

Egyptian Romanized Arabic:

A Study of Selected Features from Communication Among Egyptian Youth on Facebook

Jan Arild Bjørnsson



Thesis for the degree of Master of Arts in the field of
Arabic language (60 credits)

UNIVERSITY OF OSLO

November 2010

Table of contents

| | |
|---|-----------|
| Acknowledgements | 5 |
| Chapter one: Introduction | 6 |
| 1.1 Research question | 7 |
| 1.2 Outline of the thesis | 8 |
| Chapter two: Writing and language | 9 |
| 2.1 Writing, scripts and alphabets | 10 |
| 2.1.1 Conversion of scripts..... | 12 |
| 2.1.1.1 Transcription of Arabic | 13 |
| 2.1.1.2 Judaeo-Arabic | 14 |
| 2.1.1.3 Romanization of Turkish | 15 |
| 2.1.1.4 Maltese..... | 16 |
| 2.1.1.5 Arabic language reform | 17 |
| 2.2 The language situation in Egypt | 20 |
| 2.2.1 Arabic in Egypt | 20 |
| 2.2.1.1 Attitudes towards the varieties | 21 |
| 2.2.1.2 The continuum between the varieties | 21 |
| 2.2.1.3 Egyptian Arabic as a written language | 22 |
| 2.2.2 English in Egypt..... | 23 |
| Chapter three: Computer-mediated communication | 24 |
| 3.1 Communication situations on the Internet | 24 |
| 3.1.1 CMC – between speech and writing?..... | 26 |
| 3.2 The Internet in Egypt and the Middle East | 28 |
| 3.2.1 The Egyptian Internet user | 29 |
| 3.2.2 Arabic online | 30 |
| 3.2.3 Language choice online in Egypt..... | 31 |
| 3.3 About Facebook and communication between its members | 33 |
| 3.4 Method | 35 |
| 3.4.1 Selection criteria | 35 |
| 3.4.2 Notes on transcription | 37 |
| Chapter four: The graphemes | 41 |
| 4.1 The consonants | 41 |
| 4.1.1 Numerals representing phonemes | 42 |
| 4.1.1.1 Representation of /ħ/..... | 43 |
| 4.1.1.2 Representation of /x/ and /ġ/ | 44 |
| 4.1.1.1 Marginal representations | 46 |
| 4.1.2 Other consonants..... | 47 |
| 4.1.2.1 Representation of /w/ and /y/ | 47 |
| 4.1.2.2 Representation of /q/..... | 48 |
| 4.1.2.3 Representation of /ʔ/ as a reflex of Standard Arabic /q/..... | 48 |
| 4.1.2.4 Representation of /ʔ/ as a reflex of Standard Arabic /ʔ/..... | 49 |
| 4.1.2.5 The emphatic consonants..... | 51 |
| 4.1.6 Regional differences..... | 51 |
| 4.2 The vowels | 53 |
| 4.2.1 Vowels in the data..... | 56 |
| 4.2.1.1 Long vowels..... | 57 |
| 4.2.1.2 Short vowels..... | 60 |
| 4.2.1.2.1 <y> representing final /i/..... | 62 |

| | |
|--|-----------|
| 4.2.1.2.2 The minor representations | 63 |
| 4.2.1.2.3 The epenthetic vowel | 64 |
| 4.2.1.2.4 Short vowels and emphatic consonants | 65 |
| 4.2.3 A summary: long and short vowels | 66 |
| 4.2.4 The inconsistent use of vowels | 66 |
| Chapter five: Some selected features | 69 |
| 5.1 The definite article | 69 |
| 5.2 The feminine ending | 71 |
| 5.2.1 Realized as /a/ | 71 |
| 5.2.2 Realized as /it/ | 71 |
| 5.2.3 Realized as /ā/ | 72 |
| 5.3 Gemination | 72 |
| 5.4 The future prefix ḥa-/ha- | 74 |
| 5.5 The prepositions <i>li</i> and <i>bi</i> | 75 |
| 5.6 Some other prepositions | 76 |
| Chapter six: conclusion | 78 |
| Literature | 81 |
| Abstract | 84 |

Acknowledgements

First of all I wish to thank my supervisor Professor Gunvor Mejdell for her valuable advice, encouraging enthusiasm, and patience throughout the process of writing this thesis. I also owe her gratitude for leading me onto the subject of Romanized Arabic.

I am also greatly indebted to Unn Gyda Næss for reading through the thesis and giving constructive feedback.

Thanks to Helen Rodgers for her efforts in helping me improve the language of the thesis.

I would also like to thank my Egyptian friends who have been happy to answer all my questions about their mother tongue, and for helping me whenever I have come across something in my data I did not understand.

And lastly, thanks to all fellow students of Arabic and Middle Eastern studies who have kept me company for numerous lunch and coffee breaks.

Blindern, November 2010 – Jan Arild Bjørnsson

Chapter one: Introduction

In Cairo, when visiting one of the numerous upper class cafeterias that are scattered around the city, it is almost impossible not to notice the television sets playing music videos rather indiscreetly. During the music, text messages sent from viewers' mobile phones scroll across the bottom of the screen. Some of these are written in Roman letters, but are still a representation of Arabic. Although I became intrigued by this way of writing Arabic during my stays in Cairo, I never considered making it the theme of my master's thesis until it was suggested to me by Professor Gunvor Mejdell. At first I tried to collect data from text messages that Egyptians had written on their phones, but it proved difficult as many claimed to write their messages in English or Arabic with Arabic letters, or they were reluctant to share messages that, after all, are often private communications between two people. Instead, I collected data from Facebook, by far the most popular social networking website worldwide of the last few years.

The last decade or two have seen the Internet spread out to all corners of the world. Not only has it made all kinds of news and information, reliable as well as less reliable, easily accessible. With the spread of this vast network have come new ways to communicate. An e-mail is delivered in a matter of seconds, regardless of whether the recipient is in the house next door or on the other side of the world. Just to illustrate the superiority of the Internet where swiftness is concerned, a letter shipped by ordinary mail is now often referred to as snail-mail by experienced Internet users. Recently e-mail has perhaps been surpassed in popularity among younger Internet users by a more direct communication that most of all resembles a written conversation. While e-mails more or less follow the conventions of personal letters, another form of communication through computers, *chatting*, contains features like incomplete sentences, abbreviations and sometimes spelling that are inspired by the oral realization.

For Arabic speakers, among others, technicalities made it impossible, or at least very difficult, to write with letters of their own, Arabic, alphabet on the Internet until quite recently. Instead they had to resort to the Roman script, which for a long time was the only one widely supported by software and hardware. This resulted in Arabic written in the Roman script, the Egyptian variety of which, Egyptian Romanized Arabic (ERA), is the subject of this thesis.

1.1 Research question

One of the problems one encounters when writing Arabic with Roman letters is the fact that Arabic contains more consonant phonemes than the Roman alphabet has graphemes to represent. In academic transcription this problem can be solved e.g. by using a Roman letter that in many languages does not represent its own phoneme (*x* to represent /*x*/), placing a straight stroke above the letters (to represent long vowels like /*ā*/) or a dot below (to differentiate e.g. emphatic letters from non-emphatic, like /*ṣ*/ from /*s*/). The use of strokes and dots, however, are not possible without installing software specifically designed to enable these special signs.

The difficulty of transcribing into the Roman alphabet, without using special signs or diacritic marks, can be seen when people with Arabic names immigrate to Western countries and have to write their names so that the natives can read them. On the web site belonging to Statistics Norway, one can obtain information about the number of people in Norway who have a certain name. Here are some of the existing versions of the male name *Muḥammad*, and how many who use them, per 1 January 2010¹:

| | |
|----------|------|
| Mohammad | 3264 |
| Mohamed | 1865 |
| Mohammed | 1764 |
| Muhammad | 951 |
| Mohamad | 216 |
| Muhammed | 198 |
| Muhamed | 127 |
| Muhamad | 14 |

Learning to read and write, in most languages, goes hand in hand with learning an alphabet. For the Egyptians whose texts I have gathered for this thesis, learning a second, Western language included learning a new alphabet as well. The combination of them wanting to write to other Egyptians on the Internet, and the already mentioned technical restrictions, led to them writing their mother tongue using the alphabet they had so far only used for the western foreign language. One might then assume that they would, to the greatest extent possible, use

¹ <http://www.ssb.no/emner/00/navn/> [Accessed 19 August 2010]

the Roman letters to represent the same phonemes they do in the foreign language, and write more or less phonetically.

The main research question I want to investigate in my thesis is: “to what extent is the writing of Egyptian Romanized Arabic (ERA) influenced by Arabic orthography?” By Arabic orthography is meant both the official written language Standard Arabic, and the common standards for writing the informal Egyptian dialect. The latter, although representing the oral language and thus has a partially different lexicon and grammar, complies more or less with the official orthography.

I also want to examine whether there is any consistency in how ERA is written. Does every user write as he or she feels like doing, or have some norms started to emerge?

In my attempt at answering these questions, I will not make a complete survey of ERA. Instead, I have chosen to analyze how certain features are represented. Firstly, I will look at the phonemes of Egyptian Arabic – the consonants, and the long and short vowels. Secondly, some more features are analyzed, mainly chosen because their pronunciation does not coincide with how they are written with Arabic orthography. This means that they are well suited to reveal any discrepancy between ERA representation and speech caused by influence by Arabic orthography.

1.2 Outline of the thesis

In chapter two, I will write about writing and language in general, the relationship between the two, and provide a few examples showing that alphabets or scripts are often not designed for a specific language. The special language situation in Egypt will also be dealt with here. Chapter three is devoted to Computer-Mediated Communication in general, in the Middle East, and in Egypt. Further, I will describe here how I got the data from Facebook, and how I decided to interpret the data. In chapter four, I examine how the phonemes of Egyptian Arabic are represented, first the consonants and then the vowels. In chapter five, I look into some other features, e.g. the definite article, gemination and the feminine marker. The last chapter, chapter six, offers a summary and a conclusion to the research questions.

Chapter two: Writing and language

As this chapter, and also the thesis in its entirety, will be about writing and different writing systems, certain terms need to be defined to avoid confusion. I will use the terms as they are defined by Coulmas (2003, p. 35).

First there is the relatively broad term *writing system*, which has two distinct meanings (Coulmas, 2003, p. 35):

It refers to the writing system of an individual language and to an abstract type of writing systems. In the first sense, there are as many writing systems as there are written languages, but in the second sense the number is limited to a few types, such as logographic or word writing systems, syllabic writing systems, phonetic writing systems, or variant forms thereof.

ERA does not have any formal rules or orthography, and I will refer to it as a method of writing rather than a writing system.

The term *script* is “reserved for the graphic form of the units of the writing system” (Coulmas, 2003, p. 35). Some scripts are felt to be related to their language, like the Korean, Yi and Cambodian scripts, while others, like Roman and Arabic², can serve numerous languages (Coulmas, 2003, p. 35). A closely related term is *alphabet*. The expression ‘Latin alphabet’ can mean “the writing system of the Latin language” and “a set of 26 letters serving the writing systems of a great number of languages” (Coulmas, 2003, p. 32). Roman is often used about the latter. In this thesis I will use ‘Latin alphabet’ when writing historically about the alphabet that was and still is used to write Latin, and Roman script for any other script based on the Latin alphabet.

Orthography refers to “the standardized variety of a given, language-specific writing system” and can be used interchangeably with spelling (Coulmas, 2003, p. 35).

² Even though the Arabic script is used to write other languages than Arabic, it can be perceived as belonging to the Arabic language through its relation to Islam and the Qur’an.

2.1 Writing, scripts and alphabets

Writing can be defined as “to communicate relatively specific ideas by means of permanent, visible marks” (Sampson, 1985, p. 28). Among other things, it differs from spoken language in that speech is something anyone can learn without formal instruction, while writing must be learned through teaching and deliberate effort. Illiterate people are found all over the world today, either as individuals in a society where many can read and write, or as groups who do not have a writing system at all (Dobrovolsky & O’Grady, 1997, p. 591).

The earliest known writing was from Sumeria and probably came to existence about five thousand years ago. This writing consisted of pictograms that convey their meaning through their resemblance to physical objects, but also concepts that could be associated by them. A pictogram of the sun did not only mean “sun”, but also “light”, “heat” or “energy” (Dobrovolsky & O’Grady, 1997, p. 595). While several of the early, major writing systems were developed independently from each other in the Mediterranean and Oriental world, the alphabet was invented only once. The origin of all other alphabets is an old Canaanite alphabet and its immediate descendant, the Early Linear Phoenician alphabet (Cross, 1991, p. 77). In the first writing systems, symbols were used to represent words or syllables, or a combination of the two. The alphabet made it possible for the first time to represent each phoneme of a word, and, according to Cross (1991, p. 78), made the art of writing notably simpler, and literacy was able to spread in centuries rather than millennia. The first alphabets were used to write consonants only, and the Greek alphabet, probably first used in 776 BC, was the first to represent vowels as well (Sampson, 1985, p. 99). The importance of the Greek alphabet is summed up by Gelb (1963, p. 184):

The development of a full Greek alphabet, expressing single sounds of language by means of consonant and vowel signs, is the last important step in history of writing. From the Greek period up to the present, nothing new has happened in the inner structural development of writing. Generally speaking, we write consonants and vowels in the same way as the ancient Greeks did.

The earliest inscriptions in the Latin alphabet are believed to be from somewhere between the seventh and fifth centuries BC. The Romans acquired their writing from the Etruscans, who had an alphabet based on the West Greek script (DeFrancis, 1989, pp. 181-182). The Latin alphabet originally had 23 letters, which has eventually increased to the 26 letters the Roman script today embraces. <j> and <v> have been added, as they earlier were not distinguished

from <i> and <u> respectively. In addition, <w> was added much later as a ligature of <uu> or <vv> (Coulmas, 2003, p. 93).

The Latin language was spread all over Europe by Christian missionaries after the fall of Rome. Even though most languages in the region did not evolve from Latin, many of them ended up using the Latin script as Latin was the language of literary, philosophical and scientific discourse (Wellisch, 1978, p. 45). While the Latin alphabet spread throughout the continent, it did not come to the situation where one standardized alphabet was used. On the one side, different languages contained phonemes that could not easily be represented by the alphabet, and the alphabet had letters that did not necessarily correspond to a phoneme in every language. To adapt the alphabet to different languages, several different methods were used. Here are some as listed by Wells (2000, pp. 250-253):

- Combining two or more letters to represent a single phoneme. The sound [ʃ] is not a phoneme in Latin, and thus the Latin alphabet does not have a letter that represents it. In English this sound is mostly represented with the digraph <sh> (although other, less typical representations exist, as <ti> in “station”). Other languages employ other digraphs for the sound, like <ch> in French and <sc> in Italian, while the trigraph <sch> is used in German and <skj> in Norwegian.
- Inventing new letters that may or may not be based on already existing letters. This can be seen in Icelandic where the letter <æ> is a ligature of <a> and <e>, while <þ> has come from the Runic alphabet. In some African languages, particularly in Ghana, new letters have been added from the International Phonetic Alphabet (IPA), like <ɛ> and <ɲ>.
- Adding diacritical marks to the basic letters. A great number of diacritics are added to existing letters to create a new, accented letter. The diacritics can be classified according to where they are placed: above the basic letter, below or through it. Examples of the three are <á> from Spanish, <ø> from Norwegian and Danish, and <a̧> from Polish.

A vast number of languages today use the Roman script for their written languages, but English is one of the very few that employs only the basic 26 letters of the Latin alphabet without amendments (Wells, 2000, p. 249). In other words, English has only used the first of the three above mentioned ways to represent phonemes that do not exist in Latin. The

invention of new letters is in general not desired as it impedes the universality of the Roman script (Coulmas, 2003, p. 102).

2.1.1 Conversion of scripts

When exposed to a language other than the official majority language, e.g. in the media, we are usually supplied with a translation as well. Names, however, are not easily translated, nor is it desired. If the language is written in a script other than the Roman one, like Arabic, Chinese, Cyrillic or some other of the numerous scripts that exist, most westerners will not be able to read it, so we need to somehow write the name with Roman letters. Conversion of scripts is also a common way to make literature in other languages available in libraries and bibliographic lists (Wellisch, 1978, p. vii). Two terms are used concerning the conversion of scripts, transliteration and transcription. Wellisch explains the difference between the two (1975, p. x):

Following the established usage of ISO, the term “Transliteration” is employed for “representing the characters (letters or signs) of one alphabet by those of another, in principle letter by letter”, whereas “Transcription” is used for “the operation of representing the elements of a language, either sounds or signs, however they may be written originally, in any other written system of letters or sound signs.”

From this follows that transliteration is a direct conversion of the alphabet, which does not necessarily give any hint to pronunciation, while transcription aims to come as close as possible to the oral realization of the word. The Arabic word <كتب> could be transliterated into *ktb*, without giving heed to the short vowels that are necessarily present in the oral representation of the word, but not represented in the Arabic script. It would be more difficult to give a correct transcription if there is no context that reveals the meaning of the word, and hence which short vowels are present. *Kataba* (“he wrote/has written”), *kutiba* (“it was written”) and *kutub* (“books”) could all be correct transcriptions.

It should be noted that several conventions exist for Arabic transcription, and that transcription is often used in a broader sense than the definition above. Reichmuth (2009, p. 516) says: “Scientific transcription is, essentially, a hybrid system, based on phonetic transcription of Classical or Modern Standard Arabic with some elements of transliteration and morphophonemic representation”. Arabic for “the sun” might well be transcribed as *al-*

šams. The short vowel is included even though it is not when written in the Arabic script, while the definite article here is written *al* as in Arabic, although it would be pronounced *aš* as it is followed by a “sun letter” triggering assimilation. Also one can sometimes see a short vowel in the final position written as if it were long because it is marked long when written in the Arabic script, as in the word *fušḥā*. Both examples contain features from transliteration as well as transcription.

While transliteration and transcription can be a conversion from one script to any other, I will use the term *romanization* for the special case of using any of the two systems for conversion where Roman is the target script.

2.1.1.1 Transcription of Arabic

Writing Arabic with Roman letters is usually done to accommodate foreigners who do not read Arabic or are trying to learn it. As Arabic contains silent letters (like the <ل> inserted after the <و> in the final position of some plural verb forms), ambiguous letters like <و> and <ي>, and lack marking of short vowels, a consistent transcription might be better suited to display correct pronunciation (Beesley, 1998, n.p.). Some literature, like the dictionary of Hans Wehr, offers a transcription of every word next to the Arabic writing, while the Egyptian colloquial textbook of Woidich and Heinen-Nasr (2004), *kullu tamām!*, is in its entirety written with the Roman script.

While academic textbooks and dictionaries rely on a consistent and unambiguous transcription of words in order for the non-native reader to learn the correct pronunciation, a rather random, non-standardized method of transcription can be seen all over the Arab world on shop and road signs, often accompanied by the same word or words in the Arabic script. In this kind of transcription, the Arabic phonemes that do not have a natural representation in the Roman script may be written with a letter whose value phonetically resembles the Arabic phoneme (<h> for /ħ/), with the help of a digraph (<sh> for /š/), or by omitting it altogether or inserting an apostrophe (for /' / and /ʾ/).

The randomness of this method can easily be spotted. In Cairo, road signs only kilometres apart spell the name of the district *muhandisīn* in several different ways, among them <mohandessen>, <mohndseen> and <mohandsein>. According to Reichmuth (2009, p. 515), there are on the Internet more than thirty variants of the name معمر القذافي with the last name

spelled <Qadhafi>, <Gaddafi>, <Kadafi> etc. The lack of rules and consistency sometimes makes it impossible to reconstruct the Arabic from the Roman spelling. Palfreyman & Al Khalil (2007, p. 48) refer to this Romanized Arabic as “common Romanized Arabic.” Reichmuth (2009, p. 515) calls it “ad hoc transcription”, which is the term that will be used in this thesis.

2.1.1.2 Judaeo-Arabic

As will be explained in more detail below, in Egypt one variety of the Arabic language is used mainly for written purposes, while another, quite different, variety is used orally. According to Blau (1988, p. 10), this was not so before Islam, but a situation that evolved as the Arabic speaking areas grew bigger. The Arabs conquered new territories in a short time, and amidst the rapid changes was the emergence of *Neo Arabic* (Blau, 1988, p. 9). Blau uses the term *Old Arabic* to describe the language of *Muhammad* (Blau, 1988, p. 20), while Neo Arabic embraces the dialects that evolved side by side with Standard Arabic.

Judaeo-Arabic “refers to a type of Arabic that was used by Jews and was distinct in some way from other types of Arabic” (Khan, 2007, p. 526). The most remarkable feature of Judaeo-Arabic might be the fact that it was written in the Hebrew script. The Judaeo-Arabic that was written before the 10th century is referred to as early Judaeo-Arabic (Khan, 2007, p.526). One of its characteristics is that the spelling is phonetic based on the orthographic practices used for Hebrew and Aramaic at that time rather than those used for classical Arabic. This means that the <ﻝ> of the definite article was not written when it was followed by a ”sun letter”. Instead, it was assimilated into that letter. A final <ﺀ> was represented with *taw* whenever pronounced as /t/, also in the first word of a genitive construction where it is always written as <ﺀ> in the Arabic script (Khan, 2007, p. 527).

In classical Judaeo-Arabic, approximately from the 10th to the 15th centuries, the orthography had changed to be more similar to that of classical Arabic. The <ﻝ> of the definite article was regularly written, also when followed by a ”sun letter”. <ﺀ> was always represented with *heh* regardless of its pronunciation. Even the few words that contain /ā/ which is not represented in the Arabic script were written without it with Hebrew letters (Khan, 2007, p. 527).

The Judaeo-Arabic from later than the 15th century is called Late Judaeo-Arabic, and in this period, inspired by the orthography of the early period, the Judaeo-Arabic moved further

away from classical Arabic again (Khan, 2007, p. 528). The writers of Judaeo-Arabic texts spoke Arabic dialect, but the language in the texts is not a direct representation of these. They tried instead to produce a literary form (Khan, 2007, p. 528).

2.1.1.3 Romanization of Turkish

Although numerous languages have converted from other scripts to Roman – many European languages that today use the Roman script were earlier written with other scripts – Turkish is the unique example of such a conversion by a whole nation to have taken place in modern times (Wellisch, 1978, p. 57). By the beginning of the 11th century, most of the ancestors of today's Turks, had become Muslim. The religion was introduced to them first and foremost by peoples speaking Persian and other Iranian languages, something which is evident by the fact that most religious terms in Turkish come via these languages (Lewis, 1999, p. 5).

Although Persian had a great influence on the Turkish language, an even larger invasion of words came from Arabic. Not only because Arabic is the language of the Qur'an, and hence is the natural language for religion and theology, but also, Lewis (1999, p. 6) claims, because the importation of an Arabic word, due to most words belonging to a three-consonant root, would bring along other words stemming from the same root. Thus, when the Arabic word *'ilm* "knowledge" entered the Turkish language, its linguistic family was brought along with it, like *'ālim* "scholar", *ma'lūm* "known", *mu'allim* "teacher", *ta'līm* "instruction" and *isti'lām* "request for information".

A lot of Arabic words may have entered the Turkish language, but the Arabic alphabet was never well suited to represent Turkish. Some of the consonants in the alphabet represent sounds not present in Turkish, while the letter <ك> may represent Turkish /g/, /k/, /n/ or /y/. In Arabic the three short vowels /a/, /i/ and /u/ can be indicated through diacritical marks, but Turkish needs to distinguish eight vowels. This means that many words written in the Arabic script could represent several different words. The word *'wlw* may be read as Turkish *ulu* "great", *ulu* (Arabic) "possessors", *ölü* "dead", *evli* "married", *avlu* "courtyard" and *avli* "stocked with game" (Lewis, 1999, p. 27). After reformist Kemal Atatürk replaced the Arabic alphabet with an improvised Roman one in 1928, the new alphabet was completely phonetic and did not contain any unused or silent letters, nor did it contain any digraphs. Wellisch (1978, p. 57) attributes the success of the new alphabet to four factors: The Arabic script was never well suited for the Turkish language; the decision could be backed up by dictatorial

force; more than 90% of the population was illiterate; and the Turkic-speaking peoples in Azerbaijan, Turkmenistan and Uzbekistan had already discarded the Arabic alphabet. The impact of the change of script is summed up by Lewis (1999, p. 37):

[...] the Latin alphabet is undeniably the best that has ever been used for Turkish, and has played a large part in the rise of literacy; according to the official figures, from 9 per cent in 1924 to 65 per cent in 1975 and 82.3 per cent in 1995.

2.1.1.4 Maltese

Maltese is the only Semitic language that is written in the Roman script, and is now considered to be a language on its own rather than a dialect of Arabic due to its early loss of diglossia with any form of Arabic standard, as well as the influence European languages have had on it (Mifsud, 2008, p. 146). Maltese displays some features typical of Maghrebine Arabic, and its initial source seems to be Tunisia (Borg & Azzopardi-Alexander, 1997, p. xiii). It has, however, been heavily influenced by medieval Sicilian, mainland Italian in Tuscan form, and later English (Mifsud, 2008, p. 146). Latin, Sicilian, Italian and English have all been used as written languages in Malta. As the last two were fighting for supremacy, Maltese surprisingly emerged with an official orthography with Roman letters (Borg & Azzopardi-Alexander, 1997, pp. xiii-xiv).

Contemporary Maltese contains 24 consonantal phonemes. Over the years it has lost 11 such, but has also added seven new phonemes from non-Arabic sources. All four emphatic consonants in Arabic have in Maltese merged with their non-emphatic counterparts (Mifsud, 2008, p. 146). After the Arabs had ruled Malta for nearly two hundred years, the Norman conquest in 1090 might explain why contemporary Maltese has lost so many consonant phonemes. Here in the words of Aquilina (1961, pp. 122-123):

The reason is that the new overlords introduced a new set of sounds which were unfamiliar to the Semitic-speaking natives. It must have been at this time that the typically or distinctively Arabic sounds began to give way to sounds common to Sicilian with the result that the emphatics and fricatives of Arabic were levelled down to one sound common alike to Arabic and Sicilian. Indeed, the typically Arabic sounds must have been very difficult to the newcomers and, as often happens when the foreigners had to express such sounds, somehow they did so approximately by using the nearest, or what they thought was the nearest, sound of their own.

The standard orthography of Maltese was officially recognized in 1934 (Aquilina, 1961, p. 75), and as most of the phonemes coming from Arabic were already gone, their representation using the Roman alphabet did not present a great problem. One of the two still in common use, /ħ/, got its own letter, <Ħ, ħ>. The other, /ʕ/, is represented with <Q, q>. The only digraph in the Maltese alphabet, <GH, gh>, although corresponding etymologically to Arabic /ʕ/ and /ġ/, does in general not have a consonantal value except that it “indicates a slight pharyngalisation of the preceding and/or following vowel lengthened” (Aquilina, 1961, p. 130).

As in Egyptian Arabic, the definite article in Maltese is <il-> (or simply <l-> if followed by a word starting with a vowel). Thus, <mara> means “a woman” while <il-mara> means “the woman”. Maltese also contains some “sun letters”, letters that assimilate the /l/ of the definite article if preceding them directly. In Arabic this takes place orally although the orthography does not reveal it. In Maltese, however, the article is written as it is pronounced. In this way the indefinite <xih>³ (“an old man”) becomes <ix-xih> (“the old man”) when definite (Aquilina, 1965, p. 35). Some prepositions, like <bi> (“with”) and <fi> (“in”), are in Maltese attached to a following definite article, thus becoming <bil-> and <fil->. The /l/ will still be assimilated by following “sun letters”, forming e.g. <bis-> or <fix-> (Aquilina, 1965, p. 110). While the attachment of *bi* also happens in Arabic orthography, albeit regardless of what word follows it and not only with the definite article, *fi* is always written separately.

2.1.1.5 Arabic language reform

Even though the Arabic script and writing has hardly been subject to change at all since it was codified, some have voiced the need for modernization. The starting point for this was, according to Abu-Absi (1986, p. 337), the Arab renaissance in the nineteenth century that started in Egypt and Syria, and from there spread to the other Arabic speaking countries. Abu-Absi (1986, p. 339) identifies the three areas which were above all considered to be in need of modernization: orthography, grammar and vocabulary. I will focus here on the first as the two others are less relevant for the topic of this thesis.

Al-Toma (1961, p. 404) pinpoints two defects from which the Arabic writing system suffers. The first is the number of forms the various letters might assume. Not only do most of the

³ In Maltese, <x> represents /š/.

letters have four different shapes, depending on whether they occur as initial, medial or final in the word, in addition to a basic form for when they appear independently – these shapes also vary slightly according to which letter they follow or precede, giving a total of hundreds of different shapes. This problem, however, is more or less eliminated with today's modern computerized typesetting (Abu-Absi 1986, p. 341).

The other defect mentioned by Al-Toma is the tendency not to mark short vowels, which are usually only marked in texts for beginners and children, in addition to the Qur'an. Not only will some literate people see it as an insult to be presented with a vocalized text, but the short vowel signs, which are merely small dots, might also make the text more cumbersome to read as many of the letters themselves already have diacritical dots below or above them to distinguish them from each other. When short vowels are not written, one may have problems pronouncing a word correctly unless one actually knows the word. And since, occasionally, the same spelling is used for several words, only the context will make clear which word is meant. In other words, understanding is a vital part of reading. As Mahmoud (1980, p. 728) says:

Much of the reader's effort is expended in hunting for contextual clues and redundancies that could help him supply the missing vowels. This process usually requires the very grammatical knowledge the writer can afford not to master, but the reader cannot do without. Because of the tradition of printing Arabic without vowels, the writer is not accountable for any built-in ambiguities or vagaries his writing may lend itself to. The onus of deciphering what was written or printed falls upon the reader.

Many have been occupied with reforming the Arabic writing system, illustrated by the fact that the Egyptian Academy of Arabic language received more than 300 reform proposals between 1938 and 1968 (Abu-Absi 1986, p. 339). The attempts vary from a slight alteration of how the script is today to a total romanization. In between the two are suggestions such as replacing the alphabet with a new one that has vowels included, and another that invents new vowel signs that can be included in the main body of the word (Al-Toma 1961, p. 406).

ʿAbd al-ʿAzīz Faḥmi argued that the Roman script was used by a great number of languages throughout the world, and that it had proved to be efficient. Writing Arabic with the Roman script would not only solve the problems with the current script, but it would also, he imagined, bring the Arabic speaking nations closer to the rest of the world (Al-Toma 1961,

p.406). His suggestion in the early 1940s, in the aftermath of Turkey's successful change of script (Somekh, 1991, p. 69), was to replace every Arabic letter with a Roman one, and introduce various amendments to represent the phonemes that are not easily represented using a Roman letter (Al-Toma, 1961, p. 406), a proposal not all that different from academic transcription systems in use today. In 1961 the Lebanese poet Sa'īd 'Aql published a whole volume of poems written in his own dialect using the Roman script with added letters and diacritical signs. Even though another poet later published several volumes using a similar script, the initiative was met with controversy and failed to serve as an inspiration for the next generation of poets (Somekh, 1991, p. 69).

Apparently, all proposals for reform have been rejected with the exception of one with minor ligature changes adopted by the Moroccan government in the late 1950s (Al-Toma, 1961, p. 403). As this reform only intended to bring down the total number of ligatures in print, it is outdated with today's modern printing press. The difficulty of presenting an acceptable proposal is summed up by Abu-Absi (1986, p. 341):

The problem with most of these proposals was the fact that they represented a break with a very important tradition. Some introduced new complexities into the existing system; and others, which involved simplifications, did not satisfy the esthetic demands which the Arabs have traditionally placed on the system. The change to a Latin alphabet has been and will be viewed with extreme suspicion on both nationalistic and religious grounds.

The official romanization of the Arabic script does not, in spite of the examples mentioned above, seem to be a realistic nor popular solution to what some regard as a defective writing system. Al-Toma (1961, p. 408) argues that the success of romanization depends on how well the existing script is suited for the language it portrays, and the Arabic alphabet was, as we have seen, not well equipped to represent the Turkish language in the first place, in particular not its vowel system. He also argues that the Arabic alphabet has been in use for a long time, and a great number of books and texts have been written using it for the past thirteen centuries, a fact which would also make such a drastic change unwelcome (Al-Toma, 1961, p. 408).

2.2 The language situation in Egypt

2.2.1 Arabic in Egypt

The language situation in Egypt, as in all other Arabic countries, is one that Ferguson (1959, p. 336) termed *diglossia*. In short, he describes diglossia as a situation where one or different dialects are spoken, while the written language is a more formal variety of the same language which is never used for ordinary conversation. The formal variety, which so far in this thesis has been referred to as Standard Arabic, is called *fuṣḥā* in Arabic, while the vernacular is called *‘āmmiyya*. In general in diglossic situations, the former is referred to as the High variety, while the latter is the Low variety.

A diglossic situation is, however, not necessarily as dichotomous as the concept of two varieties living side by side may sound. As one of several who has introduced more levels than the two, Badawī (1973, p. 7) distinguishes five different levels of Arabic in Egypt: One of them is what we might call “pure” *fuṣḥā*, another “pure” *‘āmmiyya*, while the three others are mixed varieties in between the two. Even though he defines five levels, he stresses that what lies between the two “pure” or extreme variants, is a continuum where there are no clear boundaries. All Egyptians are, to some degree, capable of elevating or lowering their speech when the situation calls for it (Badawī, 1973, pp. 8-11).

As Ferguson’s definition of diglossia states, *‘āmmiyya* is the vernacular that people use in everyday situations. It is the variety that parents speak to their children, and hence becomes their mother tongue. It is also spoken in television series, movies and might also be used in talk shows. *Fuṣḥā* is used orally in television news broadcasts and documentaries among others, and its affiliation with the Qur’an makes it the natural choice in religious discourse.

In many situations only one of the two will be appropriate, and a breach of these conventions will not pass unnoticed. This is well illustrated in a story told by an informant to Parkinson (1991, p. 40). A friend of the informant, a passionate supporter of *fuṣḥā*, had decided to only speak *fuṣḥā* to his children so that they would have the advantage of having it as their mother tongue. Once the two of them and the friend’s three year old daughter got on the bus. The daughter was lost in the crowd, and all the yelling that was needed to find each other again took place in *fuṣḥā*. Inappropriate as this is, everybody in the bus burst out laughing. Even the

informant started laughing because, as he said it himself: “you don’t talk to your three-year-old daughter in *fuṣḥā*.”

2.2.1.1 Attitudes towards the varieties

The terms that are commonly used to describe the two variants might give a first hint at what attitude people have towards them. *Fuṣḥā* can be translated into something like “eloquent”, while *‘āmmiyya* means “common”. *Fuṣḥā* is the language of the Qur’an, the language of God, and is considered to be both beautiful and perfect. A foreigner trying to learn to speak the vernacular will often find it difficult to elicit words from locals. An Egyptian who is asked about a word in Arabic, will in most cases try to give an answer in *fuṣḥā*. When it comes to grammar, *fuṣḥā* is again perceived as superior. It has clear rules, while *‘āmmiyya* is perceived to be a language of anarchy where almost everything is permitted.

2.2.1.2 The continuum between the varieties

As mentioned above, Badawī speaks of five different levels of Arabic in Egypt, but recognizes that the boundaries between them are fluid. As there is a continuum between the “pure” *fuṣḥā* and the “pure” *‘āmmiyya*, there is actually no need to discuss boundaries at all. Instead, there is a gradual increase in the proportion of *fuṣḥā* features the higher up on the continuum the speech is. Holes (1995, p. 280) writes:

The descriptive difficulty is that the language levels which constitute much of this speech continuum (however one defines ‘level’ and however many levels one proposes) are probabilistic, not absolute: in other words, the levels are constructs, produced by the patterns of simultaneous choices which speakers in a community make, in a consistent and predictable way, on many linguistic variables. But there are few variables where one of the variants which constitute it always occurs in one level and never in the adjacent one: the way most work is that the likelihood of one variant rather than its rival occurring gradually increases in a given range of contexts in more or less the same proportions for all speakers.

Although the terms *fuṣḥā* and Standard Arabic are often treated as synonyms in academic texts, the term *fuṣḥā* is in Egypt used about a very specific form of Standard Arabic.

Parkinson (1991, p. 33) was told by a dean at Al-Azhar University that Egyptians had no knowledge of *fuṣḥā* whatsoever, explaining that the language in newspapers is not *fuṣḥā* at all. *Fuṣḥā*, to him, was limited to the classical period. Others use the term *fuṣḥā* also when

speaking about the modern version of the written language. Still it seems that many even see *fuṣḥā* as its own little continuum, and that a text is “more” *fuṣḥā* when it includes “a fine, somewhat archaic style, recherché vocabulary, heavy use of metaphor, and a certain expected complexity or even convolutedness.” (Parkinson, 1991, p. 34).

2.2.1.3 Egyptian Arabic as a written language

As the official written language of Egypt, *fuṣḥā* dominates newspapers, magazines and books. Formally, the vernacular in Egypt, I will refer to it as Egyptian Arabic⁴, cannot be written as there are no official rules of spelling. In reality, however, it can be done as the sounds of the dialect are mostly present in *fuṣḥā*, and therefore letters that can represent them exist in the Arabic alphabet. Although the writing of a dialect that has no formal rules presents freedom for the writer, some of the orthography of *fuṣḥā* is often applied. This happens even when a writing closer to the oral realization is possible.

Because of the perceived inferiority of the dialect, any text written in dialect would automatically be considered a text of low status, which again never has encouraged the writing of such texts (Rosenbaum, 2004, p. 282). Toward the end of the twentieth and beginning of the twenty-first century there were enough Egyptian Arabic writings around for Rosenbaum (2004, p. 283) to claim that it more or less had its own orthography. The same alphabet as in *fuṣḥā* is used, but some letters can represent other phonemes than they do in *fuṣḥā*. He says (2004, p. 285):

‘Āmmiyya orthography is, in fact, the end result of the confrontation between two opposing tendencies. One preserves as much as possible the spelling rules of fuṣḥā and its graphic patterns (in words of fuṣḥā origin or those derived from radicals which exist in fuṣḥā). The other forms a graphic representation of the pronunciation and sounds of ‘āmmiyya [...] such graphic representation may occur even when it contradicts the orthographical rules of fuṣḥā. Between these two extremes lies a wide range of intermediate possibilities, frequently giving rise to two or more ways of spelling one word.

For some time it has been common for novelists to write dialogues in dialect and the narrative in *fuṣḥā*, but of late a few books written entirely in dialect have started to appear. In his book *kitāb ma lūš ism* (“A book with no name”), which is a gathering of small essays rather than a novel, the writer Aḥmad al-‘Asīlī (2009, p. 9) writes on his choice of language:

⁴ There are different dialects also within Egypt. In this thesis Egyptian Arabic refers to the vernacular of Cairo.

ما كنتش عارف اكتب بالعامية ولا بالفصحى!.. مشكلة معضلة جدا.. الفصحى هي لغة القراءة و الكتابة.. فيه طريقة واحدة لكتابة و نطق كل كلمة ممكن تستعملها.. بس العامية مش كده.. ممكن مثلا تكتب "إنهرده" و ممكن تكتب "النهارده" و ممكن الاتنين ببقوا صح, و ممكن مايقوش. الفصحى عندها قدرة أكبر على البلاغة, بس العامية أقرب للقلب. الفصحى قد تنال إعجاب محبي اللغة العربية و المثقفين, بس العامية بتكلم كل الناس. المهم من غير رغي يعني قررت اني اكتب زي ما بفكر, بالعامية. و لو فيه حاجة استعصت عليها, اكتبها بالفصحى.

I did not know whether to write in *‘āmmiyya* or *fushḥā*. A very problematic issue. *Fushḥā* is the language of reading and writing, and there is only one way to write and pronounce every word. *‘Āmmiyya*, however, is not like that. For example, you can write “today” [*’innahar dah*] or you can write “today” [*il nahār dah*], and they can both be correct. *Fushḥā* is more capable of eloquence, while *‘āmmiyya* is closer to the heart. *Fushḥā* might offer admiration from friends of the Arabic language and the educated, but *‘āmmiyya* addresses everyone. To get to the point: I decided to write as I think, in *‘āmmiyya*, and if something proves difficult to convey, I will write it in *fushḥā*.

Al-‘Asīlī writes, something that Rosenbaum (2004, p. 283) also notes, that in dialect, there are often several ways to write a word, contrary to in *fushḥā* where the spelling of most words is regulated by clear rules. He chooses to write in dialect because it is the language in which he thinks, and it addresses everyone, not just the elite.

2.2.2 English in Egypt

Towards the end of the nineteenth century, English was only the fourth largest of the European languages in Egypt, after French, Greek and Italian. English spread rapidly in the public schools during British occupation, but French was still the first choice of the Egyptian aristocracy (Schaub, 2000, p.227). Today learning a foreign language is mandatory in schools, and English, French and German are most common. The Egyptian government is trying to promote more foreign language teaching in schools, especially English (Mahrouse, 1994, p. 1946).

Haeri (1997, p. 161)⁵, interviewing diplomats and physicians and others from the upper class, discovered that they had, with few exceptions, received all their education in a foreign language. In the language schools, maybe as little as a few hours per week is spent learning *fushḥā*. All other classes are taught in the foreign language of each school. In public schools, English is of course only taught in English classes, and the quality of this teaching is thought

⁵ There is some uncertainty as to whether this is actually the source, as the book was not available the weeks before I finished the thesis.

to be poor (Schaub, 2000, p. 230). Ever since Sadat opened up the Egyptian economy, more and more students are learning English in the hope of getting employment in a foreign company operating in Egypt (Elkhatib in Schaub, 2000, p. 228). English language schools are now much more common than French (Imhoof in Schaub, 2000, p. 230). This is one of the reasons why I chose for this thesis to assume that English forms the basis when chatters on the Internet romanize Arabic, rather than French or any other language.

Chapter three: Computer-mediated communication

On October 29, 1969, the student Charley Kline sat in front of a computer at the University of California in Los Angeles (UCLA) and attempted to send the sequence of characters "LOGIN" to a computer at Stanford Research Institute via a network connection over a distance of roughly 5000 kilometers. Though initially the transfer of data appeared to be successful, the system crashed when he had reached the letter G due to an overload. Still, that date marks the birth of the exchange of data via computer.

This is how Beißwenger (2002, p. xvii) describes the very modest start of computer-mediated communication (CMC) more than forty years ago. Although the seventies saw the emergence of electronical chat as well as mail, this was a technology that was only available to the few. After the invention of home computers and more developed networks, their use was no longer limited to military or academic circles (Beißwenger, 2002, p. xvii). The Internet has, in a relatively short time, become by far the most important electronical network. According to the Computer Industry Almanac⁶, the number of users has exploded, and in 2006 there were 1.2 billion users worldwide, up from two million in 1990, 45 million in 1995 and 430 million in 2000. It is expected that the 2 billion mark will be reached in 2011 or 2012.

3.1 Communication situations on the Internet

Language use on the Internet is not homogenous. As in "real" life, some situations will be seen as more formal than others, thus requiring more formal language. Crystal identifies seven

⁶ <http://www.c-i-a.com/internetusersexec.htm> [Retrieved 21 October 2010]

different situations that are useful to distinguish for someone interested in Internet linguistics (Crystal, 2006, pp. 11-15):

- **Electronic mail (e-mail)**

E-mail is today mostly referring to a message sent from one private inbox to another. The writer can send the e-mail to whatever e-mail address he wants, and only the recipient or recipients can read it.

- **Chatgroups**

Chatgroups are discussions that take place in particular “rooms”. The chatgroups will often be organized around topics that are discussed by those present in the room. This situation can be divided into two subcategories, depending on whether the discussion is in real time (synchronous) or in postponed time (asynchronous). In a synchronous situation the user enters a room and joins a discussion between other users who are online and present in the room at that very time. In an asynchronous situation users can read posts written by others at an earlier stage, and one cannot expect an immediate response, something which is necessary in a synchronous discussion where everything written is usually lost as it is being pushed out of the screen by newer text.

- **Virtual worlds**

A virtual world is an imaginary world where users enter the role of a fantasy character. This type of communication differs from the already mentioned situation in that the users are not talking about real topics, but the characters, events and environments of the virtual world (Crystal, 2006, p. 178)

- **World Wide Web (WWW)**

The World Wide Web consists of all computers linked to the Internet which contain documents written in the HyperText Transfer Protocol, HTTP. A web browser is needed in order to view these documents in a readable format.

- **Instant messaging**

Instant messaging (IM) allows for people who know each other to communicate synchronously in private. This differs from e-mails where the messaging is asynchronous, and from chatgroups where the involved users may not know each other.

- **Blogging**

Blog is short for weblog. A blog is a website where the owner or owners can write about whatever he or she wants. Many are personal diaries, others might write on a

certain topic. If the owner enables the possibility of readers to comment, discussions may arise. One thing they all have in common is that they are unmediated. No editor is there to correct or approve the text; the owner gets the last word.

These seven situations – synchronous and asynchronous chatgroups are counted as two – are not entirely mutually exclusive. One can find several of them combined, or one situation used within another (Crystal, 2006, p. 15). While the social network website Facebook is accessed through the World Wide Web, several of the other situations are available within the Facebook platform. The data for this thesis is gathered from groups within Facebook where members can discuss or proclaim their opinions. As this does not take place in real time, my data would best fit in to the situation called asynchronous chatgroups. I will give a more detailed description of Facebook below.

3.1.1 CMC – between speech and writing?

It is possible today to communicate with others on the Internet without typing a single word. With a microphone and a loudspeaker one can talk with each other as if it was a regular phone call. In fact, Crystal (2006, p. 16) recognizes that interactive voice dialogue soon will be (or now probably already is) a situation that can be added to the others. However, most of the communication taking place on the net is still done by typing letters and words that others can read on the screen. The newspaper language on the Web is more or less equivalent to that of printed newspapers, hence it is not perceived as being a different kind of language. This probably goes for the majority of the content on the Web. It is when we come to the other situations that the borderline between writing and speech sometimes becomes blurred.

Before going deeper into the question of whether some of the situations should be considered closer to writing or speech, it can be useful to look at some differences between the two. The perhaps most obvious differences can be summed up as done so here by Coulmas (2003, p. 11):

[...] it is by no means self-evident how an audible sound continuum produced by the human voice, which can only be perceived at the time of utterance, relates to a discrete sequence of fixed visible marks produced by the human hand, which can be perceived at any time.

Looking at it this way, all the seven mentioned situations are clearly writing. There are, however, more subtle differences as well. Writing tends to consist of longer sentences and is more structurally complex than speech, while speech contains more one-word sentences, a narrower lexicon and more slang, to mention some (Baron, 2008, p. 47).

What Crystal calls synchronous chatgroups, is, together with all other written synchronous CMC, often referred to as *chatting*, a name that itself indicates that this type of communication can be seen as some kind of “written speech”. As Storrer (2002, p. 3) points out, the chatters themselves usually refer to their chatting activity with verbs such as “say” and “hear”, rather than “write” and “read”. Nevertheless, the chatters are well aware of their ambiguous situation, something which is clearly shown in Storrer’s example where a chatter called “dr.hc” writes <könnd ih mis höen?> and another replies <dr,hc. nur schwach, es fehlen Buchstaben...> (Storrer, 2002, p. 3).

Even though the use of oral expressions and incomplete sentences can give chatting some resemblance to speech, numerous factors ensure that chatting can never replace speech. Intonation and loudness of the voice can imply a message on its own, something which is not possible to do with letters on the screen. Body language is also not easily conveyed in writing. While someone listening can hear what the other says as it is being uttered, and even interrupt if necessary, the chatter can never read what the other one is writing until the message is finished and the enter-key on the keyboard is pressed. In fact, the chatter can never be sure that he will get an answer at all, or how long it will take. If no answer is received, he does not know whether the other party did not get the message, is away from the keyboard for a moment, or simply does not want to answer (Crystal, 2006, p. 35).

According to Baron (2008, p. 69), what makes it natural to talk about an IM conversation rather than writing, is first and foremost the fact that young people during chatting, as with speech, tend to do other activities simultaneously with the chatting. Also, the conversations are usually not subject to corrections from anyone. Baron’s conclusion is nevertheless that instant messaging contains some features of both speech and writing, but not “enough speechlike elements [...] to explain why it seems so natural to talk about IM ‘conversations’ and not IM ‘letters’” (Baron, 2008, p. 69). Of Crystal’s (2006, pp. 50-51) seven situations, he identifies the Web as being by far the closest to written language, while IM is at the other extreme. Netspeak, which is the term he uses for all seven situations, “is better seen as written

language which has been pulled some way in the direction of speech than as spoken language which has been written down” (Crystal, 2006, p. 51). Frehner (2008, p. 26) summarizes that many linguists have concluded that CMC is a hybrid register that makes use of written as well as spoken language.

In a diglossic society like Egypt, the discussion about whether CMC is writing or speech has an additional dimension. While written English offers the writer the possibility to make a text less formal e.g. with the help of lexicon and contractions, Standard Arabic will always sound formal as it is a language never used in the everyday life. As most of the CMC situations are rather informal, Egyptians will find that they have to write in the vernacular, thus already leaning towards speech. An IM conversation taking place in Standard Arabic would be just as inappropriate as the father talking Standard Arabic to his daughter on the bus mentioned above.

3.2 The Internet in Egypt and the Middle East

Never before has an information technology spread as fast as the Internet. It took the radio 38 years to reach 50 million users and the television 13 years, whereas the Internet only needed four years to reach the same number (Wheeler, 2006, p. 18). It is probably no exaggeration to say that the Internet has revolutionized the way we communicate and acquire information. Instead of waiting for the newspaper in the morning, or being content with what the radio or television decides to broadcast, we can read the latest news more or less as they happen. And more importantly, with the help of a search engine we can, in a matter of seconds, find information about obscure topics, information that earlier might have been difficult to find in books where only the title is searchable. This enormous wealth of easily accessible information is what makes the Internet appreciated and, by some, feared. Transparent and democratic societies have little to fear from the information that the Internet makes available. Less democratic regimes, however, that depend on keeping government criticism away from the masses, are less enthusiastic. Many less democratic countries have been reluctant to allow Internet connections, and when they first did, it was often with direct censorship, or they would at least supervise its usage closely.

The Middle East was first connected to the Internet when Tunisia linked up in 1991. Kuwait followed the year after, while the UAE, Turkey and Egypt went online in 1993. More

reluctant were Syria and Saudi-Arabia who did not connect until later in the nineties (Wheeler, 2004, n.p.).

3.2.1 The Egyptian Internet user

According to Internet World Stats, 10.5 million Egyptians, 12.9% of the population, were using the Internet as of December 2008⁷. The number of users⁸ has doubled since 2006 when there were 5.1 million users, 7.0% of the population. Although the penetration rate is higher than most countries in Africa south of the Sahara, it is one of the lowest in the Middle East and North Africa. The penetration rate in Western European countries is usually somewhere between 60% and 90%.

The reasons for the low penetration rate should be the same for Egypt as for the rest of the region. Illiteracy is no doubt a problem for someone who wishes to access the web. According to the CIA World Factbook, 71.4% of the Egyptian population aged 15 and over can read and write⁹, thus the remaining 28.6% are more or less excluded from using the web, at least on their own. The economy might be an even bigger obstacle for many. Using a computer in an Internet café for an hour might cost around ten Egyptian Pounds, which is as much as an average Egyptian earns from a day's work. For those having a hard time just obtaining life's basic needs, the Internet is a luxury they cannot afford. The economical argument is supported by Wheeler (2006, p. 37) who notes that the connectivity in the oil rich Gulf states "is in some cases thirty times (per capita) more than it is in countries in North Africa and the Fertile Crescent". Other explanations could be the rapid spread of mobile phones, and the fact that many do not know anyone else who uses the Internet, and thus would have nobody to communicate with there anyway (Wheeler, 2006, pp. 34-35).

From this can be deducted that Egyptian Internet users must be from the wealthiest layers of the population. As most URLs are written with Roman letters – Arabic letters in web addresses are still extremely rare – knowledge of the Roman script and keyboard is a great advantage, if not a necessity, to navigate on the web. This means that most Egyptian users can also be said to be well educated. When it comes to age, there is no reason to believe that the

⁷ <http://www.internetworldstats.com/africa.htm#eg> [Accessed 9 February 2009]

⁸ Internet World Stats adopts a broad definition of a *user*, namely as "anyone currently in capacity to use the Internet." Only two conditions need to be fulfilled for a person to be an internet user, a) "The person must have available access to an Internet connection point" and b) "The person must have the basic knowledge required to use web technology" (<http://www.internetworldstats.com/surfing.htm> [Accessed 9 February 2009]).

⁹ <https://www.cia.gov/library/publications/the-world-factbook/geos/eg.html> [Accessed 9 February 2009]

distribution is different from the rest of the world. The 20-30 age group are the most active, making up twice as many Internet users as their share of the population would suggest in 2003. The most rapid growth is now seen in the below-20 group (Hofheinz, 2007, p. 62).

3.2.2 Arabic online

In the early years of the Internet, and also during the rapid spread in the nineties, English was by far the most used language worldwide. In the mid 90s, a widely quoted figure said that more than 80 percent of the Net was in English (Crystal, 2006, p. 229). The estimates for other languages have since risen, and in 1998 the number of newly created non-English websites had passed that of English (Lebert in Crystal, 2006, pp. 230-231).

As mentioned earlier, the use of Arabic in web addresses is still very rare, and it was not even possible until recently as only Roman characters were accepted. That meant that if the owner of a site wanted its URL to be in Arabic, it had to be written using the Roman script.

Numerous such sites exist today, some examples are:

- www.ahram.org.eg – (*ahrām*) An Egyptian state controlled newspaper.
- www.otlob.com – (*uṭlub*) A site that offers online ordering from numerous restaurants in Cairo.
- www.almokhtsar.com – (*al-muxṭaṣar*) A news site formed by radical Wahhabis opposed to the Saudi regime (Hofheinz, 2007, p. 72).
- www.7alim.com – (*ḥalīm*) A site dedicated to the Egyptian singer ‘Abd al-Ḥalīm Ḥāfīz.

These few sites show some of the features that characterize romanized Arabic, all of which will be elaborated on further below:

- Long vowels are often not distinguishable from short.
- Short vowels are sometimes written, sometimes not.
- Some Arabic sounds that do not have their equivalence in the Roman script are represented by numbers.

Another site worth mentioning is www.yamli.org, a search engine designed to help Arabs

who are not comfortable writing Arabic letters on the computer. The user types a word with Roman letters, and then chooses from a list of suggestions which Arabic word is meant. If a user types <3arab>, the site will give عرب as the first and most likely option. However, as this ad hoc transcription is ambiguous, other less likely options are also suggested, like عراب and عارب. Once the desired word is selected, the user can press enter to search the web for documents containing the word, or several words in a sequence.

The last few years, typing with Arabic letters online has become a widely available option, thus giving the user a choice. This means that five options can be considered in Egypt today:

- Standard Arabic with Arabic letters
- Standard Arabic with Roman letters
- Egyptian Arabic with Arabic letters
- Egyptian Arabic with Roman letters
- English

3.2.3 Language choice online in Egypt

Some studies exist on what language and script Egyptian Internet users prefer to use.

Warschauer, El Said & Zohry (2007, pp. 308-309) did a survey among 43 “young professionals”, all of whom had a bachelor’s degree and 70% a master’s or doctoral degree. Of the 31 participants who answered under the category “online chat”, 71.0% reported that they used English, 6.5% Classical Arabic in Arabic script, 9.7% Classical Arabic in Roman Script, 16.1% Egyptian Arabic in Arabic script and 54.8% Egyptian Arabic in Roman script. As the authors themselves point out, this survey has its limitations as the sample was small, non-random and non-anonymous.

Another survey is based on the answers of 502 students at the American University in Cairo (AUC) (Abdulla, 2007, p. 119). The students were asked to estimate how much of their time online they write with or read Arabic letters, the alternatives being 0%, 10%, 20% and so on. 30.3% never read Arabic letters, while 37.1% read them 10% of the time. This means that two thirds hardly ever read Arabic letters on the Internet. 77.1% never write with Arabic letters on the net, and 13.9% use it 10% of the time (Abdulla, 2007, p. 123). The answers are more evenly spread when the students are asked how much they read and write Arabic with English

(Roman) letters. Unfortunately the exact numbers are not given – they are shown as graphs. From the graphs a very rough estimation can be made, however. It seems that the average time they use to read and write Arabic with English letters is somewhere around 30-40% (Abdulla, 2007, p. 125).

Both surveys suggest that Arabic is used rather sparingly by the participants on the Internet. English is by far the most used language among the young professionals, and, although it is not stated explicitly, must be assumed to be so also among the AUC-students as English is the teaching language there. The two surveys do not, however, represent the average Egyptian Internet-user as both samples consist of users who have studied or are studying in English, and whose general capabilities of English are excellent. Also worth mentioning is that both are based on the participants' self-reporting. As Parkinson (1991, p. 60) points out, Egyptians tend to have very different opinions as to what is *fushā* and what is not. Besides, any particular person would find it difficult to give anything other than a very rough estimation of one's own language use on the Internet.

Facebook has in very few years become immensely popular among students and youth all over the world. Among Internet users in their teens and twenties, it is uncommon to come across anyone who does not at the very least have an account on Facebook. This means that samples taken from Facebook will include users from different social backgrounds. To examine language use among Egyptians on Facebook, the ten first wall-entries in fifty randomly chosen groups were counted. Although a few users have written more than one entry, this gives a sample from nearly five hundred different users.

Table 3.1 *Distribution of language among Egyptian users of Facebook*

| | Number of entries | Number in % |
|---------------------------------|-------------------|-------------|
| Egyptian Arabic (Arabic script) | 180 | 36.7 |
| English | 125 | 25.5 |
| Egyptian Arabic (Roman script) | 111 | 22.6 |
| Standard Arabic (Arabic script) | 75 | 15.3 |
| Total | 491 ¹⁰ | 100.1 |

Table 3.1 suggests that Facebook-users use considerably less English than the participants in

¹⁰ The total number of entries is less than 500 as some of the groups contained fewer than ten wall-entries.

the two surveys mentioned above. Only one quarter of the entries were written in English, the rest were in some form of Arabic. Egyptian Arabic written with Arabic script is the preferred variant with its 36.7%. 15.3% writing Standard Arabic seems like a high number, and might be due to some of the groups being devoted to religious themes. Egyptian Arabic with Roman script, the variant that will concern the rest of this thesis, constitutes 22.6% of the wall-entries.

3.3 About Facebook and communication between its members

Facebook describes itself as “a social utility that helps people communicate more efficiently with their friends, family and coworkers.”¹¹ In February 2004 Facebook was founded and opened to all students at Harvard University. From being initially available to students only, it opened up to everyone throughout the world in September 2006, reaching more than 12 million active users (active users are defined as “users who have returned to the site in the last 30 days”¹²) three months later. Recently the network has witnessed an explosion in popularity, reaching 100 million active users in August 2008 and 175 million by February 2009.¹³

Information about the age of users is not supplied, but the currently fastest growing group consists of users 30 years and older,¹⁴ which could simply mean that those younger than 30 are already members.

Facebook does not offer official statistics showing the geographical background of the users, but estimated numbers can still be found using its tool for creating advertisements.¹⁵ An advertiser can insert what country is targeted, and other more specific characteristics like gender and age, and see approximately how many users that correspond. The total number of users in Egypt 11 March 2009 is estimated to be 1 125 420. Of those 920 040, or 81.8%, are between the ages of 16 and 30. Even though my data was collected more than a year earlier, it can be assumed that most of the data will be from this age group as they were the first to join.

Every registered user of Facebook has his or her own profile. The profile can contain pictures and personal information such as address, phone number, education, work and interests. Also displayed are other users of Facebook whom the owner of the profile considers to be friends.

¹¹ <http://www.facebook.com/press/info.php?statistics#/press/info.php?factsheet> [Accessed 9 March 2009]

¹² <http://www.facebook.com/press/info.php?statistics#/press/info.php?factsheet> [Accessed 9 March 2009]

¹³ <http://www.facebook.com/press/info.php?statistics#/press/info.php?timeline> [Accessed 9 March 2009]

¹⁴ <http://www.facebook.com/press/info.php?timeline#/press/info.php?statistics> [Accessed 9 March 2009]

¹⁵ <http://www.facebook.com/ads/create/> [Accessed 11 March 2009]

Through the “privacy settings”, the user can decide whether all users should be able to see his profile, or whether it should be accessible to friends only.

The site offers several ways for the users to communicate with each other. Sending a message would be the obvious choice for two or more interlocutors to have a private conversation. The receivers will be notified of the message as soon as they log in, and can read and answer the message. Only those participating in the message exchange can read the messages sent, thus it can be said to be an internal e-mail.

A more public way to communicate is to write on someone’s “wall.” The wall is a part of the profile, and anyone who can access someone’s profile can also write on the wall and read everything that has been written there. The entries will be shown chronologically with the newest on top. Although it is possible to write on one’s own wall, it is more common to answer on the wall belonging to the other user. Often one can find several successive entries on a wall written by the same person, many of which will look as if they are responses to something not visible on the wall. A curious outsider will then have to visit both profiles and read one entry at a time for the conversation to make any sense.

A third way of communicating is through photographs that users can upload and place in albums which friends can browse. After uploading photos one can “tag” any friend appearing in them, and the friends will be notified of this. Commenting on a picture works in much the same way as writing on a wall and can develop into a conversation or friendly dispute between several friends, but the activity will often cool down in a relatively short time.

All means mentioned so far are usually restricted to communication between friends, although it is possible also to send a message to users one does not know. In addition, Facebook offers the possibility for individuals, who do not necessarily know each other, to join groups. Anyone can create a new group, and the creator can choose whether the group is open to everybody or only to whoever he decides to accept as members. When a new group is created, it must be decided whether the group should be global, meaning that anyone can join, or inside a country network. If the latter is chosen, only those who are members of the same country network can join.

Groups are created to gather people sharing an interest, whether it be supporting a sports team, political ideologies or a particular cause, listening to a type of music, or people having something in common, like living in the same place or having attended the same school in childhood. The groups offer two main alternatives for the members to communicate: the wall, which is equivalent to those on each profile, and the discussion board. Although the discussion board would seem to be the more suitable of the two in order to have a discussion, it is hardly used at all in many groups. The reasons for this are probably that the wall is more visible, and it takes fewer clicks with the mouse, and thus less time, to read and write on it. When someone writes on the wall it is usually a statement regarding the topic of the group. Sometimes the statement is an independent one, other times an answer to an entry written by someone else, and a discussion with two or more contributors could develop. Also common is people using the wall to advertise for other groups or commercial websites. As everything written on the wall in groups is open for everyone to see, as opposed to the wall on user profiles which can often only be seen by their friends, I chose to collect data for this thesis from them. This way the selection will be random and from numerous participants.

3.4 Method

3.4.1 Selection criteria

As everything written on the wall in groups is open for everyone to see, as opposed to the wall on user profiles which can often only be seen by their friends, I chose to collect data for this thesis from them. This way the selection will be random and from numerous participants.

As I wanted only Egyptian Arabic data, only groups that are inside the country network for Egypt were used in the searching for data. This was assumed to increase the number of posts written in Egyptian Arabic, as well as making it less probable that those writing would elevate their language or in any other way adapt it to make it more understandable to non-Egyptians. The majority of the posts on the walls are written in English or in Arabic (Standard Arabic or Egyptian Arabic) with Arabic script, all of which have been disregarded. The following criteria were used to decide which posts were to be included in the corpus:

- Posts must be written in ERA. Many posts written in ERA contain English, some only a word or two, others to such an extent that half the post is in English. Posts with a

few words or even a sentence in English were included in the data, but I decided to exclude posts with larger chunks of English.

- No more than one post from each user. To systematically collect several posts from different users is not possible without knowing which groups they are members of, so I decided to collect only one post from each user even in the few instances where more were available.
- A post should not contain less than twenty words. Shorter posts would reveal very little of what is sought for this thesis.
- Obvious spam and advertisements were excluded.

The corpus consists of 110 posts from as many different users. A total of 7255 words gives an average of nearly 66 words per user. Age and location of the participants is not known unless they somehow reveal it in their post. Gender can often be seen from the users' name, but some do not use their real name. It was anyhow decided to analyse the posts without taking age and gender into consideration.

I considered two different approaches before gathering the data. One was to collect a large amount of writing from a few informants, the other to get a smaller sample from numerous users. A larger sample from each user would have made it easier to look for intrapersonal consistency. That is, to unearth to what extent a user has one fixed way to spell every word, or whether he performs an oral to written, or phoneme to grapheme, transcription continuously while writing.

I chose the latter approach mainly for two reasons. Firstly, I wanted to investigate ERA as a general phenomenon. With data from only a few persons, there is no way to be sure that they are in any way representative for the writing in general. With data from a large number of chatters, in this case over a hundred, the data almost certainly gives a descriptive picture of how ERA is written. While the samples are so small that the same word is seldom written multiple times, they are large enough to demonstrate consistency, or lack thereof, on the phoneme level, even though it makes it harder to investigate how consistent every user is in his spelling.

To get more data from each user would, using Facebook, have been more or less impossible without actually involving the users. The second reason, then, is the observer's paradox, as

formulated by Labov (1972, p. 209): “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed”. Although the paradox is meant for speech, it is presumably valid also for writing as there are always several ways to compose the same statement. In a formal written language, the writer could among other things elevate the sentence in terms of syntax and vocabulary if he thinks that is what the researcher wants. In a non-normative writing system even the orthography can be influenced if the writers know that their work will be scrutinized. In the groups on Facebook, the chatters write with and for each other, thus making sure that the writing is completely “natural”. Copying all of the data used for the thesis from groups like these ensured that I got data from a community without showing that I was observing it.

3.4.2 Notes on transcription

Seeing that transcription is merely about finding one symbol to represent another symbol or a specific sound, everyone who writes about Arabic could, in theory, define his own transcription system as long as it is clarified and applied consistently. However, for the sake of simplicity and readability, most would favour using one of the established systems. In the transcription I chose to use for this thesis, where the representation of each Arabic phoneme is adopted from Woidich (2006), each phoneme is represented by one single letter or symbol.

As one of my research questions is to what extent ERA is influenced by Arabic orthography, I needed a basis saying what it would look like if it was not influenced by script whatsoever. My basis is that the users, for lack of anything better, write with a phoneme-grapheme correspondence in mind, and that without the influence from Arabic orthography they would attempt to write as they perceive their speech. To analyze the material, I had to interpret and reconstruct the spoken form the written message presumably reflects.

Egyptian Arabic is well documented as a language system, and my interpretations are based on Egyptian Arabic system as described in Woidich (2006). Thus, when vowels were omitted in the users’ text, as demonstrated in <hnekdar nkelel mn el salbeyat> (*ḥaniqdar niqallil min is-salbiyyāt*, “we will be able to reduce the negatives”) and <bn3rf ntkalm> (*bini ‘raf nitkallim*, “we know how to talk”), I was able to fill in the missing vowels in my reconstruction.

However, I soon discovered that many vowels I expected to be elided, were in fact retained in the writing. One example of such is the user who writes <ya3ny keteer awy benesma3> (*ya ʿni ktīr ʾawi bnisma*, ‘‘that is, we often hear’’). This user writes the /i/ in *ktīr* and the first /i/ in *binisma* although they would probably not be pronounced as the preceding words end with a vowel (a detailed description of the rules of elision, shortening and prolonging of vowels in Egyptian Arabic can be found in 4.2). Examples such as this could perhaps be explained by the speech pace. In slow speech elisions are less likely to occur. But then other, and more striking, examples came about. It is hardly imaginable that a speaker of Egyptian Arabic, no matter how slow his speech, would retain the /i/ of the definite article if it is preceded by another vowel, but this occurs frequently in the data. Examples from four users are <fe el kolya> (*fi l-kulliyya/k-kulliyya*, ‘‘in the faculty’’), <fe el balad> (*fi l-balad*, ‘‘in the country’’), <fy el zamn> (*fi z-zaman/z-zamān*, ‘‘in the era’’), and <ehna el sabab> (*iḥna s-sabab*, ‘‘we are the reason’’).

At first, I proceeded taking phonological processes of elision and shortening of vowels across word boundaries into consideration, but the findings mentioned above made me reconsider. It seemed that many users perhaps were writing every single word isolated. After much deliberation, I decided to do the reconstruction word-by-word instead. With this interpretation, I was able to register representations that would otherwise have gone unnoticed.

However, choosing to interpret the posts word-by-word had the consequence that, for instance, <enha t7es> was reconstructed as *innaha tiḥiss* (‘‘that she feels’’), although it would normally be pronounced *innaha thiiss*. More often, however, and as seen in the first two examples above, vowels are not marked even when they certainly would be pronounced in speech. It is thus unclear whether elisions in the data that coincide with speech are there intentionally, or merely by coincidence.

I admit that my interpretation has its weaknesses, and in particular when the definite article is preceded by a vowel, it looks rather artificial. I considered excluding the definite article from the registering of vowels, but ended up including it for the sake of consistency and the fact that many are represented nevertheless.

As much as I wanted to keep my reconstruction simple and clear, I found it necessary to make one exception from the word-by-word method. In Egyptian Arabic, the preposition *li* with a suffix often forms a unit with the preceding word, something that affects the pronunciation of the suffixed *li*. Isolated, *līha* would be used for "for her", attached to a verb it would be *laha*, or *lha* after a vowel (Woidich, 2006, pp. 41 and 138). Thus, when someone writes <yegeblaha>, I transcribed this as *yigib laha* instead of *yigīb līha* as it would be transcribing them word-by-word.

Sometimes the normally expected version is added to the word-by-word-transcription. Assimilation of the /l/ of the definite article, and the variants of the feminine marker, will always be marked in the transcription.

Another decision I made, is that the data was interpreted as reflecting Egyptian Arabic although one could sometimes wonder if the writer intended single words to be Standard Arabic. When a user consequently writes <wa> ("and") in an otherwise Egyptian Arabic environment, I registered this as the Egyptian Arabic equivalent *wi*. The same goes for those who have written the definite article <al>, which was reconstructed as *il*. One of the things I set forth to do in my analysis, was to find out how the different phonemes of Egyptian Arabic are represented in ERA. If I had interpreted these two words as *wa* and *al* based on the users' representation, I would have made a premature conclusion.

When a user writes <nadam> ("regret"), I registered this as *nidim* because it is the only form the dictionary¹⁶ of Egyptian Arabic includes. However, I cannot rule out the possibility that the user for some reason wrote this one word in Standard Arabic, or that he actually says *nadam*. One reason for doing so could be that he or she is not from Cairo, and hence that some verbs are conjugated using other vowels than in the Cairene dialect (Woidich, 2006, p. 62). In the same way I cannot exclude that someone might actually say <tob2ah> as *tub`a* even though I had to interpret it as common Egyptian Arabic *tib`a* ("it is/becomes"). The same could be an explanation for the two cases of <e> representing /u/, namely <7ekooma> (*ḥukūma*, "government") and <bey2akel> (*biyākul*, "he is eating"). Dictionaries and grammars will never be able to include all variants of words and possible pronunciations, and one native speaker might pronounce a word differently than the vast majority.

¹⁶ Badawi & Hinds' Dictionary of Egyptian Arabic was consulted whenever I was in doubt about a vowel or the meaning of a word.

/ʔ/ in initial position is only written in my transcription when it is a reflex of Standard Arabic /q/.

Chapter four: The graphemes

This chapter will take a closer look at how the Egyptian Facebook-users convert Egyptian-Arabic phonemes into Roman letters or numerals. I will start with the consonant phonemes, and continue with the vowel phonemes.

4.1 The consonants

Arabic is a Semitic language, and as Watson (2002, p. 1) says: “Semitic languages are marked by a limited vocalic system and a rich consonantal system.” In the official Arabic language, often referred to as Modern Standard Arabic, there are 28 consonants. In the modern Arabic dialects there have been changes concerning the number and pronunciation of some of these phonemes. Urban dialects spoken outside the Arab peninsula tend to be more innovative in terms of changes, while the dialects of nomads and the peninsula to a larger degree retain the classical features (Watson, 2002, pp.13-14). Cairene Arabic belongs to the former category, and its consonantal system contains 25 core phonemes. In addition there are eight marginal phonemes which are restricted to loanwords or to the speech of educated speakers (Watson, 2002, p. 20). Among these eight phonemes are what Woidich (2006, p. 11) calls the secondary emphatic consonants, /r/, /m/, /b/ and /l/. Even though minimal pairs can be found for each of them, they are not included in the table below as they are usually not distinguished from their non-emphatic counterparts in dictionaries. The only marginal phoneme that is present in the table is /q/. According to Watson (2002, p. 21), it is used only in religious and Standard Arabic lexemes, but it occurs often enough in the data not to be overlooked.

Table 4.1 *The consonant phonemes used in ERA.*

| Phoneme | ERA representation(s) | Example(s) from data | Translation of example(s) |
|----------|-----------------------|----------------------|---------------------------|
| /b/ | b | bld | country |
| /t/ | t | tany | other |
| /j/, /g/ | g | gameela | pretty |
| /ħ/ | 7, (h) | 7aga, haga | thing |
| /x/ | 5, 7', kh | 5iana, ad7'ol, okht | betrayal, I enter, sister |
| /d/ | d | nadam | he regretted |

| | | | |
|-----|--------|------------------|-------------------|
| /r/ | r | ragel | man |
| /z/ | z | ezzay | how |
| /s/ | s | lessa | not yet |
| /š/ | sh | mesh | not |
| /ṣ/ | s | masr | Egypt |
| /ḍ/ | d | 7fdl | I will continue |
| /ṭ/ | t | tab3an | of course |
| /ẓ/ | z | metzabata | in order |
| /ʿ/ | 3 | 3alam | world |
| /ġ/ | 3', gh | 3'eer, shaghleen | except, working |
| /f/ | f | fekra | idea |
| /q/ | k, q | mawkef, qura | opinion, villages |
| /k/ | k | kalemat | words |
| /l/ | l | lazem | have to |
| /m/ | m | menha | from her |
| /n/ | n | nas | people |
| /h/ | h | feha | in it |
| /w/ | w | law | if |
| /y/ | y | y5af | he is afraid |
| /ʔ/ | 2 | so2al | question |

4.1.1 Numerals representing phonemes

Table 4.1 contains the 26 different consonant phonemes that are necessary to produce Egyptian colloquial. To write Egyptian colloquial with Roman letters, the writer would have to know the Roman alphabet through knowledge of, at least the basics of, a language that is usually written with these letters. Although it might not be true for all writers present in the data, English is the most widespread foreign language in Egypt, and it will be assumed that English pronunciation forms the writers' mind when transcribing. Knowing that the Roman alphabet, as used in English, only has 21 consonant letters, there are not enough of them to simply replace every Arabic one with an English one. 15 of the Arabic consonants can easily be exchanged with an English consonant representing the same phoneme: /b/, /t/, /g/, /d/, /r/, /z/, /s/, /f/, /k/, /l/, /m/, /n/, /h/, /w/ and /y/. The remaining consonants in the English alphabet are used, in English, for sounds that do not exist, or only marginally, in Egyptian Arabic, like

<j> and <v>, or they do not represent a distinct phoneme, like <x> and <c>. Moreover, Egyptian Arabic makes use of several phonemes that do not exist in English. Some, like /ħ/ and /q/, are totally unknown sounds that are tricky for most Europeans to learn. The glottal stop, /ʔ/, does exist as a sound in some English dialects, but this is a phenomenon that is not revealed in the orthography, and hence the English alphabet is not equipped with a letter that would naturally represent it. In other words, creativity is needed to find a useful representation from the Roman script for some of the Arabic phonemes.

In English CMC, some numerals are used to form small rebuses, in which the important thing is how they are pronounced. <2> standing alone can represent the words “to” or “too”, just as <4> can represent “for”. These numbers, and others, can also be used to represent parts of words by replacing the letters that would normally be pronounced as the numeral itself. Examples are <b4> (b + four) “before”, <l8r> (l + eight + r) “later”, <2l8> (two + l + eight) “too late” and <m8> (m + eight) “mate” (Crystal, 2006, pp. 91-92). This shows that, in English CMC, it is not how a word is spelled that decides whether a numeral can be inserted or not, but how it is pronounced. In English this is possible because most of the numerals consist of very few phonemes, contrary to Arabic where they are based on the three-consonant root system, like most other words. As for the Greek alphabet, Tseliga (2007, p. 118) speaks about two different forms of transliteration, “phonetic” and “orthographic/visual”. The first case is when users employ Roman characters to represent phonemes in Greek words, while the latter is when a Greek letter is replaced by a Roman letter or number that visually resembles it. She gives an example (Tseliga, 2007, p. 118):

If a Greek user wants to type the word Αθήνα in the Roman alphabet, the two main options are “Athina” and “A8hva.” The phonetic alternative, “Athina,” should be familiar to most Greeks and foreign visitors to Greece since it is identical to the official transliteration according to ELOT standards and is found on road signs, maps, and passports. Conversely, “A8hva” would be intelligible only to a competent – if not native – Greek user who could identify the visual similarity between “8” and [θ].

4.1.1.1 Representation of /ħ/

The orthographic/visual transliteration is used to represent several different phonemes in ERA. While <ahmed> (aħmad) is understandable to everyone who can read Roman letters and knows this common Arabic name, <a7med> certainly is not. As in the Greek example above, <7> to represent /ħ/ is probably chosen because of its relative visual similarity to the

Arabic letter <ح>. Although a few of the users seem to use both <7> and <h> quite randomly to represent /ħ/, the great majority consistently write the unambiguous <7>. Names represent an exception. Examples such as <m7md> (*muḥammad*) and <a7med> occur, but the norm seems to be that they are written using letters only. In translated texts the users will be accustomed to seeing Arabic names written with Roman letters using ad hoc transcription. Users who have attended or are attending schools instructing in a European language, have probably seen names written this way throughout their lives. Thus when writing names they are not in “phonetic mode” anymore, instead they write them as they have always done.

4.1.1.2 Representation of /x/ and /ġ/

While /ħ/ more or less consistently is represented by <7>, some of the other phonemes leave the users with a choice. Whenever writing /x/, a user has three different options to choose from. One of them is <7'>. As for /ħ/, this is based on the number 7 looking like the letter <ح>, and with the apostrophe representing a diacritical dot above it, <7'> forms a <خ>. Another alternative is to write <5>, which occurs slightly more often than <7'> in the data. Unlike all the other numerals encountered, which are used to approximate an Arabic letter visually, the use of <5> is probably based on the fact that the Arabic word for five, *xamsa*, starts with a /x/. The third, and last, option is <kh> which is regularly seen in both academic and ad hoc transcription.

Table 4.2 *Distribution of representations of /x/.*

| Representation of /x/ | Number of users | % of these users |
|-----------------------|------------------|------------------|
| <5> | 30 | 43.5 |
| <7'> | 27 | 39.1 |
| <kh> | 12 | 17.4 |
| Total number: | 69 ¹⁷ | 100.0 |

Table 4.2 shows the distribution of the three variants. As all users consistently stick to one of the options throughout their post, the number of users has here been counted instead of the total number of occurrences. Of course, it cannot be ruled out that a user may vary between the variants in different posts, but it is not very likely. As we shall see, there are other inconsistencies in the posts, so it is natural to assume that it would be seen also in this matter

¹⁷ The data consists of 110 posts, but /x/ was present only in 69 of them.

if the choice was based on something else than a permanent personal preference. It is not surprising that <5> is the most popular representation considering that it is written pressing only one key. One can wonder, however, why the majority prefer one of the two digraphs when there is a simpler and faster option available.

The case of /ġ/ is similar to that of /x/, but it lacks the one-key option. The two digraphs are “related” to those representing /x/. <3> is used as a mirrored image of <ġ>. When adding an additional diacritical dot, or apostrophe, <3’> for <ġ’> is constructed in the same way as <7’> for <ġ’>. The other alternative is <gh> which is common in academic and ad hoc transcription.

Table 4.3 *Distribution of representations of /ġ/.*

| Representation of /ġ/ | Number of users | % of these users |
|-----------------------|-----------------|------------------|
| <3’> | 42 | 77.8 |
| <gh> | 12 | 22.2 |
| Total number: | 54 | 100.0 |

As table 4.3 shows, most users write the number-variety in this instance as well. <gh> is encountered in 12 posts, just as <kh> is, but in percentage terms it is still slightly more widespread. This may be due to it having less competition. Table 4.4 shows how users write both <kh> and <gh> in posts where at least one of the representations, and both phonemes, are present:

Table 4.4 *Co-occurrence of <kh> and <gh>.*

| Representation of /x/ and /ġ/ | Number of users | % of these users |
|-------------------------------|-----------------|------------------|
| <kh> and <gh> | 6 | 60.0 |
| <kh> and <3’> | 2 | 20.0 |
| <5> and <gh> | 2 | 20.0 |
| <7’> and <gh> | 0 | 0.0 |
| Total number: | 10 | 100.0 |

The two more popular combinations, <7’> and <3’> on the one hand, and <5> and <3’> on the other, are not included in the table as they are not relevant here. Out of ten posts, six users combine <kh> and <gh> while only four use one of them together with a maybe “cooler” alternative for the other. Although ten users is a very small number to conclude from, the

table indicates that using the digraph consisting of two letters is a conscious choice, and that there definitely is a correlation between the use of <kh> and <gh>. If it was not, the combination of e.g. <5> and <gh> would be higher than that of <kh> and <gh> as the number of users who write <5> is more than double that of <kh>, but instead it is notably lower. This must mean that many indeed see <kh> and <gh> as somehow being connected.

One potential disadvantage using these two digraphs is that they can be ambiguous. In addition to representing the phoneme /x/, <kh> can just as well be read as /k/ followed by /h/, as when a user writes <masakha> (*masakha*, “he grasped her”). The same goes for <gh> that can also be understood as two single phonemes in Egyptian colloquial, namely /g/ and /h/. Three users in different groups write <weghet> (*wighit*, “point of view”). These being the only examples in the data, show that the consonant combinations of /k/ and /h/, as well as /g/ and /h/, are relatively rare. And when one considers that it was easy for me to detect their function in the examples, it should not be a problem at all for Egyptian native speakers.

4.1.1.1 Marginal representations

The following is an extract from a post written in a group where Egypt is discussed:

*< lazem ay 7ad tabe3y ye7eb balado 9 lazem yeshofha 7elwa 7ata law
fiha 7agat keter awy 3'alat , di 7aga zay 7ob elwa7ed l mamto 9
babah 9 a7'wato.>*

Of course anyone should love his country and regard it as beautiful even if there are a lot of bad things there. It's like the love one has for one's mother, father and siblings.

While the visual similarity is striking, this user seems to be alone in choosing to write <9> for <و>. And contrary to how numbers are usually applied, <9> here does not consistently represent the phoneme /w/. In fact, when not isolated, /w/ is represented with <w>, as is shown in <7elwa> “beautiful”, <law> “if”, <awy> “very”, <elwa7ed> “one” and <a7'wato> “his siblings”. <9> is written only when standing alone, thus forming the word *wi* “and”. Since the data only includes one post from each person, it is not known whether this user consistently writes like this. <9> is not written by any other user in the data, but it is applied for other purposes in other parts of the Arab world. I will write more about that below.

Another striking feature can be found in one post where the user does not write any numerals at all:

<bgd ana moshtaraka fe el group da mesh ashan ana bakrah haga fe misr ana bahebaha awy bkol elly feha bshwareha bneela kolha gamela ana moshtareka ashan akool llelly bykrah haga feeha enna ehna el sabab ehna elly amleen feha keda llelasaf asl el wageb tkeel allena awy ya gamaa law kol wahed amel elly allaih hnekdar nkelel mn el salbeyat mesh hanzelha laan mafeesh balad mafefhash salabiat ya gamaa kefaya enak tkool enak masry we teftakher bkeda>

Really, I participate in this group not because I hate something in Egypt. I love her a lot with everything she has, her streets and her Nile, everything is beautiful. I participate to say to those who hate something in Egypt that we are the reason. We are the ones who do this to her, unfortunately, so we have a heavy task in front of us, people. If everyone does what he has to, we will be able to reduce the negative things. We will not get rid of all because there is no country without negative things. So, people, enough of this. Say that you are Egyptian, and be proud of it.

In this post, the user completely disregards all instances of /‘/, as we can see in <ashan> (‘*ašān*, “because”) and <amel> (‘*āmil*, “do”). Further, she does not differentiate between /h/ and /ħ/, when everyone else would use <7> for /ħ/. The former can be seen in <bakrah> (*bakrah*, “I hate”), the latter in <haga> (*ḥāga*, “thing”), <ehna> (‘*iḥna*, “we”) and <wahed> (*wāḥid*, “one”) and both in <bahebaha> (*baḥibbaha*, “I love her”). Finally, what is normally pronounced /ʾ/ as a reflex of <ق> is written with <k>, as in <akool> (‘*aqūl*, “I say”) and <hnekdar> (*ḥaniqdar*, “we will be able to”). As the data only contains one post from each user, I do not know whether this user always writes like this, or some technical issue made it difficult or impossible to write numbers at the time of writing. At least she has shown that an understandable text can be produced even without a unique representation for each consonant phoneme.

4.1.2 Other consonants

4.1.2.1 Representation of /w/ and /y/

These are the two consonants that in the Arabic script are written with letters that can represent a consonant or a vowel. The letter <و> can represent either /w/ or /ū/, while <ي> represents /y/ or /ī/. The simple solution for the chatters would be to represent /w/ with <w> and /y/ with <y>, which is also what they do. However, there are a few scattered examples

that they are omitted, most often if they are preceded by the vowel that in Arabic would be written with the same letter. That is, sometimes /w/ is omitted if preceded by /u/ as in <hoa> (*huwwa*, “he”), while /y/ can be omitted if it follows /i/, as in <hea>, <hia> (*hiyya*, “she”) and <motden> (*mutadayyin*, “religious”). In <tdeon> (*tadayyun*, “religiousness”), /y/ is omitted without the presence of an /i/. In the exceptional example <kois> (*kuwayyis/kiwayyis*, “good”), they are both missing.

4.1.2.2 Representation of /q/

While /q/ always corresponds to the Arabic letter <ق>, the letter is not always realized as /q/ in Egyptian Arabic as the normal reflex of Standard Arabic /q/ is /ʔ/ in Egyptian Arabic. However, in many lexical words, usually cultural and stylistic loan words from Standard Arabic, the variant /q/ is preserved. Four users write <q>, as in <qura> (*qura*, “villages”), <3oqad> (*uqad*, “problems”), <nentaqid> (*nintaqid*, “we criticise”) and <ll2mr el waq3> (*li il amr il wāqi*, “indeed”). Most of the time, however, it is written <k>, as in the examples <mostakela> (*mustaqilla*, “independent (f.)”), <mawkef> (*mawqif*, “stopping place”), <2ektesad> (*iqtiṣād*, “economy”) and <rakam> (*raqam*, “number”).

4.1.2.3 Representation of /ʔ/ as a reflex of Standard Arabic /q/

Rosenbaum’s (2004, p. 286) work on written Egyptian *ʾāmmiyya* with Arabic script in literature shows that <ق>, although pronounced /ʔ/ in the majority of instances, is still usually written with <ق>. Only occasionally is it written as <ء>.

The situation seems to be quite different in ERA. The data contains 400 occurrences where /ʔ/ as a reflex of <ق> would be expected in speech. 323 (80.8%) of these are represented with <2>, presumably due to its visual similarity with <ء>, as seen in <ye2ol> (*yi ʾūl/yu ʾūl*, “he says”), <delwa2ty> (*dilwa ʾti*, “now”) and <ba2a> (*ba ʾa*, “so, then”). The remaining 77 (19.2%) are not marked at all. 74 (96.1%) of these 77 are in word-initial position, thus giving a clear pattern as to when it is not marked. The three omissions in medial position are <ba> (*ba ʾa*, “so, then”), <delwaty> (*dilwa ʾti*, “now”) and <yool> (*yi ʾūl/yu ʾūl*, “he says”). These are more or less isolated instances, as the three same users mark the /ʔ/ in other words, and other users mark it in these words, as seen in <ba2a>, <delwa2te> and <ye2ol>.

The word that seems to be most consistently written without any marking of this /ʔ/ is *ʔawi* (“very”). Out of 29 users who have written *ʔawi* once or more in their post, 25 (86.2%) chose to omit the <2>. By far the most common way of writing it is <awy>. Some choose <awi>, while one single user opts to write <awe>. Among the users who mark the /ʔ/, we find <2we> twice, in addition to <2awy> and <2wyyyyyyyy >.

55 users have in their post one or more words that have an initial /ʔ/ as a reflex of <ق>. 34 (61.8%) of these never mark the /ʔ/, 13 (23.6%) always do so, while eight (14.5%) have examples of both. Some of the users in the first two groups only had one such word in their post, the last group might have been larger if there had been more data from each user. The last group is in many ways the most interesting as it demonstrates that, to them at least, not all initial /ʔ/'s are alike. The best example of this is the user that writes <ader> (*ʔādir*, “capable of, able to”) three times and <osady> (*ʔuṣādi*, “opposite of me”) once, but then again <2alby> (*ʔalbi*, “my heart”) twice in addition to <2al2y> (*ʔalʔi*, “my worrying”) once. Another writes <ma 2ademetlosh> (*ma ʔaddimit lūš*, “did not present to him”), but <alby>. The six remaining from the last of the three groups all have *ʔawi* as the word where /ʔ/ is omitted. It seems then, that *ʔawi* is more prone to be written without the initial /ʔ/ than other words, but the data is not large enough to conclude in this matter.

4.1.2.4 Representation of /ʔ/ as a reflex of Standard Arabic /ʔ/

The glottal stop as a consonant in Standard Arabic, is often not retained in its Egyptian Arabic cognates. If in the medial position, it frequently becomes a /y/ (the Standard Arabic word *bāʔin* (“evident”) is *bāyin* in Egyptian Arabic), or it is dropped while the short vowel preceding it becomes long (the Standard Arabic words *raʔs* (“head”) and *biʔr* (“well, spring”) are *rās* and *bīr* in Egyptian Arabic). In the final position it usually disappears, and the long vowel preceding it, if any, is shortened (the Standard Arabic word *samāʔ* (“heaven”) is *sama* in Egyptian Arabic). In the initial position it can often be elided, but this elision is usually optional, and it is less likely to be elided by educated speakers. Some words that in Standard Arabic have an initial /ʔ/ that can be elided in Egyptian Arabic, are *umm* (“mother”), *abyaḍ* (“white”) and *inta* (“you (m.s.)”). It is never elided in loan words from Standard Arabic, as *ʔabadan* (“never”) and *ʔagāza* (“vacation”), and when it belongs to the root of the word, as in *ʔakl* (“food”) and *ʔugra* (“rent”) (Woidich, 2006, p. 36).

There are 163 <2>'s encountered in the data that reflect a /' / and not a /q/. The single word that by far appears most frequently is *ra`y* with or without a suffix, as in <r2y> and <ra2y> (“opinion”), <r2ey>, <ra2ye>, <ra2yee>, <r2ii> (*ra`yi*, “my opinion”), <ra2yak> (*ra`yak*, “your (m.) opinion”), <ra2ioh> (*ra`yu*, “his opinion”), <r2yaha> (*ra`yaha*, “her opinion”) and <r2yko> (*ra`yuku*, “your (pl.) opinion”). Some other words present are <so2al> (*su`āl*, “question”), <mas2oleya> (*mas`uliyya*, “responsibility”), <y2s> (*ya`s*, “hopelessness”) and <tas2li> (*tis`ali*, “you (f.) ask”).

The norm in my data is that an initial /' /¹⁸ is not written: <ana> (*ana*, “I”), <el atfal> (*il atfāl*, “the children”) and <a3mel eh> (*a`mil ē*, “what should I do?”). Six users, however, more or less consistently choose to mark it. These six users share among them 69 of the total 163 occurrences. In addition four others have a few occurrences of marking an initial /' /, but appear to be less consistent in doing it. One of these users writes <wel moshkela 2en 2el 2e7bat fe kol mkan> (*wi il muškila in il iḥbāt fi kull makān*, “and the problem is that the frustration is [present] everywhere”). Here the user marks every /' / in the initial position, apart from in <wel> where the two vowels in *wi* and *il* are merged into one in *wi l* (“and the”). Others leave out the short vowel after /' /, as in <2l2nsn> (*il insān*, “one”), <2smaha> (*ismaha*, “her name”), <2w> (*aw* “or”) and <2y> (*ayy*, “any, what”). One user has 13 instances of /' /, in all but two the short vowel is omitted. The two with a short vowel after /' / are verbs: <2a2ol> (*a`ūl*, “I say”) and <2a2dar> (*a`dar*, “I am able to”). In both examples the second <2> is a reflex of /q/. Perhaps the vowel is retained to avoid two succeeding <2>'s, or simply just because it would be harder to understand without the vowel. This user does not write any verbs where the first root consonant is other than /' / for comparison.

The few users who mark /' / in all or nearly all instances apparently do not distinguish whether the /' / is pronounced or not. Especially the /' / of the definite article *il* is prone to be left out in speech if it follows a vowel. Three users have one example of a marked /' / after a vowel. One writes <2ana kman ban2eloko 2el 2e7bat 2ely galy> (*ana kamān ban`iluku il-iḥbāt illi gā li*, “I will also give you the frustration that came to me”). An expected pronunciation of <ban2eloko 2el 2e7bat> would be *ban`iluku l-iḥbāt*. Another user writes <fe 2el qura> (*fi il qura*, “in the villages”), while the third writes <mshbet2ly 2lterf 2ltany> (*mish bit`ul li il-ṭarf it-tāni*, “you are not telling me the other part”). The representation of the definite article is treated further in 5.1.

¹⁸ /' / is used here even though it may be elided.

4.1.2.5 The emphatic consonants

Seeing that only the primary emphatic consonants have special Arabic letters to represent them, <ص>, <ض>, <ط> and <ظ>, I will keep the secondary emphatics out of this. As a general rule in my data, the emphatic consonants are not distinguished from their non-emphatic equivalents. Examples for each of them are <sha5sya> (*šaxšiyya*, “personality”), <nedifa> (*niḏīfa*, “clean”), <ybtI> (*yibaṭṭal*, “he stops”) and <weghet nazary> (*wighit nazari*, “my point of view”). This means that in ERA <s> is used for both /s/ and /s̄/, <d> for /d/ and /d̄/, <t> for /t/ and /t̄/, and <z> represents /z/ as well as /z̄/. While several phonemes from the Arabic language that do not have a corresponding Roman letter have been innovatively replaced by numbers, the emphatic consonants are treated as they mostly are in ad hoc transcription: they are represented by a letter which represents the non-emphatic, otherwise similar, phoneme. This mixing does not seem to cause any confusion or problems among the chatters.

Among the 383 underlying emphatic consonants I found in the data, there is actually one case of it being marked. This user writes <6ol 3mrha> (*tūl ‘umraha*, “all her life”), using <6> to represent /t̄/. While this is not common in Egypt, it is common elsewhere, as I will write more about below. With two of the other emphatic consonants, the user follows the same pattern as the others, which means no distinction: <7fdI> (*ḥafḍal*, “I will continue”) and <sd2ony> (*šadda ‘ūni*, “believe (pl.) me”).

4.1.6 Regional differences

I have chosen to call the language written in the data Egyptian Romanized Arabic not only because the underlying dialect is Egyptian, but also because it might be possible to identify the nationality of a writer without even looking at the words or grammar, but only by how certain phonemes are represented. Egyptian Romanized Arabic, instead of Romanized Egyptian Arabic, implies that the Romanization process in this writing is done “the Egyptian way”. I do not know whether this is applicable for all countries where Arabic is spoken and a Romanized version exists in CMC, but it is certainly possible to distinguish in this way between Egypt and the two other countries I have found research from, UAE (Palfreyman & Al Khalil, 2007) and Morocco (Berjaoui, 2002).

Of the three countries, chatters in the UAE stand out as the most innovative. As in Egypt, <7> and <3> represent /ħ/ and /ʕ/ respectively. However, the diacritical dot that in Egypt follows the letter, precedes it. This means that /x/ and /ġ/, that Egyptians write <7'> and <3'>, in UAE are written <'7> and <'3>. <kh> and <gh> is not seen at all, but that might have been different had their corpus contained data from more than ten users (Palfreyman & Al Khalil, 2007, p. 54). Moroccans, on the other hand, do not make a visual representation of the diacritical dot. They write <7> and <3> just like their fellow chatters in Egypt and UAE, but for /x/ they use <x> and <kh>, while only <gh> is used for /ġ/ (Berjaoui, 2002, p. 455).

Although emphatic consonants and their non-emphatic peers are not distinguished in ERA, this difference is, as Palfreyman & Al Khalil (2007, p. 55) point out, crucial to Arabic speakers:

For example, if native English speakers who do not speak Arabic were asked to represent /s'/'¹⁹, they would probably use <s> (the closest correspondence in English), but none of the conversations in the corpus used <s> to represent this sound. For an Arabic speaker, /s/ and /s'/' are quite distinct phonemes, as distinct as the sounds at the beginning of “sing” and “thing” are for English speakers.

Chatters in the UAE differentiate between the emphatic consonants and their non-emphatic peers. This is done by using numerals that can be said to represent the letters visually, just as with most of the numerals used in ERA, and an apostrophe is added before the number when a diacritical dot is needed. Table 4.5 shows how users in the three countries write the emphatic consonants.

Table 4.5 Representation of emphatic consonants in Egypt, UAE and Morocco.

| | Egypt | UAE | Morocco |
|-----|-------|------|---------|
| /š/ | <s> | <9> | <s> |
| /ḍ/ | <d> | <'9> | <d> |
| /ṭ/ | <t> | <6> | <t> |
| /ẓ/ | <z> | <'6> | <z> |

¹⁹ /s'/' is here used for /š/.

The quotation above explains why the seemingly similar phonemes /s/ and /ʃ/ are represented with <s> and <9> respectively in the UAE: They are in fact distinct phonemes to an Arabic speaker, and sometimes replacing one with the other can create a different word that is not at all related to the original one. The lack of distinction in Moroccan Romanized Arabic is according to Berjaoui (2002, p. 454): “probably due to the fact that the use of appropriate phonetic symbols would require not only their availability on the computer, but also some knowledge of the MA [Moroccan Arabic] phonetic system, which the chatters fail to have.” One can readily agree that no symbols on the keyboard logically represent the emphatic consonants, although the UAE study shows that finding unique keys is possible. However, the fact that UAE chatters have found a way to write them means that they are fully aware of their existence. The conformity of representing consonants in all three countries rather suggests that when someone starts using this language, they write it like they have seen others do it before them. This view is supported by Berjaoui, whose data is gathered from synchronous CMC over a period of nearly two years (Berjaoui, 2002, p. 453):

The chat language under study can be referred to as a system as it interestingly displays regular patterns that all chatters obligatorily adopt in their daily computer-mediated communication with the exception of new chatters (chat beginners), who get accustomed to the regularities of the system through an indirect as well as a direct initiation from the advanced, experienced chatters.

Chance more than anything might have decided what the first chatters started using, and in Morocco and Egypt ad hoc transcription might have had the strongest influence. The ambiguity of <s>, <t>, <d> and <z> in these two countries does not seem to impair understanding between the chatters.

4.2 The vowels

As mentioned above, Arabic is a language with a fairly limited vocalic system. Standard Arabic differentiates between only three vowels, in return each of them constitutes two phonemes as they occur both short and long. The short vowel phonemes in Standard Arabic are /a/, /i/ and /u/, the long vowel phonemes /ā/, /ī/ and /ū/. The long vowels are marked in Arabic script, while the short ones are normally not, exceptions being children’s literature and the Qur’an.

While some Arabic dialects, e.g. the Maghreb dialects (Versteegh, 2001, p. 166), have fewer vowels than Standard Arabic, most, like the Egyptian one, have more. In addition to the three long vowels already mentioned, Egyptian Arabic has two more. Both of them are reflexes of Standard Arabic diphthongs, /ō/ for /aw/ and /ē/ for /ay/. However, the diphthongs are not eradicated from the dialect. Watson (2002, p. 23) lists *šaytan* (“to behave naughtily”), *dawla* (“state”), and *dawša* (“loud noise, din”) as examples of a large number of words with the diphthong intact. The existence of minimal pairs, like *‘awza* (“wanting”) and *‘ōza* (“want, need”) on the one hand, and *šayla* (“carrying”) and *šēla* (“burden”) on the other, shows that the two long vowels must be considered to be phonemes (Watson, 2002, p. 23).

According to Norlin (1987, pp. 48-49), the literature seems to disagree whether there are three or five short vowels in Egyptian Arabic. The minimal pairs supporting the existence of [o] and [e] as phonemes, however, can only be found when a long corresponding vowel is shortened through adding a pronominal suffix to a verb or a noun. One example is *bēt* (“house”). When adding the suffix *-na* (“our”), the vowel is shortened, *betna* (“our house”). The only difference between *betna* and *bitna* “we spent the night” is the first short vowel; hence this might look like a minimal pair. This shortening is, as will be explained in detail below, completely predictable, and for this reason they are usually not considered to be phonemes. Instead, [o] is considered an allophone of /u/, and [e] an allophone of /i/ (Norlin, 1987, p. 49).

In the data, vowels are used less consistently than the consonants. Users will sometimes write long vowels in the same manner as if they were short, and short vowels are sometimes not written at all. For the analysis, it is imperative to know when a vowel is normally pronounced long and when it is short. Even though this is not as simple in Egyptian Arabic as in Standard Arabic, where vowels are never shortened or elided when a suffix is added, knowing the syllabic structure, as presented by Woidich (2006, p. 21), is very helpful.

Table 4.6 *Syllables in Egyptian Arabic* (C = consonant, v = vowel, \bar{v} = long vowel)

| | Open | Closed | Double closed |
|----------|-------------|---------------|---------------|
| Short | Cv | | |
| Long | C \bar{v} | CvC | |
| Overlong | | C \bar{v} C | CvCC |

A syllable is open if it ends with a vowel, closed if the last letter is a consonant, and double closed if ending with two consonants. The double closed syllable is only found finally; from this follows that one will never have three successive consonants. C̄C is also always at the end. When dividing a word into syllables, one starts from the end of the word. A syllable always starts with a consonant, so the first consonant that follows a vowel will separate two neighbouring syllables. Woidich gives the following examples (2006, p. 21):

| | | |
|-----------------------|------------------------------------|------------------------------|
| <i>šuftuhum</i> | > <i>šuf – tu – hum</i> | "I saw them" |
| <i>ma-msikš</i> | > <i>mam – sikš</i> | "he did not grasp" |
| <i>ma-gabithalnāš</i> | > <i>ma – ga – bit – hal – nāš</i> | "she did not bring it to us" |

Sequences of three open syllables, where the one in the middle is short and contains an /i/ or /u/ that is not emphasized, will be reduced to two syllables through the elision of this vowel (Woidich, 2006, p. 22):

| | | | |
|--------------------|------------------------------|---------------------|------------|
| (a) CvCiCv > CvCCv | (b) C̄vCiCv > C̄vCCv > CvCCv | | |
| <i>wiḥiš + a</i> | > <i>wiḥša f.</i> | "bad" | |
| <i>xuluṣ + it</i> | > <i>xuṣit</i> | "she was exhausted" | |
| <i>kāmil + a</i> | > <i>kāmla</i> | > <i>kamla f.</i> | "complete" |

Elision can also take place over word boundaries (p. 22):

| | | |
|------------------------|-------------------------|--------------------------------------|
| <i>ṣāḥib ilbēt</i> | > <i>ṣaḥb ilbēt</i> | "the landlord" |
| <i>ana fi xidmitak</i> | > <i>ana f-xidmitak</i> | "I am at your service" ²⁰ |

Three successive open syllables, of which the first has a long vowel and the second a short /a/, will be reduced to two syllables by elision of the /a/ (p. 22)²¹:

| | | |
|-----------------------|------------------------|---------------|
| C̄vCaCv > CvCCv | | |
| <i>midōḥas + a</i> | > <i>miduḥsa f.</i> | "inflamed" |
| <i>bani 'ādam + a</i> | > <i>bani 'adma f.</i> | "human being" |

²⁰ Woidich (p.22) lists a number of exceptions to this rule, where elision does not occur.

²¹ In the examples /ō/ is shortened to /u/ and /ē/ to /i/. When speaking slowly, educated speakers tend to realize the shortened /ō/ and /ē/ as [o] and [e] respectively, but this does not apply when they speak at normal speed (Woidich, 2006, p. 7).

In addition to illustrating elision of short vowels, many of the examples above also include shortening of a long vowel. The syllable CṽC, as mentioned earlier, always occurs at the end of a word. If it ends up in the middle of a word through the addition of a suffix, or through the elision of a vowel, the long vowel becomes short (p. 31):

| | | | |
|-----------------------|-------------------|-----------------|-----------------|
| <i>bāb + kum</i> | > <i>babkum</i> | | “your door” |
| <i>fēn + ha</i> | > <i>finha</i> | | ”where is she?” |
| <i>yōm + ha</i> | > <i>yumha</i> | | “her day” |
| <i>ma + tišīl + š</i> | > <i>ma-tšīlš</i> | | “do not carry!” |
| <i>šāhib + i</i> | > <i>šāhibi</i> | > <i>šāhibi</i> | “my friend” |

The following examples show that the same type of shortening occurs when the accent is relocated as long vowels only can be in stressed syllables (p. 31):

| | | | |
|-------------------|------------------|-----------------|---------------------|
| <i>bēt + ēn</i> | > <i>bitēn</i> | | “two houses” |
| <i>yōm + ēn</i> | > <i>yumēn</i> | | “two days” |
| <i>šāfu + (h)</i> | > <i>šafū(h)</i> | | “they saw him” |
| <i>kātib + īn</i> | > <i>kātbīn</i> | > <i>katbīn</i> | “they have written” |

As in the case with elision, also shortening takes place over word boundaries, as already seen in one of the elision-examples: *šāhib ilbēt* (= *šāhib ilbēt*) “landlord”.

In addition to elision and shortening, prolonging of short final vowels occurs when the vowel is not the feminine marker and it is followed by a suffix (p. 34):

| | | | |
|----------------------|-------------------|--|---------------------|
| <i>abu + ya</i> | > <i>abūya</i> | | ”my father” |
| <i>šāfu + na</i> | > <i>šafūna</i> | | “they saw us” |
| <i>ma + gara + š</i> | > <i>ma-garāš</i> | | “it did not happen” |

The final vowel is also prolonged when followed by the demonstrative pronouns *da* and *di* and the prepositions *bi* and *li* followed by a suffix:

| | | | |
|-----------------------|----------------------|--|-----------------------------|
| <i>issana + di</i> | > <i>issanādi</i> | | “this year” |
| <i>bi nnisba + lu</i> | > <i>bi nnisbālu</i> | | “as far as he is concerned” |
| <i>’ālu + li</i> | > <i>’alūli</i> | | “they told me” |

4.2.1 Vowels in the data

Before analysing the vowels in the data, every long and short vowel had to be marked. The above mentioned rules from Woidich were used to determine whether a vowel is pronounced

as long or short. As mentioned in the chapter three, however, I opted to do my interpretation word-by-word. Then the vowels were counted and summed up in tables that show all representations present in the data for each of the vowel phonemes, and how many instances of each representation. I encountered some difficulties during the analysis though, and it became clear that I had to make some choices:

- Vowels that vary between short and long were overlooked. Examples of such is *kām/kam* (“how much, how many, some”) and *izzāy/izzayy* (“how”).
- Vowels of variable quality were not included. Such vowels are found in words as *dīnya/dunya* (“the world”), *miš/muš* (“not”), *šīwayya/šūwayya* (“some, a bit”) and *kuwayyis/kiwayyis* (“good, nice”). The same goes for imperfect verbs of the first stem whose mid-vowel is /u/. While nearly all other verbs have /i/ as the prefix-vowel, these verbs can have /i/ or /u/. Examples of such verbs are *yi’ud/yu’ud* (“he sits/stays”), *yirūh/yurūh* (“he goes”) and *yikūn/yukūn* (“he is”).

4.2.1.1 Long vowels

For table 4.7, I counted all instances of the five different long vowels. For each vowel there are three columns: representation, number of occurrences, and how many percent of all instances of that vowel are written using that particular representation.

Table 4.7 Representation of long vowels based on number of occurrences²²

| /ā/ | | | /ī/ | | | /ū/ | | | /ē/ | | | /ō/ | | |
|-------|-----|------|-------|-----|------|-------|----|------|-------|----|------|------|---|------|
| <a> | 752 | 97.9 | <ee> | 120 | 42.3 | <o> | 96 | 59.3 | <e> | 62 | 44.9 | <oo> | 5 | 41.7 |
| <aa> | 8 | 1.0 | <e> | 95 | 33.5 | <oo> | 41 | 25.3 | <ee> | 24 | 17.4 | <ou> | 4 | 33.3 |
| <aaa> | 5 | 0.7 | <i> | 48 | 16.9 | <ou> | 18 | 11.1 | <ei> | 19 | 13.8 | <o> | 3 | 25.0 |
| <-> | 3 | 0.4 | <y> | 8 | 2.8 | <u> | 4 | 2.5 | <i> | 12 | 8.7 | | | |
| | | | <eee> | 4 | 1.4 | <ooo> | 3 | 1.9 | <ai> | 6 | 4.3 | | | |
| | | | <iii> | 3 | 1.1 | | | | <ea> | 6 | 4.3 | | | |
| | | | <ei> | 2 | 0.7 | | | | <y> | 3 | 2.2 | | | |
| | | | <ii> | 1 | 0.4 | | | | <a> | 3 | 2.2 | | | |
| | | | <ea> | 1 | 0.4 | | | | <ae> | 2 | 1.4 | | | |
| | | | <ie> | 1 | 0.4 | | | | <eee> | 1 | 0.7 | | | |

²² The representations <aaa> for /ā/, <eee> for /ī/, <ooo> for /ū/ and <ee> for /ē/ include all representations that have three or more succeeding vowels. <-> means that the vowel is omitted.

| | | | | | | | | | | | | | | |
|--|-----|-------|-----|-----|-------|--|-----|-------|--|-----|------|--|----|-------|
| | | | <-> | 1 | 0.4 | | | | | | | | | |
| | 768 | 100.0 | | 284 | 100.3 | | 162 | 100.1 | | 138 | 99.9 | | 12 | 100.0 |

As mentioned above, every user seems to be consistent when it comes to consonants. Whenever there are several options available, he makes a choice, and then stays loyal to it. Although the picture is clearly different with the vowels, the use of long vowels also seems to have some consistency to it. In table 4.7, all representations are present as every single instance was counted and included. It does, however, not adjust for the size of the posts. Some are short and others are long, so inevitably a long post will contain more vowels than a short one. If it is so that each user shows some consistency in writing the long vowels, this will favour the representations used in the long posts.

In an attempt to compensate for this, table 4.8 shows what seems to be each user's preferred choice. This means that for each user, only the one representation for each vowel occurring the most times was counted. If a user has written a long vowel in two or more different ways, without using one of them more than the other, the user was not counted for this particular vowel. One instance is a user who has written two words that contain /ū/ twice in his post, in the words <mawdu3> (*mawdū*, "topic") and <ma2bola> (*ma'būla*, "accepted"). In the first word, <u> is used to represent /ū/, while he writes <o> in the second, thus making it impossible to determine what is this user's preferred choice.

Of course, especially for the vowels which occur infrequently, it is impossible to determine whether the user writes then consistently if they occur only once in a post. This one occurrence might be an exception if the user normally writes it differently. It is, however, believed that this error will be evened out by the fact that it does not favour any representation. Besides, only in a few instances was a long vowel used only once in a post.

Table 4.8 *Representation of long vowels based on the users' preference*

| /ā/ | | | /ī/ | | | /ū/ | | | /ē/ | | | /ō/ | | |
|-----|-----|-------|------|----|------|------|----|------|------|----|------|------|---|------|
| <a> | 110 | 100.0 | <ee> | 38 | 48.1 | <o> | 38 | 54.3 | <e> | 28 | 43.8 | <oo> | 4 | 36.4 |
| | | | <e> | 26 | 32.9 | <oo> | 21 | 30.0 | <ee> | 16 | 25.0 | <ou> | 4 | 36.4 |
| | | | <i> | 12 | 15.2 | <ou> | 10 | 14.3 | <ei> | 8 | 12.5 | <o> | 3 | 27.3 |
| | | | <y> | 3 | 3.8 | <u> | 1 | 1.4 | <i> | 4 | 6.3 | | | |

| | | | | | | | | | | | | | | |
|--|-----|-------|--|----|-------|--|----|-------|------|----|-------|--|----|-------|
| | | | | | | | | | <ai> | 3 | 4.7 | | | |
| | | | | | | | | | <ea> | 2 | 3.1 | | | |
| | | | | | | | | | <ai> | 2 | 3.1 | | | |
| | | | | | | | | | <y> | 1 | 1.6 | | | |
| | 110 | 100.0 | | 79 | 100.0 | | 70 | 100.0 | | 64 | 100.1 | | 11 | 100.1 |

The tables show that both methods yield roughly the same numbers, the main difference being that table 4.8 does not include the marginal representations as they are never the preferred choice.

/ā/ is hardly ever marked as a long vowel in the data. Every user prefers to write it as <a>. Most users also write the short /a/ in this manner, thus writing words like <banat> (*banāt*, “girls”), <kalam> (*kalām*, “speech”) and <7aga> (*hāga*, “thing”) where one cannot differentiate the long and the short vowel in writing. A very few examples of marking the /ā/ as a long vowel do exist. One can find in the data <taany> (*tāni*, “again”) and <3aalm> (*ālam*, “world”). Just as frequent, or rare, is using more than two <a>’s to exaggerate the long vowel, as in <raaaaagel> (*rāgil*, “a man”) and <sodaaaaaaa3> (*ṣudā*, “headache”), giving a more expressive look to it. The last minor exception is not writing the /ā/ at all, seen in posts written by three different users: <lkn> (*lākin*, “but”), <hnl2e> (*hanlā*’i, “we will find”) and <2l2nsn> (*il-insān*, “one”). Omitting the vowel is, as we shall see, more common with short vowels.

Where the other long vowels are concerned, they have three or four representations that are used in all or most instances. The exception is /ē/ which seems to be the trickiest one for the chatters to agree on. Although more representations are found for /ī/ than for /ē/, only four of them are among the users’ preferences while /ē/ has twice as many. The representations <ee> and <e> are, in that order, the most popular for /ī/. The two same representations, although in the opposite order, are also most used for /ē/. Nearly the same pattern emerges for /ū/ and /ō/, in that the three representations <o>, <oo> and <ou> are the most used for both. Nine posts in the data contain at least one /ū/ and one /ō/. In three of these, the same representation has been chosen for both. One user writes <yloM> (*yilūm/yulūm*, “he blames”) and <elmot> (*il mōt*,

“death”), thus using <o> for the two different phonemes. Another writes <mazbouta> (*mazbūta*, “in order”) and <oum>²³ (*yōm*, “day”), using the digraph <ou> to represent both.

Of the 56 posts containing /ī/ and /ē/, 27 have one representation that is used at least once to mean two different sounds. One user actually has two, writing <keber> (*kibīr*, “big”) and <teteer> (*tiṭīr*, “she flies”), as well as <lafet> (*laffēt*, “I turned”) and <la2eet> (*la’ēt*, “I found”), thus using both <e> and <ee> to represent /ī/ and /ē/. The same user also writes <eah> (*ē*, “what”), which means there are three representations for /ē/ in one single post. In the data there are even examples of users writing the same word in different ways. One writes <mawgoodeen> and <mawgooden> (*mawgudīn*, “present”) with only a few other words separating the two. Here it can also be noticed that the /u/ seems to be written as if it were long, even though it will be shortened in speech by the suffix *-īn*. Another user writes <salbeyat> and <salabiat> (*salbiyyāt*, “negatives”).

In general one can say that the vowels that are spelt the same way with Arabic letters are also done so with Roman letters. In Standard Arabic, /ū/ and /aw/, the diphthong from which /ō/ originates, are both written <و>, while /ī/ and /ay/, the origin of /ē/, are written <ي>.

4.2.1.2 Short vowels

<msr 6ol 3mrha gmela bs
w 7fdl 27bha 7ta elmot
w lw fe 7d mday2 mnha 2w 7ta 3ayz y5rg
w byshfm feha dh l2noh by7bha lkn
msh 3arf ylom 7d 3'erha
sd2ony bldna de 3'alya 3'lana 2we kolna
7ta wlw kabrna>

Egypt is pretty forever
I will continue to love her until death
If someone is tired of her or even wants to leave
and says bad things about her, it's because he loves her but
doesn't know who else to blame
believe me, our country is very dear to all of us
even when we grow up

When it comes to short vowels, the use is even more inconsistent. While long vowels are mostly represented with one or two Roman vowels, the short ones are written with one Roman vowel, or, as in the Arabic script, not written at all. The most extreme example of the

²³ From the context it is clear that <y> accidentally has been left out from the beginning of the word.

latter is quoted above. Nearly all vowels in the text represent either a long vowel, or a short end vowel that would have been written in the Arabic script, the only two exceptions being <l2noh> (*li innu*, “because he”) and <kabrna> (*kibirna*, “we grew up”).

Table 4.9 *Representation of short vowels*

| /a/ | | | /i/ | | | /u/ | | |
|-------|------|------|-------|------|-------|-------|-----|-------|
| <a> | 4088 | 86,4 | <e> | 2183 | 63,3 | <o> | 725 | 92,1 |
| <-> | 602 | 12,7 | <-> | 703 | 20,4 | <-> | 42 | 5,3 |
| <e> | 19 | 0,4 | <y> | 310 | 9,0 | <oo> | 10 | 1,3 |
| <o> | 11 | 0,2 | <i> | 183 | 5,3 | <ou> | 4 | 0,5 |
| <aa> | 5 | 0,1 | <a> | 49 | 1,4 | <u> | 2 | 0,3 |
| <aaa> | 4 | 0,1 | <yyy> | 10 | 0,3 | <e> | 2 | 0,3 |
| <i> | 2 | 0,0 | <ee> | 5 | 0,1 | <ooo> | 2 | 0,3 |
| | | | <ey> | 3 | 0,1 | | | |
| | | | <iii> | 2 | 0,1 | | | |
| | | | <eey> | 1 | 0,0 | | | |
| | | | <o> | 1 | 0,0 | | | |
| | 4731 | 99,9 | | 3450 | 100,0 | | 787 | 100,1 |

Table 4.9 shows that the short vowels are in fact written most of the time. In general one could say that <a> is used to represent /a/, <e> to represent <i>, and <o> for /u/. While /a/ and /u/ only have one widespread representation, <y> and <i> are used frequently enough for /i/ to be regarded as a common alternative. /i/ also stands out as the vowel being elided most often, but it is still written four out of five times. /u/, on the other side, is not written in only one out of 20 occurrences.

20.4% of all instances of /i/ are elided according to table 4.9. One should, however, take into consideration that some of these would not be pronounced anyway, but they are registered as elisions because of my word-by-word interpretation of the data. The number of actual /i/’s that are not represented is thus lower in reality.

4.2.1.2.1 <y> representing final /i/

The special case of representing /i/ with <y> needs to be mentioned. Although it makes up for less than a tenth of all /i/'s, there is a very clear pattern as to when it is used. The norm seems to be that when /i/ is in the final position, <y> is used to represent it.

Some words are exempted from this norm, though. *Wi* (“and”) is nearly always written <w> or <we>. A very few examples of <wi> also occur, but <wy> does not. <fy> is not an uncommon spelling of *fī* (“in”), but <fe> is used more frequently, and <fi> also occurs. The third and last exception is *di* (“this/that (f.)”) where also <dy> is found, but other variants such as <de>, <deh>, <di> and <dih> are more common.

In all other words <y> is the by far most widespread representation for a final /i/. *'Awī* (“very”) is already mentioned as a word frequently spelled without marking the initial *qāf*, and that <awy> is the most widespread spelling of the word in the data. The sentence <enty bet7eby hewayat eh we betmarsy eh> (*inti biḥibbi hiwayāt ē wi bitmarsī ē*, “what hobbies do you (f.s.) like and what do you (f.s.) engage in”) illustrates well how a user that writes <e> as a standard for /i/, still uses <y> at the end of words. Also the personal pronoun *-i* or *-ni* is typically written with <y>, as seen in <alby> (*'alby*, “my heart”) and <by7bny> (*biḥibbini*, “he loves me”).

In the data there is only one <y> that may represent a /i/ which is not in the final position, <mafyyhash> (*ma fīhāš*, “there is not in it”) Another user writes <mafyysh> (*ma fīš*, “there is not”) which is quite similar, except that the vowel <y> represents here is long instead of short. However, even the /i/ from the first example would often be written in the Arabic script as it originates from the word *fī* (“in”). The suffix *-š* prolongs the /i/ of *fī* in <mafyysh>, while the additional suffix *-ha* shortens it again in <mafyyhash>.

When trying to explain why a final /i/ is so consistently represented by <y>, two theories spring to mind. The first, and perhaps most likely, is that a final /i/, contrary to one placed elsewhere in the word, is always written when the Arabic script is used to write Egyptian Arabic. The Arabic letter used to represent /i/ is <ي>, which is the same one used to represent the consonant phoneme /y/. This seems to be an instance where the chatters are influenced by the Arabic script when writing with the Roman script.

The other theory is based on my assumption that English, probably the most common Western language among the chatters, is the basis for the conversion. I will elaborate on this later, but for now only note that in English the main rule is that one writes <y> in the final position of the word and before another vowel, and <i> elsewhere (Venezky, 1999, p. 88). Two succeeding vowels never occur in Egyptian Arabic, so the final position is the only place it can be used if English orthography forms the basis.

4.2.1.2.2 The minor representations

Table 4.9 lists 25 representations distributed among the short vowels. Most varieties are found only a few times in the data. This includes the use of three or more of the same vowel, mostly applied for emphasizing the word or its meaning. This, for example, gives the word *'awi* ("very") a meaning of something being "even more very", as in <fabgd fer7t awyyyyyyyyyyyy> (*fa bi gadd firiht 'awi*, "so really, I became very happy") and <ben7ebha awyyyy> (*binhibbaha 'awi*, "we really love her"). Other instances seem to be more coincidental, like <maaaaashiiii> (*māši*, "ok"), <yerooooo> (*yirudd/yurudd*, "he answers") and <bardoooo> (*bardu*, "also").

In a fair number of instances, /i/ is represented by <a>. Most of these are a victim of my attempt at being consistent. That is, as mentioned in chapter one, all words registered in an environment of Egyptian Arabic, will be registered as Egyptian Arabic words if it is not apparent that they are not. Others are more difficult to explain, such as <a7na> (*iḥna*, "we"), <a3lan> (*i lān*, "advertisement") and <yasa3edny> (*yisā'idni*, "he helps me"), words that do not fit into common Egyptian Arabic nor Standard Arabic patterns. The two first could, however, be that the quality of the first vowel is affected by the following pharyngeal consonant, which are here /ħ/ and /ʕ/.

There are relatively few short vowels in the data that are represented with two vowels, but there are some. What most of them have in common is that they would be written as long if the word was written using Arabic letters. There are two categories here: long vowels that have been shortened by a suffix, and vowels that are in the final position. Belonging to the first categories we have <youmen> (*yumēn*, "two days"), <nou3ha> (*nu'ha*, "her type"), <mawgoodeen> (*mawgudīn*, "those present") and <bashoufha> (*bašufha*, "I am seeing her"). In the second category there is <mabade2koo> (*mabādi'ku*, "your (pl.) principles"), <ya3nee> (*ya'ni*, "that means"), <hatshoufou> (*ḥatšūfu*, "you (pl.) will see") and.

These examples are an indication that shortened vowels are sometimes marked as long if they are marked as long in the Arabic script. However, as illustrated in table 3.7, the majority of the long vowels as well are in fact represented with only one Roman vowel. In other words, there could have been more examples of shortened vowels written as long if only more users would distinguish between short and long vowels. As an example, there is no way of telling whether the user who writes <nshofha> (*nišufha/nušufha*, “we see her”), and another who writes <yomeen> (*yumēn*, “two days”) actually intend the <o> as long or short, as they might use <o> to represent both.

There are only eleven instances of <o> representing /a/, but they are still interesting in that they come from several users and form a pattern. Nine instances are taken from five users who write <low> (*law*, “if”) instead of the more common <law>. The last two come from two different users who, independent from each other, write <mowdo3> (*mawḏūʿ*, “subject, topic”). That means that all occurrences of <o> representing /a/ are found when /a/ precedes /w/, thus forming the diphthong /aw/. As we have already seen under the consonants, some users omit the /w/ when preceded by /u/. These two features, writing <o> for /a/, and omitting /w/, are combined by the user who writes <lo> for *law*.

4.2.1.2.3 The epenthetic vowel

The epenthetic vowel, or helping vowel, is usually inserted in speech to prevent more than two successive consonants. This is done when a word ending with two consonants either gets a suffix starting with a consonant, or the following word starts with a consonant. This vowel is only present in speech, and never written when Egyptian Arabic is written in the Arabic script. It is also rare in my data, but I found nine instances from seven different users. All are caused by the negative suffix -š: <mashoftesh> (*ma šufiʃš*, “I did not see”), <makonteish> (*ma kuniʃš*, “I was not”), <ma7addish> (*ma ḥaddiʃš*, “no one”) and <makadabtish> (*ma kadabiʃš*, “I did not lie”). Many users do not have any examples of this kind of negation in their post, so writing the epenthetic vowel in this particular position might be more common than it looks here. A further indication of that is that only seven users had examples of the same construction without marking the vowel. Some are <ma7dsh> (*ma ḥaddiʃš*, “no one”), <m2oltsh> (*ma ʿultiʃš*, “I did not say”), <maro7tsh> (*ma ruḥtiʃš*, “I did not go”) and <mafkrtshtsh> (*ma fakkartiʃš*, “I did not think”).

The epenthetic vowel is never written when, in speech, it is inserted between words: <koll youm> (*kull yōm*, “every day”, possible pronunciation *kullⁱ yōm*), <ba3d kol marra> (*ba ‘d kull marra*, “after each time”, possible pronunciation *ba ‘dⁱ kullⁱ marra*) and <kont katbt 7aga> (*kunt katabt ḥāga*, “I had written something”, possible pronunciation *kuntⁱ katabtⁱ ḥāga*).

4.2.1.2.4 Short vowels and emphatic consonants

An important characteristic of the emphatic consonants is that they affect the quality of vowels next to them. Some may even say it is the most important characteristic: “There is an argument in Egyptian Arabic phonological circles about whether it is not the vowel that carries the pharyngealized features rather than the consonant” (Parkinson, 1985, p. 14). As mentioned above, the chatters do not distinguish emphatic consonants from their non-emphatic counterparts in the data, so I decided to see whether they somehow mark them via vowel quality. In table 4.10 are listed all markings of short vowels adjacent to an emphatic consonant. Only vowels found directly before or after an emphatic consonant, and belonging to the same word, have been included.

Table 4.10 *Representation of short vowels adjacent to primary emphatic consonants.*

| /a/ | | | /i/ | | | /u/ | | |
|-----|-----|-------|-----|----|-------|-------|----|------|
| <a> | 204 | 81.3 | <e> | 40 | 65.6 | <o> | 40 | 88.9 |
| <-> | 43 | 17.1 | <-> | 18 | 29.5 | <-> | 2 | 4.4 |
| <e> | 2 | 0.8 | <y> | 2 | 3.3 | <oo> | 2 | 4.4 |
| <i> | 1 | 0.4 | <i> | 1 | 1.6 | <ooo> | 1 | 2.2 |
| <o> | 1 | 0.4 | | | | | | |
| | 251 | 100.0 | | 61 | 100.0 | | 45 | 99.9 |

The pattern is very similar to that of the short vowels in general, and nothing indicates that the chatters represent them differently even if their quality is affected by the emphatic consonant. One could, of course, argue that even if they want to distinguish between different qualities, the Roman alphabet does not offer the characters to do so. I did not do a similar counting for long vowels as the number of occurrences was very low.

4.2.3 A summary: long and short vowels

The analysis of the vowels shows that there is great inconsistency in how they are represented. A pattern seems to be that whenever one encounters two successive Roman vowels in ERA, they represent a long vowel, while no representation at all means a short vowel. One Roman vowel, however, is a common representation for both long and short vowels. In table 4.11, I have used the numbers from tables 4.7 and 4.9 to calculate how many Roman letters are used on average to represent each vowel. All representations consisting of three or more vowels were counted as consisting of three.

Table 4.11 Average number of Roman letters used to represent long and short vowels.

| Long vowels | | | Short vowels | | |
|-------------|-------------|----------------|--------------|-------------|----------------|
| Vowel | Occurrences | No. of letters | Vowel | Occurrences | No. of letters |
| /ā/ | 768 | 1,02 | /a/ | 4731 | 0,88 |
| /ī/ | 284 | 1,49 | /i/ | 3450 | 0,80 |
| /ū/ | 162 | 1,40 | /u/ | 787 | 0,97 |
| /ē/ | 138 | 1,43 | | | |
| /ō/ | 12 | 1,75 | | | |

With long and short vowels together, it becomes very clear that /a/ and /ā/ are nearly treated as equals in ERA. The average number of letters for /ī/, on the other hand, is nearly twice that of /i/, but inconsistent use among the chatters calls for context in order to determine which of the two is meant. While /ō/ is being represented by two letters considerably more often than the other long vowels, the number of occurrences is too low to decide whether this is a stable pattern.

4.2.4 The inconsistent use of vowels

In this chapter I have shown that while the representation of consonants is more or less consistent, the representation of vowels is a lot more variable. Most of the vowel phonemes undeniably have one representation which is used more than the others, but there is still a lot of variation both between the users and often also within the post of a single user. The variation is to be expected in a written language form that officially does not exist and does not have any normative rules. The chatters are left to themselves, and to copying others, when trying to convey their speech in writing.

One obvious reason why vowel phonemes are represented less consistently than the consonant phonemes, is to be found in the nature of vowels. Consonants are formed in a distinct way, while vowels have fewer features by which they can be distinguished (Brinton & Brinton, 2010, p. 36):

Vowels are articulated not by putting the articulators into discrete configurations, but by shaping the tongue in the mouth. Hence, there are theoretically infinite different vowel sounds, forming a continuum with no distinct boundaries.

Even though Egyptian Arabic has few vowel phonemes, the number of vowel sounds is greater. Speakers of a language tend not to do a lot of linguistic analysis while using the language in everyday life, so they are likely to consider the sounds instead of the phonemes when converting to a different alphabet. Woidich (2006, pp. 7-8) identifies five different realizations of /a/, and three of each of /u/ and /i/. This means that these three vowel phonemes are realized with at least 11 different sounds. After the chatter has identified the sound, he or she needs to find an appropriate Roman letter to represent it. In doing so, other problems reveal themselves.

The English language includes a great number of vowel sounds, a lot more than the six vowels in the alphabet would indicate. “The letter <o> corresponds to at least 17 different sounds, <a> to 10, <e> to nine, and the combined group to 48. When the morphemic structure and consonant environment of the words in which these units appear are considered, however, a single major pattern emerges, with a bevy of subpatterns” (Venezky, 1999, p. 173). According to Venezky, each of the five primary vowel units (the single-letter spellings <a>, <e>, <i/y>, <o> and <u>), basically has two different pronunciations, one checked and one free. Which pronunciation a vowel gets in each instance is determined by “the morphemic structure of the word of the word in which it occurs and the consonant and vowel units that follow it” (Venezky, 1999, p. 173).

So in order to know how an English vowel is pronounced, one must either simply know how the word it occurs in is pronounced, or one has to know the patterns and subpatterns that Venezky describes. The free alternate of <e> is /i/, as in “athlete”, while the checked alternate is /ɛ/, as in “athletic” (Venezky, 1999, p. 174). The first corresponds to the Arabic /i/, while the second is close to the Arabic /a/, hence hinting at a reason sometimes <e> is used for /a/ as

well. The inconsistent use of vowels might, among other things, occur because the morphemic patterns in Arabic words do not necessarily exist in English, so even if one knew all the patterns, the pronunciation of a vowel could still be unclear.

Coulmas (2003, p. 98) says about the letters of the Latin alphabet:

Their usage is determined not just by the phonetic interpretations of individual letters but by higher-level units, morphemes and words. In spite of the persistent notion that letters are associated with sounds or sets of sounds, it is impossible to construct an algorithm for the spelling of the words of a language like English on the basis of a list of all, or even the most commonly used, graphemic representations of the phonemes of English.

Examples of the polyvalence are numerous. The English schwa can be represented by all the five vowels of the English alphabet: <a> in “about”, the second <e> in “rebel”, <i> in “compatible”, <o> in “oblige” and <u> in “circus” (Coulmas 2003, p. 98). Coulmas (2003, p. 99) also lists up 14 ways of spelling the English phoneme /u:/, some of them are <u> in “truly”, <o> in “do”, <oe> in “shoe”, <ue> in “true” and <ui> in “lawsuit”. Of the more obscure, but still perfectly valid, examples, are <ewe> in “jewel”, <oeu> in “manoeuvre”, <ough> in “throughout” and <oups> in “coups”. The /u:/ of English does not correspond to /ū/ in Arabic, but is still an example of the polyvalence in English writing which might make it hard for Egyptian chatters to agree on representations for vowel phonemes.

Chapter five: Some selected features

In the last chapter I analyzed the data on the lowest level. That is, I tried to find out how every phoneme in Egyptian Arabic is represented in ERA, and to what extent the users are consistent in their choices. In this chapter I will examine constructions that are often different in speech and writing, or constructions that in Arabic are written according to certain rules that would not have to be followed with a different alphabet. In this way I hope to find out whether the chatters are influenced mainly by writing or speech, or whether it is an indefinable mix of the two.

First I will consider the definite article. Even though there are many ways to pronounce it, decided by the following phoneme, there is only one way to spell it using the Arabic script. The feminine marker can also be pronounced in different ways although this does not always show in the Arabic script. Special attention is also given to the prepositions *li* and *bi*. Although considered to be words on their own, they are always connected to the following word when written in the Arabic script.

5.1 The definite article

The definite article in Standard Arabic is *al*, and in the Arabic script it is always written using the two letters *alif* and *lām* together, <ل>, regardless of how it is pronounced. The /a/ is omitted in speech if preceded by a vowel, and the /l/ is assimilated if followed by a "sun letter", which constitutes roughly half of the consonants in Standard Arabic. In Egyptian Arabic the definite article is *il*, and it undergoes the same changes in pronunciation as in Standard Arabic. Even though the writing of Egyptian Arabic with Arabic letters is not restricted by official rules like Standard Arabic, and numerous dialect words can be seen written in several different ways, the norm is that the definite article is written <ل>.

Of the total 507 occurrences of the definite article in the corpus, as many as 469 (92.5%) are written <el>. Then there are 13 (2.6%) instances of <al>, 11 (2.2%) of <2l> and 4 (0.8%) of <2el>. In the rest the vowel is missing due to the article being squeezed between words, as in <ll2asaf> (*li il-asaf*, "unfortunately") and <fl a7'r> (*fi il-āxir*, "in the end"). The data does not include a single example of the /l/ being assimilated, the chatters write instead <el sob7> (*iṣ-ṣubḥ*, "the morning") and <eldonya> (*id-dunya/id-dinya*, "the world"). This seems to show

that, in the case of the definite article, the users write it using the orthography of written Arabic instead of writing what they actually say. Another possible explanation, however, is written below.

Contrary to Arabic orthography, the definite article in ERA is mostly written separated from the following word. In 377 (74.4%) of the occurrences, a space separates the definite article from the following word. In the remaining 25.6% it is written as a part of the word itself, as in <2lterf 2ltany> (*iṭ-ṭarf it-tāni*, “the other part”) and <elmoshkila> (*il-muškila*, “the problem”). With a very few exceptions, like <fein gam3eit el dowal el 3arabeya we el etefa2 al gama3y wa tarabot al 3araby> (*fēn gam ‘it id-duwal il- ‘arabiyya wi il-ittifa’ il-/ig-gamā ‘i wi tarabbuṭ il- ‘arabi*, “where is the League of Arab Nations and the collective agreement and the Arabic unity”), the users are consistent in how they write the definite article. It seems, however, that they are more likely to fuse it with the following word if it is preceded by the prepositions *li* and *bi*. Several users who otherwise write the article separated from the following word, provide examples like <bellel> (*bi il lēl*, “at night”), <l2asaf> (*li il-asaf*, “unfortunately”), <beltari2a deh> (*bi iṭ-ṭarī’ a di*, “in that manner”) and <l27sn> (*li il-aḥsan*, “for the better”).

As the numbers show, the most common way to represent the definite article in ERA is <el> separated from the following word. The fact that the /l/ is never assimilated may prove that some conventions from the Arabic script are incorporated to such an extent that they are followed even when the Roman script, and an otherwise quite oral and informal language, is used.

The space often seen between <el> and the following word is more difficult to explain as the definite article is always integrated in the word in the Arabic script. It may, however, point to an influence by ad hoc transcription. Although I have no empirical data to refer to, it is my impression that the definite article typically is written independently in ad hoc transcription. In addition, the /l/ is hardly ever assimilated. Thus, it cannot be excluded that ad hoc transcription actually is what influences the chatters rather than Arabic orthography. They are influenced by what they are used to see in their daily life. The decisive question, then, might be to what extent ad hoc transcription is influenced by Arabic orthography.

5.2 The feminine ending

The feminine ending can, in Egyptian Arabic, be pronounced in three different ways. So even though it has a letter of its own in Arabic, <ة>, there is not one grapheme used specifically to represent it in ERA, hence its inclusion in this chapter.

The feminine ending is only found in the final position of a word, and will as a general rule be pronounced /a/. It can, however, be followed by a suffixed pronoun which in Egyptian Arabic alters the pronunciation to /it/. The same happens when the word containing the feminine ending forms a genitive construction with the following word. In the former situation the <ة> will be replaced by a <ة> in Arabic writing. In the latter situation, even though the pronunciation is altered, it is still written the same. The third possible pronunciation is /ā/. This occurs most frequently when the word is followed by the demonstrative pronouns *da* or *di*, or the prepositions *li* and *bi* with a suffix, or the word containing the feminine ending itself is a suffixed participle (Woidich, 2006, pp. 33-35). As this pronunciation does not exist in Standard Arabic, it is never written in formal Arabic.

There are 479 occurrences of the feminine ending in the corpus. They will be dealt with below according to their realization.

5.2.1 Realized as /a/

The feminine ending, when the oral realization /a/ is expected, occurs 424 times in my data. 416 (98.1%) of these are represented with <a> as in <fekra> (*fikra*, "idea"), <egaba> (*igāba*, "answer") and <7aga tanya> (*hāga tanya*, "something else"). The last eight represent it with <ah>, as in <sanah> (*sana*, "year") and <mokhtalifah> (*muxtalifa*, "different (f.)"). One user consistently uses <ah>, while three users have written it only once among their preferred choice <a>. One of the three has written <3ayzah> (*'ayza*, "she wants"), while the two others have written <7ayah> (*haya*, "life").

5.2.2 Realized as /it/

There are 48 occurrences of the feminine ending where it would be pronounced /it/. As mentioned above, in Arabic orthography the feminine marker <ة> itself is only altered when a pronoun suffix is added. It stays the same when its pronunciation is affected by the following noun. The chatters, however, always involve a <t> in the representation of both. Examples

from the data involving a suffixed pronoun are <7ayate> (*ḥayāti*, “my life”), <mo3aksetha> (*mu’aksitha*, “her harassment”) and <zemlty> (*zimilti/zamilti*, “my colleague (f.)”). Examples with a genitive construction are <bsoret el fard> (*bi šūrit il-fard*, “on an individual basis”), <mo3ammilt el 7ywanat> (*mu’amlit il-ḥayawanāt*, “the treating of animals”) and <re7et el samak> (*rīhit is-samak*, “the smell of fish”).

The last two examples show that when the pronunciation of the feminine ending is altered, but its writing in Arabic orthography is not, the chatters choose to represent it as they realize it orally.

5.2.3 Realized as /ā/

Only seven instances when the feminine ending is realized as /ā/ occur in the data. As elsewhere, the distinction between /a/ and /ā/ is not marked: <3arfaha> (*’arfāha*, “I (f.) know her”), <fahmahom> (*fahmāhum*, “I (f.) understand them”) and <belnesba leh> (*bi in-nisbā li*, “for my sake”).

5.3 Gemination

The doubling of a consonant is normally not marked with the Arabic script, except for in fully vocalized texts where the *shadda* symbol <◌◌◌> is put above the doubled letter. The few instances where both a double and single consonant is possible, like /y/ in *izzayy/izzāy* (“how”) were disregarded. In total there are 1044 instances in the data as a whole. The great majority are written as if no doubling was present, like <atmana> (*atmanna*, “I wish”), <7ora> (*ḥurra* “freedom”), <7ata> (*ḥatta*, “even”), <talawos> (*talawwus*, “pollution”) and <7ad> (*ḥadd*, “someone”). A marked gemination was found in 132, or 12.6%, of the instances. Some examples are <2otta> (*’uṭṭa*, “cat”), <ba3addy> (*ba’addi*, “I pass”) and <ezzay> (*izzayy/izzāy*, “how”). Three users excel in that they more or less always mark the doubling. In fact, these three alone account for more than a third, 45, of the marked geminations in the corpus.

One word also stands out, seemingly getting its doubled consonant marked more than other words. This word is the relative pronoun *illi* (“that, who, which”), not surprisingly, as it is written with two <◌◌>’s also with the Arabic script. Table 4.1 shows how *illi* is written in the data. To compare, three other words occurring frequently and containing a doubled consonant

are also included: *lamma* (“when”), *kull* (“all, every”) and *bass* (“only, but”). Whenever *kull* appeared with a suffix, like <kolna> (*kullina*, “all of us”), I shortened it down to the basic form for the sake of simplicity (that is, <kolna> was registered as <kol>). Every user writing one or some of these words more than once consistently wrote it in the same manner, thus the number of users is given.

Table 5.1 Representations of *illi*, *lamma*, *kull* and *bass*.

| <i>illi</i> | Users | <i>lamma</i> | Users | <i>kull</i> | Users | <i>bass</i> | Users |
|--------------------|-------|--------------|-------|-------------|--------------------|-------------|-------|
| <elly> | 16 | <lama> | 12 | <kol> | 49,5 ²⁴ | <bas> | 27 |
| <ely> | 9 | <lma> | 7 | <koll> | 2,5 | <bs> | 21 |
| <eli> | 8 | <lamma> | 2 | <kool> | 1 | <bass> | 2 |
| <ele> | 3 | | | | | <bss> | 1 |
| <2ely> | 2 | | | | | | |
| <ell> | 1 | | | | | | |
| <aly> | 1 | | | | | | |
| <elle> | 1 | | | | | | |
| <elli> | 1 | | | | | | |
| <el> ²⁵ | 1 | | | | | | |

The following sums up how many users mark the gemination in each word:

- *illi* – 19 of 43 users – 44.2%
- *lamma* – 2 of 21 users – 9.5%
- *kull* – 2,5 of 53 users – 4.7%
- *bass* – 3 of 51 users – 5.9%

Due to the spelling of these words in the Arabic script, it is probably no coincidence that the gemination in *illi* is marked five to ten times as often as in the other words. Contrary to the case of the definite article, *illi* is a word that seldom, if ever, appears in ad hoc transcription. So for this particular word it seems that Arabic orthography definitely influences some of the users.

²⁴ One user writes <kol> as well as <kollena>, thus counting as half a user on both <kol> and <koll>.

²⁵ Presumed to represent *illi* in <ana 3aref en el mashakel el fe el balad deeh> “I know that the problems that are in this country”.

Another word that has a marked gemination in Arabic script is *allāh*. It occurs in 17 posts, alone or as part of an expression, and the double /l/ is marked in ten (58.8%) of them. Standing alone it is written <alah> and <allah>, as well as <Allah>. Most frequent is the expression *w-allāhi* (“by God”), usually written <walahy>, but also <wllhy>, <wallahi> and <wlahi>. Three users write the expression *il-ḥamdu li-llāh* (“praise be to God!”) once, with some variation: <el 7amd le allah>, <el hamd lellah> and <el7amdolelallah>. Unlike *illi*, *allāh*, being an internationally known word, is frequently written using ad hoc transcription. That might explain, at least partially, the high occurrence of double /l/ when the users write *allāh*.

Even though the marking of double consonants is scarce, there exist a few examples of writing the same consonant twice when only one is expected. This is displayed in three posts. One user writes <bettgama3> (*biyitgama* ‘ or *biyitgamma* ‘, “to be gathered”). Another writes <mo3ammla> (*mu* ‘*amla*, “treatment”) as well as <7aykkbar> (*ḥayikbar*, “it will grow”). The third writes <malall> (*malal*, “boredom”) and <mallal>, seemingly convinced that a gemination is in there somewhere.

5.4 The future prefix ḥa-/ha-

The future prefix is according to Mitchell (1956, p. 36) *ḥa-*, while both Abdel-Massih (1975, p. 95) and Woidich (2006, p. 278) place *ḥa-* and *ha-* on equal terms. A friend from Cairo once told me that she uses both without giving much thought to which one she uses. An Egyptian colloquial teacher advised me not to use *ḥa-* because it would make me sound as if I were “from the countryside”. Another said that *ḥa-* by far is the most common among Cairenes.

The users are also divided on the issue. The marginally more popular is *ha-*, which is always represented with <h>. 14 users resort to this solution, as seen in <hatkallim> (*hatkallim*, “I will speak”) and <hy2ol> (*hay’ūl*, “I will say”). One less, namely 13 users, write *ḥa-*, represented with <7>. Examples are <7tb2a> (*ḥatib’a*, “it will become”) and <7atefdaly> (*ḥatifḍali*, “you (f. s.) will continue”). Two users have one example of each in their posts. This number might have been higher had more users had more than one example of the future prefix in their post. It appears that the thoughts of my Cairene friend can be applied in general.

5.5 The prepositions *li* and *bi*

I group these two prepositions together as they are both written integrated with the following word in the Arabic script. There are a total of 115 occurrences of *li* in the data, and 92 of *bi*.

Li is in general connected to the preceding word, as in <lél bent> (*li il-bint*, “for the girl”) and <lenafsy> (*li nafsi*, “to myself”). In only 17 (14.8%) of the 115 instances is *li* separated from the following word, as in <le sbab> (*li sabab*, “because of”), <le nas> (*li nās*, “for people”) and <l ahmed> (*li aḥmad*, “for Ahmed”). The other example of *li* followed by a name is written in the same manner. In none of these 17 instances, where *li* is standing alone, is the preposition followed by the definite article. In other words, *li* is, in the data, always connected to a following definite article.

Contrary to Arabic orthography, in ERA it is quite common to attach *li* also to the preceding word. This seems to happen most of all when the preceding word is a verb: <yegeblaha> (*yigib laha*, “he brings to her”), <7ayekteblek> (*ḥayiktib lik*, “he will write to you (f.s.)”) and <hygelha> (*ḥayīgi lha*, “he will come to her”). This construction is particularly common with the verb *’āl* (“to say, tell”): <hay2olaha> (*ḥay’ul laha*, “he will tell her”), <2ltly> (*’ālit li*, “she told me”), <2ololy> (*’ulū li*, “tell (pl.) me!”), and <ye2olena> (*yi’ul/yu’ul lina*, “he tells us”). As with gemination in general in ERA, the two /l/’s that often succeed each other in this construction, are only written as one. In all the examples above *li* is followed by a suffix. The data contains 14 instances of a verb followed by *li* which again is followed by a noun or another word, and in all of them there is a space separating *li* from the preceding verb: <kont ba2ol le s7aby> (*kunt ba’ul li soḥābi*, “I used to tell my friends”), <a2oul lenafsy> (*a’ul li nafsi*, “I say to myself”), <barga3 le rabbena> (*barga’ li rabbina*, “I return to our Lord”) and <yerga3 le7’atbto> (*yirga’ li xaṭibtu*, “he returns to his fiancée”). *Li* is also found connected to the preceding word when it is a participle, as in <methya2ly> (*mithayya’ li*, “it appears to me”), or the negation *ma*, as seen in <malhash> (*ma lhāš*, “she does not have”) and <malhomsh> (*ma lhumš*, “they do not have”).

As with *li*, also *bi* is as a rule connected to the word or suffix that follows it. This is found in 78 (84.8%) of the 92 occurrences of *bi*. Sometimes the /i/ is elided, as in <bsoret el fard> (*bi šūrit il-fard*, “on an individual basis”) and <bsara7a> (*bi šarāḥa*, “frankly”). Other times it is marked, as in <betary2a nedifa> (*bi ṭarī’a niḍīfa*, “in a decent manner”) and <beshola> (*bi suhūla*, “with ease”). If the word following *bi* starts with an /i/, the two /i/’s are merged into

one, as in <be7teram> (*bi ihtirām*, “with respect”) and <bellel> (*bi il-lēl*, “at night”). *Bi gadd* (“seriously”) is written in numerous ways, but always as one word: <begad>, <bgd> and <bgad>. There are 17 instances of *bi* following a verb²⁶, and they are always written separately. That is also true for the few instances where *bi* is followed by a suffix: <ht7se behom> (*ḥathissi bīhum*, “you (f.s.) will feel them”) and <yt3aml beha> (*yit ‘amal bīha*, “he is dealing with her”).

5.6 Some other prepositions

As a general rule, *fi* is written separately. It is, however, always connected to its suffixed pronoun, as in <yfakar feha> (*yifakkar fīha*, “he thinks about her”) and <el balad feeha neẓam> (*il-balad fīha niẓām*, “there is proper order in the country”). It also occurs connected to a following definite article in 11 (30.6%) of 36 instances, as seen in <fel share3> (*fī iṣ-šāri ‘*, “in the street”) and <fl a7’er> (*fī il-āxir*, “in the end”). Otherwise *fi* and *il* are written separately, as demonstrated in <fe el gm3a> (*fī il-/ig-gam ‘a*, “at the university”). *Fi* is, in Egyptian Arabic, not only used as a preposition, but also as a particle in the meaning “there is, there exists”. Negated, this kind of *fi* is in ERA connected both to the preceding *ma* and the following suffix *-š*, as seen in <mafeesh 7aga> (*ma fīš ḥāga*, “there is nothing”) and with a suffixed pronoun, as in <mafehash> (*ma fīhāš*, “it (f.) does not have”).

Although there are numerous exceptions, the norm in ERA seems to be that the prepositions *li* and *bi* are written integrated with the following word, while *fī* is not. This correlates with Arabic orthography, and hints to the chatters being influenced by the underlying form.

The prepositions *‘ala* (“upon, above, at”) and *min* (“from”) can, in Egyptian Arabic, form a contracted unit with the definite article in rapid speech. Egyptians may say *‘al* instead of *‘ala il* and *mil* instead of *min il*. It is not common to write this contraction in the data, but it does occur. One of the users write <3al akal> (*‘al-aqall*, “at the least”) as well as <a3ayat mel nas> (*a ‘ayyaṭ min-nās*, “people make me cry”). Another writes <3alcomputer> (*‘al-computer*, “on the computer”) while a third writes <3al 7’areeta> (*‘al-xarīṭa*, “on the map”). These three users are the only ones to write the contracted forms. These examples are also the only instances in their posts of the two prepositions followed by the definite article, so there is no way to establish whether they write it this way consequently or not.

²⁶ Excluding *bi gadd* as it is an expression that does not in any way belong to the verb.

In Egyptian Arabic, a function word has been created by the merging of *'ala* and *ša'n* (“affair, matter”). The word can be pronounced *'alašān* or simply *'ašān* (“in order to, because (of)”). This word is spelled in many different ways in the data, with the main difference being whether an <l> is present or not. It is written without the <l> 42 times out of 59 occurrences in total, like <3shan>, <3ahsn> and <3ashan>, and with <l> the remaining 17, as in <3alshan>, <3lshan> and <3alashan>. It is always written as one word.

Chapter six: conclusion

When I started to write this thesis, there were above all two matters I wanted to explore. The first was to investigate the extent of the influence of Arabic orthography on the users of ERA. The other was to find out whether some norms are starting to emerge in this kind of writing that has evolved without the guidance of any language authority. In order to answer these questions, I analyzed samples of writing with regard to how the consonant and vowel phonemes and some selected grammatical features of Egyptian Arabic were represented.

In general, the users do not seem to perceive ERA as a system. While the consonant phonemes are relatively consistently represented, the users struggle to find common ground when it comes to the vowel phonemes. Although the long vowels in average are represented with a higher frequency of graphemes than the short vowels, using a single Roman vowel grapheme is the most popular option for both. In addition, most of the vowels have more than one common representation. This inconsistency can, in the most extreme cases, lead to the same user writing one word in different ways. The lack of a fixed spelling forces the user to make choices for every word he or she writes.

In spite of the instability, all vowels have a few representations that are used in the great majority of instances. These are the most common representations for the long vowels, written according to their popularity: /ā/ is written <a>, /ī/ is written <ee>, <e> and <i>, /ū/ is written <o>, <oo> and <ou>, /ē/ is written <e>, <ee> and <ei>, while /ō/ is written <oo>, <ou> and <o>. The most common representations for the short vowels are: /a/ is written /a/, /i/ is written /e/, and /u/ is written /o/. All three short vowels have omission as their second most common representation. At least 76% of all instances of each vowel phoneme are covered above. The instability is clear when one considers that the most used representation for the different short vowels are the same that are most used for the corresponding long vowels. It is also noteworthy that of the six listed representations for /ē/ and /ō/, five are listed for /ī/ and /ū/ as well.

As the application of the Roman alphabet varies from language to language, I have assumed that the users of ERA have English in mind when they try to find a suiting representation for an Egyptian Arabic phoneme. The assumption is based on English being the most popular foreign language in Egypt nowadays, but also the fact that the users write the typical English

digraph <sh> for /š/. English is a language with a low phoneme-grapheme correspondence, and most vowel phonemes have numerous spellings in different words. It is not unlikely that this polyvalence, at least partially, is to blame for the lack of stability in representing the vowels in ERA. When a user of ERA wants to write /ī/, he might think of e.g. <ee> as in “keen” or <e> as in “athlete”. This might not explain all representations of vowel phonemes in ERA, but it seems likely to be at least part of the explanation.

There are some features that may, but do not necessarily, demonstrate influence by Arabic orthography. One of them is the definite article, which in ERA is always written with <l> even when the /l/ is assimilated in speech. This, however, may just as well be influenced by ad hoc transcription, which the users probably come across every day. The fact that the definite article is always written separately supports the ad hoc influence, as in Arabic orthography it appears as eclitic to the following. The lack of marking gemination in ERA can also be explained by Arabic orthography as well as ad hoc transcription.

The relative pronoun *illi* stands out in being written with marked gemination in ERA five to ten times more often than other and similar words. This can hardly be traced back to anything else than writing in the Arabic script. As a word not commonly being part of names, it is rarely, if ever, written in ad hoc transcription. The Arabic script also seems to be influential in how the users write the three prepositions I have included in the analysis. While the majority writes *li* and *bi* attached to the following word, *fi* is mostly written separately, as in Arabic orthography.

English orthography, Arabic orthography, and ad hoc transcription all seem to have some influence on ERA. There should, however, be no doubt that the users assert a lot of influence on each other as well. As soon as a newcomer starts typing with others for the first time, he or she will undoubtedly quickly adopt to how the others are writing. Even though the vowel phonemes far from have standardized representations, there are only a few representations for each that can be regarded as common.

The future development of ERA looks uncertain for two reasons. Firstly, the users are already a minority among Egyptian Internet users. In order to write ERA one needs not only to be a speaker of Egyptian Arabic, but also to know the Roman alphabet, which again is reserved for those who know a language which is written in the Roman script. The other reason is that

technical solutions for using other scripts on computers and on the Internet have abolished the need to write Arabic in the Roman script. It is probable that the younger generation, even those who know other languages, are accustomed to Arabic keyboards and will be able to type Arabic letters just as quickly as Roman ones. ERA has, however, survived thus far, and it may continue to live its own life in certain circles for a long time still.

Literature

- Abdulla, R.A. 2007, *The Internet in The Arab World: Egypt and Beyond*, Peter Lang Publishing, New York.
- Abu-Absi, S. 1986, "The Modernization of Arabic: Problems and Prospects", *Anthropological Linguistics*, 28, pp. 337-348.
- Al-Toma, S.J. 1961, "The Arabic Writing System and Proposals for Its Reform", *Middle East Journal*, 15, pp. 403-415.
- Aquilina, J. 1961, *Papers in Maltese Linguistics*, The Royal University of Malta, Valetta.
- , 1965, *Teach Yourself Maltese*, The English Universities Press Ltd, London.
- ʿAsīlī, A. 2009, *Kitāb Malūš Ism*, Dār Al-šurūq, Cairo.
- Badawī, E.M. 1973, *Mustawayāt al-luġa al-ʿarabiyya al-muʿāšira fī mišr*, Cairo, Dār al-maʿārif.
- Badawi, E. and M. Hinds 1986, *A Dictionary of Egyptian Arabic*, Librairie du Liban, Beirut.
- Baron, N.S. 2008, *Always On: Language in an Online and Mobile World*, Oxford University Press, Oxford.
- Beesley, K.R. 1998, *Romanization, transcription and transliteration*. [Online] <http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/romanization.html> [Accessed 23 March 2009].
- Beißwenger, M. 2002, "Introduction", in Beißwenger, M. (ed.), *Chat-Kommunikation: Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation*, ibidem-Verlag, Stuttgart, pp. XVII-XXIII
- Berjaoui, N. 2002, "Aspects of the Moroccan Arabic Orthography with Preliminary Insights from the Moroccan Computer-Mediated Communication", in Beißwenger, M. (ed.), *Chat-Kommunikation: Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation*, ibidem-Verlag, Stuttgart, pp. 431-468.
- Blau, J. 1988, *Studies in Middle Arabic and its Judaeo-Arabic Variety*, The Magnes Press, The Hebrew University, Jerusalem.
- Borg, A. and M.A. Azzopardi-Alexander 1997, *Maltese*, Routledge, London.
- Brinton, L.J. and D.M. Brinton 2010, *The Linguistic Structure of Modern English*, John Benjamins Publishing Company, Amsterdam.
- Coulmas, F. 2003, *Writing Systems: An Introduction to Their Linguistic Analysis*, Cambridge University Press.
- Cross, F.M. 1991, "The Invention and Development of the Alphabet", in Senner, W.M. (ed.) 1991, *The Origins of Writing*, University of Nebraska Press, Lincoln.
- Crystal, D. 2006, *Language and the Internet*, 2nd edition, Cambridge University Press, Cambridge.
- DeFrancis, J. 1989, *Visible Speech: The Diverse Oneness of Writing Systems*, University of Hawaii Press, Honolulu.
- Dobrovolsky, M. and W. O'Grady 1997, "Writing and Language", in O'Grady, W., M. Dobrovolsky and F. Katamba, *Contemporary Linguistics: An Introduction*, 3rd edition, Longman, London.
- Ferguson, C.A. 1959, "Diglossia", *Word*, 15, pp. 325-340.
- Frehner, C. 2008, *Email – SMS – MMS: The Linguistic Creativity of Asynchronous Discourse in the New Media Age*, Peter Lang, Bern.
- Gelb, I.J. 1963, *A Study of Writing*, 2nd edition, The University of Chicago Press, Chicago.
- Haeri, N. 1997, *The Sociolinguistic Market of Cairo: Gender, Class, and Education*, Kegan Paul International, London.

- Hofheinz, A. 2007, "Arab Internet Use: Popular Trends and Public Impact", in Sakr, N. (ed.), *Arab Media and Political Renewal: Community, Legitimacy and Public Life*. I. B. Tauris, London.
- Holes, C. 1995, *Modern Arabic: Structures, Functions, and Varieties*, Longman Group Limited, Essex.
- Khan, G. 2007, "Judaean-Arabic", in Versteegh, K. et al. (eds.), *Encyclopedia of Arabic Language and Linguistics*, Volume 2, Brill, Leiden, pp. 526-536.
- Labov, W. 1972, *Sociolinguistic Patterns*, University of Pennsylvania Press, Philadelphia.
- Lewis, G. 1999, *The Turkish Language Reform: A Catastrophic success*, Oxford University Press, Oxford.
- Mahmoud, Y. 1980, "On the reform of the Arabic Writing System", *Journal of Reading*, Vol. 23, No. 8 (May, 1980), pp. 727-729.
- Mahrouse, M. E. 1994, "Egypt: System of Education" in Husén, T. and T. N. Postlethwaite (eds.) 1994, *The International Encyclopedia of Education*, 2nd edition, Volume 4, Pergamon Press, Oxford, pp. 1939-1947.
- Mifsud, M. 2008, "Maltese", in Versteegh, K. et al. (eds.) 2008, *Encyclopedia of Arabic Language and Linguistics*, Volume 4, Brill, Leiden, pp. 146-159.
- Mitchell, T.F. 1956, *An Introduction to Egyptian Colloquial Arabic*, Oxford University Press, London.
- Norlin, K. 1987, *A Phonetic Study of Emphasis and Vowels in Egyptian Arabic*, PhD-thesis, Lund.
- Palfreyman, D. and M. Al Khalil 2007, "A Funky Language for Teenzz to Use", in Danet, B. and S.C. Herring (eds.), *The Multilingual Internet: Language, Internet, and communication online*, Oxford University Press, New York, pp. 43-63.
- Parkinson, D. 1985, *Constructing the Social Context of Communication: Terms of Address in Egyptian Arabic*, Walter de Gruyter, Berlin.
- 1991, "Searching for Modern Fuṣḥa: Real-life Formal Arabic", *Al-‘Arabiyya*, 24, pp. 31-64.
- Reichmuth, P. 2009, "Transcription", in Versteegh, K. et al. (eds.), *Encyclopedia of Arabic Language and Linguistics*, Volume 4, Brill, Leiden, pp. 515-520.
- Rosenbaum, G.M. 2004, "Egyptian Arabic as a Written Language", *Jerusalem Studies in Arabic and Islam*, 29, pp. 281-340.
- Sampson, G. 1985, *Writing Systems: A Linguistic Introduction*, Hutchinson & Co, London.
- Schaub, M. 2000, "English in the Arab Republic of Egypt", *World Englishes*, 19, pp. 225-238.
- Somekh, S. 1991, *Genre and Language in Modern Arabic Literature*, Otto Harrassowitz, Wiesbaden.
- Storrer, A. 2002, "Sprachliche Besonderheiten getippter Gespräche: Sprecherwechsel und sprachliches Zeigen in der Chat-Kommunikation", in Beißwenger, M. (ed.), *Chat-Kommunikation: Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation*, ibidem-Verlag, Stuttgart, pp. 3-24.
- Tseliga, T. 2007, "It's All Greeklish to Me!", in Danet, B. and S.C. Herring (eds.), *The Multilingual Internet: Language, Internet, and communication online*, Oxford University Press, New York, pp. 116-141.
- Venezky, R.L. 1999, *The American Way of Spelling: The Structure and Origins of American English Orthography*, The Guildford Press, New York.
- Versteegh, K. 2001, *The Arabic Language*, Edinburgh University Press, Edinburgh.
- Warschauer, M., G.R. El Said and A. Zohry 2007, "Language Choice Online: Globalization and Identity Online", in Danet, B. and S.C. Herring (eds.), *The Multilingual Internet:*

- Language, Internet, and communication online*, Oxford University Press, New York, pp. 303-318.
- Watson, J.C.E. 2002, *The Phonology and Morphology of Arabic*, Oxford University Press, Oxford.
- Wellisch, H.H. 1978, *The Conversion of Scripts: Its Nature, History, and Utilization*, John Wiley & Sons, New York.
- , 1975, *Transcription and Transliteration: An annotated bibliography on Conversion of Scripts*, Silver Spring, Maryland.
- Wells, J.C. 2000, "Orthographic Diacritics and Multilingual Computing", *Language Problems and Language Planning*, 24, pp. 249-272.
- Wheeler, D.L. 2004, *The Internet in the Arab World: Digital Divides and Cultural Connections*. [Online] Lecture presented at Royal Institute for Inter-Faith Studies on 16 June 2004.
http://www.riifs.org/guest/lecture_text/Internet_n_arabworld_all_txt.htm [Accessed 9 February 2009].
- , 2006, *The Internet in the Middle East: Global Expectations and Local Imaginations in Kuwait*, State University of New York Press, Albany.
- Woidich, M. 2006, *Das Kairenisch-Arabische: Eine Grammatik*, Harrassowitz Verlag, Wiesbaden.
- Woidich, M. and R. Heinen-Nasr 2004, *kullu tamām!: An Introduction to Colloquial Arabic*, The American University in Cairo Press, Cairo.

Abstract

Egyptian Romanized Arabic (ERA) is the Egyptian variety of the writing that arose when Arabic speakers joined the domain of computer-mediated communication. At a time when almost all text on the Web was in English, and only basic Roman letters were supported, they had to use the Roman script to communicate with each other in Arabic. Based on data from more than one hundred users, I hoped to draw a picture of how ERA is written in general. As there is no official orthography or spelling rules in ERA, the basis for my analysis is that they would attempt to write as they speak, from phoneme to grapheme, but perhaps with interference from Arabic orthography. In addition to finding out to what degree Arabic orthography influences on the writing, I wanted to examine whether ERA is a stable writing system with emerging norms.

I claim that ERA primarily is a transcription from the users' speech to writing, but it certainly seems to be influenced by Arabic orthography as well, although the degree of influence varies between different features. Additionally, it seems to be influenced by ad hoc transcription, the non-standardized transcription that is common on road signs and in general when Arabic names are written in the Roman script, and by English orthography. The writing systems is quite stable when it comes to representation of the consonant phonemes, but less so where the vowel phonemes are concerned.