

Glocal Arabic online: The case of 3arabizi

Robert Michael Bianchi

Virginia Commonwealth University, Qatar

rmbianchi@qatar.vcu.edu

Abstract

The term *glocal* has been used to describe phenomena that simultaneously blend both global and local elements (see Featherstone, Lash, & Robertson, 1995, p. 101). Nowhere is this more evident than in the existence of 3arabizi, itself a blended language composed of English and Vernacular Arabic, written in Latin letters but using arithmographemes, that is, numerals as letters to represent hard-to-transliterate sounds because they do not exist in English (see Bianchi, 2012).¹ As part of a doctoral study investigating online language choice involving Arabic and English, this paper examines the unique stylistic and topical functions of 3arabizi when compared with its linguistic forbears, that is, Arabic and English in a multilingual web forum. The findings indicate that 3arabizi is used for more informal, intimate and phatic communication than either Arabic or English, though these latter two languages or codes are not entirely formal in form and purpose either.

Keywords: Arabic, English, script, CMC, glocal

¹ The name *3arabizi* itself reflects a fascinating peculiarity of this blended language; its frequent use of arithmographemes. In this case, the 3 represents Arabic's voiced pharyngeal fricative /ʕ/, traditionally written as ξ in the Arabic script. Notice the visual similarity between 3 and ξ (cf. Tseliga, 2007, for arithmographemes in Latinized Greek).

In computer-mediated communication (CMC) contexts, 3arabizi has developed as a unique hybrid language consisting of Vernacular Arabic (VA) written in Latin script interspersed with English. Using corpus and discourse analysis methods, this report discusses the stylistic and topical differences between 3arabizi, Arabic, and English as encountered in the mahjoob.com corpus of web forum messages.

Background

Modern communications technologies such as personal computers and mobile phones have spread so quickly that they have not always easily adapted to local linguistic realities and conventions. This has occasioned an increase in linguistic diversity in electronic contexts such as script-switching in CMC environments (Palfreyman & Al Khalil, 2003). The most common form of CMC script-switching has been Latinization of a non-Latin-scripted language (see Palfreyman, 2001). Crystal (2001) attributes the source of this trend to the fact that the Latin script was forced upon early CMC adopters even though it was not their native script because earlier computer encoding systems such as ASCII were Latin-script based. This situation resulted in several Latin script-based "makeshift" orthographies such as Latin-scripted Greek (see Koutsogiannis & Mitsikopoulou, 2003), Latin-scripted Japanese (Nishimura, 2003), and Latin-scripted Arabic (Palfreyman & Al Khalil, 2003; Warschauer, El Said, & Zohry, 2002).

In recent years, the apparent necessity for Latinization in CMC has diminished due to multilingual and script support for most CMC applications (Androusoopoulos, 2007; Palfreyman & Al Khalil, 2003). Despite this, Latinization of non-Latin-scripted languages continues (Al Share, 2005; Palfreyman & Al Khalil, 2003), posing interesting questions about code choice and code use. Indeed, Latinization, which began as a response to a constrained orthographic choice, is now a bona fide linguistic resource for its users (Lee, 2007; Pavlenko & Blackledge, 2004). Within the present study, 3arabizi is a prime example of such a new linguistic resource.

The Data

Mahjoob.com is a website owned by Emad Hajjaj, a popular Jordanian political cartoonist based in London. The website is Jordan-based and its users appear to be mainly Jordanian and/or Palestinian as well. Nevertheless, the site's popularity extends far beyond Jordan and Palestine as its advertising shows. Structurally, Mahjoob.com contains an Arabic website and an English one. As of November 2008, the Arabic featured 35 forums, 1,330,999 posts, 58,855 threads, and 28,025 members while the English site featured 41 forums and subforums, with 982,084 messages (or posts), and 13,724 members. The 41

forums range in content from professional ones focused on engineering, architecture, health and studies for example, to forums on hobbies such as cooking and TV, to forums on relationships, jokes, local culture and politics. In terms of poster profiles, circumstantial evidence suggests that the majority of posters are teenagers and young adults. Linguistically, despite such an official division of the website by language, even the most superficial browsing of the English forums makes it clear that forum posters freely post in both Arabic-scripted Arabic and Latin-scripted Arabic, that is, 3arabizi within the English forums whereas the Arabic forums are far more homogeneously Arabic-script in content.

Method

Once the English forums had been selected for further analysis, a purposive sample of all messages posted between March 2007 and May 2008 was collected and compiled into a corpus containing 460,220 messages, spread across 21,626 discussion threads, within 41 topical forums.

In order to categorize each message as containing a particular code, wordlists based on the Arabic Gigaword and the British National Corpus (BNC) were used to scan each message and classify it as being written in Arabic, English, or a mixture of the two. Later, a third wordlist was created to annotate messages written in 3arabizi. While other codes were detected in this process, that is, mixed script codes, a "Muslim English" code (see Mujahid, 2009) and a Non-BNC English code, they accounted for less than 15% of all messages combined. Consequently, they will not be dealt with further here.

Once annotation of messages had been completed, a computation of messages revealed that, overwhelmingly, messages were composed in Arabic (32.3%), BNC English (17.5%), and 3arabizi (35.5%). In other words, these three codes alone accounted for over 85% of all messages in the corpus. Thus, these three dominant codes in the corpus were selected for a thorough cross-linguistic stylistic and topical comparison. In order to carry out this comparison, the ten most frequently encountered lexical words in each of these three languages were identified using the frequency list function of WordSmith 5.0 corpus analysis software.

To sum up, the method adopted in this study is to first identify the ten most frequent open class lexical items in each of the three main monoscriptal codes in the entire corpus, that is, Arabic, BNC English, and 3arabizi. This "first brush" gives an overall sense of what the topical foci of each of these codes might be. Next, the top ten frequent words of each of these codes words are hand-checked using a 100-line concordance in order to establish their respective usage patterns in the corpus, suggesting, in turn, broad stylistic differences between the codes themselves.

Methodological Limitations and Other Considerations

This study describes only the broadest salient topical patterns associated with each code as suggested by the top ten frequent lexical items along with examination of random samples of 100 concordance lines of each of these frequent items. These highly frequent words are used as the measure in determining to what extent each code resembles or differs from the others in terms of topical content (Baker, 2006). Where specific topics, references, and functions are cited for the concordance line of a specific lexical item, it is important to bear in mind that these were determined solely by inferring them from the immediate context of the item within the boundaries of its concordance line of between 10-15 words. This was done because the time-consuming process of referring back to the original message for each of the 3,000 concordance lines in order to specify beyond doubt the topic of each concordance line would have proven highly unfeasible. However, additional clues as to topic, reference, and function of an item were provided by the presence of smileys and other stylistic features such as standard grammar and formal vocabulary. And while several lines appeared relatively ambiguous in terms of topic, such lines still exhibited stylistic features such as smileys or discursive functions such as criticisms. Admittedly, in several cases, an utterance could have been construed as belonging to more than one topic, for example wearing hijab as a form of female dress or as a form of Islamic practice. Callahan (2004) notes that such overlapping and blurring of boundaries between topics is often apparent in corpus discourse analysis, although in the present study consistency in making categorizations and judgments was of utmost importance.

Another important limitation should be mentioned here. The method involved working with data from a single website at a specific juncture in time, that is, the 14-month period between March 2007 and May 2008. Thus, generalizations beyond the data about the functions of 3arabizi, Arabic, and English on other websites or in other contexts are unwarranted. Indeed, especially with reference to topics, it is clear that local, regional and world events will likely have played a role in shaping the content of the corpus so that data collected at the present time from this same website might yield very different results for 3arabizi, Arabic, and English.

A couple of final notes on the citation of example lines from the frequent lexis concordances are in order. First, rather than writing Line 1 in full I employ the shorthand L1 here. Second, in each boxed concordance line, the concordance word has been bolded to set it apart from the other words in the concordance line. Third, where present, smileys are indicated by italics. Fourth, English words written in Arabic script are bolded and italicized. These same conventions are used for the translations of concordance lines provided below the original boxed concordance lines where necessary.

Identifying the Top Ten Open Class Lexical Words in the Main Codes

Once wordlists for each of these codes had been compiled, it was decided to identify the top ten open-class lexical words in each wordlist as a means of detecting topical content, following the lead of Baker (2006). Baker showed that by focusing on only the most frequent lexical items in a given code, it would be possible to generate initial hypotheses about the topical focus of that code. For instance, in the current data set, if the word *Allah* occurred frequently in 3arabizi, it would be worthwhile to explore whether 3arabizi texts might be used to talk about God or religion. Pragmatically, this relatively small number of items also made it easier to compare the surface topical similarities between the main codes and provide a deeper level of analysis for each of these. Clearly, a number greater than ten lexical words could have been selected, but given the vast number of items in each wordlist, a cut-off point had to be selected especially since a certain amount of repetition was observable among frequent lexical items in each wordlist such as Arabic's top ten frequency items *yawm* يوم 'day' and *al-yawm* اليوم 'the day' and BNC English's THANKS (wordlist item no. 93) and THANK (wordlist item no. 116).

In light of the above, the claims made about the topical and stylistic features for each of these codes cannot be taken as absolute or exhaustive for each code. At best, they are an indication of salient themes and styles associated with each code in the context of its most frequent open class lexis. Nevertheless, the in-depth analyses of the top ten lexical items from each of the three main codes did in fact reveal certain distinctive characteristics of each of these three codes.

In order to select the top ten open class lexical items for each code, the UCREL CLAWS7 Tagset² was used as a measure of determining whether a given lexical item was an open class one. In the case of 3arabizi, several homograph cases were encountered in which a word could have been either English or Latin-scripted Arabic. These ambiguous items were hand checked to determine whether they functioned as open class items or closed class items such as prepositions. If an 3arabizi item functioned as an open class noun, adjective, or verb in 50% or more of all cases, it was kept in the 3arabizi top ten list. Once all ambiguous items had been discarded, the remaining top ten lexical items for all three codes were compiled into a table for comparison (see below) and annotated in terms of language (vernacular/standard, formal/informal), topic (sports, religion, relationships, etc.), discursive function (rhetorical question, assertion, exclamation, etc.) (see Callahan, 2004), level

² UCREL stands for University Centre for Computer Corpus Research on Language, a corpus linguistics research centre at Lancaster University, UK. The original CLAWS (Constituent Likelihood Automatic Word-tagging System) is a part-of-speech tagging system developed at Lancaster, Oslo, and Bergen Universities in the early 1980s (Garside, 1987).

of involvement of text composer and/or addressee with the text (involved for first and second person references, noninvolved for third person references), and stylistically with respect to whether it contained smileys or not.

Findings: Stylistic and Topical Functions of Arabic and BNC English

In order to provide a general sense of the kinds of words featured in Arabic, BNC English, and 3arabizi, Table 1 below displays the top ten lexical words for all three codes (note that grammatical or closed class words such as pronouns, articles, determiners, modal verbs, auxiliary verbs, conjunctions and prepositions are not included in the table). Instead, the focus here is on open class content words (lexical nouns, adjectives, verbs and adverbs), which help to reveal more about topics:

Table 1 Top ten lexical words across Arabic, BNC English, and 3arabizi³

Rank	Arabic	BNC English	3arabizi
1	الله 'God'	KNOW	ALLAH 'God'
2	قال 'He said'	THINK	KNOW
3	والله ⁴ 'By God'	GOOD	THINK
4	الناس 'People'	PEOPLE	LOVE
5	يوم 'Day'	LOVE	TIME
6	صلى 'He blessed'	TIME	GOOD
7	بدي 'I want'	SEE	WALLAH ⁵ 'By God'
8	اليوم 'The day'	GO	MAN
9	وسلم 'And he saved'	THANKS	PEOPLE
10	طيب 'Good'	WANT	WAY

Table 1 reveals a number of interesting lexico-semantic similarities across the codes. For instance, Arabic, BNC English, and 3arabizi share one semantically related highly frequent concept in common: People. This indicates that in all three codes references to people are common, suggesting that

³ Words in the BNC English and 3arabizi lists are given in capitals reflecting the WordSmith 5.0 convention of displaying frequency wordlist items in capitals.

⁴ This word can be translated as 'and God' according to context.

⁵ This word can also be translated as 'and God' according to context.

perhaps the topic of People or generalizing statements employing the word *people* may be prevalent across all three codes. Another concept that these codes have in common is Good (see Arabic: item 10, BNC English: item 3, and 3arabizi: item 6). Again, on the surface, these words imply that something (or someone) is frequently described in a positive manner.

Other sets of similarities are discernible between these three codes. For instance, in addition to the concept of People, Arabic and 3arabizi also show the concept of Allah/God to be highly frequent as both codes feature the words ALLAH ('God') and WALLAH ('by God' or 'and God').⁶ Such surface lexical similarities in word list items suggest that perhaps the topic of God or religion may be commonly discussed in both of these codes. When the wordlists of Arabic and BNC English are examined in conjunction, again, considerable overlap is apparent. BNC English and 3arabizi also share a number of lexical items in common. In fact, these codes feature a total of six identical words within the top ten frequent words in their respective subcorpora. In addition to the words PEOPLE and GOOD (which also had semantic counterparts in Arabic), BNC English and 3arabizi have four other top ten words in common: KNOW, THINK, TIME, and LOVE. The words KNOW, THINK, and LOVE suggest that personal viewpoints, opinions, and feelings may often be expressed frequently in BNC English and 3arabizi. As an aside, the fact that 3arabizi shares semantically related concepts in common with both Arabic and BNC English serves to underscore 3arabizi's code-mixed nature as a "fused lect" between Arabic and English (cf. Auer, 1998; McLellan, 2005).

An important cautionary note needs to be borne in mind: Surface similarities should not be taken too uncritically, and without further evidence from samples drawn from specific concordance lines, it would be premature to conclude that these three codes employ the common concepts cited here in the same manner. Indeed, when concordance line data is presented below, ample evidence will be offered to highlight that such seemingly similar lexis is in fact often employed in different ways by users of these three codes.

Having provided a brief overview of the similarities between the three codes, 3arabizi's top ten frequent lexis will now be highlighted and contrasted with both Arabic and BNC English. Table 2, which summarizes the findings for Arabic, reveals some very interesting points about the use of this code in the mahjoob.com forums. For instance, it is clear that Arabic is stylistically heterogeneous, that is, that it ranges from very formal and standard usage indicated by the Modern Standard Arabic (MSA) labelling, to highly informal, involved and nonstandard usage as indicated by

⁶ The only difference between these items is the script in which they are composed in each code, that is, in Arabic the words are written in Arabic script as الله and والله, while in 3arabizi these same words are written in Latin script as ALLAH and WALLAH respectively.

the presence of VA forms within the list. Interestingly, the more formal elements show a clear link to the topic of religion especially to Islam and to the Prophet Mohammed in particular. More of this will be discussed in below. But first, a summary of BNC English's topical and stylistic features will be given.

Table 2 Arabic top ten lexical words showing topical and linguistic features

Rank	Arabic top ten lexical words	Transliteration	Meaning	Language: MSA vs. VA	Recurrent topics	Involved vs. informational	Smileys
1	الله	<i>allāh</i>	'Allah/God'	MSA 66%	Religion (Islam), Christianity, Palestine	Informational 60%	30%
2	قال	<i>qāla</i>	'He said'	MSA 64%	Religion (Islam, mostly Hadith) 64%, humour 36%	Informational 100%	8%
3	والله ⁷	<i>wallāh</i>	'By God'	VA 69%	General, religion (15%)	Involved 67%	32%
4	الناس	<i>al-nās</i>	'The people'	MSA 55%	General, religion (Islam 35%)	Informational 60%	10%
5	يوم	<i>yawm</i>	'Day'	VA 60%	Islam (25%), romance, narratives, politics, jokes, food	Involved 61%	19%
6	صلى	<i>Salla</i>	'May (God) bless (him)'	MSA 100%	Prophet Mohammed (100%)	Informational 100%	0%
7	بدي	<i>baddī</i>	'I want'	VA 100%	General (no religion), songs, food, clothing, relationships	Involved 100%	47%
8	اليوم	<i>al-yawm</i>	'The day' i.e., 'today'	VA 50%	General, news, religion	Involved 66%	32%
9	وسلم	<i>wa-sallam</i>	'And may (God) save (him)'	MSA 100%	Prophet Mohammed (100%)	Informational 100%	0%
10	طيب	<i>Tayyib</i>	'Good'	VA 97%	Jokes, love, food, songs, well-wishing	Involved 80%	74%

As demonstrated in Table 3, in contrast to the Arabic top ten lexical items featured above, the BNC English highly frequent items reflect a much greater topical spread. Additionally, a high percentage of utterances reflect a more involved style of discourse where either first person *I*, *we*, or second person *you* are found as this example from the BNC English concordances illustrates:

32

lol..tough one! i dont know what i would do..*contented*

⁷ This word can also be translated as 'and God' according to context.

Table 3 BNC English top ten lexical words showing topical and stylistic features

Rank	BNC English top ten lexical words	Topics	Involved vs. Informational	Smileys
1	KNOW	General, relationships, songs/lyrics, politics, religion, food, clothing	Involved 90%	25%
2	THINK	General, posting, songs/lyrics, local and social issues, relationships, homosexuality	Involved 98%	7%
3	GOOD	Varied: food, health, politics, fashion, religion, songs/lyrics, hobbies	Involved 58%	32%
4	PEOPLE	Qualified groups, e.g., Jordanians, Jews, Palestinians, etc., generic references, lack of jokes	Informational 59%	8%
5	LOVE (noun or verb)	General: affection, romance, songs/lyrics, hobbies	Involved 84%	19%
6	TIME	General: relationships, food, work, songs/lyrics	Involved 62%	19%
7	SEE	General: songs/lyrics, local culture, posting, relationships, religion	Involved 93%	17%
8	GO	General: songs/lyrics, health and fitness, internet, places	Involved 91%	18%
9	THANKS	General: well-wishing, health/fitness	Involved 100%	58%
10	WANT	General: songs/lyrics, posting, relationships, family/children, local culture, sex	Involved 80%	13%

Findings: Stylistic and Topical Functions of 3arabizi

3arabizi is the most linguistically unconventional of the three codes by virtue of its mixed nature, featuring both English and Arabic lexis, the latter written in Latin script often with numerals. Its linguistic hybridity is observable in its top ten frequent lexical items seen in Table 4 below:

Table 4 3arabizi top ten lexical words showing topical and stylistic features

Rank	3arabizi top ten lexical words	Language: VA vs. English	Topics	Involved vs. Informational	Smileys
1	ALLAH 'Allah/God'	VA 100%	General: relationships, child-bearing, health/fitness, condolences, (no religion)	Involved 100%	42%
2	KNOW	Mostly English	General: marriage/marital status/relationships, fashion, family	Involved 100%	21%
3	THINK	Mostly English	General: Palestine/Israel, gender roles, Islamic practice, posting	Involved 94%	17%
4	LOVE	Mostly English	General: male/female romantic behaviour, romance/marriage, songs/lyrics, food, hobbies	Involved 80%	26%
5	TIME	Mostly English	General: posting, gender issues, Islam, Middle East, business/work	Involved 80%	20%

19 a pink or a blue? blue bi ezr Allah tab3an. How r ur prepar

... a pink (girl) or a blue (boy)? Blue Allah/God willing of course. How r ur prepar...

Others topics concerned health and illness, condolences, food, cars, and specific countries such as Canada, Jordan, and Kuwait.

Although it occurred in seventh place, it is opportune to discuss the lexically and semantically similar term WALLAH 'by Allah/God' at this point. As with its Arabic counterpart, WALLAH functioned mostly as an intensifier in 3arabizi. However, while a few lines of its Arabic counterpart were found to mean 'and Allah/God,' no such usage was detectable for WALLAH in 3arabizi. Stylistically, WALLAH was used almost exclusively in involved utterances while its Arabic counterpart occurred in noninvolved utterances roughly 33% of the time. Regarding smileys, compared to its Arabic equivalent, WALLAH exhibited almost twice the number, that is, 62 of 100 lines, suggesting a comparatively more personalized and light-hearted use of WALLAH in 3arabizi. Further, virtually all lines contained VA as opposed to MSA, underscoring the informal connotation of WALLAH:

86 wallah saba2teeni *stickingtongueout*

Hey, you beat (Fem. Sing.) me to it *stickingtongueout*

56 o5te fa 7adret janabha will stay in amman till aug!!! wak wallah gaharatne

...my sister, so Her Royal Highness will stay in Amman till August!!! Anyway, she really used to boss me around

In terms of topics and functions, a whole range was apparent: school subjects, food, mobile phones, money, sports, smoking, posting to mah-job.com, jokes, shopping malls, summer vacation, cars, downloading CDs, references to the Middle East such as places, and people such as Jordanian girls and Saddam Hussein. Others lines concerned wearing hijab, friends, family, marriage including choosing a wife, romantic relationships, and relationship advice. Discursively, self-disclosure statements and personal narratives were very common as were well-wishing statements, exclamations, questions, opinions, and assertions. Briefly, WALLAH was similar to its Arabic counterpart in terms of topics but had apparently no connection to the theme of religion.

The next set of 3arabizi items discussed here are the stative verbs KNOW and THINK, both also found in the BNC English top ten list. Interestingly, in clear contrast to their BNC English counterparts, both KNOW and THINK were frequently accompanied by Latin-scripted Arabic items such as discourse markers e.g. *ba3den* 'and then', *5ala9* 'that's enough' or the Arabic subordinate conjunctions *inno*, *eno*, and *enno* 'that he/she/it is' or *eny* 'that I am'. As ostensibly English-language items, perhaps it is not surprising that their respective concordance lines contained relatively little Latin-scripted Arabic

compared to both ALLAH and WALLAH, which featured such items in almost each line of their concordances. Nevertheless, sporadic use of Vernacular Latin-scripted Arabic appeared to underscore text-producers' attempts to forge a direct link to local, popular Arab culture. Some noteworthy Latin-scripted items present in the concordance lines were Arabic proper names such as *Tattar*, cultural terms such as *a7maq* 'fool', *ashkaljeyeh* 'trouble-maker', *fay3a* 'hip, cool', *9atyat* 'rude girls' or very short phrases and exclamations like *(ma) 2dert 2adal sakta anymore* 'I couldn't keep quiet anymore'.

Topically, like their BNC English counterparts, KNOW and THINK exhibited a variety of themes: video clips, food, gender issues, relationships, single life and marriage, health, (female) dress and clothing, children and family, friends, music and Arabic-language songs, Islam and Muslims, morality including terrorism, career/work, studying, politics including references to Arab-related places and politics especially Palestine and Israel, and forum posting. Stylistically, the vast majority of lines revealed involved style with *//i* or *you/u* as the most frequent subjects. Both concordances exhibited a mix of formal and informal English alongside 3arabizi, especially Netspeak features. Discursively, both KNOW and THINK were similar, featuring assertions, opinions, and self-disclosure statements, as well as various types of questions, though THINK also revealed several statements of intent. Regarding smileys, KNOW had 21 lines with smileys while THINK had only 17 lines, suggesting that more serious discussion often took place with these words as seen here:

93 It is completely illogical to think that blowing yourself up

In brief, KNOW and THINK behaved similar to their BNC English counterparts with the exception that VA elements occurred, typically highlighting Arabic cultural content such as names, expressions and exclamations.

LOVE was the fourth most frequent 3arabizi item. It should be noted that as with BNC English, in 33 lines LOVE was found to function as a smiley (see Footnote 84 above). And in three more lines, LOVE was part of an author ID, that is, Happy Love. Consequently, as was done for its BNC English counterpart, the 33 concordance lines containing the smiley LOVE were eliminated from the concordance and a randomized sample of 33 new concordance lines containing valid cases of LOVE was collected and appended to the original concordance in order to carry out a fuller analysis. And as with BNC English, the 3arabizi top item LOVE featured several topics in common with its BNC English counterpart as well as topics observed across other BNC English and 3arabizi frequent item topics: social commentary and critique, posting, personality types, well-wishing, a Qur'anic verse translated into English, and references to music, both Arabic and English-language as seen here:

47

ek o3'neyeh ismha ever lasting love *blushingface*(Do you) have a song named 'Everlasting Love'? *blushingface*

Unsurprisingly, as with its BNC English counterpart, a large number of LOVE's lines dealt with topics related to love: male and female romantic behaviour and across cultures, falling in love, relationships and advice, and marriage and proposals. Regarding discursive function, LOVE was often used phatically toward an addressee, that is, in *love u* and *love ya* with or without a term of endearment such as *sís*. Further, stylistically, 80 concordance lines exhibited involved discursive features while 26 contained smileys, indicating an overall personalized style in the use of the word LOVE. Other utterances featured narratives, personalized questions, assertions, and especially positive evaluations of places such as Jordan-specific people, for example: *i love her outgoing personality* or *I just love this guy huggingfriend*, and even local food such as the popular Middle Eastern vegetable stew, *Molokhia*:

94

I love el mlo5eyyeh *lovefilled*I love molokhia *lovefilled*

Notice this use of Latin-scripted Arabic content for local cultural references as seen with the other frequent 3arabizi lexis. Again, Arabic discourse markers, expressions, and exclamations were also observed.

The next 3arabizi item was TIME, also found in BNC English. In terms of topics, the same kinds of themes were discovered as with its BNC English counterpart and elsewhere in the other concordances: forum members and their posts, sports like football and games, cooking and food, photography, gender issues and differences, and female rights, for instance not wearing hijab, Islam, its teachings, religious leaders and followers, Middle East politics, rulers, and wars involving Palestine, Israel, Lebanon, and Afghanistan, playing songs such as English songs as well as Latin-scripted Arabic references to Arabic songs and singers, business, work, study, time management, vacations, friendship, relationships, marriage, motherhood and child rearing, and health and skin care.

Beyond specific references to Arabic proper nouns such as *3olama2* (i.e. *ulama* علماء 'Ulema,' Islamic religious scholars), in the TIME concordance lines, Latin-scripted Arabic items, while relatively infrequent, served similar functions as seen before: exclamations, untranslatable expressions, and discourse markers. Stylistically, 80% of TIME's lines were involved. However, only 20 lines contained smileys, suggesting that most utterances were more serious than frivolous. This was seen in several utterances featuring self-disclosure statements, serious questions, criticisms, personal narratives, warnings, assertions, and advice, using expressions containing

TIME: "at the same time," "any time," "at this/that time," "some time," "from time to time," "the first/last time," "a long time ago," and "it's time to."

The sixth most frequent 3arabizi item was GOOD, which also occurred in BNC English, and it resembled its BNC English counterpart in several ways. First, GOOD in 3arabizi had a similar number of lines containing smileys to its BNC English counterpart (36 and 32 respectively). Next, both concordances featured a majority of involved utterances. However, GOOD in 3arabizi exhibited substantially more involved lines than in BNC English, that is, 87% vs. 58%. Nonetheless, both concordances featured either positive or negative, that is, "not good," evaluations of people (e.g., mahjoob posters) and things. Moreover, personalized greetings (e.g., "good morning/evening/night"), well-wishing (e.g., "good luck"), compliments (e.g., "good job/one"), questions about quality (e.g., "is it good?"), and advice (e.g., "a good way") were common in both 3arabizi and BNC English uses of GOOD. 3arabizi topical similarities to the BNC English concordance of GOOD were evidenced by references to sports such as football, food and cooking, health and fitness, work, study and careers, pastimes such as songs, art and photography, posting to, reading, and moderating the forums as well as discussing or addressing specific posters, and Middle East politics including anticorporatism. Curiously, there were no obvious references to religion. Other common references involving GOOD in 3arabizi were to love, marriage, relationships, parents, children and childrearing, clothing, cars, and shopping.

The next most frequent 3arabizi item was MAN, which had no counterpart item in the top ten lists of the other two codes. Hand checking of the concordance revealed that in 85 lines it was used to refer to males. The remainder of instances were either references to author IDs, for example, "K_man," football clubs, for example, "man city" for Manchester City, or Latin-scripted Classical Arabic where "man" means 'who' or 'whoever'.¹⁰ Also, three more lines were examples of the MSA relative pronoun *man* من 'who' that had been transcribed using Latin script. In each of these cases, quotations in Classical Arabic were evident. In terms of discursive function, references to MAN were found in 46 lines to consist of vocatives and/or exclamations rather than as subjects or objects of verbs:

42 specially for zalmate offersflower Welcome Back, man huggingfriend Weenak

...specially for my man offersflower Welcome Back, man huggingfriend where've you been?

In this example, notice the semantic redundancy of Jordanian VA *zalamate* 'my man' and the vocative use of English *man* later on. Such utterances underscore the

¹⁰ Two more lines were excluded because they appeared to have been wrongly identified as 3arabizi due to verse numbers being attached to the first word in each verse, creating pseudo-arithmographic Latin-scripted Arabic items such as "8For man did not come from woman...."

use of MAN to express peer relationships between males. In this regard, there were no less than 11 occurrences of the awkward-sounding hybrid English cum Latin-scripted Arabic expression *ya man* 'hey, man,' which combines the Arabic vocative marker *ya* meaning 'hey' or 'yo' with the English word *man*, as exemplified here:

40 *eyeswatering* wallah ya man kolo tamam bs zae ma zolt enta elsho'3ol fo2 rasi

eyeswatering really, man, everything is fine but as you said, work is over my head

In fact, involved utterances using MAN were evident in 74 lines out of the 85 lines where MAN occurs meaning 'male.' Stylistically, VA and Netspeak were very frequently encountered and mixed throughout the concordance in over 50 lines, further suggesting informal communication. Among these utterances, complimenting, greeting, inviting, well-wishing, and thanking were very common. Moreover, 35 of the 85 lines contained smileys as the above examples illustrated, indicating informality, playfulness, and affection. In this last connection, the public expression of affection and emotion between males, which is very acceptable in Arab culture, was frequently observed here as suggested by the *huggingfriend* and *eyeswatering* smileys. This is so despite pressure on males to project a virile heterosexual image toward others as seen here:

39 man ana shaaab mish benet *beatinghandwithbatbeatinghandw*

Man I am a guuuyyy not a girl *beatinghandwithbat beatinghandwithbat*

96 i love you man (7ub akhawi bas) *offersflower offersflower*

I love you man (but only brotherly love) *offersflower offersflower*

77 hate it when you see a nickname of a man that says something like, Strawberry

This last example appears to have been written by a female. Nonetheless, it underscores expectations for men to be macho on mahjoob.com (see references to gay friends in BNC English above).

Recurrent topics were marriage, divorce, and relationships including desirable qualities in a male partner, women's rights vis-à-vis men, harassment, and male-dominated society. Topics common to the other top ten concordances were computers, food, TV and movies, money (e.g., "*money can't buy u...a decent man*"), American politics, Middle East government and politics involving Jordan, Palestine, and Israel, Islamophobia and Anti-Shi'ism, and childrearing. Discursively, several lines were parts of narratives or jokes. The remaining utterances consisted of assertions, self-disclosure, and questions often expressing incredulity such as "*Man get a grip, what the hell are you talking about?*"

PEOPLE was the next item in the 3arabizi list, also found in BNC English. In contrast to MAN, PEOPLE featured in fewer lines with Latin-scripted Arabic, that

is, 29 out of 100. Apart from discourse markers, Latin-scripted Arabic here tended to consist of hard-to-translate expressions and proper names such as *majlis nowab* 'assembly of deputies.' Stylistically, while involved usage was found in 72% lines, smileys were found in only nine lines. Further, Netspeak was found in less than one third of the concordance. Forty percent of lines concerned people in a general sense, indicating that generalizations were relatively common. Combined, these features suggest generally involved but serious discussion, as was the case with BNC English PEOPLE. This observation was confirmed by the relatively weighty topics frequently encountered: relationships and marriage, gender issues, warning and criticisms about posting to mahjob.com, study and careers, appearance and dress, nonhumorous narratives, politics and economics, especially of Palestine and Jordan, social issues such as war, injustice, corruption, poverty, and unskilled social classes, and, related to these previous themes, Islam at the centre of a theological and social debate including references to jihad and de facto religious police as seen here:

55 does sharee3a allows people to become ameer by force too?

Does Sharia (Islamic law) allow people to become rulers by force too?

93 religious groups to run wild in the country and apply islam on poor people....

In terms of discursive function, references to specific kinds of people were usually part of generalizing assertions about "other people," "few people," "some people," "many people," "lots of people," "most people," and "people you know." More descriptive references were to "old people," "Muslims," "people in Jordan," "Maan and Zarqa people," and "our people." Briefly, assertions and opinions followed by questions were the most typical types of utterances involving PEOPLE, as was the case in BNC English.

The final 3arabizi item was WAY. Typically, this word occurred in expressions describing a manner or method of doing or being as in "a timely and prompt way," "the same way," "is no way to treat...," and "a sane way." Other examples were as parts of discourse marker expressions such as *by the way*, *any way*, or as an amplifier, for example *way better* and "no way u can compare." Occasionally, WAY preceded prepositional phrases as in "Islam is the way of life" and "your twisted way of thinking." Stylistically, 74% of utterances were involved, though smileys were found in only 18 lines. As with PEOPLE above, WAY appeared to be featured most often in serious topics: health and fitness, gender and equality, family issues, marriage and relationships, and heated discussions about moderating and freedom of speech in posting:

4 u are distroying this site by your way, and treat us as ur childs and u are the fathers here,

5 follow up on your word and keep this thread. Freedom of speech is a two way road, after

Other more serious topics¹¹ were politics and social criticism, especially involving Arabs in general as well as Palestine, Israel, Lebanon, and political parties like Hamas, Fatah, Hezbollah, and the Muslim Brotherhood, and religion, especially Islam including this attack on Islam presumably by a Christian poster:

94 christians to adopt islam, but not the other way around? shu el islam msh le3beh?

...(ok for) Christians to adopt Islam, but not the other way around? What?! Islam isn't a game (and Christianity is)?

Less serious topics were also present such as food and cooking, songs, especially Arabic ones featuring Latin-scripted Arabic singers and song titles, jokes, cars, and computers.

As for Latin-scripted Arabic, as seen in the rhetorical question above, Arabic language expressions were often employed in order to add emphasis to an assertion, a question, or a suggestion. Other utterance types were statements of self-disclosure, narratives, advice, descriptions, and compliments.

Summary and Conclusions

Table 6 summarizes the main features of 3arabizi when compared with both Arabic and English within the corpus:

Table 6 Topics and stylistics of 3arabizi, Arabic, and BNC English

Feature	3arabizi	Arabic	BNC English
Lexical items	People, God	People, God, want, know	People, want, know
Topics	Broad range of topics: mostly light, especially humorous with cultural references	More restricted range of topics, strong focus on Islam and politics	Broad range of topics: mostly light, but several serious social and taboo issues such as homosexuality and pre-marital sex raised
Stylistics	Highly involved and informal style including a very high proportion of smileys, 3arabizi most common in discourse markers	Often informational and formal, but also some involved, informal language, smileys rare	A mix of formal and informal with more involved style and smileys than Arabic but less than 3arabizi

It can be concluded that Arabic exhibited the closest link to the topic of religion, especially Islam, as evidenced by several of its items and its numerous stylistically Classical and MSA utterances. Surprisingly, though, Arabic also fea-

¹¹ Generally, topics touching on politics, religion, nationality, and social issues were labelled *serious*, whereas those pertaining to hobbies and pastimes were dubbed *nonserious* or *light*.

tured numerous vernacular forms, signalling a clear break from accepted practice when writing Arabic perhaps due to the online environment. Remarkably, Arabic diglossia between MSA/Classical Arabic as the high language and VA as the low language seemed to be reproduced in CMC texts examined here. Indeed, VA lexis was used for more mundane and frivolous topics, underscoring the role of vernacular style as a common feature of humorous style while Classical/MSA lexis appeared primarily in religion-related lines. This seems to concur with Bentahila's (1983) findings based on spoken contexts about the functional and topical distribution of Classical Arabic and VA in Morocco.

In contrast to Arabic, BNC English featured a more diverse variety of topics ranging from hobbies to work and study, from computers to cooking, and from religion and politics to cars, with a range of styles from formal English grammar, spelling, and punctuation to informal Netspeak-style English. Interestingly, BNC English, in particular, revealed references to relatively taboo and sensitive topics such as homosexuality, sex, and women's rights, perhaps indicating that such "Western" topics and issues were better expressed in a language other than Arabic.

3arabizi exhibited a similar range of style and was also topically closer to BNC English with a diffuse range of topics overall. However, in contrast to BNC English, the frequent samples of Latin-scripted Arabic in 3arabizi helped to draw a clear link between it and local VA culture as typified by the frequently phatic use of Latin-scripted Arabic lexis such as ALLAH and WALLAH. As with Arabic, 3arabizi vernacular use often betrayed humour and levity. That several items in 3arabizi were identical to items in both Arabic and BNC English emphasized that it is a linguistically-mixed code (cf. McLellan, 2005; Smedley, 2006).

In brief, 3arabizi, when compared to the other two principal codes in the corpus, appears to serve more phatic functions especially when its Arabic items such as ALLAH and WALLAH are used. However, its relatively frequent English content underscores its status as a mixed code reflecting a glocal reality in which English script (i.e., Latin script) and lexis link its users to the wider world while its Arabic lexis and discourse markers help these same users to maintain connections to their local Arabic roots.

There are several implications of this study for further research in the field. First, the demonstrable existence of vibrant hybrid forms of language such as 3arabizi in CMC contexts invites further research into such mixed codes that clearly reflect glocalness. Second, in terms of literacy, it is evident that the development of new user-driven written genres in the absence of institutional or educational support is not only possible, but may even be widespread. Third, the phenomenon of script-switching and borrowing implied by the existence of 3arabizi poses important questions about the cognitive processes entailed when such borrowing occurs. Fourth, for the field of corpus linguistics,

the method used here has shown that a multilingual corpus can be profitably annotated and compared for lexical, topical, and stylistic features across codes.

Ultimately, the very existence of 3arabizi as a unique glocal linguistic phenomenon suggests that in an ever shrinking world, the seemingly futile aspirations for expression of cultural autonomy and individuality in the face of globalizing and homogenizing forces can in fact be realized in the form of hybrid codes such as 3arabizi, providing fascinating sociolinguistic compromises that straddle and bridge the global-local divide.

References

- Al Share, B. (2005). *A sociolinguistic analysis of Jordanian Netspeak (JNS)* (Unpublished master's thesis). Jordan University of Science and Technology, Amman, Jordan.
- Androutopoulos, J. (2007). Language choice and code switching in German-based diasporic web forums. In B. Danet & S. C. Herring (Eds.), *The Multilingual Internet: Language, culture, and communication online* (pp. 340-361). Oxford: Oxford University Press.
- Auer, P. (1998). From code-switching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *Interaction and Linguistic Structures*, 6.
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Bentahila, A. (1983). *Language attitudes among Arabic-French bilinguals in Morocco*. Clevedon: Multilingual Matters.
- Bianchi, R. (2012). 3arabizi – When local Arabic meets global English on the Internet. *Acta Linguistica Asiatica*, 2(1).
- Callahan, L. (2004). *Spanish/English codeswitching in a written corpus*. Philadelphia: John Benjamins.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Featherstone, M., Lash, S., & Robertson, R. (1995). *Global modernities*. London: Sage.
- Garside, R. (1987). The CLAWS word-tagging system. In R. Garside, G. Leech, & G. Sampson (Eds.), *The Computational analysis of English: A corpus-based approach*. London: Longman.
- Koutsogiannis, D., & Mitsikopoulou, B. (2003). Greeklsh and Greekness: Trends and discourses of *glocalness*. *Journal of Computer-Mediated Communication*, 9(1).
- Lee, C. K.-M. (2007). *Multilingual, multiscriptal, and multimodal text-making: Instant messaging among young people in Hong Kong* (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.
- McLellan, J. (2005). *Malay-English language alternation in two Brunei Darussalam on-line discussion forums* (Doctoral dissertation). Curtin University of Technology, Perth and Sydney. Retrieved from <http://adt.curtin.edu.au/theses/available/adt-WCU20060213.092157/>
- Mujahid, A. M. (2009). Can there be Muslim English? Retrieved April 7, 2009, from <http://www.soundvision.com/Info/education/muslimenglish.asp>
- Nishimura, Y. (2003). Linguistic innovations and interactional features of casual online communication in Japanese. *Journal of Computer-Mediated Communication*, 9(1).
- Palfreyman, D. (2001). Informal Latinized orthographies. *Linguist List*, 12(12-2760). Retrieved from <http://www.ling.ed.ac.uk/linguist/issues/12/12-2760.html>

- Palfreyman, D., & Al Khalil, M. (2003). "A funky language for teenzz to use:" Representing Gulf Arabic in instant messaging. *Journal of Computer-Mediated Communication*, 9(1).
- Pavlenko, A., & Blackledge, A. (2004). Introduction: New theoretical approaches to the study of negotiation of identities in multilingual contexts. In A. Pavlenko & A. Blackledge (Eds.), *Negotiation of identities in multilingual contexts*. Clevedon: Multilingual Matters.
- Smedley, F. (2006). *Code-switching and identity on the blogs: An analysis of Taglish in computer mediated communication* (Master's thesis). Auckland University of Technology, Auckland, Australia. Retrieved from <http://repositoryaut.ionz.ac.nz/theses/378/>
- Tseliga, T. (2007). "It's all Greeklsh to me!" In B. Danet & S. C. Herring (Eds.), *The multilingual Internet: Language, culture, and communication online* (pp. 116-141). Oxford: Oxford University Press.
- Warschauer, M., El Said, G. R., & Zohry, A. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication*, 7(4).