
Arabic Documentation Standards

Mohammed Elzubeir, Arabeyes Project <contact (at) arabeyes dot org>

\$Revision: 1.4 \$

This document will hopefully set some standards for those involved in Arabic technical documentation and translation.

Table of Contents

Introduction	1
Acknowledgments	1
The Problems	1
Difference Between MA and RA	2
Word-for-Word in Computer Software	2
Merging Spoken and Written Arabic	3
Standards	3
The Balance of SA and RA	3
The Use of Harakat (Ligatures)	4
Verb Forms	4
Foreign City and Person Names	4
Borrowing Latin Words	4
Interchangeable Terms	5
Abbreviations and Acronyms	5

Introduction

Acknowledgments

This paper could not have been written without Chahine Hamila (hchahine (at) magic dot fr) and Mustafa Merza (merza (at) leb dot net). In fact, this draft is a compilation of exchanged postings between Chahine, Merza, and myself, edited and formatted.

The Problems

The Arabic language is *diglossic*, which is a sign of richness and complexity. If it has not been for the preservation of *Mother Arabic*, the Arabic language would have been as diverse as the European languages today¹.

Language syntactic, phonological, and semantic change is a natural law and is unavoidable. The metaphor *mother tongue* could be taken literally, as it ages and dies while its daughters go to become languages, and eventually evolve to be mothers.

The *elixir formula* is clear in the case of diglossia in the Arabic language, as it preserved its integrity. It also provided a medium ground that can accommodate change. The true elixir in the case of Arabic is the *Quran*².

¹True bilinguals, as research into the cognitive aspects of language acquisition shows, tend to show better problem solving skills.

We will define the relationships of the Arabic language classes in the following manner:

- MA - Mother Arabic
- SA - Standard Arabic - the medium ground where heritage and language laws negotiate.
- RA - Regional Arabic

To better demonstrate the above concepts, have a look at the list in The Balance.

There are roughly two major *intelligibility* areas in the Arab World:

1. Maghreb - covering the majority of the Western areas.
2. Middle East - covering the rest of the Arab world.

Although the Arabic language is preserved, it is threatened to be a purely liturgical language, refusing the evolution to our modern needs. In this new era of mass communications, many languages face the threat of disappearing.

Some of the factors that may influence the survival of a language are:

1. Broad geographical native use
2. Alphabets - for example, Chinese does not have alphabets.

Difference Between MA and RA

There are several differences between *Mother Arabic* and *Regional Arabic*. They can be categorized in three levels, syntactic, phonological, and semantic.

- Syntactic - One of the major differences between the MA and RA's is that RA's drop the *case*. The *case* defines whether the sentence is *nominative* or *accusative*.

Had it not been for the *elixir* formula, RA's would have been even further polarized to model the Latin language and its daughters ³.

- Phonological
- Semantic

Word-for-Word in Computer Software

Unfortunately, much of the available Arabized software seems to have a different idea of what translation entails. Although one loses certain aspects of the meaning and context, word-for-word translations look and sound unnatural and drive users away.

~~This phenomenon~~ is evident in the younger generation who refuse to use Arabized software because of ²Regardless of your religious belief, the Quran is the heart of why this formula exists in the first place. Not only was it at the heart of maintaining the link between the daughters (RA's) but it also had its impact on other languages (such as Urdu, Farsi, etc) ³Latin and Arabic are not governed by the same standards. Indo-European languages depend on the vowels to carry its semantic load, unlike Semitic languages.

the awkwardness of the terminology.

Merging Spoken and Written Arabic

The Question of Imports

Importing new words into a language is a dangerous task. It opens the door for the corruption of the language's integrity. However, it also can enrich the language and accelerate its evolutionary process. Translators are to treat this with extreme care.

Resistance from the Arab academia and intellectuals has accompanied the care-free approach that is taken toward importing foreign words. This practice has lessened the acceptability of Arabized equivalents.

However, in certain instances the use of borrowed terms is most appropriate.

Current Arabic Computer Literature

Current computer literature has both odd and awkward terminology, rendering it often unintelligible. This is a result of two factors:

- Arabic Romanticism - Arabic has been considered (and some still argue) to be a language only suitable for fiction. One of the elements which persuade this line of thought is the *School of Mahjar*.
- Post-colonialism - Arab countries have suffered a long and enduring colonial period. This has resulted in drastic shifts in RA's.

Fortunately, the lack of legacy among the currently available computer Arabic literature is to our advantage to design a wiser organized standard.

Standards

In order to avoid the creation of a new hybrid Arabic, which would somehow include an intersection of RA's with SA, we will bring the gap between the two a little closer.

The Balance of SA and RA

In the process of Arabic technical documentation, a balance between both tiers (SA and RA) is wiser than emphasizing one over the other. In our technical context, we must examine the elixir formula, deducing solutions based on its responsibility for the survival of Arabs as a culture.

Technical Arabization falls under the media usage. Please refer to Table 1 for a list of Arabic classifications. This should not be an obstacle to simplification, in fact it is a vehicle toward simplification. In effect, the function of the SA is to act as the intelligible medium to Arabic in any given environment and at any level of education.

<i>Type</i>	<i>Example</i>
Mother Arabic	Quran
Standard Arabic	Media
Regional Arabic	Conversations

Another issue to consider is the user's familiarity with the English language and prior knowledge of existing computer applications. The borrowing (discussed in more detail later) is here advised, as it accelerates the user's ability to read and understand unarabized computer literature (e.g. man pages, etc.)

A good balance of SA, RA, and borrowing, as well as providing the user with the proper tools to independently accumulate a larger vocabulary from the target language (English, French, etc.) is the recipe for technical Arabization.

The Use of Harakat (Ligatures)

Harakat are generally omitted from literature in Standard Arabic. However, there are special cases where they are mandatory, especially in translations of UI's (User Interfaces).

- If the 'haraka' is a shadda then it is mandatory to place. Consider the shadda to be the equivalent of a character itself.
- If the omission or addition of a haraka can change the meaning of the word. For example, the word ##### and ##### are fundamentally different.

Verb Forms

In the translation of a UI, the verb forms play a fundamental role. In a survey of what the users would like to see, the overwhelming majority prefer the noun form as opposed to the imperative form.

For this reason, the following rule is to be followed: If the user is performing a task that will hopefully result in feedback from the application or system, the noun form is to be used. However, if the system is the one expecting the user to provide input, the imperative form is used.

For example, in the case of a "File->Open" operation, it would be in noun form. However, if the system is asking the user to "Enter Username", the imperative form is to be used.

Foreign City and Person Names

Foreign city and person names should be translated as commonly known, referenced against the following references:

- QamMoose [<http://qamoose.arabeyes.org/>]

If the name is not found then it is to be transliterated and inputted into QaMoose for future references.

Borrowing Latin Words

Arab scholars have a tendency to be strongly conservative against the borrowing from other languages. In contrast, the English language has gained more words than any other existing language today⁴.

Borrowing should be limited to an absolute minimum and left as a last resort.

The Common Popular Law

This law says that borrowing should only occur when the resulting translation would go against the *common popular*. That is to avoid the awkwardness which plagues our current Arabic technical literature.

For example, the English words *CD*, *Internet*, *disk* are so commonly used that it would be a more likely

⁴A fact that is conveniently omitted during our public education. English has about 700,000 words whereas Arabic only has about 200,000.

candidate than an inferior translation such as ##### #####.

Interchangeable Terms

In the case where two common terms are used to describe one thing (e.g. [##### and #####], it is left up to the translator. The scarcity of Arabic computer literature legacy allows for a greater degree of freedom.

However, translators are advised against the abuse of this option. Although interchangeability can help when reading different material, it can lead to a lot of unnecessary confusion. When unsure, put the additional terms between parenthesis to clarify the usage.

Abbreviations and Acronyms

There are many acronyms and abbreviations in computer and technical terminology, which makes it difficult to introduce Arabic equivalents. To resolve this issue the technical writer is to do the following:

- Product names should always be transliterated in Arabic characters. This is regardless of whether they are an acronym or an abbreviation.
- All other acronyms (e.g. HTML, SGML, HTTP) should be transcribed in their Latin form (using English characters), but their translated antonyms could be added between parenthesis for clarification.