

Why the Abjad Suits the Arabs

Guy Boursier

Abstract. The initiators of the Wikipedia Egyptian Arabic edition strove to differentiate it from the Standard Arabic edition, notably via the script chosen. Some proposed a script adapted from the Latin alphabet, thus repeating previous attempts to systematically transcribe vowels, which the traditional Arabic abjad does not. Oddly enough, most Egyptian Arabic Wikipedians, despite their desire to cultivate their difference from Standard Arabic, chose the Arabic script used as an abjad for writing their encyclopedia's articles. Besides the weight of habit, the diglossic status of Arabic may explain this choice. Yet, they have managed to adapt the script to give it a specific "Egyptian" flavor. This article is inspired from some of the results of the research presented in the author's master's thesis, entitled *Écrire l'arabe égyptien, le cas d'un Wikipédia 'dialectal'*, defended at UFR ALL-Nancy, Université de Lorraine, France, in June 2025.

1. Standard Arabic and Egyptian Arabic

In his seminal 1959 article entitled "Diglossia," Charles A. Ferguson takes Egyptian Arabic as an example of a "low" variety in this type of situation where "two varieties of a language exist side by side through the community, with each having a definite role to play" (Ferguson, 1959). Egyptians, therefore, speak what is often described as a dialect of Arabic in their daily lives, while writing is primarily in the "high" variety, in this case "Classical Arabic," or its modern version, thence referred to as "Standard Arabic".

It is this variety that children learn to write in school, in which masterpieces are composed, such as the works of Naguib Mahfouz, laureate of the Nobel Prize for Literature in 1988. Most instances of written Egyptian Arabic have traditionally been confined to marginal items such as intimate correspondence, captions on cartoons and comic literature¹. However, recently there have been increasing attempts to cross

Guy Boursier  0009-0007-0445-376X
Aix Marseille Univ, CNRS, IREMAM, Aix-en-Provence, France
E-mail: guy.boursier@etu.univ-amu.fr

1. For an exhaustive study of diglossia in Arabic, see Diem (1974).

the border, with novels written entirely in Egyptian Arabic² or even a work translated into this variety from Classical Arabic³.

In 2008, an edition of Wikipedia in Egyptian Arabic was created (thence shortened to Wikipedia maṣri, from an adjective meaning “Egyptian,” used in the website’s language menu), parallel to the 2004 Arabic edition (Wikipedia al-‘arabiyya). This constituted a further step towards the dialect’s legitimation as a language in its own right, capable of dealing with all the fields of human knowledge to be found in an encyclopedia. However, as with any emergence of a written corpus, many questions needed to be resolved.

2. How to Write Egyptian Arabic?

A major question is: how to write? Of course, the traditional way of writing Arabic is the Arabic script used mainly as an abjad, a system in which mostly consonants are represented, leaving vowels to be deduced by the reader, as seen in Table 1. For sure, vowels can be indicated with diacritic marks, but these are rarely used. The Arabic script can also be used so as to write consonants and vowels phonographically, essentially in loan words such as *dimuqrāṭiyya*—ديمقراطية (democracy), but this is marginal in Standard and Egyptian Arabic. One of the *raisons d’être* of abjad is to make the writer gain in writing speed, at the cost of placing the burden of comprehensibility of the written text on the shoulders of the reader. In any case, the abjad has regularly been judged deficient in rendering the phonology of Arabic.

2.1. Resorting to the Latin Script

A drastic solution is the recourse to the Latin script used phonographically for consonants and vowels.

2. A list of novels written in Egyptian Arabic is to be found in Davies (2005).

3. In 2016, the reference work of Arabic literature *Risālat al-ḡufrān* ‘The Epistle of Forgiveness’ by Abū al-‘Alā’ al-Ma‘arrī (973-1057) was translated from Classical Arabic into Egyptian Arabic by Nārīmān al-Šāmīlī (al-Kutub Ḥān lil-naṣr wa al-tawzī‘, Cairo). In her introduction, she explains that as a young reader, she was unable to understand this classic text, which then motivated her to translate it into Egyptian Arabic, in order to make it accessible to Egyptian readers (Rosenbaum, 2017, pp. 59–61).

TABLE 1. Abjad writing

Non-vocalized abjad writing	#chars
بدون الديمقراطية لا حرية ولا حقوق ولا مساواة. ⁴	45
bdūn -ldīmqrāṭya lā ḥrya wlā ḥqūq wlā msāwā-. ⁵	45
Vocalized abjad writing	#chars
بدون الديمقراطية لا حرية ولا حقوق ولا مساواة.	74
bidūni l-dīmuqrāṭiyyati lā ḥurriyyata walā ḥuqūqa walā musāwā.	62
Without democracy, there is no freedom, no rights, no equality.	

Some contributors to Wikipedia maşri suggested using what they call “al-frānkō ‘arab,” similar to what is known as “arabizi,” with the use of Latin letters for the corresponding sounds in Arabic, and numbers, or even digraphs, for the others. This is illustrated in Table 2. They highlight the ease of writing on a computer or mobile phone keyboard that does not have an Arabic alternative.

TABLE 2. Wikipedia maşri examples of experimental articles⁶

“frānkō ‘arab” / “arabizi”
Botros Botros 8aly howa diplomasy Ebti Masry [...]. Kan sekerter 3aam el Omam el Motta7eda el sades [...], we howa 7afed Botros Nayruz 8aly, ra2ees wozara2 Masr fe awa2el el karn el 3eshreen. [...] 5ad el doktorah fel Kanoun el Dawly men Gamey3et Paris [...].
Boutros Boutros-Ghali is an Egyptian Coptic diplomat [...]. He was the sixth Secretary-General of the United Nations [...], and is the grandson of Boutros Nayrouz Ghali, Prime Minister of Egypt in the early twentieth century. He got a doctorate in international law from the University of Paris [...].
“phonetic writing”
El Malti heyya-l loġa-l wataneyya betāeet Malta, we heyya kamān wahda m’n el loġāt er rasmeyya fe-ll Ettehād el Oropi. El loġa-l Malti mettaxda m’n el Ǝarabi w’ be-z zāt el lahgāt betāeet Camāl Afrikya [...].
Maltese is the national language of Malta and one of the official languages of the European Union. The Maltese language is derived from Arabic, particularly North African dialects [...].

4. This is a quote from Syrian poet and essayist Adonis, pronounced at a symposium entitled “Towards a New Religious Discourse” held on February 4, 2015, at the Cairo International Book Fair.

5. The Arabic transliteration system used here is based on the one recommended by the journal *Arabica*.

Others tested in experimental articles a “phonetic” writing with Latin letters, some of them enriched with diacritical marks. The goal is to write Egyptian Arabic by representing the pronunciation more precisely than with the abjad mode of Arabic.

This method is even presented as a “liberation of the Egyptian language from Arabic letters incapable of representing Egyptian speech”⁷.

2.2. Transcribing the Pronunciation of Egyptian Arabic

This attempt to transcribe Egyptian Arabic using graphs from the Latin alphabet is nothing new. In 1897, Daniel Willard Fiske (1831-1904), an American academic, published *An Egyptian alphabet for the Egyptian People*. In this booklet, he presented a transcription model for writing the Egyptian dialect in the Latin alphabet (see Table 3).

TABLE 3. One of Fiske’s example sentences for his Egyptian alphabet

Fiske	wâgib ‘ala kull wâhid jeħibb maşr, in jisâ‘id ‘ala migy el jôm elly fyh kull en nâs es sâkinyn ‘alal nyl jimkinhum jiqru we jiktibu el lisân elly jtkallimuh we jifhamuh. (Fiske, 1897, p. 4)
IPA	wɑːgib ʔala kullə wɑːhid jɪħibbə mɑsˤrə ʔin jisɑːʔid ʔala mɪgi l joːm illi fiː kullə nnaːs issakniːn ʔala nniːl jimkinhum jiʔru wə jiktibu lliːsɑːn illi jtkallimuː wə jifhamuː
English	It is the duty of everyone who loves Egypt to help bring about the day when all the people living on the Nile can read and write the language they speak and understand.

The goal was nothing less than to have this alphabet adopted in Egypt for the teaching of written Egyptian Arabic, which would replace Classical Arabic in everyday use (Zack, 2014, p. 4). Despite the distribution of a whole range of promotional material to Egyptian teachers, this alphabet was never adopted.

Many among such Western scholars linked the high prevalence of illiteracy in Egypt at the time to the diglossic situation of the Arabic language. Since the Arabic abjad was intrinsically associated with Classical Arabic, which was considered complex to master, abandoning it in favor

6. Wikipedia, ويكيبيديا:طريقة الكتابة/ارشيف 1 – [Wikipedia talk: Writing method/Archive 1 – arz.wikipedia.org].

7. Wikipedia, ويكيبيديا:طريقة الكتابة/ارشيف 1 – [Wikipedia talk: Writing method/Archive 1 – arz.wikipedia.org]

of a phonographic system, where vowels would be systematically written, seemed an effective way to allow a greater number of ordinary people to access reading and writing. The association of the Latin alphabet with modernity, and of the Arabic abjad with the past, was a common view at the time (Al-Toma, 1961).

Many other transcription systems have been created since the 19th century to give the most accurate phonographic representation of the pronunciation of Egyptian Arabic, mostly for the needs of foreign learners of the language.

For some of the phonemes, a one-to-one correspondence between Egyptian Arabic phonemes and Latin letters is straightforward and indeed used by all the transcription systems consulted (see Table 4).

TABLE 4. A selection of transcription systems with single graphs

Arabic letters	IPA	Spitta 1880	Gairdner 1916	Mitchell 1956	Louis 2008	Abdel-Massih 2009	Al-Tonsi 2010	Boutros 1993
ب					b			
ت					t			
ج					g			
د					d			
ر					r			
ز					z			
س					s			
ف					f			
ك					k			
ل					l			
م					m			
ن					n			
ه					h			
و					w			

But for other phonemes, a basic Latin graph may be used with a diacritic. This is particularly the case for the emphatic consonants that have their own graph in Arabic and a few other phonemes with no equivalent in the target language of the transcribers. We even find instances of the use of the Arabic graph ‘ayn (see Table 5).

Finally, for a certain number of phonemes, transcriptions vary between basic Latin graphs, sometimes with a diacritic on the one hand or digraphs (and even a trigraph) on the other hand. The first solution allows for a nice isomorphic correspondence to be maintained, but, in addition to the lack of consensus on the choice of diacritics, it forces modern scribes to resort to rare graphs that are beyond the range of the common keyboard. The second solution is the digraph. It is easy to im-

TABLE 5. A selection of transcription systems with symbols and diacritics

Arabic letters	IPA	Spitta 1880	Gairdner 1916	Mitchell 1956	Louis 2008	Abdel-Massih 2009	Al-Tonsi 2010	Boutros 1993
ص	s ^ʕ	ʂ	s	ʂ	ʂ	ʂ	ʂ	ʂ
ض	d ^ʕ	ḍ	d	ḍ	ḍ	ḍ	ḍ	ḍ
ط	t ^ʕ	ṭ	t	ṭ	ṭ	ṭ	ṭ	ṭ
ظ	z ^ʕ	ẓ	z	ẓ	ẓ	ẓ	ẓ	ẓ
ح	h	ḥ	h	ḥ	ḥ	ḥ	ḥ	ḥ
ع	ʕ	ʕ	ʕ	ʕ	ʕ	ʕ	ʕ	ʕ
ء	ʔ	ʔ	ʔ	ʔ	ā ē ō ū	ʔ	ʔ	ʔ

plement and already common in European languages (ch, th, gh, ...) (see Table 6).

TABLE 6. A selection of transcription systems with symbols and digraphs

Arabic letters	IPA	Spitta 1880	Haggenger 1892	Willmore 1901	Gairdner 1916	Mitchell 1956	Louis 2008	Girod 2016
خ	x	ḫ	ch	kh	x	x	kh	ḫ
ش	ʃ	ś	sch	sh	ʃ	ʃ	sh	ś
غ	ɣ	ġ	ġ	gh	ɣ	ɣ	gh	ġ
ا	aː	â	a	â	aː	âa	a	â
و	uː	û	u	û	uː	ûu	uu	û
و	oː	ô	o	ô	oː	ôo	oo	ô
ي	iː	y	i	î	iː	îi	ii	î
ي	eː	ê	e	ê	eː	ée	ee	ê

But the digraph is ambiguous, given that a sequence of two letters can correspond to one phoneme or two phonemes.

In the latter case, the solution adopted by Willmore (1901), is to place a hyphen between the two letters, which looks like an unwieldy quick fix, establishing a separation right in the middle of a word (see Table 7).

TABLE 7. disambiguizing digraphs (Willmore, 1901, p. 1)

Digraph	One phoneme	Two phonemes
<gh>	/ɣ/ ṣughaiyar ‘small’	/gh/ ag-har ‘albino’
<kh>	/x/ âkhir ‘last’	/kh/ dik-ha ‘this’
<sh>	/ʃ/ mishi ‘he left’	/sh/ yis-ha ‘he is distracted’

No consensual solution has therefore really emerged to handily transcribe Arabic into the Latin alphabet. Still, even for some Arabs, the recourse to the Latin script has been seen as a desirable solution, notably to write the vowels of Arabic on the same level as the consonants.

3. Writing the Vowels of Arabic

One of the most famous attempts at using Latin graphs for writing Arabic was the project of ‘Abd al-‘Azīz Fahmī (1870-1951). A member of the Academy of the Arabic Language in Cairo, he was no less than the chairman of the committee on writing and grammar when he presented his project in 1944. One originality was to mix Latin graphs, some with a stroke, and Arabic graphs, a few of which were reversed and linked in a single form to the Latin graphs. But his main concern was writing the vowels, whose absence he saw as the main drawback in the Arabic writing system.

Fahmī	<i>al sayfu aṣḍuq inbāsan min l kutubi</i> <i>fay raddi hi l raddu bayna l riddi wa l laḥibi</i>
Arabic	السيف أصدق إنباء من الكتب في حده الحد بين الجد واللعب
English	The sword is more truthful in informing than books, In its edge is the limit between seriousness and play.

FIG. 1. A sentence in Fahmī's proposed alphabet (Fahmī, 1944, p. 156)⁸

Fahmī admitted that the same verse written in Arabic script was much shorter than when written in the one he suggested, but he found it preferable to transfer the burden of explicitation from the reader to the writer. As expected, Fahmī's proposal created a scandal, and it was never implemented, but it is still remembered as a symbol of blatant hostility toward Islam and Pan-Arabism⁹.

8. This is the first verse of an ode by Abū Tammām (803-845) to Abbasid caliph al-Mu‘taṣim for his victory in the 838 battle of Amorium over the Byzantines.

9. For a detailed account of Fahmī's campaign for the use of Latin graphs, see Shraybom-Shivti (1998, pp. 189–193). For a recent example of a hostile reaction to Fahmī's and others' proposals to write Arabic with Latin letters, see Al-Ṣāwī (2017)

3.1. Chasing the Perfect Phonemic Script?

Yearning for a perfect phonographic script is an understandable undertaking, but probably a vain one. As Florian Coulmas cautions, even a system of phonetic transcription like the International Phonetic Alphabet should not be taken to mean that it incorporates “the ultimate solution to the problem of rendering a continuous stream of sound into a discrete array of graphic signs” (Coulmas, 2003, p. 32) .

This is apparent in the differences that exist between Egyptian Arabic and Standard Arabic. One of the three vowels of the latter is the front unrounded open vowel [a] to which some descriptions of Egyptian Arabic add its back twin [ɑ]. They provide the minimal pair *bāba* ‘the second month of the Coptic calendar’ and *bāba* ‘pope, pontiff, patriarch’ (Badawi and Hinds, 1986, pp. 48–49). Other descriptions prefer in this case to speak of a single vowel and of a series of “marginal” pharyngealized consonants noted by a dot under the Latin graph (Watson, 2002, p. 21).

TABLE 8. back [ɑ] or pharyngealized consonants?

	Badawi, Hinds 1986	Watson 2002
2nd month of Coptic calendar	bāba	bāba
pope, pontiff, patriarch	bāba	ḥāḥa
take care of	barr	barr
be faithful (to a promise)	barr	baṛṛ
my dead	mayyiti	mayyiti
my water	mayyiti	mayyiti

The hesitation in choosing which element of the context, consonant or vowel, should be elevated to a phoneme testifies to the difficulty of finding the holy grail of the “perfect” phonographic script. This should remind us that there is more to speech than a mere sequence of consonants and vowels. An abjad, with its focus on the consonants in the syllable, may offer a better ambit for suprasegmental elements, closer to that of a syllabary.

3.2. Why Bother With Vowels?

The obsession with vowels is also strange for a language like Standard Arabic that basically has just three. Quite regularly, the language hesitates between one or the other, while retaining the same meaning. In any case, the writing does not differ.

TABLE 9. Vowel fluctuation in Standard Arabic

Arabic script	Pronunciation		Translation
	Standard Arabic		
يُحَسِّبُ	/jahsabu/	/jahsibu/	he considers
جَنَازَةٌ	/ʒanaːza/	/ʒinaːza/	funeral
جِصٌّ	/ʒasˤsˤ/	/ʒisˤsˤ/	gypsum
بَزٌّ	/buzz/	/bizz/	teat
يُحْرِقُ	/jahruqu/	/jahriqu/	it burns

Similarly, between Standard Arabic and Egyptian Arabic, the differences are often limited to the pronunciation of vowels.

TABLE 10. Vowel fluctuation between Standard Arabic and Egyptian Arabic

Arabic script	Pronunciation		Translation
	Standard Arabic	Egyptian Arabic	
كَبِيرٌ	/kabi:r/	/kibi:r/	big
صَغِيرٌ	/sˤayɪ:r/	/sˤuɣajjar/	small
تَعِبَ	/taʕiba/	/tiʕib/	he was tired
مَشَى	/maʕa:/	/miʕi/	he walked
يَكْتُبُ	/jaktubu/	/jiktib/	he writes

Writing as an abjad thus allows some room to be left for different interpretations of the written word. This makes it possible to accommodate the various pronunciations of vowels in the different regiolects of Arabic and in the standard variety of the language. The principle of the abjad appears indeed to be particularly compatible with the situation of diglossia that exists in Arabic. By highlighting the consonants, it leaves some considerable slack for the reader to interpret the written word. So what could be felt as a deficiency by a few may have turned out to be a useful compromise for many.

4. The Traditional Abjad Chosen for Wikipedia Maṣri

In any case, for Wikipedia maṣri, the suggestions to write in a phonographic alphabet were not retained, and the abjad mode was preferred. The main arguments were that Egyptians are much more accustomed to

writing and reading it, and the suggested change was perceived as being too bold a step¹⁰.

For the contributors to Wikipedia maṣri, rejecting the Latin alphabet and adopting the Arabic graphs implies the recognition that the language of their project is strongly related to Arabic, despite their desire to prove the contrary. Pragmatism is here the main reason for this decision, because most Egyptians have a better command of the Arabic abjad than of the Latin alphabet. Although this choice may seem logical, it is in contradiction with the stated goal of many early contributors to Wikipedia Maṣri, which was to “liberate” their language from its submission to Arabic, felt as an outside influence from the Arabian peninsula.

This stance is particularly illustrated by the evolution of the terms chosen early on to describe the language of Wikipedia maṣri. First, it is al-luḡa al-‘āmmiyya al-maṣriyya—اللغة العامية المصرية ‘the Egyptian colloquial language’, then it becomes kalām maṣri—كلام مصري ‘Egyptian talk’ and the pejorative term “colloquial” disappears. Finally, it becomes al-luḡa al-maṣriyya al-ḥadīsa—اللغة المصرية الحديثة, that is to say ‘the modern Egyptian language’¹¹, thus placing it in an idealized continuum with the ancient (hieroglyphic) Egyptian language. The goal is to minimize its proximity to Arabic while laying claim to more ancient and prestigious (Pharaonic) roots. The preference for the same abjad mode as Arabic is thus a major concession to the recognition of a certain kinship between this “modern Egyptian language” and Arabic.

5. Writing Arabic Differently

However, the early contributors to Wikipedia maṣri have devised ways to differentiate the script they use from the traditional Arabic script.

5.1. Yāʾ and ʾalif Maqṣūra

In Arabic, the graph yāʾ <ي> most often corresponds to the semi-consonant /j/ or the long vowel /iː/. A similar graph but without the two dots <ى>, used exclusively in the final position, corresponds to the long vowel /aː/. Arabic grammarians call it ʾalif maqṣūra. Egyptians are known for merging the two graphs in one, in the final position, most often the undotted ʾalif maqṣūra.

10. Wikipedia, ويكيبيديا:طريقة الكتابة/ارشيف 1 - ويكيبيديا [Wikipedia talk: Writing method/Archive 1—arz.wikipedia.org]

11. اللغة المصرية الحديثة - ويكيبيديا - Modern Egyptian Language—arz.wikipedia.org

TABLE 11. Fusion of yā' and 'alif maqṣūra in Wikipedia maṣri

	يَاء – yā'	ألف مقصورة – 'alif maqṣūra
Standard Arabic	ي علي /ʕali:/ Ali (proper noun)	ي علي /ʕala:/ on top of
Wikipedia maṣri		ي علي

This national specificity was therefore seized upon by the initiators of Wikipedia maṣri to write any final yā' as a 'alif maqṣūra,¹² thus exploiting the treatment of these two graphs in Egypt¹³.

5.2. 'alif Al-waṣl and 'alif Al-qaṭ'

Another subtlety of Arabic is the difference between 'alif al-waṣl <أ> and 'alif al-qaṭ' <أ>, which could be translated respectively as “fusion 'alif” and “separation 'alif”. These are the two main statuses of the 'alif at the beginning of a word. The 'alif al-qaṭ' is a 'alif that indicates a glottal stop /ʔ/ (the separation) generally marked with a hamza <ء>, as a diacritic above or beneath. The 'alif al-waṣl is a token graph that indicates a liaison (the fusion) with the last vowel of the preceding word. It is not pronounced except at the beginning of a sentence, where it is pronounced as a 'alif al-qaṭ' with a glottal stop. In most texts, the 'alif al-waṣl has no diacritic mark <|>, which differentiates it from the 'alif al-qaṭ' with its hamza <أ>.

Getting one's 'alifs wrong is one of the most common errors in contemporary Arabic (Attia et al., 2015, p. 13). This may be one reason why the general rule set by the initiators of Wikipedia maṣri is to almost systematically write the initial 'alif without a hamza¹⁴.

12. ويكيبيديا: طريقة الكتابة – ويكيبيديا: How to write—arz.wikipedia.org

13. This Egyptian tendency not to dots the final yā' is illustrated in an article by Egyptian author 'Ādil 'Aṭṭia: “The big problem—for me—started when I began writing. Whenever I sent an article or a story that had a final “yā” without its two dots, the non-Egyptian publisher would put them! And if I put the dots, the Egyptian publisher would delete them!” (Aṭṭia, 2010).

14. ويكيبيديا: طريقة الكتابة – ويكيبيديا: How to write—arz.wikipedia.org

TABLE 12. Fusion of ʾalif al-waṣl and ʾalif al-qatʿ in Widiopedia maṣri

	ʾalif al-waṣl	ʾalif al-qatʿ
	fusion ʾalif	separation ʾalif
Standard Arabic	آ ابن ibn son	أم أم ʾumm mother
Wikipedia maṣri	ابن ا	أم ا

5.3. Tāʾ Marbūṭa

A third contextual variant of an Arabic letter is the tāʾ marbūṭa (linked tāʾ) <ة>. In Standard Arabic, it is a graph, always placed at the end of the word, which takes the form of the graph hāʾ <ه>, topped by the two dots of the tāʾ <ت>. It is necessarily preceded by the vowel /a/ (short or long) and is pronounced /t/ followed by the vowel of declension, except at the pause (mostly at the end of a sentence), where it is only pronounced /a/.

TABLE 13. Treatment of tāʾ marbūṭa in Widiopedia maṣri

	general rule	ʾiḏāfa (genitive construction)
Egyptian Arabic	/il-luɣa l-ʔiŋgilizijja/ the English language	/luɣit isʕsʕiʔi:d/ the language of Upper Egypt
Wikipedia maṣri:	اللغة الإنجليزية	لغة الصعيد

In Egyptian Arabic, tāʾ marbūṭa is generally pronounced /a/, but it becomes /(i)t/ at the end of the first element of an ʾiḏāfa (genitive construction). To reflect this difference, the initiators of Wikipedia maṣri have established the following rule: “The tāʾ marbūṭa is written according to its pronunciation: with two dots when it is pronounced /t/, and without dots in the other cases”¹⁴. The omission of the two dots on the tāʾ marbūṭa is a phenomenon that has been observed by Refaat El-Farnawany in his thesis (in German) on Egyptian Arabic as a written language, in which he notes the following: “As far as I know, there is no rule for these dots of—ه. When and where they should be written depends on the writer’s writing habits” (Farnawany, 1981, p. 127). It is equally instructive to note that at no point does he mention the two dots under the final yāʾ, which he systematically writes without dots in the examples of his thesis, as if this rule went without saying.

Thus, the contributors to Wikipedia maṣri have managed to give the script used on their project a specific appearance that helps to differentiate it from Wikipedia al-ʿarabiyya.

6. Graph Frequency Comparison Between Wikipedia Maṣri and Wikipedia Al-ʿarabiyya

In the light of these recommendations, it is interesting to observe the frequency of use of graphs on Wikipedia maṣri (also referred to as “arz”) and compare it with that of Wikipedia al-ʿarabiyya (referred to as “ar”)¹⁵. The files containing the latest version of each article were downloaded for each edition from <https://dumps.wikimedia.org> to obtain `arwiki-20240201-pages-articles.xml` (3.5 GB) and `arwiki-20240720-pages-articles.xml` (11 GB) respectively. The 010 Editor software was used to search for each of the graphs. The graphs are ranked from least represented to most represented on arz in relation to ar in Table 14 that lists occurrences that are under-represented in arz relative to ar, and in Table 15 that lists occurrences that are over-represented in arz relative to ar.

Unsurprisingly, among the least frequent graphs on Wikipedia maṣri are the sukūn, which is used to mark the absence of a vowel between two consonants, and tanwīn diacritics, which are used to mark the flexional ending of indefinite nouns and adjectives. The latter system is inoperative in Egyptian Arabic, except for the /an/ <◌ْ◌> flexion of common adverbs such as /ʔabadan/ ‘never’ or /taʔriban/ ‘approximately’. The other flexional endings appear in arz mainly in quotations from the Qurʾān. This /an/ <◌ْ◌> inflection is the most frequent in absolute numbers on Wikipedia maṣri, but proportionally the least used in comparison with Wikipedia al-ʿarabiyya. This relative rarity probably results from a bot’s intervention to make it disappear¹⁶, which is not without problems of comprehensibility. The three short vowel diacritics <◌َ◌>, <◌ِ◌>, and <◌ُ◌> remain relatively less rare in arz, in particular to disambiguate the pronunciation of a word, as is also customary in fuṣṣḥā.

The graphs <ث> and especially <ذ> are among those less frequently used on arz compared to ar. This is because they transcribe phonemes not used in Egyptian Arabic. However, the relative rarity of <ذ> compared to <ث> is surprising since the recommendations are to keep it if it is pronounced /ذ/. In addition, <ث> and <ذ> can be used to transcribe

15. Wikipedia differentiates its language editions by the codes assigned to them by the ISO 639-3 standard, i.e., “ar” for Arabic and “arz” for Egyptian Arabic.

16. Bots (short for “robots”) generally make automated changes or actions to maintain Wikipedia’s articles and other pages on Wikipedia.

TABLE 14. Graph frequency comparison between arz and ar—Under-representation in arz

graphs	arz			ar			proportion
	Number	Proportion	Rank	Number	Proportion	Rank	ar/arz
ا	17,263	0.0022%	43	4,027,633	0.1433%	35	233.31
آ	611	0.0001%	47	129,168	0.0046%	44	211.40
أ	1,575	0.0002%	46	235,193	0.0084%	43	149.33
إ	15,490	0.0020%	44	2,063,197	0.0734%	40	133.20
أ	4,596	0.0006%	45	451,382	0.0161%	42	98.21
أ	348,229	0.0444%	36	19,730,202	0.7018%	25	56.66
أ	169,397	0.0216%	38	9,507,807	0.3382%	31	56.13
أ	53,702	0.0069%	40	2,616,609	0.0931%	37	48.72
أ	51,656	0.0066%	41	2,382,296	0.0847%	39	46.12
أ	43,806	0.0056%	42	1,159,890	0.0413%	41	26.48
أ	2,275,450	0.2904%	28	54,292,275	1.9312%	16	23.86
أ	152,032	0.0194%	39	2,868,430	0.1020%	36	18.87
أ	5,646,607	0.7206%	23	100,960,306	3.5911%	10	17.88
أ	1,008,905	0.1288%	31	12,334,500	0.4387%	30	12.23
أ	221,304	0.0282%	37	2,488,371	0.0885%	38	11.24
أ	1,573,673	0.2008%	29	94,74,864	0.3370%	32	6.02
أ	5,017,158	0.6403%	25	28,551,540	1.0156%	23	5.69
أ	1,268,363	0.1619%	30	6,591,327	0.2345%	33	5.20
أ	7,213,905	0.9206%	22	34,717,311	1.2349%	22	4.81
أ	3,226,804	0.4118%	27	15,515,668	0.5519%	27	4.81
أ	9,334,034	1.912%	21	44,379,696	1.5786%	19	4.75
أ	972,875	0.1242%	32	4,427,141	0.1575%	34	4.55
أ	63,034,942	8.0443%	3	275,542,547	9.8009%	2	4.37
أ	12,080,564	1.5417%	19	52,283,490	1.8597%	17	4.33
أ	20,309,171	2.5918%	14	87,907,105	3.1268%	11	4.33
أ	3,299,101	0.4210%	26	13,122,675	0.4668%	28	3.98
أ	23,302,609	2.9738%	12	86,999,405	3.0945%	12	3.73
أ	28,206,045	3.5995%	10	102,078,112	3.6309%	9	3.62
Corpus	783,599,634	100%	47	2,811,387,118	100%	47	3.59

the voiceless [θ] and voiced [ð] dental fricative consonants of foreign words, particularly those containing the sequence <th> in English.

The graph <ظ> is slightly less frequent on arz than on ar, This is probably the result of a balance between the tendency to phonetically render the replacement in Egyptian Arabic of the pharyngealized alveolar fricative [zʕ] by the pharyngealized alveolar occlusive [dʕ] for example with <ظهر> /dʕahr/ 'back' and the opposite phonetic trend with <ظابط> /zʕar-betʕ/ 'officer'.

TABLE 15. Graph frequency comparison between arz and ar—Over-representation in arz

graphs	arz			ar			proportion
	Number	Proportion	Rank	Number	Proportion	Rank	ar/arz
ي	71,586,498	9.1356%	2	255,872,222	9.1013%	3	3.57
و	46,553,694	5.9410%	6	161,904,072	5.7589%	5	3.48
ر	45,472,582	5.8030%	7	156,650,474	5.572%	6	3.44
ن	41,468,036	5.2920%	8	141,732,511	5.0414%	8	3.42
ع	23,291,652	2.9724%	13	79,194,474	2.8169%	14	3.4
ا	47,052,142	6.0046%	5	155,964,726	5.5476%	7	3.31
ز	118,929,723	15.1774%	1	391,184,188	13.9143%	1	3.29
س	5,613,067	0.7163%	24	18,326,115	0.6519%	26	3.26
م	27,247,443	3.4772%	11	85,551,920	3.0431%	13	3.14
ط	55,993,480	7.1457%	4	174,238,132	6.1976%	4	3.11
ك	9,366,993	1.1954%	20	25,908,792	0.9216%	24	2.77
ل	20,292,778	2.5897%	15	55,471,216	1.9731%	15	2.73
ص	16,558,167	2.1131%	16	45,069,246	1.6031%	18	2.72
ح	15,821,354	2.0191%	18	36,108,823	1.2844%	21	2.28
هـ	31,729,113	4.0491%	9	38,498,437	1.3694%	20	1.21
ى	16,225,512	2.0706%	17	12,808,495	0.4556%	29	0.79
يس	493,200	0.0629%	34	29,921	0.0011%	45	0.06
ج	598,199	0.0763%	33	21,100	0.0008%	46	0.04
	456,134	0.0582%	35	14,114	0.0005%	47	0.03
Corpus	783,599,634	100%	47	2,811,387,118	100%	47	3.59

The various spellings of hamza <ؤ>, <ئ> and <ء>, as well as those using ʾalif as a carrier: <أ>, <إ>, as well as the madda <آ>, are relatively less used on Wikipedia maṣri compared to Wikipedia al-ʿarabiyya. This corresponds to the virtual absence of pronunciation of the “glottal stop” /ʔ/ originating from a hamza in the middle and end of words in Egyptian Arabic, and therefore to the avoidance of the etymological spelling in favor of a more phonetic spelling. On the other hand, this points to a lack of popularity of the phonetic solution to mark the “glottal stop” /ʔ/, that in Cairo Arabic replaces /q/ from Arabic fuṣḥā, and the strong tendency to retain the etymological spelling <ق>. Incidentally, this grapheme is almost as frequent on arz as on ar.

Concerning the ʾalif, the rule requiring it to be written without hamza (either above, or below) seems to explain the relative rarity of <ا> and <أ> on arz compared to ar and consequently the slightly higher frequency conversely for the “naked” ʾalif <ا> on Wikipedia maṣri compared to Wikipedia al-ʿarabiyya. If we consider the total occurrences of these graphs on each of the two Wikipedia editions, we obtain a ratio of 3.83

occurrences on ar for each occurrence on arz, close to the ratio of 3.59 between the totals of occurrences of all the characters studied.

Among the relatively much more present graphs on Wikipedia maṣri is <◌◌> which, in this case, corresponds not only to the hā' of the fuṣḥā, but also to the tā' marbūṭa <◌◌>, which the recommendations enjoin to be written without its two dots when it is not at the end of a word in a state of annexation, and thus is pronounced /a/.

As for 'alif maqṣūra <◌◌>, it is the only canonical Arabic graph that has numerically more occurrences on Wikipedia maṣri than on Wikipedia al-'arabiyya, despite an overall ratio of 3.59 in favor of the latter. This reflects the tendency prevalent in Egypt of writing any final yā' <◌◌> without its two dots, making final <◌◌> an Egyptian grapheme which, in addition to its value in fuṣḥā of 'alif maqṣūra pronounced /a/, takes on that of final yā' pronounced /i/ or /j/ in Egyptian Arabic.

With the deletion of the hamza over and under 'alif and of the two dots of the tā' marbūṭa and of the final yā', the contributors to Wikipedia maṣri have succeeded in giving the script used in their project a specific appearance that helps to differentiate it from that of Wikipedia al-'arabiyya. This tendency to simplify the script is not unlike the common style of handwriting, which regularly omits these marks to increase writing speed. This is reminiscent of the popular roots of Egyptian Arabic, which was originally written mainly by hand. It is also visible in caricatures that have long illustrated Egyptian newspapers.

The rule concerning the 'alif al-waṣl and the 'alif al-qat', as well as the one concerning the tā' marbūṭa were already suggested by Faṭḥi Imbābi (born in 1949) in a sort of post-face entitled "Toward a modern Arabic language" written at the end of his 1994 novel *Marā'ī al-qatl* 'Killing grounds'. He even suggested another simplification by removing the two dots above the qāf <◌◌>, a graph that is pronounced /q/ in Standard Arabic, to clarify its pronunciation as a glottal stop in Cairene Arabic by the use of this graph: <◌◌> (Imbābi, 1994, p. 414). He did not, however, go as far as apply this innovation to the Egyptian Arabic dialogues in his novel, and I have found no trace of it elsewhere. The fact that this graph is deprived of a dedicated key on modern keyboards probably does not help its popularization. What remains to be seen is whether getting rid of all these dots and diacritics will not interfere with the readability of the texts.

7. Al-Zāhir Wa Al-Bātin

In a way, deleting elements of the script harks back to the origins of Arabic writing, when dots were not used to discriminate between letters of the same shape. This contrasts with modern versions of the Qur'an, which not only include dots but are also entirely vocalized.

This contrast may be related to the opposition between the concepts of al-Zāhir and al-Bāṭin mentioned in the Qur'an (57:3) that could be translated as "the Manifest," and "the Unseen". It is an opposition that permeates many aspects of Arab society, found, for example, in the crescent moon taken up as a symbol of the Red Crescent with the lit side of the moon standing out from its shaded part.

The solution of the abjad reflects these two facets: the possibility to write with a minimum of graphs, and let the reader make his own deductions, or with all the diacritics to provide a maximum amount of information. Paradoxically, the designers of Wikipedia maṣri are thus resorting to traditional strategies of Arab culture, while seeking to differentiate themselves from it.

The script of Wikipedia in Egyptian Arabic, therefore, reflects the tug-of-war between the contributors' desire to differentiate the language used from Standard Arabic and the need to offer texts that remain accessible to Egyptian readers. As in many cases where a script is adapted to an essentially oral language, the trade-off is between respect for the etymology or morphology of the word and the need to transcribe its actual pronunciation. Wikipedia maṣri is a major opportunity to write Egyptian Arabic in many fields to which it has rarely been applied before. It is therefore a particularly interesting candidate in the mission to help establish an orthographic standard for Egyptian Arabic.

References

- Abdel-Massih, Ernest T., Zaki N. Abdel Malek, and El-Said M. Badawi (2009). *A Reference Grammar of Egyptian Arabic*. Washington, DC: Georgetown University Press.
- Attia, Mohammed et al. (2015). "Arabic Spelling Error Detection and Correction". In: *Natural Language Engineering* 25.
- Atṭia, 'Adil (2010). "Goose Yā". In: *Al-ḥiṣār al-mutamaddin* 3145. عادل عطية، ياء الوزّة!، الحوار المتمدّن
- Badawi, S.M. and M. Hinds (1986). *A dictionary of Egyptian Arabic: Arabic-English*. Librairie du Liban.
- Boutros, Wadie (1993). *Ablan wa sablan, méthode d'arabe égyptien du Caire*. Le Caire: Éditions Dar el-Nashr Hatier.
- Coulmas, F. (2003). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Davies, Humphrey T. (2005). "Dialect Literature". In: *Encyclopedia of Arabic Language and Linguistics* 1. Ed. by Lutz Edzard and Rudolf Erik de Jong, pp. 597–604.
- Diem, Werner (1974). *Hochsprache und Dialekt im Arabischen: Untersuchungen zur heutigen arabischen Zweisprachigkeit. Abhandlungen für die Kunde des Morgenlandes*. Wiesbaden: Franz Steiner Verlag.

- Fahmī, ‘Abd al-‘Azīz (1944). *Latin letters for writing Arabic*. عبد العزيز فهمي، الحروف اللاتينية لكاتب العربية. Hindawi.
- Farnawany, Refaat (1981). “Ägyptisch-Arabisch als geschriebene Sprache, Probleme der Verschriftung einer Umgangssprache”. Inaugurale Dissertation sur Erlangung des Doktorgrades der Philosophischen Fakultät II. Friedrich Alexander Universität.
- Ferguson, Charles A. (1959). “Diglossia”. In: *Word* 15, pp. 325–340.
- Fiske, Willard (1897). *An Egyptian Alphabet for the Egyptian People*. Landi Press.
- Gairdner, William Henry (1916). *Egyptian Colloquial Arabic. A Conversation Grammar*. Oxford University Press.
- Girod, Alain and Emad Aziz (2016). *Grammaire de l’arabe égyptien d’aujourd’hui*. Paris: Édition Ellipses.
- Haggenmacher, K. (1892). *Grammatik des ägyptisch-arabischen Vulgärdialektes*. Boehme & Anderer.
- Imbābi, Faṭḥi (1994). *Killing Grounds*. فتحي إمبابي، مراعى القتل. Al-nahr lilnašr wa l-tawzī‘.
- Louis, Samia (2008). *Kallimni ‘Arabi Bishweesh*. Cairo: The American University in Cairo Press.
- Mitchell, T.F. (1956). *An Introduction to Egyptian Colloquial Arabic*. Oxford University Press.
- Rosenbaum, Gabriel M. (2017). “Discussing the use of dialect versus standard Arabic in Lenīn al-Ramlī’s Ḥaṣāwī-skits and other texts”. In: *Jerusalem Studies in Arabic and Islam* 44, pp. 37–66.
- Al-Ṣawī, Muḥammad (2017). *Writing Arabic in Latin letters*. محمد الصاوي، كتابة العربية بالحروف اللاتينية. Noor Publishing.
- Shraybom-Shivtiel, Shlomit (1998). “The Question of Romanisation of the Script and the Emergence of Nationalism in the Middle East”. In: *Mediterranean Language Review* 10, pp. 179–196.
- Spitta, Wilhelm (1880). *Grammatik des arabischen Vulgärdialectes von Aegypten*. Hinrichs.
- Al-Toma, Salih J. (1961). “The Arabic Writing System and Proposals for Its Reform”. In: *Middle East Journal* 15.4, pp. 403–415.
- Al-Tonsi, Abbass, Laila Al-Sawi, and Suzanne Massoud (2010). *Kalaam Gamiil, An Intensive Course in Egyptian Colloquial Arabic*. Vol. 1. Le Caire – New York: The American University in Cairo Press.
- Watson, Janet C. E. (2002). *The Phonology and Morphology of Arabic*. Oxford University Press.
- Willmore, John Selden (1901). *The Spoken Arabic of Egypt*. London: Dalcassian Publishing Company.
- Zack, Liesbeth (2014). “Key to Mass Literacy or Professor’s Hobby? Fiske’s Project to Write Egyptian Arabic with the Latin Alphabet”. In: *Al-‘Arabiyya* 47, pp. 1–19.