# Embrace Franco

### saadnpq

### 2019-05-12

# Contents

# 1   abstraction

7afez was here

## 1.1   the nature of the language

humans like any other animal have a way to send signals to each other by the means of sound. our brain utilizes our ears and mouth to form a communication channel by encoding entities in nature to a conventional verbal sounds. this whole thing we call it now "the language" in the English language, or more precisely "the spoken language".

## 1.2   the nature of writing systems

between 3400 and 3300 BC in sumar humans invented the most important invention in the history (its actually the invention that defined the beginning of history), the writing system. this is yet another communication channel in which our brain this time utilizes our eyes and hands to form it.

the writing system is very different than the spoken language in its nature. our brain treats those two communication channels differently. we often falsely (but understandably) include the writing system in the definition of the language. it's known that learning speaking and listening in a new language is a different skill than writing and reading it and require different training. due this natural distinction, the writing systems of languages evolved independently. actually so many languages didn't have any writing system until relatively recently (like Arabic for instance).

what i want to emphasis from this section is one thing, the writing system is **not** the language itself, and they are mostly independent from each other, some languages have multiple writing systems (e.g. Hindi and Urdu). some writing systems are used in multiple languages, as the Arabic writing system used for writing languages other than Arabic like Persian, Kurdish, and Urdu (Arabic was also used to write Spanish until the 16th century).

## 2   categories of writing system

### 2.1   phonetic based

in phonetic based systems every letter represent a speech sound, like alphabetical languages (e.g. all Latin languages) or abjadic languages (e.g. Arabic and Hebrew).

### 2.2   syllabaries or logographies based systems

in those systems each character represents a word or part of a word (e.g. Chinese and Egyptian hieroglyphs)

### 2.3   mixed

some writing systems has both types. for instance Japanese has letters that represents sounds and also uses some Chinese characters in its writing system.

## 3   character set and encoding

before we continue we have to agree on the meaning of some words that will come up often in this discussion.

### 3.1   character set

a character set is the collection of symbols that's used in a writing system. for example, a to z is character set of 26 symbols. while a to z + A to Z is another character set with 52 symbols. some writing systems uses character sets of thousands symbols like Chinese. different writing systems can share one character set or borrow some symbols from another writing system's character set.

### 3.2   encoding

long time after humans invented writing systems. it was the time for inventing the second most important invention in history, computers. how computers work demands another layer of encoding, encoding the symbols in a character set to ones and zeros. this layer forms a communication channel between we humans and computers.

## 3.3 ASCII

ASCII is character set and encoding that consists of 128 symbols {a-z + A-Z, 0-9, punctuations}, i.e. every thing you can see on English keyboard. it's the most simple and supported character set and encoding.

## 3.4 the Arabic character set

the character set of Arabic can consists of (at least) the following symbols:
! " # $ [U+066A] & ' () [U+066D] + [U+060C] - . / [U+0660] [U+0661]
[U+0662] [U+0663] [U+0664] [U+0665] [U+0666] [U+0667] [U+0668] [U+0669]
: [U+061B] <=> [U+061F] [U+0621] [U+0622] [U+0623] [U+0624] [U+0625]
[U+0626] [U+0627] [U+0628] [U+0629] [U+062A] [U+062B] [U+062C] [U+062D]
[U+062E] [U+062F] [U+0630] [U+0631] [U+0632] [U+0633] [U+0634] [U+0635]
[U+0636] [U+0637] [U+0638] [U+0639] [U+063A] || [U+0640] [U+0641]
[U+0642] [U+0643] [U+0644] [U+0645] [U+0646] [U+0647] [U+0648] [U+0649]
[U+064A] [U+064B] [U+064C] [U+064D] [U+064E] [U+064F] [U+0650] [U+0651]
[U+0652] [U+0671] [U+0670] [U+066C] " " Arabic charset shares some symbols with ASCII which is the source of many problems regarding reordering in mixed text. the Arabic charset has another point of complexity, which is how symbols are displayed on screen. for example the letter [U+063A] is never displayed as you see it right now in a typical Arabic text, basically it changes its shape according to where it appeared in the word (initial, medial, final) and also according to the letter before it and after it. Some letters would have more than eight glyphs (shapes) in some writing styles. and almost every character in Arabic behaves this way. the total number of glyphs in traditional printing exceeds four hundred in many typefaces.

# 4 the problem

## 4.1 mixed text

since writing mixed English and Arabic is inevitable for Arabic speakers in the world we live in now, the inherently problematic bidirectional text scenario arises more often bringing with it all kinds of headaches.

every text is stored electronically as a string of characters, writing RTL (Right To Left) or LTR doesn't change the order those characters are stored internally. the software displaying the string is responsible for correctly computing the relevant position of the next character on the screen based on whether or not this characters and the character preceding it belongs to an RTL character set.

if you are an Arabic speaker you know the frustration of reading or editing bidirectional text on nearly any software. some of the ways things can go wrong are:

- numbers inside RTL (specially if there is a number at the end or the beginning of a line or there is a number at a direction switch)

- highlighting text with mixed directions

- wrapping text on multiple lines

- adding a comma on a direction switch

- enclose text in parenthesis (or any mirrored characters like "<,[,{")

- specifying the base direction

you can see this for an introduction to the bidi algorithms and its limitations.

another problem is keyboard mappings, if you are a keyboard oriented person as most power users, changing keyboard layout while editing will disable all your keyboard shortcuts making editing more and more and more difficult.

## 4.2 TODO necessity of lots of diacritics

## 4.3 TODO not suitable for modern diacritics

# 5 solutions

there are two solutions i have seen people migrating to.

## 5.1 all English

this solution is popular among Arabs with a weak connection to the Arabic culture, mainly due to their foreign education. those people just use English whenever they write on a computer or a phone regardless of the formality of the writing – even if they speak Arabic in their normal life.

this solution is of course not practical for most people and so harmful on many levels. eventually those individuals will totally lose their connection to their culture and think in English. some people will argue that this as a good thing, i can understand this view, but for reasons beyond this discussion it's not considered a solution for a lot of people.

## 5.2 Writing Arabic Using Latin Characters

this solution is traced back to the time email and cell phones didn't support Arabic. users back then used to send Arabic written in English alphabet as an ad hoc solution. this system has many names: Araby, Arabizi, Mu'arrab, and Franco-Arabic. i will refer to it in this article simply as Franco as we call it in Egypt.

but way before that there was proposals to replace the Arabic letters with Latin ones. the first recorded one was by Wilhelm Spitta in 1880 in his book "Vulgardialectes von Agypten Grammatik des Arabischen [The Rules of Slang Arabic in Egypt]" in which he suggested using Latin characters to write the Egyptian slang. There were many other supporters for this proposal during the following forty years. Among them was Abdul Aziz Fahmi who proposed a full practical scheme in response to a competition organized by the Academy of the Arabic Language in Cairo. Other supporters of Fahmi's proposal, including Sa'id 'Aqil and Anis Freha developed their own proposals. They found additional support in Turkey, where Kamal Ataturk ordered the adoption of English characters to write Turkish instead of the Arabic characters in 1928.

those academic proposals didn't have any influence over the design and adoption of Franco, but rather it's developed organically over time by society, it got a lot of adoption among new generations to the point where google added Franco as one of its supported input methods in 2012 (but dropped it later due to Anti-Franco campaigns). it has been, in formal and informal daily use communication, entertainment, social media, signs and flyers, or even in students' writing books. i must say also that Franco wasn't a way of romanization of Arabic (which is a way of writing Arabic names and titles in Latin alphabet to be pronounced by English speakers). Franco meant to be a system for Arabic speakers to communicate with each other in Arabic.

### 5.2.1 characteristics of common Franco.

1. Franco only uses the ASCII charset(i.e. does not use Latin symbols as é or á).

2. the Franco system is contextual.

3. The traditional Arabic vocalization marks are substituted by vowels. The "fatha" by a; the "kasra" by i or e; and the "damma" by u, ou, or o.

4. The Franco system differs for every Arabic country, depending on the local dialect. for example, in countries which french is its second languages, "ch" is used to write the Arabic letter "sheen" instead of "sh".

5. numbers are used to represent Arabic characters that can't be written using the ASCII. numbers are chosen based how similar the number is to the letter represented. for example, the letter [U+062D] represented by 7; the letter [U+0621] represented by 2 (the same shape mirrored); the letter [U+0639] represented by 3.

# 6   war against Franco

Franco has always been fought and severely criticized by Arab nationalists and Muslim enthusiasts, who considered it as a direct attack on the Arabic identity. They also saw it as a threat to the Quran, which is written in classical Arabic using a writing system that has lasted for more than fourteen centuries. Islamic and Arabic scholars started a series of Anti-Franco campaigns around the Arab world. Anti-Franco was the main theme of World Arabic Language Day 2014 . Google as well stopped from supporting Franco and removed it from their language tools. Arabic experts reactions to Franco are typically quite negative, even when they themselves employ it sometimes. They emphasize the use of the traditional writing system and marginalize the efforts of students to use alternative forms of writing Arabic.

most Anti-Franco arguments has political, religious and racial components new generations are free of. they also come from a misunderstanding of the difference between a language and its writing system. i see the adoption of Franco is a healthy thing for the Arabic language. i love Arabic; i think in Arabic, and i will continue to think in Arabic to the end of my life. but Franco for me is a way to express myself **in Arabic** without all the hassle traditional writing brings in.

## 6.1   rational criticism

rational people who are not used to read or write Franco will criticize Franco as it's hard to read, of course using a new writing system wouldn't come without a learning curve. its not fair to compare the ease of Franco to a writing system people taught in school for years. suppose Franco taught in schools and someone introduced you to the traditional system, would you find it easier?.

i didn't see any criticism outside that it is hard to read. if you have another point against Franco please send it to me and i will include it in this section.

# 7 the future of Franco

until now i think Franco refuses to die despite everything, Arabs are still using it heavily in chats, but the public domain doesn't reflect this internal adoption. we need new studies to measure Franco use in different geographic areas. i came across some studies done in Saudi Arabia with ranging results from medium to high usage among teens. i claim that if similar studies made in countries like Egypt, Morocco, Lebanon or Jordan the studies will have significantly different results towards very high usage.

like any organically grown idea, most of the design choices in the common Franco we write today has some form of a hidden wisdom inside of it, but has some weak aspects as well. if we are serious about developing a writing system, there must be an intelligent driving force that maintains some standard or a guideline. i didn't find one yet, but i am working on one here. i am planning also to visit a linguistic college and see how people there think about this and maybe try to convince some students to maintain the project further in a scientific manner. if you are interested or want to contribute any thing in any form (even a criticism would help in some decisions) please contact me.

# 8 References

- Yaghan, M. A. (2008). "Arabizi": A Contemporary Style of Arabic Slang. Design Issues, 24, 39-52. `http://dx.doi.org/10.1162/desi.2008.24.2.39`

- Kenali, A. M. S., Yusoff, N. M. R. N., Kenali, H. M. S., & Kamarudin, M. Y. (2016). Code-Mixing Consumptions among Arab Students. Creative Education, 7, 931-940. `http://dx.doi.org/10.4236/ce.2016.77097`

- Egyptian Romanized Arabic `https://www.duo.uio.no/bitstream/handle/10852/24250/Bjoernsson_master.pdf?sequence=1&isAllowed=y`

- `http://www.bbc.com/arabic/scienceandtech/2012/12/121220_arabic_language_internet_arab_days.shtml`

- https://www.arabamerica.com/google-adds-arabiziarabic-translation-to-its-input-to
- https://en.wikipedia.org/wiki/Arabic_chat_alphabet
- https://en.wikipedia.org/wiki/Writing_system