samples were taken from public pages on the social media website Facebook, in particular the “I love Damascus” fan page. This page was useful not just because of the abundance of short posts containing Arabic EMC in both the Roman and Arabic scripts, but because its commenters reliably used the Levantine dialect spoken in Syria. The small corpus of data reveals an electronic code that challenges the divisions between literary and spoken Arabic, and reveals a user-generated Romanized orthography for the Arabic language.

Levantine dialect, spoken in Syria, Lebanon, Jordan, Palestine and Israel, is claimed to be the most similar to MSA of all Arabic dialects. However, like all dialects, it contains certain derivational, inflectional, and phonetic differences that distinguish it from MSA, as well as freer word order and lexical items specific to the dialect (see Section 2.2.1). Syrians are avid users of mobile technology, and nearly every student encountered during the research period owned a prepaid cell phone. There are only two mobile carriers in the country, MTM and SyriaTel, and together they exert a monopoly over the telecommunications business, charging for “units” purchased at convenience stores that allow users to send texts and make calls. Minute and texting rates are relatively high for a country with a low cost of living; in June 2010, the cost for a minute of calling time was approximately eight lira, which is roughly equal to sixteen cents.



Texting rates are slightly lower. The high cost of using these ubiquitous devices made texting a desirable and cost-effective option. One Syrian university student commented, “When I'm calling somebody it will take maybe two or three minutes on the phone... but when I'm texting somebody it will be just eight units, so it will be cheaper” (Personal communication, June 24, 2010).

Facebook is very popular among students of the University of Aleppo, and among the student population, and a profile on the site was nearly as ubiquitous a commodity as a mobile phone. The Syrian government blocked access to Facebook and other social networking sites until very recently, when political unrest in Egypt prompted a lift of the ban in order to appease citizens and increase online monitoring (Mackey, 2011). During my fieldwork, however, students regularly and publicly browsed the sites using proxy servers in the many internet cafes that line the campus. An NPR reporter recently corroborated this observation in Damascus, observing that “tech-savvy young Syrians know how to get around the bans and openly browse their Facebook sites” (Amos, 2010). Despite the restrictions of cost and censorship, mobile and online communication remains wildly popular among the younger segment of Syrian society, which is quietly transforming and transcribing the language of everyday interaction in Levantine dialect. Recent research on instant messaging in the United Arab Emirates, and texting in Algeria, Jordan and Kuwait suggest that this processes is a region- and language-wide phenomenon (Palfreyman and Al Khalil, 2008; Mostari, 2009; Haggan, 2007; Al-Khatib and Sabbah; 2008).

All of the EMC observed in this research was composed in dialect. In a personal

interview, a University of Aleppo student scoffed at the notion of composing online or SMS messages to friends in any other way; EMC is clearly the realm of informal registers of speech, and MSA never once made an appearance, except in one solicited example.

### 3.1 Features of Arabic EMC in Its Original Orthography

#### 3.1.1 Representations of Prosody and Phonetic Shifts

As in English EMC, the features of spoken language are indeed approximated and transcribed in Arabic EMC, departing from the conventions of written Arabic to conveying more informal types of speaker meaning. One example of this was the repeating of vowels within words, to convey stressed and elongated pronunciation. In a public Facebook site for fans of the city of Damascus, a user expresses his feelings for the city both linguistically and typographically (with the noted phenomenon in bold).

(4) < و ما أحلى ناس الشام >

[aʃa:m na:s aħla ma wa]

“And there are none sweeter than **the people of Damascus**”

The repeated vertical lines are the Arabic long vowel *alef*. It never appears consecutively in standard Arabic, but the user types it five times in the middle of the words for “people” and “Damascus” in order to create the illusion of effusive, elongated speech. Another common refrain that uses vowel repetition is the intensifier كتير, meaning “a lot” in which the long vowel *ya*, ي, may appear many times. The following constructed example is representative of several online comments observed on Syrian Facebook pages:

- (5) بحبك كتبيبيبيبي  
 [kateɪr bəhɪbak]  
 “I love you a lot”

Despite this approximation of the sound of spoken language, phoneme-wide dialectal shifts don't seem to be expressed often in the spelling conventions of the original alphabet Arabic EMC. The letter *qaf*, ق, for example, is usually pronounced as a glottal stop in Levantine dialect. The word [qalb], “heart,” would therefore be pronounced [ʔalb] Levantine Arabic, but transcribed in Arabic EMC using the original *qaf*, rather than the Arabic character signifying a glottal stop, *hamza*, ء. (b) exhibits a straight phonetic transcription of the Levantine word, which is assumed to be ill-formed. This shift never appears in Arabic orthography in the corpus of Syrian data, nor in any of the numerous online forums and pages observed during data collection.

- (6) (a) قلب  
 (b) ألب\*

However, phonemic shifts Levantine Arabic do not always lack representation. One common exception is the phoneme [θ], which changes to [t] in Levantine pronunciation. The character ت, which represents [t], is often employed in place of the character ث, which represents [θ].

- (7) (a) <كتير> (b) <كثير>  
 [kətir] [kəθir]  
 “a lot” “a lot”

However, this particular instance appears to be isolated, with a great majority of the

phonemic shifts unrepresented in the Arabic orthography.

### 3.1.2 Dialectal Lexical Items

While there is some instability in transcription on the phonemic level, other grammatical and lexical features of dialect receive regular representation in EMC. Users seem to be particularly adept at spelling lexical items unique to a dialect in a consistent manner even though they are part of an unwritten register of speech. Interestingly, the transcriptions of these samples show awareness of the etymological and phonetic roots of the dialect, and include characters representing phonemes whose pronunciations become modified in the pronunciation of the dialect. For example, the following text sent by a University of Aleppo student features one of these items.

(8) <انا قاعدة قدام المعهد>

[Imahad ʔadam ʔaʕIda ana]

“I am sitting in front of the institute.”

The word *قدام*, [ʔadam] is rarely used in MSA but functions primarily Levantine dialectal lexical item. The beginning letter *qaf*, *ق*, pronounced [q] in MSA, is pronounced as a simple glottal stop in the dialect. However, the spelling of the word doesn't reflect the phonetic change, suggesting that the author is aware of the phonological roots of the lexical item despite the relative lack of written dialect outside the EMC sphere.

Other strictly dialectal terms regularly appear in written form in EMC. A selection from a comment on the “I love Damascus” facebook page displays a widely used dialectal term [ʃwe], “a little.”

- (9) <عنا شوي خصوصية>  
 [xas<sup>h</sup>us<sup>h</sup>ie ʃwe ʕana]  
 “we have **a little** privacy”

Another widely used dialectal term, [ʃu], appears in Arabic script on the same page:

- (10) <شو اخبار الاستفتاء يا جماعة؟>  
 [dʒamaʕah ya listiftaʔ axbar ʃu]  
 “What’s the news of the referendum, group?”

Such terms, while they would never appear in any formal print sources, make regular appearances in the EMC of Arabic speakers when they choose to use their native alphabet.

The regular spelling of these dialectal terms reveals an essential trend in Arabic EMC orthography: word-specific phonetic changes appear to be represented much more faithfully than phoneme shifts, such as [θ] to [t]. One salient example is the set of nominative pronouns, which vary phonetically in the dialect, and receive regular phonetic transcription in the Arabic alphabet. Table 3 demonstrates these pronoun spellings observed in the corpus.

Pronoun	MSA		Levantine	
	1PL	نحن	[naħnu]	نحننا
2PL.M	أنتم	[antum]	أنتو	[antu]
3PL.M	هم	[hum]	هنن	[ħmɪn]

**Table 3:** *Pronouns in MSA and Levantine Dialect*

Differences in dialectal verb conjugation also conditions spelling differences. One regularly transcribed feature is verbal prefix [b], ب, which appears only in dialect and is used to indicate ongoing action in the imperfect verb tense.

- (11) بحبك يا شام  
 [ʃam ya b-əhɪb-ak]  
 Damascus oh ongoing-1SG.IMP.love-2SGM  
 “I love you Damascus”

While the examples above are only a small segment of the observed dialectal features transcribed in Arabic EMC, they are indicative of the regularity with which the grammatical and lexical items make their way from spoken dialect to electronic text.

### 3.1.3 Collapsed Phrases

Levantine dialect also features fossilized phrases that consist of words that are condensed phonetically and pronounced as if they were one lexical item. These phrases are therefore written as one word, as they are pronounced. This process is analogous to the use of English EMC greetings such as *'sup* (“what's up”) and *cya* (“see you”). In a text received from the same Syrian student, the commonly used phrase *good evening*, consisting of two words, is written as one, reflecting its pronunciation as a single lexical unit. Example (a) shows the spelling in the text, while (b) shows the standard spelling and pronunciation of that phrase in MSA:

- (12) (a) <مسالخير> (b) <مساء الخير>  
 [masalxeyr] [lxeyr masaʔ]  
 “Good evening”

Note that the final letter of the first word, *hamza*, ء, is elided in the condensed version of the phrase. Another example can be found in a statement of well-wishing, found on a public Facebook fan page for Damascus, Syria. Again, the direct quote is compared to the same phrase in MSA followed by a literal and idiomatic translation.

- (13) (a) <الحمد لله على السلامة> (b) <الحمد لله على السلامة>  
 [ʕassalame alʕamdillah] [ssalama ʕalɐ lilah alʕamdu]

“The will of God unto peace (Have a safe journey)”

These spelling changes represent more than the mere phonetic shortening that occurs in Levantine dialect, however. The preposition *على*, [ʕalɐ], which translates in different contexts as *upon* or *to*, is shortened to the single-syllable *ع*, [ʕa] in Levantine Dialect. This change is apparently a morphological one as well, because it often appears fused to its object as though it were a prefix in Arabic EMC, even though its MSA counterpart is always an independent lexical item. Another text from the same Syrian student shows the [ʕa] feature both independently and prefixed.

- (14) <نشالله ما تعبروا سلمي ع رفقاتك و تصبجي عخير  
 [ ʕa-xeyr tusbʕi wa rafiʔatIk ʕa salmi taʔburu ma enshallah

“Godwilling you get there safely, say hello **to** your friends and wake up **to** goodness (have a good night)”

Together, these features indicate that the orthography of Levantine EMC demonstrates understanding of morphological, as well as phonetic, differences between the dialect and MSA.

Indeed, the language of text messages and online media in Syria represents a transition from a spoken dialect to a written code. Arabic EMC demonstrates that literate dialect speakers are able to conceptualize and transcribe many of the features of their spoken dialects, and a broader look at its use in the Arab world demonstrates that this process is playing out largely in a foreign alphabet.

### 3.2 Romanized Arabic Orthography

Among the most fascinating and productive features of Arabic EMC is the use of the Roman alphabet to approximate the sounds of spoken dialect. Although hardly unique in the world of EMC, as we have seen in the Chinese, Japanese and Greek examples, Arabic users go about transcribing their language in a unique and surprisingly regular fashion. This system employs Arabic numerals in an improvised substitution for Arabic sounds without a Roman equivalent, with the stand-in characters generally resembling the visual form of the original Arabic characters. Other sounds with counterparts in the Roman alphabet are simply represented phonetically. A number of these numeric substitutions emerged in my own field work as well as in the work of Palfreyman and Al Khalil (2008) in the UAE, Mostari (2009) in Algeria, Al-Khatib and Sabbah (2008) in Jordan, and Haggan (2007) in Kuwait. According to Haggan, “[t]o date, there is no research on how widespread the Romanisation of Arabic is in electronic communication, whether there are inter-regional variations in the use of numerals and whether there may also be inter-media variations” (2007, p. 442). However, a current look at the research indicates that at the very least, speakers of Algerian, Gulf and Levantine Arabic have been recorded as using this innovation, spanning a large swath of the Arabic speaking world.

The table below presents a compilation of substitutions found in my work, and the work of Palfreyman, Mostari, Al-Khatib and Haggan. It therefore includes substitutions found across a variety of dialects. Interestingly, several of these substitutions feature a preceding apostrophe, which approximates the dot found in the form of the corresponding Arabic letter. Because there is variety in the phonology of these different dialects, certain



substitutions appear in one dialect and not in another. However, wherever these sounds appear, the substitution appears regular with a few exceptions. *Kha*, خ , pronounced [x], may be represented with '7 or 5 (and Levantine samples displayed uses of both), and *qaf*, ق , pronounced [q], was described by an informant in Aleppo as corresponding to 9, although Haggan found that Kuwaiti users transcribed it using the numeral 8 (2007: 441). Despite these differences, the general stability of these substitutions across the geographic distance that separates research sites suggests that the orthographic phenomenon appears throughout the Arabic speaking world.

Arabic Letter	IPA Equivalent	Romanized Transcription
ء	[ʔ]	2
ع	[ʕ]	3
ط	[tʕ]	6
ح	[ħ]	7
ق	[q]	8,9
ص	[sʕ]	9
خ	[x]	5,'7
غ	[ɣ]	'3
ظ	[ðʕ]	'6
ض	[dʕ]	'9

**Table 4:** *Compiled list of numeric substitutions in various dialects of Romanized Arabic EMC from original research, Palfreyman, 2007 and Haggan, 2007.*

A typical example of Levantine Romanized EMC, taken from the “I love Damascus” Facebook page, makes use of these numerals.

(15) <3l fkra bt3ref a7'e amgd ano alktar b6l3>

[ʕal fɪkra btaʕɪf axi amdʒad anu lkɪtar btʕlaʕ]

“by the way you know, my brother Amjad, that the train is coming”

Another example from a Kuwaiti text (Haggan, 2008: 440):

- (16) <Gowah, schoonech? Tra7a6eet al ketab 3end el 7aris>  
 [gowah, stʃunɛtʃ traɦatʕet alkItab ʕɛnd lɦarIs]  
 “Hi, how are you? I left the book with the security man”

And from an Algerian text (Mostari, 2009: 383):

- (17) <rni nekr3 fk>  
 [ɣani nqɑʕɛ fik]  
 “I am waiting for you”

Both personal interviews and other research in the area confirm that the decision to Romanize Arabic messages is one of monetary and space efficiency as well as convenience. “Each Short message is up to 160 characters in length when Latin alphabets are used and 70 characters in length when non-Latin alphabets such as Arabic and Chinese are used,” so Romanized texts convey more information per text, even taking into account that more Roman letters are generally required to transcribe a word than exist in its Arabic form (Segerstad, 2002, p. 187, quoted in Al-Khatib, 2008, p. 49). Furthermore, one University of Aleppo student noted that Romanized EMC was easier to enter on computer keyboards, and that she regularly uses it on MSN and Yahoo social media sites as well as on her cell phone (Personal communication, June 24, 2010).

This system isn't the first to use the Roman Alphabet to approximate Arabic sounds. Palfreyman and Al Khalil note that both the Library of Congress and the *Encyclopedia of Islam* have developed regularized transliteration systems, in which capital letters indicate emphatic or pharyngealized sounds such as [s<sup>ʕ</sup>] and [d<sup>ʕ</sup>], and the [x] phoneme is represented by <kh> (2007, p. 49). This numeric symbolism, however, is

a product of the digital age. Such orthography “did not exist before the advent of the Internet, and it can be noticed that there is a heavy use of this new form of written communication among students to such a degree that the traditional way of writing Arabic is counted out” (Al-Khatib, 2008, p. 48).

Although most research on this orthography has occurred since 2007, there is some suggestion that it has evolved over time; in 2002 researchers found the same phenomenon of numeric substitution in a corpus of Egyptian emails, but they were limited to 2,3 and 7 (Haggan, 2007, p. 442). Furthermore, exploiting the graphical similarity of numerals doesn't appear to be a personal or aesthetic choice, as it is with Greeklish users. Rather, the numerals serve to distinguish phonemic contrasts that remain unrepresentable with English characters. It would be ambiguous, for example, to use *d* for both the pharyngealized letter *dawd*, ض, and unpharyngealized *dal*, د, when the difference between the two is essential to text comprehension.

Recent research, meanwhile, attests to the popularity of this orthography; in Mostari's review of Algerian texting, 59 out of 60 respondents chose Latin script in their texts, even though their phones allowed



**Figure 1:** A storefront sign in the Souk Al-Hamidiyeh in Damascus

them to enter Arabic script (2009, p. 382). Palfreyman and Khalil note in their survey of Gulf Arabic in instant messaging that subjects who composed in the Arabic language were split evenly between Arabic and Latin script (2007, p. 49). Because of its popularity among the

younger segment of the population that more frequently uses electronic media, this type of orthography has spread beyond the EMC sphere. In a personal interview, a University of Aleppo student remarked, “You know they use [Romanized Arabic] to appeal to the younger generation... It's cool” (Personal interview, June 24, 2010). Electronics and mobile phone stores often display this sort of writing in their storefronts in Syria, presumably to convey a young, trendy image to their customer base (see Figure 1). Observing the rising popularity of the form in the United Arab Emirates, Palfreyman and Khalil emphasize that “an analysis of the form and use of [Romanized Arabic] illustrates some of the interrelations between language, literacy, technology and globalization” in a generation whose competence in Arabic, a foreign alphabet, and electronic media allows them to combine these skills in novel ways (2007, p. 44).

### 3.3 Transliteration and Transcription

The orthography of a language with such a distinct phonetic system and alphabet presents several challenges to the user of Romanized Arabic EMC. Beesley (1998, n.p., quoted in Palfreyman, 2007, p. 48) makes the crucial distinction between transcription and transliteration in the adaptation of languages into foreign scripts. Transcription attempts a phonological recording of the sounds of the language, while transliteration employs a system of substitution for the characters of the original alphabet, and takes into account the spelling conventions of the target language. Several studies, as well as original research in Aleppo, have confirmed that the orthography of Romanized Arabic EMC is unstable and that spellings vary widely. Furthermore, this system may be considered phonetic transcription in some respects and transliteration in others. A University of Aleppo student noted this variability, commenting on the writing of her friends: “Maybe they write it in their way, but I can understand it” (Personal interview,

June 24, 2010).

### 3.3.1 Romanized EMC as Phonetic Transcription

Straight phonetic transcription tends to be the norm with consonants in Romanized orthography, and much more so than in original Arabic typography. Furthermore, as with the original Arabic alphabet, features of spoken languages may be approximated. Similarly to English EMC, letters are generally lowercase and capitals are used “mainly for emphasis” (Palfreyman, 2007, p. 52). In an “I love Damascus” page comment, a user employs an orthographic representation of prosodic features to emphasize his greeting.

- (18) <aahleeeeeeeen>  
 [ahle:n]  
 “greetings”

Mostari encounters an exaggerated spelling of the Algerian dialectal term [xafi], “wait up.”

- (19) <KHFYYYY KHFYY>

More significantly, Romanized EMC features an accurate transcription of the consonant phonetic changes that occur in dialect. Several samples demonstrate that the letter *qaf*, represented as *q* in Romanized orthography, hardly ever appears in Levantine EMC, either as its numeric stand-in or the more traditional *q* transcription. As noted above, this emphatic velar stop generally becomes a simple glottal stop the dialect, and is most often expressed only by what would be the following vowel in Romanized EMC. In fact, the only collected instance of *q* in Levantine EMC was from a directly solicited sample, in which the subject chose to write in MSA (presumably so the author, who is

more proficient MSA, could understand).

(20) <Anna mutashawi9a jidan>

[ana mutashaiqa jIdan]

“I am very excited.”

(N. Al-Jamous, personal communication, Sep. 10, 2010)

The numeral 9 can also be used to express the letter *sawd*, ص, and so does appear in Levantine MSA. However, usage of the numeral to represent *qaf* in Levantine is not present in the corpus. A song lyric posted on a public Facebook page displays a more typical interpretation of the *qaf* phoneme:

(21) <albi bytlob ya 7abibi tnsa 7obi w 7obak tnsa>

[ʔalbi byətʰlɔb ya ʰabibi tənsa ʰɔbi u ʰɔbak tənsa]

“My sweetheart you forget my love and forget your love”

The word for “heart,” or in this case “sweetheart,” is pronounced in MSA as /qalb/. In the song lyric, the first person possessive suffix /i/ also appears at the end of the word. Here, the author here transcribes the word as it is pronounced, and the glottal stop is expressed, as in a similarly pronounced English word “alter,” with simply a word-initial *a*. There is a certain instability in expressing this phoneme, because while it is possible for it to occur in a word-initial position without numeric transcription, in other environments it requires numeric substitution. In the above case, the author relies on English orthographic rules, which stipulate that a soft glottal stop occur before many word-initial vowels.

*Qaf* appears again in word-medial position, this time represented by the numeral 2, also used to denote the glottal stop represented by *hamza*. The MSA word for *approximately*, [taqriban], is pronounced in Levantine dialect as [taʔriban]. In a comment

in the “I love Damascus,” page:

(22) <y3ne alsyara 7aln bt2tlk **t2rebn** 200 gram/km>

[yaʕni as:iyara ʕaln btʔtɪlk **taʔriban...**]

“So the car kills **about** 200 gram/km” (check translation)

Because English orthography does not allow for a word-medial glottal stop, the use of a numeral is warranted here.

Phonetic shift is also seen in the case of the [θ] phoneme, which changes to [t] in the Levantine dialect (see section 2.2.1). On the same webpage, a user transcribes the phonetic change in the word “more,” which would be pronounced [akθar] in MSA.

*(Note: the caps are the original author's own.)*

(23) <7abet el-sham **AKTAR**>

[ ʕabet aʕam **aktar**]

“I loved Damascus **more**”

Furthermore, the phonemic shift from [ð] to [z] is regularly represented in Levantine EMC. For example, the conjunction *if*, pronounced [Iðe] in MSA, is pronounced [Ize].

(24) <ma ba3ref **eza** halal or haram bes ana ma habaito>

[ma baʕrɪf **ize** halal or haram bɛs ana ma habetu]

“I don't know **if** it's haram or halal but I didn't like it”

In general, the consonantal phonemic changes of Levantine dialect are represented faithfully in EMC, suggesting that at least in this regard, its orthography can be classified as transcription according to Beesley's criteria.

### 3.3.2 Romanized EMC as Transliteration

Despite the tendency of phonetic transcription, there is ample evidence of transliteration in Romanized EMC. Arabic nouns feature an affixed definite article *al*, [al], whose pronunciation changes depending on its phonetic environment and position in the sentence, and may be pronounced as “[əs], [ər], [əθ] or other pronunciations depending on the initial sound of the word which follows” (Palfreyman, 2007, p. 48). In more traditional transliteration conventions, such as that of the *Encyclopedia of Islam*, this particle is written as <al> no matter what its pronunciation, and this practice has carried over into Romanized EMC as well, where it is often seen as <al> or <el> regardless of its pronunciation (p. 48). Sometimes, the elided [a] or [e] may be excluded, but the [l] appears quite tenacious in several Levantine examples. Palfreyman and Al Khalil suggest that it is the essential role of this particle in determining the “underlying grammatical identity” of the word that gives it typographic representation in Romanized text (2007, p. 48). Several comments on the “I love Damascus” page makes use of this convention, spelling out the name of the city with the fully transliterated article, a part of the name, even though the [l] sound is not pronounced. In a survey of all of the posts on the page since early 2007, all twelve instances of the words include at least an <l>. A typical spelling hyphenates the article and noun, but they also appear together unhyphenated.

- (25) <el-sham>  
 [eʃam]  
 “Damascus”

However, this is a rather isolated instance of non-phonetic consonant transcription based



on grammatical and orthographic significance of the <ﺍ> article.

### 3.4 Trends in Vowel Deletion and Change

With the exception of the definite article, consonants in Romanized EMC closely resemble the phonology of the spoken dialect. However, vowels in Romanized EMC present another case entirely. While the lack of suitable Roman alphabet consonants create a constraint that must be resolved using numerals, vowels provide a relative abundance of options, with “a wider range of letters and digraphs than.. [the] Arabic script” (Palfreyman, 2007, p. 56). Complicating the picture, widespread vowel deletion and change occurs in in all the different dialectal varieties examined in this corpus, including Levantine, Gulf, and Algerian Arabic, all of which have a greater number of vowel sounds than MSA. Therefore, the transcription of vowels in the orthographic system is rather unpredictable.

#### 3.4.1 Orthographic Factors

Because of the modest amount of data, is it beyond the scope of this examination to devise a unified theory of vowel deletion and change in Romanized EMC. However, there are several factors that appear to condition these processes. For example, vowel deletion can sometimes be regarded as an effect of character-to-character transliteration of Arabic; short vowels are typically absent from the original orthography, and so they are similarly left out of the Romanized EMC. As reviewed in Section 2.2, short vowels are written as diacritical marks above characters, and are generally omitted in handwriting and print meant for popular consumption. The corpus indicates that the influence of Arabic orthography is indeed at play here, based on patterns in vowel deletion. For example, the three-letter Arabic word *كتب*, *books*, is pronounced [kʊtʊb] but

only features characters for the three consonants in the word.

It is likely that the lack of orthographic representation in the original script makes these short vowels less likely to receive vowels in Romanized EMC, leading users to leave them out even when pronounced and adhere to the conventions of Arabic spelling while making use of a foreign alphabet. However, the decision to include or omit vowels in this type of orthography is far from simple, and while vowel deletion in various dialects transcribed in EMC is unstable, certain trends suggest the influence of the original Arabic orthography and dialectal phonology, as well as a vowel's position within the word.

#### 3.4.2 The Influence of Stress and Length in Vowel Transcription

In Arabic, vowels with orthographic representation are also stressed and long, so it is difficult to distinguish between the influence of orthography and phonology in vowel deletion. Although rates of vowel deletion may vary across users and dialects, deleted vowels are much more likely to be short, and thus unstressed. Long vowels receive primary stress in Arabic words, making them more salient in pronunciation, and therefore more likely to receive representation in Romanized EMC. A word like [kabir], “big” shown below in original orthography, letter isolates, and phonemic translation, has both a long and a short vowel (both in bold).

(26)      كبير  
             ك ب ي ر  
             [ r **eː** b ka]

Common EMC spellings: kbeer, kbir, kber \*kabr

A selection of words from previous examples, with their Arabic script counterparts, further demonstrates this tendency.

- (27) y3ne “it means”  
 يعني  
 ي ن ع ي  
 [e: n ʕ ya]
- (28) t2rebn “approximately”  
 تقريبا  
 ت ق ر ي ب ا  
 [an b e: r ʔ ta]

Previous research indicates that the issue of vowel deletion appears in English EMC as well. Kuil's (2007) investigation of a small corpus of English texts corroborates Dressler's (1996) semiotic principle of “figure and ground” in which consonants, as “figure features,” tend to be foregrounded while vowels, or “ground features,” tend to be backgrounded (2007, p. 45). In her analysis of text messaging spelling, she found that the ratio of deleted vowels to deleted consonants was 9:1 (p. 48). This tendency certainly holds true in Romanized EMC, in which users go to great lengths to faithfully represent Arabic consonants but leave vowel transcription highly unstable.

Kuil further predicts that “the letters which represent vowels in unstressed positions should be deleted within the domain of the instressed syllable” (2007, p. 46). However, this turns out not to be true in her research, possibly because of the high rate of deletion in single-syllable words. It appears, however, that the co-occurrence of stress and length in Arabic vowels provides a stronger imperative for these vowels to be maintained at higher rates than short, unstressed vowels.

#### 3.4.3 Position and Morphological Significance in Vowel Deletion

The corpus reveals that the morphological makeup of a word further influences whether its vowels are retained or deleted. Vowels tend to remain in the Romanized

orthography at the ends of words ending in *ta marbuta* (see Section 2.2), despite the fact that the [a] endings in these words are always short, unstressed and generally unrepresented in Arabic script. This is likely because the vocalic ending signifies the grammatically feminine status of the word, which is significant given the high rate of gender agreement in Arabic. The following Levantine sentence posted on a public Facebook page contains three examples of retained short vowels at the ends of feminine nouns. The words in question are bolded while an interlineal gloss distinguishes the feminine endings.

- (29)           bs **almshkle**           ano n7na 3na           **b6ala**           **mokn3ah**  
                   [bəs **almuʃkil-e**           anu naɦna ʕana           **btʕal-a**           **moqniʕ-a**]  
                   but **DEF.problem- F** that we           with-1PL **unemployment-F** **hidden-F**  
                   “but the **problem** is that we have **hidden unemployment**”

Additionally, many suffixes in Levantine Arabic end in long vowels, including several possessive and object markers. Because of their orthographic representation, length, stress, and the grammatical information they contribute to the word, these suffixes have a particularly strong incentive for transcription in Romanized EMC. The possessive suffix [i], ى, is always retained for these reasons.

- (30)           <sho 3m           tdros           **27'ee**           3bd>  
                   [shu ʕam           tədrus           **ʔax-i**           ʕabd]  
                   what ongoing 2SGM.IMP.study brother.**1SG** Abd  
                   “what are you studying **my brother** abd”

#### 3.4.4 Dialectal Influence

The pronunciation of various dialects also provide an impetus for widespread

vowel deletion and alteration. Levantine samples from the corpus are rife with examples of faithful representation of the vowel changes that occur in the dialect. One such instance is the lowering of the final short vowel [a] to [e] at the ends of feminine nouns ending in *ta marbuta*, as seen in (p. 30). This change is conditioned by the penultimate consonant and only occurs with certain sounds in that position. However, where it occurs, the change is regularly seen in Romanized vowel transcription. In the following example, all of the vowels are short and the Roman vowels are only inserted in the definite article <al> and the feminine ending, both of which give information about the grammatical significance of the word.

- (31) <almshkle>  
 [almuʃkile]  
 “the problem”

Palfreyman and Al Khalil note in their Gulf Arabic samples that feminine endings tend to take the form <ah> or <eh>, and this is common in Levantine samples as well. In the following example the feminine endings are bolded.

- (32) <elmoder**ah** mb3arfeh shai 3n assalf**eh**>  
 [almodera mabʃarfe ʃai ʃan asalf**e**>  
 “the headmistress does not know anything about the story”  
 (Palfreyman, 2007, p. 58)

Interestingly, they find that the <eh> ending correlates with more vernacular terms, while the <ah> ending is more commonly found in formal terms and those taken from MSA.

The researchers also note, however, a trend of vowel deletion that outpaces the deletions that occur in vernacular varieties of speech: “As in Arabic script, short vowels

are often left out entirely” whether or not they are pronounced (Palfreyman, 2007, p. 56).

In the following constructed Levantine example, representative of language found on public Facebook pages, many of the pronounced vowels have been deleted in the Romanized translation. Unrepresented vowels appear in bold in the phonetic transcription.

- (33) <3m ts2lni 3nnak>  
 [ʕ**am** tɪsʔəl**ini** ʕ**anək**]  
 “She is asking me about you”

And another example from the “I love Damascus” page:

- (34) <kiefk ya 7beeb sho 27'bark>  
 [kif**ak** ya **hab**ib shu **ax**barak]  
 “How are you dear what is your news (what's up?)”

Mostari notices the same phenomenon in Algeria, as seen in a previously mentioned example.

- (17) <rni nekr3 fk>  
 [ɣ**ani** nq**a**ʕ**e** f**ik**]  
 “I am waiting for you” (2009, p. 383)

Nearly all the vowels are deleted, leaving only the consonantal structure of the words.

It is difficult to define where these deletions are motivated by the phonology of the dialect, however. The tendency for condensed pronunciation and large vowel clusters in Arabic dialects makes it unclear whether a vowel indeed exists in certain places. This is especially true in North African varieties, in which linguists have trouble distinguishing between vowels, unstressed schwas, and vowel deletions between

consonants. Palfreyman and Al Khalil further point out that there are more vowels in the UAE dialect than in MSA, and that this is likely the reason for an observed instability in vowel orthography. For example, the same Arabic word *يمكن*, *maybe*, is alternately produced as <yumkin,> <yemken,> and <yemkin> in the data (2007, p. 57). Haggan meanwhile notices individual differences in Romanized EMC spelling in her Kuwaiti corpus, and points out that the letter *e* may be used to represent an array of sounds in Kuwaiti dialect: <wench> for [waynitʃ], “how are you;” <kent> for [kunt], “I was;” <men zeman> for [mɪn zeman], “a long time ago,” and <el ketab> for [al kitab], “the book” (2007, p. 442).

Palfreyman and Al Khalil record certain preferences for vowel representation in their UAE corpus that don't necessarily conform to typical English spelling; the phoneme [ɪ] is more often represented by the vowel <e> than by <i>, and [e] is more often represented by *ai* than <ei> (2007, p 57). This sometimes relates to spelling conventions of other languages; UAE text messages only used <y> to represent [i] at the ends of words, rather than word-initially or -medially, which follows English spelling trends (Palfreyman, 2007, p. 57). As the researchers explain, vowels in Romanized orthography can demonstrate the second language knowledge of a population: “The sound /u/ in Arabic tends to be represented as <ou> in Moroccan Arabic, on the basis of French spelling, and as <oo> in the UAE, where English is the main foreign language” (2007, p. 50). For the most part, however, vowel instability in Romanized EMC appears to be a result of the great variety of vowel sounds in Arabic dialects and the overabundance of Roman vowels. While vowel stress, length, and word position all provide a coherent framework for the vowel transcription, the number of vowels graphs combined with

dialectal phonetic diversity only serve to muddle it.

#### 4. Arabic EMC and the Conventions of Spoken Arabic

Much of the research surrounding EMC in recent years has dealt with the status of short messages composed online and on cellphones, and whether they can more accurately be considered speech or writing. The brevity, informality and rapid response rate is more reminiscent of speech, while the restriction to text, ability to edit, and space between turn taking suggest writing. While this debate requires nuanced definitions of the two categories, and we cannot make a definitive argument on where Arabic EMC falls on the speech-writing continuum, it is interesting to observe how the two categories intersect through the use of Romanized EMC in Arabic in particular. Mostari notes that in Romanized SMS messages, “we see nothing of the literary style or content so evident in the purely Arabic messages” (2007, p. 443). This observation hints that by leaving the Arabic script behind, users may also be divorcing themselves from the elaborate literary conventions typical of Arabic literature, and entering into a realm of literacy that is directly influenced by the sounds of speech. On the other hand, the very fact of transcribing vernacular spoken dialects is an act of regulating, and in a sense, of legitimizing them. Mostari asserts that “stigmatised colloquial varieties are being used in new technologies,” and that this “represent[s], *a priori*, a formal means of communication” (2009, p. 385).

While Romanized EMC may lack the regularity and prestige of literary Arabic, the language of electronic media has an expressiveness all its own. Three of the four cited researchers of the orthography also note examples of typographic symbols representing extralinguistic features of communication, including emoticons and <xxx> to signify



kissing (Mostari, 2009, p. 383). Although these features have been amply shown to exist in English, Japanese and Chinese EMC, it remains to be seen whether such typographic practices are a universal phenomenon (Crystal, 2008, p. 40). The existence of these same features in such a different linguistic and cultural environment as the Arabic speaking world suggests that, even though the constraints of the Arabic language require some different orthographic processes than those of other language, the communicative goals of EMC users are common across peoples.

In fact, the process of becoming proficient in offers more insight into the speech-like qualities of Romanized EMC. When asked how they learned the conventions of the orthography, both of my Syrian interview subjects appeared puzzled. They simply “learned it from [their] friends” outside of any formal teaching process (Personal interview, June 24, 2010). While one student was initially confused by the Romanized texts she was receiving, she quickly caught on simply from exposure to the system. This way of learning is perhaps more reminiscent of spoken language acquisition, an organic process that requires large amounts of input, than of the explicit instructions that facilitates literacy in written language. Of course, access to Romanized EMC requires at least basic literacy in Arabic and English (or French, in some varieties), but users appear to familiarize themselves with the system in an intuitive way.

#### 4.1 Language Mixing and the Significance of Culturally-Bound Expressions

It is similarly useful to note the ways in which the conventions of Arabic texting and online discourse compare to the practices and politeness maxims of spoken Arabic. Many of the subjects in studies of Arabic EMC are young, technologically literate college students, second languages feature prominently in communication between them.

According to Al-Khatib and Sabbah:

...English has become an integral part of the professional repertoire of the [Jordanian] students. This could be due to the fact that English is the medium of instruction at the university; all classes, exams and assignments are given in English. In addition, the students are constantly exposed to the language through the use of the Internet and other means of communication” (2008, p. 45).

Palfreyman and Al Khalil note the same generational momentum toward English biliteracy at Zayed University in Dubai, where they performed research on the language of instant messaging between students: “ZU also aims to produce students who are bilingual and biliterate in English and Arabic, and the majority of the courses in the university are conducted in English (2007, p. 44).

In the studies of EMC examined here, language-mixing is indeed norm. Mostari found that the largest proportion of his Algerian corpus, 41.5%, were composed in French and Arabic. Furthermore, many texts that I received in Aleppo demonstrated familiarity not just with English, but with English texting conventions (as noted by Crystal in 2.1), as well as taste for combining English and Arabic pleasantries:

(35) <Sweetly i'm very glad 2 meet u... ديري بالك ع حالك [take care of yourself]>  
(U. Aleppo student, personal communication, June 2, 2010)

However, Sabbah and Al-Khatib's work demonstrates that the mixing of languages in EMC goes beyond mere linguistic novelty. In Arabic, certain fossilized phrases and sentences, often paired with a necessary response from the addressee, are crucial to carrying out polite conversation. This is of significance in the speech versus language debate because such phrases are essential to appropriate Arabic verbal interaction. The prevalence of these formalized “call and response” conversational

elements often carry over to electronic exchanges, whether they take place in English or Arabic. The following sample taken from a Jordanian college student work displays the influence of these phrases.

- (36) <Thx god the exam was very good I'll pass **enshalla** [God willing].>  
(Al-Khatib and Sabbah, 2008, p. 54 – 55)

The phrase *enshalla* is uttered by either the speaker or addressee when discussing anticipated or desired events. Even though the author of this text composed her message largely in English, cultural politeness maxims still make it necessary to add this expression in the original Arabic. As another student in the same study concludes, “...we inherited such expressions from our ancestors. They have become an important part of our linguistic repertoire; therefore, I cannot say anything without saying *Inshallah* [‘God willing’]” (2008, p. 51). The text I received while on a trip in Syria further demonstrates the necessity of such phrases:

- (14) <نشالله ما تعبروا سلمى ع رفاقك و تصبجي عخير  
[ **ʕa-xeyr tusbʕi wa rafiʔatk ʕa salmi taʔburu ma enshallah**  
“**God willing** you arrived safely at your friends and have a good night”

As with the phrase *good evening*, *enshallah* is so commonly used that its two words are commonly collapsed into one.

Culturally-bound phrases often invoke religious sentiment, and include:

- (37) *Inshalla* ‘God willing’  
*Allah yes3idek/yes3idak* ‘May God bless you’  
*enshalla tkoun/tkouni b5air* ‘God willing, you are ok’  
(Al Khatib, 2008, pp. 50 – 51)

The cultural significance of these phrases are in some ways untranslatable, and “[i]t is highly likely that by using such culturally-bound expressions the message writers often express emotional attitudes that cannot be expressed in English” (Al-Khatib, 2008, pp. 54 – 55). A University of Aleppo student commented on the emotional weight of texting in her mother tongue, despite her knowledge of English: “when [my friends] are really happy or sad, I think Arabic is better; ... our Arabic words are more emotional. We understand them better than English” (Personal interview, June 24, 2010). The need to insert these phrases in texts, considering the space constraints of the medium, is a strong indication of the conversational, and thus speech-like, nature of texting. While Romanized EMC may lack the stylistic complexity of literary Arabic, this lack of complexity lends itself to greater efficiency and expressiveness.

##### 5. Conclusions and Suggestions for Further Research

Arabic EMC, and Romanized orthography in particular, has shown itself to be remarkably rich and rule-governed. It features rather consistent phonetic transcription in the case of its consonants, and trends in vowel deletion indicate a sensitivity to the sound of dialect, the grammatical functions within the language, and the rules of spelling in MSA. Furthermore, users navigate these influences within varieties of speech that have not previously appeared in text. Indeed, as users adapt to new technologies and the opportunities and restrictions they impose, they must resort to a sophisticated set of literary and communicative skills that allow them to mesh alphabets and languages. Much work has been done on the supremacy of English in the technological sphere, but the intricacies of orthographic and linguistic mixing in EMC remain a topic largely unexplored, and if the research contained here is any indication, it is one richly

informative about the processes by which bidialectal and bilingual users encode their thoughts into text.

Because this research has just scratched the surface of this ingenious, user-generated orthographic system, a more systematic and large-scale study of Romanized Arabic EMC is warranted. With a large body of text from a single dialectal group of EMC users, researchers may expand upon Kuil's figure-ground principle and develop a more complete theory of vowel deletion and change EMC transcription. A rigorous treatment of this topic is absent from most research on the subject, and a more comprehensive account of the phonological, grammatical, morphological and orthographic factors that affect it would help clarify the composition process in Arabic EMC.

Although the linguistic community may be years away from answering definitively where EMC stands along the continuum of speech and writing, the complexities of its orthography among Arabic speaking peoples demonstrates that sound and text remain in constant dialogue. Furthermore, the topic presents a rather unique case of users gaining fluency in this manner of writing not through any formalized process, but in the way that they learned to speak in the first place: observation of other users. This is perhaps the most compelling case for EMC as a mode of speech, and a deeper investigation of the cultivation of EMC literacy would add greatly to our understanding of it. The development of Romanized orthography is indeed an impressive exercise in wide-spread consensus-building, and has created a relatively standard system so soon after the birth of the media in which it is used.

The language of EMC encompasses, as David Crystal puts it, “a set of linguistic

adaptations being introduced by youngsters, on their own, spontaneously, rapidly, and without professional tuition” (2008, p. 148). The results, as seen in the case of Romanized Arabic EMC, are both ingenious and practical. Ultimately, this orthographic system is a response to a changing perspective on what it means to be a competent user of language. The implications of this are not limited to the linguistic sphere; as the recent events of the “Arab Spring” indicate, such internet-savvy users are a key component in the organization of large-scale social and political movements, and dozens of Facebook pages encouraging protest have sprung up since the onset of the Syrian uprisings. Constant textual communication has indeed created a veritable linguistic revolution among a new generation of technology users worldwide, and they will continue to grapple with the task of writing as they speak.

## Works Cited

- Al-Khatib, Mahmoud and A. Enaq H. Sabbah, (2008). Language Choice in Mobile Text Messages among Jordanian University Students. *SKY Journal of Linguistics*, 21 37 – 65.
- Amos, Deborah, (2010, Sep. 15). New Media Strain Government Tolerance In Syria. *NPR*. Retrieved from <<http://www.npr.org/templates/story/story.php?storyId=129706102>>
- Barakat, Dima, (2009). *Yasmeen Al-Sham*. Damascus: Counsel on Arabic as a Foreign Language.
- Baron, Naomi S., (2008). *Always On: Language in an Online and Mobile World*. New York: Oxford University Press.
- Borochofsky-Bar-Aba, Esther, (2010). The nature of SMS discourse: the case of Hebrew. *Folia Linguistica*, 44(1), 1 – 33.
- Crystal, David, (2008). *Txtng: the Gr8 Db8*. New York: Oxford University Press.
- Jones, Ryan (2010). *NoSlang.com*. 2005. Retrieved from <<http://www.noslang.com>>
- Mackey, Robert, (2011, Feb. 9). Update on Day 16 of Egypt Protests. *The Lede Blog. The New York Times*. Retrieved from: <<http://thelede.blogs.nytimes.com/2011/02/09/latest-faceupdates-on-day-16-of-egypt-protests/>>
- Melbourne, Lucy Laurretta, 2009). *An American in Morocco*. Rabat: Marsam.
- Mohammad, Ibahrine (2008). Mobile communication and sociopolitical change in the Arab world. In Katz, James E. (Eds.), *Handbook of mobile communication studies* (257 – 272). Cambridge: MIT Press.
- Palfreyman, David and Muhamed Al Khalil, (2007). “A Funky Language for Teenzz to Use:” Representations of Gulf Arabic. In Danet, Brenda and Susan C. Herring (Eds.), *The Multilingual Internet : Language, Culture, and Communication online* (43 – 63). New York: Oxford University Press.
- Haggan, Madeline, (2007). Text messaging in Kuwait. Is the medium the message? *Multilingua*, 26(4), 427 – 449.

Kuil, Malgorzata (2007). Phonology in Text Messages. *Poznan Studies in Contemporary Linguistics*, 43(2), 43 – 57.

Mostari, Hind Amel, (2009). What do mobiles speak in Algeria? Evidence from language. *Current Issues in Language Planning*, 10(4), 377-386.

Shea, Ammon, (2010, Jan. 22). The Keypad Solution. *The New York Times*. Retrieved from <<http://www.nytimes.com/2010/01/24/magazine/24FOB-onlanguage-t.html>>

Tseliga, Theodora, (2007). It's All Greeklish to Me! In Danet, Brenda and Susan C. Herring (Eds.), *The Multilingual Internet: Language, Culture, and Communication Online* (116 – 141). New York: Oxford University Press.

Yvon, Francois, (2010). Rewriting the Orthography of Text Messages. *Natural Language Engineering*, 16(2), 133 – 159.

*Cover photo credit: Eric Verdeyen, Damascus, 2010*