Latinised Arabic and connections to bilingual ability

Mariam Aboelezz
Lancaster University

Abstract

As software support for non-Latin scripts is becoming more readily available, the continuing use of Latinised forms in online discourse highlights an interesting phenomenon. This paper focuses on Latinised Arabic (LA) as one manifestation of this trend. While there appears to be significant variation in the conventions used to Latinise Arabic in association with regional vernaculars, there is evidence that LA is now being used for more than just online communication.

Thus far, researchers have dealt with LA as a form of script-switching. In this paper, LA is examined as a form of code-switching to investigate links to bilingual ability. A statistical comparison between emails from two emailing groups indicates a link between the use of English and LA. A close examination of code-switching sites reveals a number of associated trends, while email content suggests a number of factors that influence language and script choice.

Introduction

The last few decades have seen a rapid diffusion of technology in many communities across the globe. Computers became commonplace and quickly evolved, mobile phones soon followed, but it was perhaps the Internet that marked the technological revolution of the age. For quite some time, English, with its Latin (or Roman) character set, was the patron language of these innovations (Crystal, 2006). A person with no knowledge of these characters stood little chance of gaining technological literacy. Even as software support for non-Latin scripts became more accessible, technology has hardly lost its association with the English language – or more precisely, with Latin script.

As the languages of the world came under pressure to meet the urgent demands presented by rapid, English-dominated technological advancements, an apparent trend began to develop: the Latinisation of the characters of many non-Latin script languages (Danet & Herring, 2007). This paper steers attention towards this latter phenomenon with a focus on Latinised Arabic.

Latinised Arabic: An overview

What is it?

Latinised Arabic is a written form of Arabic that uses Latin or Roman characters as an alternative orthographic form of Arabic language, which normally employs Arabic script. Being a relatively new phenomenon that has featured in very little research to date, researchers have still to agree on a common term for it. So far, the process of Latinising Arabic has also been referred to as 'Romanization' (Beesley, 1998), 'ASCII-ization'[1] (Palfreyman and Al Khalil, 2003), as a form of 'transcription' and even as 'transliteration' (Language Analytics, 2007), although this last designation is criticised as being 'inaccurate' by Beesley (1998). More recently, Latinised Arabic has been given names such as 'Arabizi' (Yaghan, 2008), 'Frankoarab' and 'Arabish' (e-magazeen: see section 3).

Yaghan (2008) notes that proposals for the Latinisation of Arabic date back to 1880, with the first comprehensive scheme for Latinised Arabic proposed in 1940. In both instances, Egypt was at the centre-stage of these proposals, and the Egyptian vernacular was the subject of what became a highly politicised debate (Khayat, 2004). Faced with severe opposition from conservatives, such calls were soon abandoned. On the other hand, Palfreyman and Al Khalil (2003) point out that established conventions of Latinising Arabic have existed for some time, and coin the term "Common Latinised Arabic (CLA)" to refer to the 'standard' form influenced by these conventions. The purpose of these has been to make Arabic more accessible to non-speakers of Arabic.

---

[1] In reference to ASCII (American Standard Code for Information Interchange)

One of the most common and authoritative of these conventions is the *ALA-LC* Romanisation scheme developed by the Library of Congress (Barry, 1997: 10-19).

However, the form of Latinised Arabic discussed here (henceforth referred to as LA) is very different from its predecessors. The largely voluntary use of this volatile form is a clear indication that it is very unlikely to be backed by a political agenda of any sort. Unlike older standards, LA was not developed with a specialized purpose in mind. It is not aimed at making Arabic accessible to foreign speakers, but is rather used for communication between speakers of Arabic. The widespread use of LA, and the unusual manner in which some characteristic Arabic sounds are denoted, e.g. turning CLA names such as 'Khaled' and 'Ghada' into '7'aled' and '3'ada', set LA apart from the more traditional Latinisation conventions known to date.

There is little agreement on exactly when this contemporary form of LA came into existence, although its rise has been inextricably linked to that of the Internet. This may account for what appears to have been a steady growth of this phenomenon over the past 15 years or so. Indeed, it is suggested that LA first emerged as a reaction to the domination of Latin script in the world of technology, as is the case with a number of other non-Latin script languages such as Greek and Japanese (Bianchi, 2006). Even as support for non-Latin scripts became more readily available, LA has still remained widely popular. This is possibly because Arabic users were accustomed to typing on an English (QWERTY) Keyboard (Language Analytics LLC., 2007) which, for Arabic-speakers living abroad, may sometimes be the only typing option (see section 2.4). The association of LA with a younger generation of technology users also makes it something of an icon for a new teen identity that actively contributes to its viability (cf. Palfreyman and Al Khalil, 2003).

Initially, LA was bound to technology-mediated communication (e.g. through text messages on mobile phones, web chat, emails, etc), but there is increasing evidence that its use has spread to other domains which are not restricted by Latin script. Today, LA continues to be seen in text messages and in internet blogs, discussion forums and chat (Palfreyman and Al Khalil, 2003), while simultaneously spreading to offline mediums, including cartoons (Palfreyman and Al Khalil, 2003), handwritten notes and writings on walls (Yaghan, 2008), and even some printed magazines.

One graphological feature of LA, which is similar to a parallel trend in Latinised Greek cited by Androutsopoulos (2006), is the use of numbers and digraphs (consisting of a number and a diacritic apostrophe) which resemble Arabic letters to represent sounds that are not present in the English language. Palfreyman and Al Khalil (2003) list these as shown in Table 1.

Sound (IPA)	Arabic character	ASCII Representation	
		In Palfreyman & Al Khalil (2003)	In Present Study
/ʔ/	ء	2	2
/ħ/	ح	7	7
/x/	خ	'7 / 5	7' / kh
/ʕ/	ع	3	3
/ɣ/	غ	'3	3' / gh
/tʼ/	ط	6	t
/ðʼ/	ظ	'6	z / th
/sʼ/	ص	9	s
/dʼ/	ض	'9	d

Table 1 Difference in graphological features of LA based on a sample of Gulf Arabic vs. a sample of Egyptian Arabic

Arabic Diglossia and Latinised Arabic

We cannot hope to acquire an understanding of the context of LA without some insight into the language situation in the greater part of the Arabic-speaking world. Arabic is the official language of 25 countries. Broadly speaking, there are two main classes of Arabic: Modern Standard Arabic (MSA), a descendant of Classical Arabic (which lives on in the Qur'an today), and regional Arabic vernaculars or Colloquial Arabic. A diglossic relationship exists between the two, where MSA is considered the pan-Arab written "high" variety of education and religion, while colloquial Arabic is the spoken, "low" variety which varies from one region to another (Ferguson, 1959; Holes, 1995).

This presents a complexity when dealing with LA, as the Latinised form of Arabic is often the spoken form, which essentially reflects the regional variety that the user/speaker is accustomed to (Bianchi, 2006). This means that LA is not based on one standard variety of Arabic. Differences in regional varieties extend to differences in the graphological depiction of "characters" in LA. The consonant representations from Palfreyman and Al Khalil (2003) in Table 1 are representative of one possible variety of LA, namely Gulf Arabic. However, if the same sounds were to be represented in LA according to a different spoken variety of Arabic then the results may be significantly different. The same table illustrates the LA representations found in the samples analysed in this paper, which are all based on Egyptian vernacular. For the sake of simplification, no distinction is made between colloquial Arabic and standard Arabic in the remainder of this paper.

The Arabic language has four pairs of plain/emphatic consonants as displayed in Table 2. In the Latinised representations based on Gulf Arabic cited by Palfreyman

and Al Khalil, the emphatic sounds had their own representations. However, in the sample of Egyptian Arabic studied, these Latinised emphatic consonants were completely absent. Instead, these sounds shared the same representations of their non-emphatic (plain) counterparts. A possible explanation for this may be that emphatic sounds are more characteristic of spoken Gulf Arabic than they are of Egyptian speech, where emphatic consonants are sometimes replaced by other sounds (for a detailed discussion of the phonological features Arabic dialects, see Holes, 1995).

	Dental				Inter-dental			
	Plain		Emphatic		Plain		Emphatic	
Fricative	/s/	س	/s′/	ص	/ð/	ذ	/ð′/	ظ
Plosive	/t/	ت	/t′/	ط				
	/d/	د	/d′/	ض				

Table 2 Plain and emphatic pairs in the Arabic Language

Another variation was seen in the position of the apostrophe in a Latinised digraph, which sometimes appeared to the right of the number rather than to its left. Moreover, the Arabic sounds (خ) and (غ) are sometimes represented as <kh> and <gh> respectively. This suggests the influence of CLA on how users represent Arabic online (cf. Palfreyman and Al Khalil, 2003).

More than just transcription

Although technology has come a long way in accommodating non-Latin scripts, the domination of Latin script can still be seen even today. One area where this is particularly apparent is in the Domain Name System (DNS), where the ASCII code still has the upper hand (Huston, 2008). It is of little surprise then that Latinised Arabic domain names such as *yallakora* (Go Football), *masrawy* (into Egypt), *otlob* (order) and *yallabina* (let's go) exist, even when the website content is entirely in Arabic script (or sometimes entirely in English). More interesting for this study are domain names such as *fann3arabi* (Arabic art), *6arab* (music) and *a7la* (better) with the distinctive numerical markers of contemporary LA.

The fact that some of the domain names above are used for websites where the content is entirely in English raises an important question: If the website was clearly designed to address speakers of English, then why is the domain name in Arabic? Clearly, LA here has a purpose which extends beyond transcribing Arabic words due to technological limitations. A website like www.otlob.com for example represents an online ordering service based in Egypt. The website content is entirely in English, possibly to appeal to a wider audience of multinational companies as well as non-Arabic speakers living in Egypt. However, the domain name is in (Latinised) Arabic, which hints at the location of the service. The domain name is obviously designed to catch the eye of someone familiar with Arabic. Another example is www.6arab.com,

where the domain name suggests to the Arabic internet user that this is a website for Arabic music, setting it apart from other music websites on the internet. This would seem to imply that LA serves a function that extends beyond facilitating access to technology for Arabic users.

LA in emailing groups: A case study

Following the overview of LA in the previous section, the discussion now proceeds from description to analysis. This section describes a statistical analysis carried out on emails from two emailing groups in order to discover any links between bilingual ability and the frequency of use of LA. An attempt is also made to detect any code-switching trends by examining code-switching sites in the two groups, and explore the influences governing code/script choice. Therefore, the questions addressed in this section are:

- How frequently is LA used? (Section 2.2)
- When is LA used? (Section 2.3)
- What influences code/script choice? (Section 2.4)

The terms *code* and *script* are central to the discussion in this section. *Code* is used here in a broad sense to mean a language as well as any language varieties (standard or colloquial) and style variations (formal or informal) within that language (Romaine, 1994: 121). *Script* is used to refer to a writing system consisting of characters which represent elements expressible in language (Huston, 2008). Throughout this section, reference will be made to two codes (Arabic and English) and two scripts (Arabic script and Latin script). It is worth noting that the main concern in this section is not LA as a kind of script-switch (cf. Bianchi, 2006) between Latin script and Arabic script, but rather as a kind of code-switch between Arabic (whether in Latin or Arabic script) and English. Based on the emails studied, the possible combinations of these codes and scripts (henceforth referred to as *format*) are listed below:

- Arabic in Arabic script only
- Arabic in Latin script only
- Mixture of Arabic in Arabic script and English
- Mixture of Arabic in Arabic script, Arabic in Latin script and English
- Mixture of Arabic in Latin script and English

All references to names and places in the examples cited have been omitted to ensure the confidentiality of the participants, but any typing errors or emoticons have been preserved. Interestingly, proper names were always in CLA, demonstrating its clear influence in this aspect.

Description of the two mailing groups

The two groups referred to in this paper are designated Group A and Group B. Both groups consist mainly of Egyptian university students. Both groups include male and female members, and all the members are native speakers of Arabic. The two groups are 'listed' as English according to the formal description on their websites, however, the email content varies considerably as will be shown. All the emails reviewed from these groups were exchanged between 2003 and 2004.

Group A consists of 50+ members aged 18 to 25. It is associated with a simulation of the Arab League organized and run by university students. The mailing group is designed to plan and discuss upcoming meetings and events. This study included 168 emails from this group, representing all the emails exchanged between 2003 and 2004 after forwarded messages and automated notifications have been excluded. Although the group is listed as English, the activity it is associated with does not require the members to be bilingual. Some of the members are comfortably bilingual in English and Arabic but the majority of members are only *incipient bilinguals* (Diebold, 1964); they can understand English and can probably produce a few words in it, but cannot communicate in English as comfortably as they could in Arabic.

Group B consists of 20+ members aged 16 to 21. It is associated with a club within an English teen magazine. The mailing group is designed to coordinate and discuss article-writing and to plan meetings and events. This study included 85 emails from this group, representing all the emails exchanged between 2003 and 2004 after all forwarded messages and automated notifications have been excluded. The magazine this groups is associated with requires its members to be fluent in English. Members are required to produce quality written material in English, and so are only allowed to join the magazine on the basis of their competence in English. Being also native speakers of Arabic, all the members of Group B are comfortably bilingual in Arabic and English. It might be argued that the members of Group B do not necessarily have native-like control of English, and accordingly do not satisfy the requirements for Bloomfield's (1933: 56) definition of bilingualism. However, in contrast to Group A, all of the members of Group B are at least bilingual according to Haugen's (1953: 7) definition; they are able to produce complete, coherent statements in English. The following section investigates whether this difference in bilingual capacity between the two groups links to the choices of language varieties.

How frequently is LA used? A statistical analysis

The language varieties used in Groups A and B were recorded and statistically analysed to note trends. Figure 1 illustrates the percentage of emails in each format in the two groups.
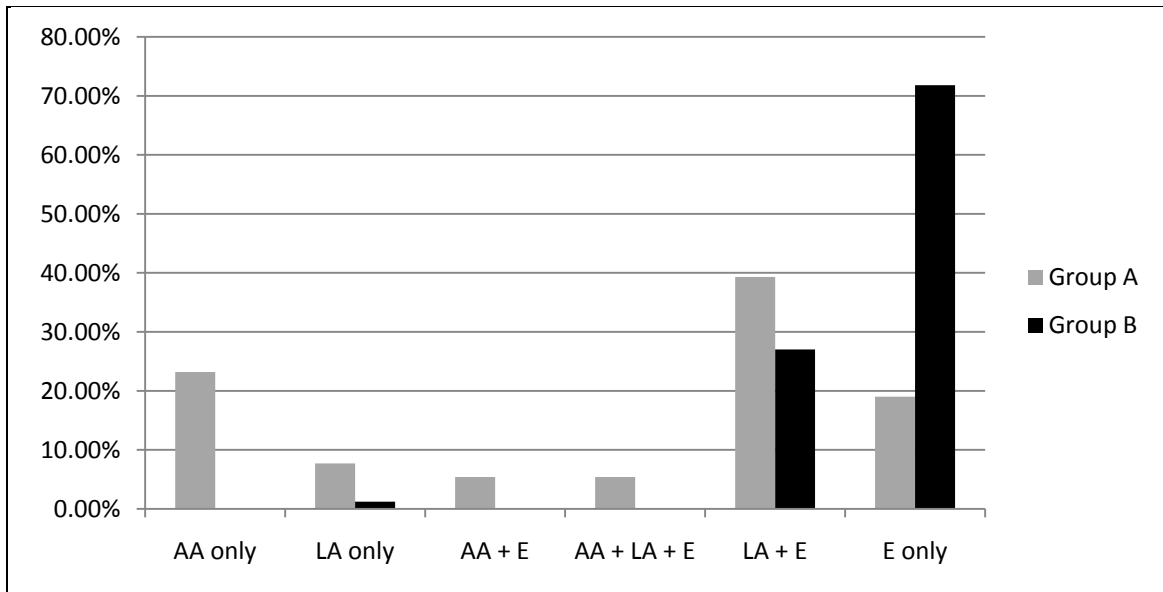
Figure 1 Percentage of emails displaying each format in Groups A and B

As Figure 1 illustrates, the email formats in Group A vary between Arabic in Arabic script (AA), Latinised Arabic (LA) and English (E), and on occasion a hybrid combination of two or more of these forms. Out of 168 emails written by 21 distinct authors, the majority of emails (81%) include Arabic (whether in LA or AA), while 69% include some amount of English. Exactly half of the emails (50%) are written in a combination of English and Arabic (whether LA or AA), while a little more than half the emails (52.4%) display instances of LA. In most cases where LA is one of the language choices used, it amounts to at least 60% of the written text.

Table 3 is a further breakdown of the language choices by member. The names of 16 distinct authors have been coded into alphabets (*a-p*) after excluding contributors of less than 3 emails due to non-significance. Mapping out the results in this manner highlights a number of interesting points. Apparently, there is a great tendency to mix and alternate between formats, with only one member (*n*) using a single format throughout. In fact, two members alone, *(c)* and *(d)*, account for half the total number of emails in the LA+E category. Except for (*n*), those who use AA are just as likely to alternate between formats. However, those who use AA alone never use LA alone. In addition, except for (*g*) and (*n*), all the members use some form of LA at least once and except for (*n*), all of the members use some E at least once. In sum, the more a person uses E, the more likely they are to use LA.

Email Format	Number of emails per distinct author																Total
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	*o*	*p*	
AA	9			1	8		9	2						8			39
LA		6	4						1						2		13
AA+E				1	3	1						2	2				9
AA+LA+E	1							1				6				1	9
LA+E			19	15	3	8			1	2	5		8		1	2	66
E	1	1	3	3	4	2	8	1	2	2	1		3				32
Total	11	7	26	19	16	13	18	4	4	4	6	8	13	8	3	3	163

Table 3 Breakdown of the email formats in Group A by member

In Group B, most of the emails exchanged are entirely in English (71.8%). However, a significant number of emails contain occasional instances of LA (27%). None of the emails contain any Arabic script. 11 of the 17 distinct authors (64.7%) alternate between using English and a combination of LA and English in their emails, while the remaining 6 (35.3%) use English only. It is worth noting that even where LA occurs it does not amount to more than 5% of the written text in most emails. The only exception is an email comprised of one sentence which is a formulaic greeting on a religious occasion written entirely in LA. This one-line email and its translation are shown in Excerpt (1) below.

> (1) kol sna w anto tybeen mould nbwi mubark
> 'Season's greetings. Wishing you blessings on the Prophet's birthday.'

When is LA used? An examination of code-switching sites

Code-switching can be defined as a shift within the same speech exchange from one language (code) to another, where a language is identified as a distinct grammatical system or subsystem (Gumperz, 1982: 59). Singh (1985: 34) also refers to *code-mixing* to indicate a language switch that occurs within the same sentence. According to Singh, code-switching refers only to cases where the language switch marks a distinctly different unit of speech. In this paper, the term code-switching is used in a sense that is inclusive of the notion of code-mixing.

Code-switching in Group A

Extensive code-switching between Arabic and English occurs in many of the emails in Group A. In many cases, the writers alternate between code and switch choices so frequently that it is impossible to assign a base (matrix) language (cf. Sankoff *et al.*, 1990) which could be treated as the main language of the email. Although much of the code-switching in this group does not appear to be governed by any clear rules, some instances fall under recognisable code-switching categories.

a) Borrowing

At one end of the continuum, all code-switching can be perceived as a form of borrowing from another language, where the borrowed items can range from single words to whole clauses, sentences and or even chunks of discourse (Gumperz *et al.*, 1975; Romaine, 1994). Some researchers make a distinction between borrowing and code-switching, although this is not always possible (Romaine, 1994). As a general rule, if the morphological and syntactic features clearly belong to one language, then any word that is not native to that language must be borrowed. However, if the foreign word is accompanied by foreign morphology and syntax, then it is a case of code-switching (Sankoff *et al.*, 1990). The following are examples of how single English borrowed words are used in LA text. Arabic words are written in italics while English words appear in normal type.

(2) e3melo search bel3araby
     'perform a search in Arabic'
(3) 3ANDY ANNOUNCEMENT BE7'SOS EL DIRECTOR BETA3NA
     'I have an announcement about our director'
(4) ha7awal ab3at el assignments fi a2rab forsa
     'I will try to send you the assignments as soon as possible'
(5) KOL SHEWAYA EL TEAM BEYE7LAW
     'The team keeps getting better and better'
(6) el group kol showaya bey2arab min ba3do
     'The group keeps coming closer'

Examples (3)-(6) display a special kind of borrowing known as *nonce-borrowing* (Poplack *et al.*, 1988; Sankoff *et al.*, 1990; Poplack, 2001). According to Poplack, nonce borrowing "tends to involve lone lexical items, generally major-class content words, and to assume the morphological, syntactic, and often, phonological identity of the recipient language" (Poplack, 2001). The words 'director', 'assignments', 'team' and 'group' are all preceded by the Arabic definite article *el* (ال), which is an Arabic suffix that appears as part of the word in Arabic script. This suggests that the borrowed English words have been partially syntactically integrated into the Arabic text, qualifying them as nonce-loans. Example (11) (below) is particularly noteworthy, where only the loan words are written in Latin script while the rest of the sentence is in Arabic script. Interestingly, the words in the examples above recur regularly in the emails of various members. Over time, nonce-loans occurring frequently in the recipient language can gain the status of established loanwords (Romaine, 1994) such as the English words in examples (7)-(11) below. It is worth noting how the authors of examples (7)-(10) appear to associate technological terms such as 'emails', 'e-group', 'site' and 'pc' with English.

(7) fain el e mails
     'where are the emails'
(8) bgad mabsoota mel e-group
     'I'm truly happy with the e-group'

(9)   e7na ben7awel ne7oto 3al site
      'We are trying to upload it to the [web]site'

(10) mafesh pc 3aleh windows Arabic
      'There is no pc with Windows Arabic on it'

(11) prejudice ال و stereotyping و الاضطهاض الثقافى وصف فى الخوض المفيد غير من
ال

الذى يعانون منه

      'It would be futile to go into the cultural oppression, stereotyping and prejudice that they suffer from'

Arabic equivalents for these words do exist, but are generally regarded as highly formal. The English terms 'prejudice' and 'stereotyping' in example (11) are perhaps the most obvious loanwords, where no widely-used immediate equivalent exists in Arabic. This example is also interesting because it involves a script-switch at the code-switching sites. This singles this instance out as a true case of code-switching according to Angermeyer (2005) who suggests that lexical borrowing occurs if a foreign word is written in the host script, but if the word retains its native script then single-item code-switching has occurred. This is particularly important because it also suggests that, in writing, "speakers have a choice between treating a given lexical item as borrowed or switched," and that "even after a word has been borrowed, it is still available for codeswitching" (Angermeyer, 2005: 525).

b)   Culture-bound expressions

The influence of the native culture is perceived in how authors switch to Arabic when they use expressions that are specific to the local culture (cf. Warschauer *et al.*, 2002). This suggests that the authors feel more at home expressing their native culture in their native language. It may also be said that the authors resort to Arabic here in order to reduce social distance.

(12) sorry for not writing for so long bas mozakret ramadan rabenna mawareeko
      'Sorry for not writing for so long, but [I have been overwhelmed with] studying [and   fasting] in Ramadan, may God never put you through this'

(13) we used to say you are the only secretariat who talks to us. ezzaher e7na 7asadna nafsena
      'We used to say you are the only secretariat who talks to us. It seems that we have given ourselves the evil eye'

c)   Personalisation vs. objectivisation

Code-switching can be used to indicate degree of personal involvement, to distinguish fact from opinion or to lend authority to speech (Gumperz, 1982: 80). Wardhaugh (1998: 103), refers to this as an "affective dimension" to metaphorical code-switching where "you change the code as you redefine the situation – formal to informal, official to personal, serious to humorous, and politeness to solidarity" as in the following

examples. The switch to English in example (14) appears to be intended to demonstrate authority. Similarly, in example (15), the addressee is specified in English, but the message itself is in LA, establishing its informal, joking tone. The author then restores formality (and authority) by switching back to English in the next sentence.

(14) ma3ad el session el gaya 6th of september. Assignment due then. NO EXCUSES!
'Our next session is on the 6th of September. Assignment due then. NO EXCUSES!'

(15) to our football team: hangeeb loko ta2m mo7tarafeen min bara 3ashan el match el gay lazem neksaboh!
to people who missed last session: call me or there will be problems!
'To our football team: we will get you a team of [football] professionals from abroad because we must win the next match!
To people who missed last session: call me or there will be problems!'

d) Reiteration

Gumperz (1982) identifies repetition as one possible form that code-switching might take. Saying something in one code and then repeating it in another can provide clarification or emphasis, such as in example 16 below. It can also have a comic or sarcastic purpose such as in example (17). This effect would not be achieved by simply repeating the same phrase in the same code.

(16) I found no one. YES MAL2ETSH WALA WA7ED RA7
'I found no one. Yes I found that no-one went.'

(17) We, as your people, ‫رعايا سيادتك‬…
'We, as your people, your people [subjects]…'

e) Tag-switching

Tag-switching (Poplack, 1980) involves inserting a 'tag' in one language into a phrase that is entirely in another language. Gumperz (1982: 77) refers to these as 'sentence fillers'. Tags or sentence fillers are words or phrases that are bound by minimal syntactic constraints and therefore can usually be inserted with ease at a number of possible points in a sentence. The following are examples of such tags. It is to be noted that the LA tags *ya3ny* and *ba2a* in examples (20) and (21) are loosely translated here since their meaning varies widely according to how they are used and where they are inserted in a sentence; they have minimal referential value in monolingual speech.

(18) Keep it up ya shabab
'Keep it up, guys'

(19) please everyone try to be there begad
'Please everyone try to be there really'

(20) I won't miss the chance to see u for a louzy midterm *ya3ny*!
'I mean, I won't miss the chance to see u for a lousy midterm!'

(21) ba2a [*name*] in the middle for all of us

'Seriously! [*name*] is in the middle for all of us?'

Code-switching in Group B

Code-switching in Group B occurs much less frequently than in Group A, with English serving as the base language of all the emails in this group. While code-switching in Group A was highly unregulated and not always possible to account for, code-switching in Group B seems to fall distinctly under one of the categories mentioned below. Interestingly, nonce loans, which occurred frequently in Group A, do not occur at all in Group B, and neither does reiteration.

a)   Culture-bound expressions

As in Group A, cultural expressions always appear in Arabic in Group B (see excerpt 1). These are often fixed expressions and range from greetings on cultural and religious occasions (22 & 23), good wishes (24), sympathies (25) and condolences (26).

(22) happppyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy   feast   for   all   of
      uuuuuuuuuu matakloosh ka7k...
      'Happy feast for all of you don't eat [a lot of] traditional pastries'
(23) Kol sana wentom tayyebeen.:-))) Takabbal Allah menna wa menkom
      seyamana isA... I hope you're all doing well
      'Season's greetings :-))) May Allah accept our fasting... I hope you're all
      doing well'

(24) have a nice time [*name*] we tege belsalama
      'have a nice time [*name*] and have a safe journey home'
(25) You seem quite tied up with school work - Rabenna ma3aki :-)
      'You seem quite tied up with school work – May God help you :-)'
(26) First of all, I would like to say "Elba2eya fe 7ayatek" to [*name*]
      'First of all, I would like to say sorry for your loss to [*name*]'

b)   Personalisation vs. objectivisation

This is another function of code-switching which commonly occurs in Group B. As in Group A, the authors appear to use English when they wish to lend authority to what they say, and switch to Arabic to be less formal and more personal. In examples (27) & (28), the author's switch to Arabic signals a more personal reproach, while at the same time informally appealing to the members' sense of duty. In examples (29) & (30), the author's switch to Arabic clearly marks a decline in formality, and sharply contrasts the sarcastic remarks made in Arabic against the serious tone of the messages in English. It is the code-switching in these examples which signals this 'changing of hats' (Romaine, 1994: 173).

(27) I wanted to know if there would be any more volunteers for both topics… (ya3ni 3aib 3alaina wa lamo2a7'aza lamma neb2a 20 members we mosh 3arfeen ne2assem el sho3'l!)

'I wanted to know if there would be any more volunteers for both topics… (I mean, it's shameful that we are 20 members and still can't divide the work [between us]!)'

(28) those who are absent for 3 consecutive meetings without a reasonable excuse are AUTOMATICALLY dismissed from the club. (men 3'air za3al ya gama3a, but seriously, this is getting WAY out of hand!)

'Those who are absent for 3 consecutive meetings without a reasonable excuse are AUTOMATICALLY dismissed from the club. (no hard feelings guys, but seriously, this is getting WAY out of hand!)'

(29) As for outings, I personally recommend them whenever we're all disposed for such plans – since outings encourage social interaction between members away from the 'official' air about meetings… (wa lawenni shakka enno 7ad feekom yekoon far2a ma3ah mawdoo3 el official air da 7'ales :-P)…

'As for outings, I personally recommend them whenever we're all disposed for such plans – since outings encourage social interaction between members away from the 'official' air about meetings… (although I doubt that this 'official air' issue makes the slightest difference to any of you :-P)…'

(30) I hope that you're all fine, doing well in your studies and beginning to prepare for the upcoming mid-year finals (kal 3ada, zayyi kedda tab3an :-P)

'I hope that you're all fine, doing well in your studies and beginning to prepare for the upcoming mid-year finals (as usual, like me of course :-P)'

c) Qualifying a message

Another form of code-switching which appears in the discourse of Group B is introducing a message in one language and then switching to another language to provide clarification or elaboration (Gumperz, 1982). In the following examples, the authors initially introduce the topic in English, but they switch to Arabic to qualify their message. It is interesting that such elaborations are consistently provided in Arabic, although the members are comfortably bilingual in English and Arabic. This suggests that the native code is regarded as the simpler, more personal or perhaps more easily understood code.

(31) There's this special display of a children's animated movie in [*name*] theatre (elli 3and el Opera)

'There's this special display of a children's animated movie in [*name*] theatre (the one by the Opera house)'

(32) Monday is fine ya [*name*].. go for it.. bass 3andy moshkela.. I finish college at 2:30

'Monday is fine [*name*].. go for it.. but I have a problem.. I finish college at 2:30'

(33) I couldnt attend the last meeting we msh fahma 7aga
'I couldn't attend the last meeting and I don't understand anything'

d) Specifying an addressee

This is another function of code-switching identified by Gumperz (1982). Here, the switching is used to designate a specific recipient of the message. This can be particularly useful in emailing groups like the ones studied here where the message goes out to the whole group indiscriminately.

(34) I really want to know what you think about these two ideas (eih ra2yek ya *[name]*)
'I really want to know what you think about these two ideas (what do you think, *[name]*?)'

(35) I hope you're all fine, and enjoying your summer vacations (I assume that you're all done with studies and exams) - tab3an ya *[name]* I know that you're an exception
'I hope you're all fine, and enjoying your summer vacations (I assume that you're all done with studies and exams) – of course, *[name]*, I know that you're an exception'

e) Tag-switching

Like Group A, Arabic tags are frequently seen in Group B. Some of the tags cited below are common to both groups. Again, it is to be noted that some of the tags do not have a fixed referential meaning.

(36) how is everybody ya gama3a
'How is everybody [you] guys'

(37) miss u all awy
'Miss u all very much'

(38) i forget names all the time fe3lan
'I forget names all the time, really'

(39) shofty we will have to send *[name]* (the smallest cat) away
'See? We will have to send *[name]* (the smallest cat) away'

(40) ma3lesh i know it's a bit too late
'Sorry, I know it's a bit too late'

f) Flagged code-switching

This is a marked form of code-switching which is not subject to syntactic restrictions (Poplack and Sankoff, 1988). A means of 'flagging' such as a pause, hesitation, repetition or a commentary is used to draw attention to the switch. Example (41) is an illustration of how hesitation signals such a switch. Parentheses and quotation marks may also be said to function as flags, in which case much of the switches in this group could be considered flagged switches (e.g. 25-31).

(41) I wasn't actually yelling, I was … baharrag.
'I wasn't actually yelling, I was … kidding around.'

What influences code/script choice? Motives and attitudes of members

It is not entirely possible to account for authors' choice of one code or script over another, but a number of clues are present. For instance, English appears to invoke authority in both groups. In Group B, English is favoured because it represents the language of the activity that the group is associated with. The fact that all the members are bilingual, and crucially are known to be so by their interlocutors, makes English an almost automatic choice, accounting for the minimal use of Arabic in Group B. On the other hand, members of Group A are not expected to be bilingual. Hence, the level of fluency of the addressees' plays an important role in determining the language used, as can be inferred from excerpt (42).

تساءلت [*name*] **عننا فى ال** email. **أنا هنا ولكن لاتتخيلى معاناتى فى الكتابة العربى على الكمبيوير.**
(42)
It was easier for me to write in English, but since not all of us are fluent in English, I wrote it that way.
رجاء : أنا أجد صعوبة كبيرة فى قراءة رسائلكم لأننى لا أفهم العربى المكتوب بالنجليزى و العكس.
Plz write the emails in either Arabic or English, not this strange language. Or you can send me a key to help me read them.

'*[name]* asked about me in her email. I am here but you can't imagine what a hard time I have typing Arabic on the computer.
It was easier for me to write in English, but since not all of us are fluent in English, I wrote it that way [Arabic in Arabic script].
Please: I find great difficulty in reading your messages because I don't understand the Arabic written in English and vice versa.
Please write the emails in either Arabic or English, not this strange language. Or you can send me a key to help me read them.'

Excerpt (42) which appears at the end of an email which was otherwise written entirely in Arabic script at once highlights a number of factors which govern code and script choice in Group B. Firstly, the bilingual ability of other members: Some members are reluctant to write entirely in English because they know that not all of the members are fluent in English. Secondly, is the ease of typing English script vs. the difficulty of typing Arabic script: The fact that most members find it easier to type in Latin script encourages them to write in English or Latinised Arabic. Next, the unintelligibility of Latinised Arabic: There is a relative degree of difficulty associated with reading LA (probably due to the use of numerals and the abundant possibilities in which one word can be represented). This is further illustrated in excerpt (43) which features in another email by the same author. The fact that this member expressly mentions her dislike for LA must have influenced others to curb its use. This is interesting in light of the already substantial amount of LA used in Group A, as it suggests that LA might have been even more prevalent.

(43) Sorry for writing in English, but it is the easiest way. I hate writing the newly invented language of ..... i don't know what to call it! I don't know how to write it and i can't even read it sometimes. I'm verrrrrrrry slow in writing Arabic on the computer, and i have a class right now.

Another factor which seems to influence members' choice of script is lack of technical support for Arabic software, especially for Arabic users abroad (cf. Warschauer *et al.*, 2002), as suggested by excerpt (44).

(44) momken bas le al mas2ol 3an el assinments we el 7agat
de(director) ... yeb3t email ye2ol eih el assinment tane..:)))) ..
we talab tane ya rat balash be el 3arbe. 3alshn ana mosh fe misr we mafesh
pc 3aleh windows arabic ...
'Would the person in charge of assignments and these things (the director) send an email explaining what the assignment *is again*? :)))
Another request, preferably not in Arabic, because I'm not in Egypt and there is no pc with Windows Arabic on it.'

It is to be noted that the evidence discussed here marks the attitude and motives of members at a given point in time (between 2003 and 2004) and may not necessarily be representative of the present situation. Technology has since become more accommodating of non-Latin scripts so the problem put forth in excerpt (44) may no longer be a hindrance. In addition, LA has spread more widely in recent years and fewer people are finding it difficult to comprehend. To verify this, the author of excerpts (42) and (43) was contacted to investigate her current attitude towards LA. Her response was that she "learnt" LA shortly after these emails were written, and that she is now a proficient user this variety. This suggests that an 'in-group' / 'out-group' construct with self-selecting membership is at work here (cf. Palfreyman & Al Khalil, 2003). The in-group are those who know/use LA, and the out-group are those who do not. The growing popularity of LA made the author mentioned here feel pressured to learn this variety for fear of being left out or left behind. In her case, learning LA was at once a response and a contribution to the spread of this variety.

Summary of findings

Evidence from both groups emphasizes the eminence of English in online discourse, even where all the participants are native speakers of Arabic and therefore no perceivable need for English exists. This suggests a generally favourable attitude towards English.

In Group A, bilingual members take full advantage of their bilingual ability, flexibly alternating between formats and using code-switching as "a mode of bilingual performance which allows the bilingual to display his or her full communicative competence" (Romaine, 1994: 173).

In Group B, English appears to be the dominant base language and most switches to Arabic are preceded by a punctuation flag. This suggests that the more fluent the participants are in English, the more frequently it is used. However, the cultural significance of Arabic is still apparent with a clearly informal connotation (which is perhaps emphasized by the use of the Latinised form), as opposed to the formal use of English. This assigning of contrasting tones allows the author to change hats by alternating between the two languages. Thus, code-switching here serves a stylistic function, as highlighted by Romaine (1994: 143):

> Mixing and switching for fluent bilinguals is … in principle no different from style shifting for the monolingual. The bilingual just has a wider choice – at least when he or she is speaking with bilingual speakers.

Indeed, the perceived bilingual ability of other users appears to be central to the choice of one code or the other. This is perhaps the key to understanding why fluent bilinguals choose to minimize the use of English when in doubt about the bilingual ability of their addressees (Group A), whereas English seems to be the automatic code of choice when no doubt exists (Group B). In other words, the choice of code and script is governed by the ability of the addressee to understand the chosen code or script.

In addition, the evidence studied indicates that lack of software support for Arabic script and the comparative ease and speed of typing in Latin script influence users to write in English or in Latinised Arabic.

Limitations

The difference between the size of the email corpora in Groups A and B undermines the comparability of the results to some extent. It was not possible to produce evidence of statistical significance because the sample size was too small after classifying emails into individual format categories. The total in one of these categories (LA+E) in Group A was greatly skewed by the contributions of only two authors. As original contributions by the authors, the lengths of the emails in the two groups varied to a great extent, and in effect so did the proportion of the varieties studied. It was beyond the scope of the present study to take this into account, but it may be worth investigating in the future how these varieties measure as a proportion of the individual word count of each email.

It must be noted that the evidence studied in this paper is only representative of the groups studied. If anything, this evidence illustrates great individual variation in choices, motives and attitudes. Some clear differences between the orthographical tendencies in the examples cited here and those found in Palfreyman and Al Khalil (2003) indicate that such conventions vary by region, and emphasise that the examples here are not representative of the online discourse of the wider community of Arabic users. In addition, the emails studied here are a 'snapshot' reflecting a situation at a particular point in time (2003-2004) and cannot be considered representative of present-day online discourse. Finally, it is important to note that the purpose of this

study was not to generate generalisable results, but to provide insight into a very specific, yet interesting, situation.

This paper shows that the dichotomy in the sort of online communication studied is not between Arabic in Arabic script and Arabic in Latin script (LA), since these hardly feature together. The real dichotomy lies between English and LA, which are very closely linked, as the analysis in Section 2.2 suggests. It is the bilingual ability of the addresser and the perceived bilingual ability of the addressee which apparently determine just how much English vs. LA is used. In the future, it would be interesting to investigate the role that English plays in online communication between native speakers of Arabic.

The very fact that it is possible to analyse online communication using criteria of conversational code-switching highlights how much these emerging genres resemble spontaneous spoken speech. Indeed, it is Arabic vernaculars that are more commonly found in online communication – the so-called 'less-than-standard' spoken varieties of Arabic. However, Standard Arabic is not completely absent from online communication. While this was not the focus of this study, the evidence studied here indicates a tendency to write the standard form in Arabic script, while colloquial Arabic is almost always Latinised. In this sense, it could be claimed that Latinisation 'empowers' Arabic vernaculars, marking this as an area worthy of further research. Some further questions this research raises concern which the online genres standard Arabic is more likely to feature in. Also, whether the use of Arabic script more associated with Standard Arabic and Latin script with vernacular Arabic and with the spread of LA outside of computer-mediated communication, what effect, if any, is this having on the perceived power, scope of use and attitudes towards Arabic vernaculars within Arabic-speaking communities?

Recent developments

With software support for Arabic script more readily available today than it was in 2003-2004, more recent developments which have potentially increased the applications of LA are offering a whole new perspective on its growing use. In September 2007, *Orascom Telecom*, an international telecommunications company, launched *Onkosh*[2] ('unearth' in Arabic), a search engine allowing users to look for Arabic content on the Internet using LA. Although users also have the option to type in English or Arabic in Arabic script, the company showcases the transliteration feature in particular, saying that, "for the first time ever on the Web, users are able to search for Arabic words by typing in English characters, through the patented 'Bel3araby' [in Arabic] transliteration feature" (Orascom Telecom, 2007). As the user types in Latin characters, a drop-down list of possibilities in Arabic script appears (Figure 2). If the user does not manually select one of the possibilities, the first option is automatically chosen in a manner similar to how predictive texting works on mobile phones.

---

[2] The search engine can be found at: http://www.onkosh.com/

Figure 2 Onkosh in use

Shortly afterwards, in early 2008, another online solution which is expected to "revolutionize the Arabic web" was launched by *Language Analytics* (Sergie, 2007). Like Onkosh, the web-based application, *Yamli*[3] ('dictate' in Arabic), enables Arabic users to type in LA and converts what they type into Arabic script. The application features the LA phrase *2oktob 3arabi* (write in Arabic) in its logo. Yamli is different from Onkosh in that it appears to employ a slightly improved technology to produce more accurate results and a more exhaustive list of possibilities. It also gives users more time to type in the full words before converting what they type into Arabic script. Unlike Onkosh, Yamli can be used as an independent conversion application and not just as a search engine.

Simultaneously, in January of the same year, *LinguArabica* released a software package, *Eiktub*, which enables users to type Arabic script using a QWERTY keyboard and to save what they typed. In addition to an online pad, there is also a free downloadable version of the software available from the website[4]. In a PR release, the founders of Eiktub say that "anyone who is used to the English keyboard can now type beautiful Arabic fast, without having to switch keyboards". Beautiful Arabic here refers to Arabic with diacritics (*tashkeel* in Arabic), a feature of Arabic script which, though largely ornamental to the native speaker, is indispensible in certain genres, such as poetry and Quranic verses. From a linguistic perspective, what is striking about Eiktub is that it offers an entire transcription scheme: the Bikdash Arabic Transliteration Rules or BATR©, named after its proprietor, Dr. M. Bikdash.

Another interesting development in the online career of LA occurred in February 2008, with the piloting – and subsequent official launch in April 2008 – of *e-magazeen*, an online magazine targeted mainly at teenagers which makes extensive use of LA. On the magazine's Facebook group[5], its developers make the unprecedented and bold claim that it is "the first Egyptian online magazine written in frankoarab style to make it easy for all of you to read". To date, it has been claimed that LA is easier to *type* than Arabic in Arabic script, but never before has it been claimed that it is easier to

---

[3] The application can be found online at www.yamli.com
[4] The software can be found at www.eiktub.com
[5] http://www.facebook.com/inbox/?ref=mb#/group.php?gid=8195399453 (accessed 25/10/08)

read. A more recent entry on the magazine's website[6] describes e-magazeen as "the first Egyptian ONLINE magazine written in all commonly used languages English, Arabic and Arabish". This description is perhaps more revealing as it highlights the fact that the magazine is not comprised entirely of LA (Frankoarab or Arabish), it is rather a blend of English and Arabic (both in Arabic and Latin scripts) which could only mean that it is aimed at a bilingual readership. Nevertheless, this daring step at once highlights how much currency LA has gained among Egyptian teenagers, and raises the question of whether LA is truly easier to read than Arabic in Arabic script.

Conclusion

Wider support for Arabic script online does not appear to curb the increasing popularity of Latinised Arabic, with growing evidence of its diffusion to non-computer-mediated communication. For the time being, the use of LA remains highly unregulated with clear regional variation. This paper suggests that there is a link between bilingual ability and the frequency with which users use English and LA in their emails. Bilinguals appear to possess the widest set of communicative options in online discourse, with LA serving as one stylistic possibility.[7]

References

Androutsopoulos, J. (2006). Greeklish: Transliteration practice and discourse in a setting of computer-mediated digraphia. In Georgakopoulou, A. and Silk, M. (eds.) *Standard Languages and Language Standards: Greek, Past and Present*. Retrieved 15 April 2008 from: http://www.archetype.de/texte/2006/LOGOS_Greeklish_2006.pdf

Angermeyer, P. (2005). 'Spelling Bilingualism: Script Choice in Russian American Classified Ads and Signage'. *Language in Society* 34 (4), 493-531.

Barry, R. K. (ed.) (1997). *ALA-LC Romanization Tables: transliteration schemes for non-Roman scripts approved by the Library of Congress and the American Library Association*. Washington: Library of Congress.

---

[6] www.e-magazeen.com

Beesley, K. R. (1998). Romanization, transcription and transliteration. Retrieved 10 April 2008 from http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/romanization.html

Bianchi, R. (2006). Revolution or Fad? Latinised Arabic Vernacular. In Davidson, P., Al Hamly, M., Aydelott, J., Coombe, C., & Troudi, S. (eds.) *Proceedings of the 11th TESOL Arabia Conference: Teaching, Learning, Leading*, (10). Dubai: TESOL Arabia, 329-344.

Bloomfield, L. (1933). *Language*. New York: Holt.

Crystal, D. (2nd ed.). (2006). *Language and the Internet.* Cambridge: Cambridge University Press.

Danet, B. & Herring, S. C. (2007). Introduction: Welcome to the Multilingual Internet. In Danet, B. & Herring, S. C. (eds.) *The Multilingual Internet: Language, culture and communication online*. New York: Oxford University Press

Diebold, A. R. (1964). Incipient Bilingualism. In Hymes, D. (ed.) *Language in Culture and Society*. New York: Harper & Row, 495-511.

Ferguson, C. (1959). 'Diglossia'. *Word, 15*, 325-340.

Gardner-Chloros, P. (1991). *Language Selection and Switching in Strasbourg*. Oxford: Oxford University Press.

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

Gumperz, J. J. and Hernández-Chavez, E. (1975). Cognitive Aspects of Bilingual Communication. In Hernández-Chavez, E., Cohen, A. & Beltramo A. (eds.) *El Lenguaje de 10s Chicanos*. Arlington, VA: Center for Applied Linguistics, 154-64.

Haugen, E. (1953). *The Norwegian Language in America: A Study in Bilingual Behavior*. Philadelphia: University of Pennsylvania Press. [Reprinted in 1969. Bloomington: Indiana University Press].

Holes, C. (1995). *Modern Arabic: structures, functions and varieties*. London: Longman.

Huston, G. (2008). Internationalizing the Domain Name System. *The Internet Protocol Journal* 11(1): 2-22.

Khayat, M. H. (4th ed.). (2004). في سبيل العربية *(Towards Arabic).* Cairo: Wahba Library. [Arabic]

Language Analytics LLC. (2007). Yamli.com Allows Arabic Speakers to Write Arabic Using English Alphabet (press release). Cambridge, MA: Language Analytics LLC. Retrieved 1 April 2008 from http://www.yamli.com/pr/

Orascom Telecom. (2007). Orascom Launches Onkosh (press release). Dubai, UAE: Orascom Telecom. Retrieved 1 May 2009 from http://www.orascomtelecom.com/files/media_Files/143749826_Onkosh.pdf

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). Retrieved 29 March 2008 from http://www.ascusc.org/jcmc/vol9/issue1/

Poplack, S. (1980). 'Sometimes I'll start a sentence in English y terminó en español: toward a typology of code-switching'. *Linguistics* 19: 581-16.

Poplack, S. (1985). Contrasting patterns of code-switching in two communities. In Wrkentyne, H. J. ed. *Methods V: Proceedings of the V International Conference on Methods in Dialectology*. Victoria: University of Victoria Press.

Poplack, S. (2001). Code-switching (linguistic), in SMELSER, N. & P. BALTES (eds.), *International Encyclopedia of the Social and Behavioral Sciences,* Elsevier Science Ltd., 2062-2065.

Poplack, S., Sankoff, D. (1988). Code-switching. In Ammon, U., Dittmar, N. and Mattheier, K. J. eds, *Sociolinguistics. An International Handbook of Language and Society*. Berlin: Walter de Gruyter.

Romaine, S. (2nd ed.). (1994). *Bilingualism*. Oxford: Blackwell.

Sankoff, D., Poplack, S. and Vanniarajan, S. (1990). 'The case of the nonce loan in Tamil'. *Language Variation and Change* 2: 71-101.

Singh, R. (1985). 'Grammatical constraints on code-switching: Evidence from Hindi-English'. *Canadian Journal of Linguistics* 30: 33-45.

Yaghan, M. (2008). ''Arabizi': A Contemporary Style of Arabic Slang'. *Design Issues*, 24 (2): 39-52. USA: MIT Press.

Wardhaugh, R. (3rd ed.). (1998). *An Introduction to Sociolinguistics*. Oxford: Blackwell.

Warschauer, M., El Said, G. R., & Zohry, A. (2002). 'Language choice online: Globalization and identity in Egypt'. *Journal of Computer-Mediated Communication, 7 (4)*. Retrieved 29 March 2008 from http://www.ascusc.org/jcmc/vol7/issue4/warschauer.htm