

**Arab English: The Case of 3arabizi/Arabish on Mahjoob.com**

Robert Michael Bianchi

Virginia Commonwealth University in Qatar

rmbianchi@qatar.vcu.edu

**Abstract**

This article describes a doctoral study on the variety of Arab English known as 3arabizi/Arabish, which occurs commonly in computer-mediated communication (CMC). The research context is a series of discussion forums on mahjoob.com, a Jordanian website where, alongside 3arabizi/Arabish, English and Arabic are both used. The study explains the process of developing a multi-lingual corpus from the forum data. It then focuses on the occurrence and use of 3arabizi/Arabish, English, and Arabic according to text type, topic, and poster profile. It also discusses the stylistic and topical features of these three languages. The study concludes with an examination of an on-line debate where the three languages are used strategically by posters to portray distinct identities and ideologies. 3arabizi/Arabish, English, and Arabic are found to carry out distinct functions and roles for their users.

**Key words**

Arabic, English, corpus, writing, CMC

**Introduction**

Ever since the advent of the Internet, English has played a dominant role in computer-mediated communication (CMC) (Crystal, 2001). This situation was partly caused by the use of the ASCII encoding system, which was composed of the Latin letters and numerals as featured on the QWERTY keyboard ( Palfreyman & Al Khalil, 2003). As a result, other language groups were at a disadvantage in the early days of CMC, especially if their languages were written in a non-Latin script. With regard to Arabic-speaking CMC users, this meant that initially they were forced either to use English (or French) or forego using the Internet to communicate with one another due to the lack of support for Arabic script. However, over time, a solution emerged: to hybridize the two languages by writing Arabic using the letters and numerals of the QWERTY keyboard. The linguistic offspring of this on-line hybridization is 3arabizi/Arabish, a mixed code containing English elements interspersed with Latinized Arabic strings. Nowadays, Unicode has largely replaced ASCII as the most widespread encoding system, allowing users to type in their own native scripts. Yet despite access to Arabic script, many Arab CMC-users are still using 3arabizi.

It is worth noting that Arabic is not the only non-Western language to be affected by such keyboard-based technical limitations; Latinization has been forced upon other languages, too, such as Greek (see Tseliga, 2007), Chinese (see Su, 2003), Japanese (see Nishimura, 2003), and Russian (see Mironovschi, 2007). Regarding Arabic specifically, a number of small-scale studies have been carried out to investigate its use in e-mail (Warschauer, El Said, & Zohry, 2002), SMS messaging (Palfreyman & Al Khalil, 2003), and IRC chat (Al Share, 2005). However, to date, no study has

attempted to rigorously compare and contrast the co-occurrence of 3arabizi/Arabish with other languages such as Arabic and English in the on-line context on a single multilingual, multi-scriptal website. Further, outside this study, no other study has collected such a vast corpus of data, just under half a million forum messages. In this light, the current study is unique in its ability to present salient findings from such a large data set.

The website that was used to compile the data set is mahjoob.com, owned and maintained by Emad Hajjaj, a popular Palestinian-Jordanian political cartoonist whose work focuses on Middle Eastern issues (Cagle, 2009). Hajjaj’s website, mahjoob.com has several thousand posters and, at the time of data collection in May 2008, featured 41 topical forums divided into 21,626 threaded discussions. Figure 1 below shows the main portal of the website:



Figure 1: Screenshot of the Mahjoob.com website main portal

From the screen shot of the portal, it is clear that the website is multilingual; Arabic-scripted items are found alongside English items. A bit less conspicuous in this example, but much more prevalent within the discussion forums themselves is the use of 3arabizi as seen under the “Discussions” list on the bottom right of the screen shot. Here we have the 3arabizi forum entitled “e7ke wfadfed”, meaning ‘speak and get it off your chest’. Clearly, as this small example indicates, this website offers a rich context to examine code and script choice as will be seen below.

In order to contextualize the study further, it will be helpful to present the four original research questions that underpinned the research. These questions built successively upon each other. In other words, by addressing the first question, the necessary data was collected to approach the second question and so on. Here are the four questions:

1. What are the various types of code and script-switching featured on the English website of Mahjoob.com?
2. How are script-code pairings distributed across certain types of postings in terms of
  - a. text type (title, first message, following messages, quoted content)?
  - b. overarching forum topic?
  - c. thread length?
  - d. author posting frequency?
3. What do frequent lexical items in the main script-code pairings in the corpus reveal about the topical content and stylistic features of those pairings?
4. How do forum posters use the available script-code pairings to construct identities?

These questions were explored using a mixture of quantitative methods primarily from corpus linguistics and qualitative methods from discourse analysis. Key theoretical terms relevant to this study were the notions of linguistic code, script, orthography, spelling, and identity. The pairing of a specific script, i.e. writing system, with a specific lexicogrammar such as Arabic or English is taken to constitute a linguistic code. Thus, Arabic written in Latin script represents a new linguistic code to be contrasted with Arabic written in Arabic script. Orthography provides a further nuance to the classification of codes. Based on Sebba (2007, p. 56), it is argued here that orthography and spelling, like other aspects of discourse (see Wodak et al. 1999; Fairclough, 2003), are social practices and that acts of spelling, especially systematically unconventional ones, are acts of identity (see Pavlenko and Blackledge, 2004).

### **Literature Review**

At this juncture, it will be opportune to discuss some of the previous works that inform this study. Auer (1998, 2008) deplores a monolingual bias in code-switching research which makes *a priori* assumptions about the existence of distinct and discrete linguistic systems which are then mixed in the speech of bilinguals to produce code-switching. Instead, Auer (1998) posits the possible existence of mixed codes or “fused lects” as the normative code of interaction among certain groups. Such a categorization naturally blurs the lines between discrete linguistic varieties. Indeed, prior research into Latin-scripted Arabic in CMC contexts seems to point toward the existence of such a hybridized form of language which, while using the Latin script, incorporates lexicogrammatical elements from both Arabic and English (Abdallah, 2008; Al-Tamimi & Gorgis, 2007; Al Share, 2005; Palfreyman & Al Khalil, 2003; Warschauer et al., 2002). Further, most of these studies also point to a unique feature of this hybrid language, namely the use of numerals as graphemes in

order to represent Arabic sounds which have no ready or widely agreed upon equivalents in the Latin alphabet, a phenomenon which has been labeled “arithmoglyphemes” (Bianchi, 2005). Despite such earlier work, to date, almost no studies have systematically contrasted the use of Latin-scripted Arabic with English and Arabic-scripted Arabic within a corpus of texts (but see Warschauer et al., 2002). And apart from the present one, studies that focus on code-switching between Arabic and English within web forums have yet to be carried out, though studies involving code-switching between English and other languages such as Bruneian-Malay exist (see McLellan, 2005).

Beyond language mixing, language use and identity feature prominently in the present research. Butler (1990) and Rampton (1999, 2007) explore the notion that language users copy the styles of others from out-groups in order to produce distinct identities. In this vein, Fairclough (2003, p. 161) introduces the notion of “characters”; each culture possesses a set of “characters” or character types which Fairclough describes as “its culturally most salient identities” (ibid.).

Now that some of the relevant preliminary research has been surveyed, the data and methodology featured in the study will be presented in the next section.

### **Data and Methodology**

To address the above questions, the help of corpus expert, Sebastian Hoffmann, was enlisted to collect a purposive sample of all messages posted between March, 2007, and May, 2008, on the English website of mahjoob.com (Hoffmann, 2009). Consequently, 460,220 messages, found within 21,626 discussion threads, were downloaded and compiled into a corpus. The discussion threads, in turn, were found within 41 topical forums. Once this had been completed, threads from each forum were purposively sampled and read in order to determine the main topic of each forum. These main topics were then collated into eight over-arching topics. Table 1 below shows the final eight overarching topics to which each of the 41 topics was assigned.

Next, in order to determine the languages that were featured within the corpus, language-specific wordlists were used based on the Arabic Gigaword and the British National Corpus (BNC) wordlists. This meant annotating each message to indicate whether it was composed in Arabic, English, or a mixture of the two. To achieve this, a Perl script was applied to the corpus. A third wordlist was developed to indicate messages written in 3arabizi. To illustrate the results of the annotation process, Figures 2 and 3 below show a bi-scriptal message before and after annotation respectively.

**Table 1: Eight over-arching topics and all 41 forums in the data set**

| No | Topic  | Forum(s)   |
|----|--|--|
| 1  | Humor and jokes                                  | <i>Joke Zone</i>   |
| 2  | Poetry   | <i>A7la Alkalam, Copied material, San7 w Le3b W Jadd w 7obb, and Soul Retreat</i>  |
| 3  | Fields of work and study                         | <i>Engineering, Graphic Design and Architecture, Health and Science, and Politics Business and Economy</i>   |
| 4  | Friends and family                               | <i>Family Matters, Pink and Blue, and Wishes and Greetings</i>   |
| 5  | Local/regional culture, nationality and politics | <i>Arab Arts Symposium, Iraq under Occupation, Kuluna Al Ordun, Palestine, Patriotic and Spiritual Art, and Religion Forum</i>   |
| 6  | Hobbies and pastimes                             | <i>Art Gallery, Automotives, Chef's Corner, Entertainment, Home Decoration, Humanities, Male Chef Corner, Mobiles, SIG (Special Interest Groups), Sports, Tech Talk, TV and Movies, and World Talk</i> |
| 7  | Gender and age                                   | <i>Girls Talk, Men's Corner, and Teens Thoughts</i>  |
| 8  | General discussion/opinion                       | <i>Announcements, Dababees and Hala 3ammi, e7ke w fadfed, General Forum, On Focus, Sentiments, and Suggestions and Complaints</i>  |



**Figure 2: Screenshot of a biscriptal message on mahjoob.com (Hoffmann, 2009, p. 6)**

```

<message n="10">
<date id="22nd April 2007, 10:16:21 PM">
<author id="amar^14">
<location id="jordan">
<script id="mixed" ratio="1.16037735849057" bias="balanced">
<AG id="yes">
<code_9>
<arabic_s>
<content>
<new> it depends on 2 things ان الصداقة او الصداقة بين الاثنين
طبيعة العلاقة او الصداقة بين الاثنين وطبيعة المشكلة المشكلة ان كانت كبيرة قد تتغلب على صداقة سنوات و الصداقة ان
كانت حميمة قد تتغلب على اكبر المشكلات <smiley>foolishsmacksownhead</smiley> complicated !
<smiley>stickingtongueout</smiley> never happened to me though , usually me and my frnds get over our fights directly ,
el7amdolellah <smiley>blushingface</smiley> </new>
</content>

```

**Figure 3: Corpus-annotated version of the biscriptal message shown in Figure 2 above<sup>1</sup>**

<sup>1</sup> In Figure 2, tagged elements such as "<smiley>blushingface</smiley>" indicate graphical smiley emoticons (e.g. 😊).

Despite their obvious differences, the actual message text shown in Figure 2 has been captured in Figure 3 (this text is enclosed by the tags <new> and </new>). Nevertheless, Figure 3 clearly resembles the kind of HTML source code one might expect to underpin the webpage data in Figure 2 minus the boilerplate elements. However, additionally, Figure 3 features several annotational tags that would not be found in the original HTML source code such as <code\_9> which refers to the linguistic code in which the message was composed and <arabic\_s> (i.e. Standard Arabic) which refers to the type of Arabic found in the message.

In the process of annotating Arabic, BNC English, and 3arabizi messages, several more types of messages were discovered: those that contained varying mixtures of Latin script and Arabic script (Mixed Script messages), those that contained typically single Islam-related Arabic items transliterated without arithmographemics (Salafi English messages), and those that contained English items alongside items that could not be categorized as either English or Latin-scripted Arabic (Non-BNC English messages). In total some 16 combinations of Arabic, English, or both written in either Latin script and/or Arabic script were identified. But given the paucity of several of these combinations, a conflation of smaller combinations of mixed script texts was effected, resulting in the following six linguistic codes for text-based messages in the corpus:

1. Arabic-scripted Arabic (Code 1)
2. BNC English (Code 2)
3. 3arabizi (Code 3)
4. Mixed Script (i.e. Latin script with Arabic script elements) (Codes 4-9 & 11-13)
5. Salafi English (Code 10)
6. Non-BNC English (Code 14)

Messages were also annotated to indicate other features such as forum, thread and thread code, order within a thread, author ID, and date of posting. Once this had been done, a second corpus was built in SPSS so that various features of the messages could be compared and tabulated for statistical significance. As shown in Table 1 above, given the relatively large number of forums (41 in total), it was decided to code these into eight overarching topics: 1) Humor, 2) Poetry, 3) Work and study, 4) Friends and family, 5) Local culture 6) Hobbies, 7) Gender and age groups, and 8) General discussion. Next, several SPSS-based cross-tabulations were carried out to determine code distribution across thread titles, seed posts, main messages, and messages with quoted material. Subsequently, code occurrence was determined across forum, forum topic, thread length, and prolific posters vs. non-prolific posters. These results were all tested for significance using chi-squared. The results of these cross-tabulations will be presented in the “Findings” section below.

In order to address distinct topical and stylistic differences within the corpus, the three predominant codes in the corpus, Arabic-scripted Arabic, BNC English, and 3arabizi were analyzed by focusing on their ten most frequent open-class lexical

content items, i.e. semantic nouns, verbs, and adjectives. Table 2 shows the list of top ten lexical items for each of these three codes:

**Table 2: Top 10 lexical words across Arabic-scripted Arabic, BNC English, and 3arabizi<sup>2</sup>**

|    | Arabic-scripted Arabic       | BNC English | 3arabizi                          |
|----|------------------------------|-------------|-----------------------------------|
| 1  | الله<br>God                  | KNOW        | ALLAH<br>God                      |
| 2  | قال<br>He said               | THINK       | KNOW                              |
| 3  | والله <sup>3</sup><br>By God | GOOD        | THINK                             |
| 4  | الناس<br>People              | PEOPLE      | LOVE                              |
| 5  | يوم<br>day                   | LOVE        | TIME                              |
| 6  | صلى<br>He blessed            | TIME        | GOOD                              |
| 7  | بدي<br>I want                | SEE         | WALL<br>AH <sup>4</sup><br>By God |
| 8  | اليوم<br>The day             | GO          | MAN                               |
| 9  | وسلم<br>And he saved         | THANKS      | PEOPLE                            |
| 10 | طيب<br>Good                  | WANT        | WAY                               |

In this study 100-line random concordances were compiled for each top-10 open-class lexical item, resulting in 3,000 lines for analysis. Each line was hand-checked for topical references and stylistics by adapting Biber, Conrad, and Reppen’s (1998) method for categorizing utterances for function. This involved reading each concordance line to determine whether it was stylistically informal or formal based on its lexis and the presence of vernacular items and involved (1st and 2nd person pronouns) or informational (3rd person pronouns) markers. Additionally, the presence of smileys was investigated as these were seen to be further markers of informal style. Here is an example of one of the concordance lines taken from the 100-line concordances of the 3arabizi word *Allah*; the translation is provided immediately underneath:

|    |  |
|----|--|
| 19 | a pink or a blue? blue bi ezn <b>Allah</b> tab3an. How r ur prepar |
|----|--|

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

In the example above, it is evident that the utterance is involved because of its inclusion of a second person singular “ur”. It is informal stylistically because of its

<sup>2</sup> 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.

<sup>3</sup> This word can also be translated as “and God” according to context.

<sup>4</sup> This word can also be translated as “and God” according to context.

use of Netspeak forms such as “r” for *are*. The presence of 3arabizi forms is also clear in the phrase “bi ezn Allah tab3an” meaning “God willing of course”. Analyzing concordance lines such as the one above for Arabic-scripted Arabic, BNC English, and 3arabizi helped to indicate which topics were most frequently discussed in each of these codes in the corpus and what kinds of stylistic and communicative functions tended to be performed using each code.

The final research question regarding the strategic use of codes for identity-construction was addressed by conducting a discourse analysis of a lengthy and linguistically-heterogeneous discussion thread. The chosen thread was entitled “Masha2allah Masha2allah Masha2allah” (lit. “What God has willed, what God has willed, what God has willed”). This thread was selected because of its sheer length, i.e. 322 messages long; its linguistic diversity, since it featured examples of Arabic, English, 3arabizi, and Salafi English; and its uniqueness: it had a greater number of Salafi English messages than any other thread. The thread in question featured a heated debate concerning whether it was permissible within Islam to question and admonish the governmental authorities for blatantly un-Islamic behavior. For this analysis, the notion of “linguistically realized styles” was employed (see Fairclough, 2003, p.162).

Four principal debaters were discovered in the argument, i.e. Kharoof Tayeh, Guillotine, Snipe\_aac, and Muslim4. They were then grouped into anti-establishment and pro-establishment sides and their choice of language was examined in light of their respective ideological positions and the salient “characters” (2003, p.174), i.e. readily recognizable stereotypes they assumed through their writing. The argument starts off with the following 3arabizi message by the anti-establishment poster, Kharoof Tayeh:

Message 1:

- 1 wbeejo begoolooly enno elhai2a btenteqed elmashayekh.  
*And they come and tell me that the organization criticizes the Sheikhs?*
- 2 6ayyeb...does anyone know the phone number for the hai2a?  
*Fine then organization*
- 3 I'd like to inform them that 300 billions of gulf money is helping the economy of  
"their enemy" as they claim. 🙄<sup>5</sup> *whistling smiley*<sup>5</sup>

This message is interesting because it shows the anti-establishment tone of Kharoof Tayeh who writes in 3arabizi as evidenced by the mixture of English with Latin-scripted Arabic items featuring arithmographemes such as ‘elhai2a’ (the moral police) and ‘6ayyeb’ (fine). Also, Kharoof Tayeh uses smileys, which adds informality to his message.

---

<sup>5</sup> Here the use of the *whistling smiley* 🙄<sup>5</sup> underscores Kharoof Tayeh’s position that the religious authorities feign ignorance when it comes to Gulf countries sending billions of dollars to purportedly anti-Islamic Western economies such as the US, the UK, and the rest of Europe who recognize and support Israel while waging war against Muslim countries, such as Iraq and Afghanistan, as alluded by Kharoof Tayeh in Message 1.

Now that the data and methodology have been briefly discussed, the key findings of the study will now be presented.

### **Findings**

The following bulleted list summarizes the most salient findings with regard to each of the original research questions:

1. The code and script switching found on the English website of mahjoob.com:
  - The most common form of script-switching is the use of Latin script with arithmographemics to transliterate (mainly) Vernacular Arabic, resulting in 3arabizi.
  - The English-website contains more 3arabizi than Arabic-scripted Arabic and more of the latter than it does BNC English messages, suggesting that English is not always the code of choice in all CMC contexts.
  - Messages featuring script switching (i.e. both Arabic script and Latin script together) are surprisingly rare, suggesting that changing between scripts may be inconvenient and/or impossible for many posters.
  - Numerous messages contain only imported content or non-linguistic content.
2. The distribution patterns of the codes:
  - A. Text type findings:
    - Discussion thread titles are composed in BNC English and Arabic-scripted Arabic rather than 3arabizi, perhaps because 3arabizi and BNC English are more informational in function.
    - Of the three main codes, 3arabizi is the most used code in following messages while BNC English is the least used, suggesting differences in function
  - B. Overarching topics:
    - Arabic-scripted Arabic is used for Poetry, Humor, and Local culture forums
    - BNC English is used for Work/Study related forums
    - 3arabizi is used for General discussion forums
  - C. Thread length:
    - Longer threads are composed in Arabic-scripted Arabic, shorter threads in 3arabizi
  - D. Posting frequency:
    - Top 10 posters compose in Arabic-scripted Arabic, non-top 10 posters in 3arabizi
    - Top 10 posters have a tremendous impact on code use, accounting for 20% of all following messages and 12% of all Arabic-scripted Arabic messages.
    - Top 10 posters mostly post in Joke Zone and Poetry-related forums
3. The topical and stylistic features of Codes 1, 2, and 3 as indicated by top 10 lexis
  - Arabic-scripted Arabic is topically and stylistically-dichotomous, where Arabic vernacular is used for humorous content and phatic functions while

MSA/Classical Arabic is used for religion, highly reminiscent of Arabic diglossia in face-to-face contexts.

- BNC English is topically diverse and contains references to taboo topics.
- 3arabizi is mainly English linguistically, and its Latin-scripted forms are usually discourse markers, expressions, and Arab cultural references.
- 3arabizi is more phatic than Arabic-scripted Arabic and BNC English, with frequent use of smileys.

#### 4. The use of codes to construct identities

- 3arabizi indexes solidarity with other mahjoobians and young Arabs
- Arabic-scripted Arabic indexes scholarship and authenticity
- Salafi English indexes Salafist ideology
- Posters choose which codes to use based on topic, level of formality, and interlocutor

The most salient findings overall:

- the English-website is highly heterogeneous and script-switching, though rare, does occur in fact occur.
- although codes do overlap in topical content and function to a certain extent, Arabic, English, and 3arabizi appear to have distinct functional uses.
- 3arabizi is the most common code in the corpus, but is mainly English with a smaller amount of actual Latin-scripted Arabic.
- mahjoob.com provides a space for Arabic bilinguals to engage in a wide range of topics including politically-charged and taboo ones such as political and religious dissent, sex, and sexual orientation.

It is worthwhile to discuss these findings in more detail. First, the considerable linguistic heterogeneity of the corpus was clearly established, highlighting the overall dominance of the mixed code, 3arabizi in the corpus followed by Arabic-scripted Arabic, and then BNC English. Evidently, mono-scriptal codes are by far the more popular choice among the website posters, a hypothesis posited by Al Share (2005), which she attributes to the inconvenience involved in changing keyboard sets while composing messages based on the results of her own study of IRC involving Jordanian Netspeak and English (see Al Share, 2005), an assumption that has been hard to test on a large-scale basis until now. In this regard, certain codes were found to associate more clearly with certain forums such as Arabic-scripted Arabic with poetry, humor, and religion-related forums. cursory observation of these forums suggests that this is so because, on several occasions, content in these forums may have been copied and pasted from other Arabic-scripted Arabic websites on the Internet. As a corollary, English-dominant forums such as Health and Science perhaps reflect a similar process of large-scale importation of texts such as health-related news articles. Consequently, 3arabizi-dominant forums appear to be those that feature the most original content composed primarily by Mahjoobians themselves.

Regarding the findings related to the second research question, a number of observations can be made here: first, the 14 textual codes of the corpus differ in both quantity and distribution across the corpus. Second, mono-scriptal codes are by far the most common type of code in the corpus, especially Arabic-scripted Arabic, BNC English, and 3arabizi. Third, topic has a certain relationship to code choice where Humor, Poetry, and Local culture-related forums reveal more Arabic-scripted Arabic use, while Work/study forums favor BNC English, and General discussion, Age/gender-related, Friends/family-related, and Hobby-related forums exhibit more 3arabizi use. In this regard, to a certain degree, the findings seem to parallel Bentahila's study (1983) highlighting the functional distribution of Classical/Modern Standard Arabic, Vernacular Moroccan Arabic, and French in Morocco, where Classical and Standard Arabic are used for religious and classical literary domains, Vernacular Arabic for intimate and informal settings, and French for professional and academic purposes. Lastly, in terms of mahjoob.com authors' code use patterns, prolific authors favor Arabic-scripted Arabic over both BNC English and 3arabizi which are preferred by the bulk of non-prolific authors for composing their following messages. This concurs with the finding that prolific authors post very frequently to the Joke Zone forum in Arabic-scripted Arabic, which in itself may betray frequent importation of humorous material composed in Arabic from other humorous websites, though such an assertion is difficult to test given the problems associated with determining ultimate authorship on the web (Hoffmann, 2007).

The findings stemming from the third research question reveal that Arabic-scripted Arabic, BNC English, and 3arabizi, while exhibiting degrees of topical overlap, were clearly distinct in several ways. For instance, of the three codes, Arabic-scripted Arabic seemed to be most closely connected to the topic of religion and to exhibit a formal and non-involved style. On the other hand, BNC English and 3arabizi were much more involved in style and featured a much broader range of topics which paralleled the major topics of the website's forums. Yet despite certain topical and stylistic similarities between these two codes, in contrast to BNC English, 3arabizi was most clearly connected to Vernacular Arabic and local Arab culture, featuring a relatively light, humorous, and playful style as evidenced by the preponderance of smileys found in the concordances of its top 10 frequent lexical items. The awareness of such discursive function and stylistic differences between the three major codes of the website helped provide the necessary background knowledge in order to carry out the micro-level analysis of identity-creation as required by the fourth main research question of the study. Indeed, the step-by-step investigative process guided progressively by the first, second, and third research questions made it possible to then examine whether the distinct linguistic codes in the corpus could be used strategically in support of identity-construction. The results in this regard were positive. Specifically, it was found that, whether written in Arabic-scripted Arabic or Latin-scripted English, stylistically-formal utterances as marked by distinct lexis and orthography were used systematically to establish erudite and pro-establishment identities that contrasted rather sharply with the vernacular and informal lexis and orthography of the anti-establishment posters. In this regard, perhaps due to the relative novelty and flexibility of its sociolinguistic context, mahjoob.com appears to

provide an example of bilingualism without diglossia insofar as its main codes often overlap in their topics and discursive functions despite observable trends for specific codes to be used for certain topics, levels of formality, and communicative functions (see Fishman, 1967).

### **Implications of the Research**

These findings question the notion that English is the preferred language of the Internet since 3arabizi and Arabic-scripted Arabic were actually more prevalent than English-only messages in the corpus derived from this English-language website (cf. Danet & Herring, 2003). Of particular interest, was the finding that Vernacular Arabic has become a written variety in CMC contexts, challenging the primacy of Modern Standard Arabic as the only legitimate form of written Arabic. Furthermore, the widespread use of 3arabizi when compared with the more conventional code-script pairings represented by Arabic-scripted Arabic and BNC English was highly salient. Perhaps, the potential value of 3arabizi, as a pedagogical aide in the development of literacy (even bi-literacy) among young Arabs should not be ignored (cf. Maamouri, 1998).

On a separate note, the mahjoob.com corpus itself represents a valuable resource for cross-cultural discourse analysis in the post-9/11 era. Concordances of terms such as “the West”, “America”, “men”, “women”, “Islam”, etc., both in Arabic and English may yet reveal very interesting findings.

From a sociolinguistic perspective, the fact that 3arabizi was found to be the most common code for composing following messages in the corpus, suggests the vitality of a form of language that is at once innovative, hybrid, unplanned, unofficial, and unsupported. In this light, it is fascinating to consider whether a grassroots, youth-driven 3arabizi movement across the Arabic-speaking world might be capable of challenging the primacy of Modern Standard Arabic in a way akin to Demotic Greek’s supporters’ struggle against, and ultimate triumph over, Katharevousa’s supporters. 3arabizi may represent a sort of linguistic “Arab Spring”.

The originality of the thesis in terms of content lies in the fact that, to date, there appear to have been no studies done on script and code-switching involving Arabic and English within discussion forums, and certainly not with a corpus of almost half a million messages. In this sense, the data have helped to provide yet another angle from which to investigate Arabic-English code-switching and specifically, script-switching involving the Arabic and Latin scripts in particular. Nevertheless, despite the interesting findings noted above, it is not warranted to generalize these beyond the data presented here. This is because code-switching is a situated practice, reflecting the peculiarities of a given context. For instance, it may be the case that a large number of “Mahjoobians” (i.e. posters to mahjoob.com as they often style themselves) are based outside the Arab world and consequently have to use Latin script even though they would prefer to use Arabic script. Indeed, the lack of direct feedback from forum posters on their code use represents a major drawback of this

study, though such author feedback on code use would naturally reflect self-perceptions of the authors' behavior rather than their actual behavior.

### **Conclusions**

This research has sought to investigate areas to which scant attention had been paid in the field of sociolinguistics until relatively recently: written code-switching and on-line code and script choice in support of identity creation. In this regard, the study has highlighted several important aspects of such linguistic behavior. First, humans as complex linguistic beings are capable of appropriating pre-existing linguistic practices and re-molding them to suit their own communicative and identity-related purposes. In so doing, they are able to create highly innovative linguistic varieties such as 3arabizi which then acquire their own stylistic and communicative values and functions separate from their putative linguistic forbears, i.e. Arabic-scripted Arabic and Latin-scripted English. Then, in similarly creative fashion, language users further innovate by strategically taking up (or shunning) these linguistic varieties and their features in order to create distinct and salient identities which are readily recognizable to their interlocutors. All this occurs in a medium devoid of physical voice and sound. In some sense, one might be led to believe that the physical world is somehow being reproduced in the virtual world by authors where orthography takes the place of accent. However, the possibility for script-switching adds an element of style which has no parallel in the spoken world. As a result, it is clear that the on-line medium allows for certain nuances of identity expression that are unavailable in face-to-face environments. As such, mahjoob.com forums appear to constitute an on-line community practice, an amalgam of influences and inputs that represent a different kind of reality, no matter how similar they are to face-to-face reality, resembling as they do face-to-face clubs, pubs, and meeting rooms. Similarly, the written codes of mahjoob.com, exemplified best by the hybrid code of 3arabizi, are an amalgam of influences that, while deriving much input and structure from face-to-face forms of language, are clearly distinct from these in both form and function. Indeed, the on-line identities created by Mahjoobians through their skilful use of written codes, though certainly inspired by real-life character types, are nevertheless only made possible by virtue of the on-line medium in which they are situated. Thus, mahjoob.com, perhaps above all else, is a reminder of the fascinatingly complex interplay between language user and medium, and may be a harbinger of potentially greater technologically-influenced language use, language change, and language-related identity creation in the future.

## References

- Abdallah, S. (2008). Online Chatting in Beirut: Sites of Occasioned Identity-Construction. *Ethnographic Studies*, 10, 3-22.
- Al-Tamimi, A., & Gorgis, D. T. (2007). Romanised Jordanian Arabic E-messages. *The International Journal of Language Society and Culture*, (21). Retrieved from <http://www.educ.utas.edu.au/users/tle/JOURNAL/>.
- Al Share, B. (2005). *A Sociolinguistic Analysis of Jordanian Netspeak (JNS)*. Unpublished Master's Thesis. Jordan University of Science & Technology. Amman.
- Auer, P. (1998). From Code-switching via Language Mixing to Fused Lects: Toward a Dynamic Typology of Bilingual Speech. *Interaction and Linguistic Structures*, 6.
- Auer, P. (2008). The monolingual bias in bilingualism research - or: Why bilingual talk is (still) a challenge for linguistics Retrieved from [www.germanistik.uni-freiburg.de/auer/?download=monolingual\\_bias\\_in\\_bilingualism\\_research.pdf](http://www.germanistik.uni-freiburg.de/auer/?download=monolingual_bias_in_bilingualism_research.pdf).
- Bentahila, A. (1983). *Language attitudes among Arabic-French bilinguals in Morocco*. Clevedon, Avon, England: Multilingual Matters.
- Bianchi, R. (2005). *Revolution or Fad? Latinized Arabic Vernacular*. Paper presented at the 11th TESOL Arabia Conference, Dubai.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Cagle, D. (2009). Daryl Cagle's Political Cartoonists Index Retrieved from <http://www.cagle.com/politicalcartoons/PCcartoons/hajjaj.asp>.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Danet, B., & Herring, S. C. (2003). Introduction: The Multilingual Internet. *Journal of Computer-Mediated Communication*, 9(1).
- Fairclough, N. (2003). *Analysing Discourse Textual Analysis for Social Research*. London and New York: Routledge.
- Fishman, J. A. (1967). Bilingualism with and without diglossia; diglossia with and without bilingualism. In L. Wei (Ed.), *The Bilingualism Reader* (2 ed., pp. 47-54). London: Routledge.
- Hoffmann, S. (2009). *Corpus linguistics and the Internet*. Paper presented at the English Language Research Seminar, University of Liverpool.
- Maamouri, M. (1998). *Language Education and Human Development: Arabic Diglossia and its Impact on the Quality of Education in the Arab Region*. Paper presented at the Mediterranean Development Forum, Marrakech, Morocco.  
<http://www.worldbank.org/mdf/mdf2/papers/humandev/education/maamouri.pdf>.
- McLellan, J. (2005). *Malay-English language alternation in two Brunei Darussalam on-line discussion forums*. (PhD), Perth and Sydney, Curtin University of

- Technology. Retrieved from <http://adt.curtin.edu.au/theses/available/adt-WCU20060213.092157/>.
- Mironovschi, L. (2007). Russian SMS compliments. [Article]. *Written Language & Literacy*, 10(1), 53-63.
- Nishimura, Y. (2003). Linguistic Innovations and Interactional Features of Casual Online Communication in Japanese. *Journal of Computer-Mediated Communication*, 9(1).
- 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 Ltd.
- Rampton, B. (1999). Language Crossing and the Redefinition of Reality. In P. Auer (Ed.), *Code-switching in Conversation*. Florence, KY, USA: Routledge.
- Rampton, B. (2007). Language crossing and the problematisation of ethnicity and socialisation. In L. Wei (Ed.), *The Bilingualism Reader* (pp. 177-202). Oxon: Routledge.
- Sebba, M. (2007). *Spelling and Society: The Culture and Politics of Orthography Around the World*. Cambridge: Cambridge University Press.
- Su, H.-Y. (2003). The Multilingual and Multi-Orthographic Taiwan-Based Internet: Creative Uses of Writing Systems on College-Affiliated BBSs. *Journal of Computer-Mediated Communication*, 9(1). Retrieved from <http://www.ascusc.org/jcmc/vol9/issue1/>.
- Tseliga, T. (2007). "It's All Greeklish 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).
- Wodak, R., de Cillia, R., Reisigl, M., & Liebhart, K. (1999). *The Discursive Construction of National Identity* (A. Hirsch & R. Mitten, Trans.). Edinburgh: Edinburgh.