

Simplification of the Arabic Script: Three Different Approaches and their Implementations

Yannis Haralambous

Atelier Fluxus Virus
187, rue Nationale
59800 Lille, France
<http://pobox.com/~yannis>

Abstract. In 1945, the Cairo Academy for the Arabic Language opened a contest to find the best project for simplification of the Arabic writing system. They received about 200 replies. We have chosen three of these projects (a preliminary project by an Academy subcommittee, and projects by Ahmed Lakhdar-Ghazal and Yahya Boutemene) and have implemented them via the Ω typesetting system. In this paper we describe and discuss these projects and their implementations. A text sample is presented both in simplified and in regular form.

Showing new aspects of the Arabic script, these systems can be useful for Arabic typesetting, as well as in providing new directions for Arabic type design.

1 Introduction

In 1918, Lenin simplifies the Russian alphabet by abolishing 4 letters considered unnecessary from the phonetic point of view. In 1928, Kemal Atatürk simplifies the writing of Turkish by switching from Arabic to Latin script—a drastic change motivated also by political and historical factors; Latin script was supposed to bring Turkey closer to the “western world.” In 1941, the Fraktur script is abolished in Germany: besides some pseudo-historical racist Nazi argumentation, the main argument against Fraktur was low readability. In the sixties Mao simplifies Chinese writing by replacing a certain number of ideograms with simpler ones. In 1980, Greek writing system is simplified by abolishing breathings and all but one accents: like in the case of Lenin, the argument is that they play no rôle anymore in the phonetic representation of language.

These are “official” (grammatical) simplifications. There have been also purely typographical ones: for example, sans serif typefaces which appeared in the beginning of this century are simplifications of the traditional typographical forms of the first 450 years of typography. Such typographical simplifications have occurred also in other scripts: Cyrillic and Greek—of course, since they are direct relatives of the Latin script—but also Armenian, Georgian—which are quite “latinocentric,” typographically speaking—Hebrew, Arabic, Thai, CJK, etc. There are also examples of the contrary (scripts for which simplification can be only minimal to avoid esthetic disasters); notorious examples being Amharic and Tibetan.

1.1 Simplifications of Arabic Script

Until now, the Arabic script has undergone only typographical simplifications. Indeed, traditional Arabic typography, as described in [2] and [3], has some rather uncomfortable aspects:

1. too many glyphs (hundreds of ligatures) make it difficult both to typeset and to read;
2. diacritics (when used) are hard to position correctly and often require manual adjustment, which is expensive in time and effort;
3. big differences in size between letters make small size typeset Arabic unreadable;
4. big ascenders and descenders require a bigger baseline skip than other scripts: less lines fit on a page and hence more paper is necessary.

Another, less important, but still characteristic problem is that traditional Arabic typography is a “closed world”: the variety of Western typographical styles (bold, italic, etc.) can hardly be applied or find analogies. This can be a handicap both in translations of Western books (for example scientific ones, where often italic, bold and small caps have special meanings) or when mixing Arabic and other scripts.

In the last 30 years, thanks to new typesetting techniques (phototypesetting, computer) there have been “modern” Arabic typefaces (see for example the sample of text on fig. 9) resolving some—if not all—of the problems mentioned, and this is indeed a “typographical simplification” of Arabic. In this paper we will present some new approaches (which in fact are not new historically, since they all have been proposed in the fifties, but new because they have not yet been implemented¹) which, we hope, will get better known through our public domain implementation and can be useful both to Arabic writers and readers of today as to Arabic type designers of the future.

1.2 Some History . . .

Besides providing a courageous and refreshing alternative vision of the Arabic writing system, the simplification proposals we present have also an historical value. Here are some details on their historical context, the reader will find more information in [7].

In 1936, at a meeting of the Cairo Academy of the Arabic Language, one of its members, Ahmad al-Awamiri, proposes the establishment of a commission to study the problem of foreign words written in Arabic. After two years of discussions, the Academy comes to the conclusion that the problem to solve is more general, and that the solution had to be a simplification of the Arabic script. On January 23, Mansur Fahmi moves that a commission be established, to . . . *work, by all possible means, towards the simplification of the writing of Arabic letters, by inventing [a new system], and to make correct Arabic reading easier, but without abandoning the fundamental principles of the language.* The motion is accepted.

For years the commission makes no proposal. In 1942, it is repopulated. Finally in 1944, two projects are presented: the one by Abd al-Aziz Fahmi is basically a switch to

¹ Except, of course, the LAKHDAR system (see section 3) which has been widely used in Morocco, and has been adapted to a multitude of printing techniques.

the Latin alphabet, where a few letters (the *heh*², *sad*, *ain* families and the *hamza*) keep their original shapes (in two sizes, corresponding to upper and lower case); the second one, by Ali l-Garim, keeps the original shapes of consonants but adds new letters shapes for short vowels, noonation and *madda*. A month later, the Academy decides to drop both projects, but nevertheless to publish them, as well as the minutes of the discussions.

A year later, on october 15th 1945, the Academy announces throughout the Arabic world an international contest to find the *best project of simplification of the Arabic writing system*, with a prize of 1,000 Egyptian pounds. The deadline for the contest is november 1st 1946. The deadline is postponed to march 31st 1947. On may 19th 1947, a new commission is established to study the 200 projects that have been received by the Academy.

Five years later, on october 27th 1952, the commission declares that no project has met the necessary requirements, and so nobody wins the £1,000 prize.

In march 1959, the commission forms a sub-commission which elaborates a project of *typographical* simplification of Arabic. This project is the first of the three we are presenting in this paper, we will refer to it as the “ACADEMY project” (see 2 for more details). On december 19th, the commission thanks the subcommission and approves the project, and on january 14th 1960, the Academy thanks the commission and approves the project. Part of number XXIV (january 1969) of the Academy’s Journal is typeset using this new typeface.

The second project (3, [6], [8]) we present is the one by Ahmed Lakhdar-Ghazal (LAKHDAR) which is perhaps the only system that has been widely used (road signs in Morocco are written in his typeface) and has already been adapted to several printing methods.

Finally we present a project (4) by Yahya Boutemène, sent to the commission after the end of the contest (in november 1955). This project has a very innovative approach: it uses characters from the Latin and Greek alphabet, eventually a little transformed, to approach Arabic letters graphically. We will refer to this project as BOUTEMENE.

2 The ACADEMY Project

2.1 Description

The ACADEMY project proposes typesetting of Arabic with as few as 72 types (the basic traditional system was using about 470 types). These types are displayed on fig. 1. On the last row of this table we see five graphical elements (appendices) used in combination with other characters; we have numbered them to make the description easier.

On the first row, we see the ligature for ‘God’ (Allah), without the vertical *fatha*.

On the second row we see the types for the isolated *hamza*, the *alef*, and letters *beh*, *teh*, *teh marbuta*, *theh*. Nothing extraordinary for the moment, except for the fact that the dots of letters *beh*, *teh*, *theh*, are slightly offset to the left, and that the isolated and final forms are missing. These forms are constructed by concatenating them with appendix 5.

² Throughout this paper, the names used for Arabic characters are taken from the Unicode v2 standard.

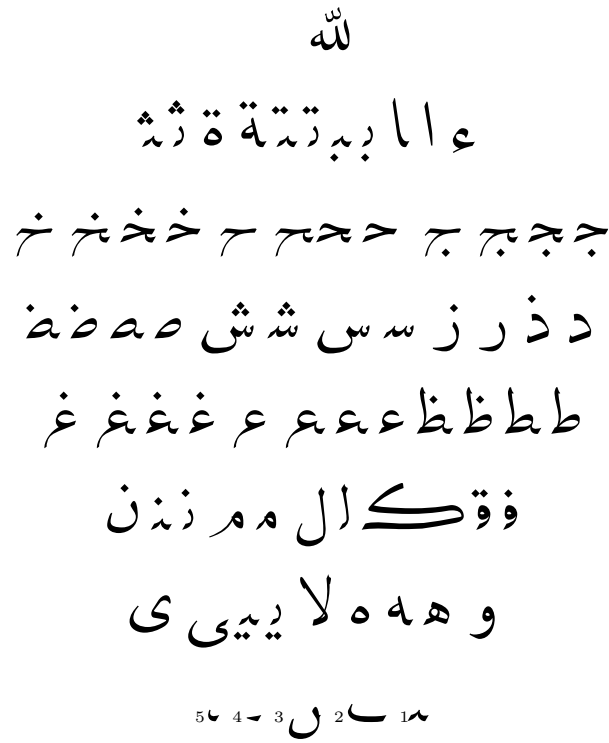


Fig. 1. Table of typographical types of the ACADEMY project

On the third row we see the types for letters *jeem*, *hah*, *khah*. Here, finally, we see something revolutionary (in the context of Arabic typesetting): the descenders of isolated and final letters *jeem*, *hah*, *khah* have been *cut off*. Instead, we find a small descender part, sufficient to distinguish isolated from initial, final from medial form.

On the fourth row we find types for letters *dal*, *thal*, *reh*, *zain*, *seen*, *sheen*, *sad* and *dad*. Notice that once again a certain number of contextual forms are missing: there are no final forms for 2-form letters and no medial and final forms for 4-form letters. What is very interesting is that the forms corresponding to letters *sad* and *dad* are “incomplete,” in the sense that they cannot be used by themselves for *any* contextual form: to produce an isolated or final form, you need to concatenate it with appendix 3, and to produce an initial or medial form you need to do the same operation with appendix 1.

On the fifth row we find types for letters *tah*, *zah*, *ain* and *ghain*. To produce the final and isolated forms of *tah* and *zah*, you need to concatenate the existing types with appendix 4. The letters *ain* and *ghain* have been drawn in a very special and unorthodox way: instead of using three different graphical forms as usually (see fig. 2) only one is used, namely the one of the initial letter. And, of course, the descenders of the final and isolated forms have been cut off like in the case of the *jeem* family of letters.

	ACADEMY	Traditional typeface
<i>jeem family</i> ^{1,2}	ج ج ج ج	ج ج ج ج
<i>ain and ghain</i> ^{1,2}	ع ع ع ع	ع ع ع ع
<i>feh and qaf</i> ¹	ف ف ف ف	ف ف ف ف
letter <i>kaf</i> ³	ك ك ك ك	ك ك ك ك
letter <i>meem</i> ¹	م م م م	م م م م
letter <i>heh</i> ⁴	ه ه ه ه	ه ه ه ه
<i>lam-alef lig.</i> ¹	لا بلا	لا بلا

¹ No difference between contextual forms design. ² Broken descenders.

³ New form. ⁴ No difference between initial and medial form.

Fig. 2. Main design differences between traditional Arabic typeface and the ACADEMY project

On the sixth row, we find glyphs for letters *feh*, *qaf*, *kaf*, *lam*, *meem*, *noon*. With letters *feh* and *qaf*, we obtain the *nec plus ultra* of typographical economy: all forms are constructed out of a single type, all you need to obtain a final or isolated form is to add appendix 2 (the same one used for the *beh* family of letters).

Letter *kaf* is very special: instead of using the regular four forms of *kaf*, a special glyph is taken, used primarily in the typesetting of classical texts (initial and medial forms of it can be found in the Quran, together with another ‘gigantic’ letter: the *noon*, final and isolated forms only). This letter, in a somewhat shorter form, is also used in ordinary Sindhi.

For the letter *meem*, the final/isolated form is drawn with a very oblique (almost horizontal) descender.

On the seventh row, we find types for letters *waw*, *heh* and *yeh*, as well for the *lam-alef* ligature. There seems to be no difference between the final and isolated forms of letter *yeh* and the *alef maksura*. A remarkable design decision seems to have been taken for the medial form of letter *heh*: as we know, there are (at least) two possible variants for this glyph: ‘ه’ and ‘ه’; these are both so important that in the Urdu language they have been chosen to have different grammatical rôles: the former is used as a letter (phonetically close to *heh*) and the second one is written after a certain number of consonants to indicate consonantic diphthongs. Well, ACADEMY goes a step further and abolishes *both* variants, in favour of a glyph similar to the initial form of letter *heh*.

There is only one type for the *lam-alef* ligature, so this one has to be used for both isolated and final form.

Finally, on the last row, we find the graphical elements used to construct supplementary contextual forms. The first one is the base of medial *beh*, *teh*, *theh*, etc. It is nec-

essary to form the medial and initial forms of *sad*-like letters. The remaining ones are ‘tails’, necessary to form a certain number of final and isolated contextual forms.³

2.2 Discussion

The ACADEMY project mainly aims to solve two problems:

- (a) the large number of types needed for traditional Arabic typesetting, and
- (b) the large vertical space needed by the ascenders and descenders of traditional types.

Both problems are solved without abandoning the classical design of types (don’t forget that all this happened in the fifties!). From the overall design of the typeface we see that a supplementary goal has been fulfilled:

- (c) most of the types of the ACADEMY typeface, either exist already in the traditional one, or can be obtained by cutting off parts of types of the latter.

For example, to obtain the new final/isolated *jeem*, one only needs to cut off part of the type of the traditional typeface. This method even works for ‘tails’: to obtain the tail needed for the *sad*-like letters (appendix 3), just take an isolated *noon* and cut the dot off.

It would be interesting to know if one can obtain the final/isolated *meem* type, by *bending* the descender of the traditional *meem*.

Still, there is one case where one has to draw new types: the *ain* and *ghain* letters. One cannot obtain the medial and final ACADEMY forms of these letters by cutting or bending traditional types.

And this brings us to a very interesting fact: if we restrain ourselves to the main goals of this project (reducing the number of types, reducing the baseline skip, using only available glyphs), there is no reason to invent new design forms for the medial and final *ain/ghain*. And not only there is no reason for them to be, but their use contradicts principle (c) [these glyphs have to be specially drawn!].

This brings us to the conclusion that the project ACADEMY *actually wanted to be more than just the design of a new “economic” typeface*. It had to bring changes which would (in the long term) simplify the Arabic writing system, and make it more “logical.”⁴

Another example illustrating this tendency is the one of medial letter *heh*: by taking the same form as for the initial letter, we obtain a more “logical” way of writing. Of course, this could hardly be applied to the final *heh* (the design bears no resemblances anymore to the archetypal form of letter *heh*) but why hasn’t this been applied to the isolated *heh*? Maybe the project didn’t wanted to be *too* radical? Or was it because a

³ It should be noted here that two forms seem to be missing: the final and isolated *lam*. In fact, ACADEMY is not clear on the construction of these: the only logical choice would be to use the initial/medial forms of *lam* followed by appendix 3. But these don’t seem to fit. For this reason (and until a satisfactory solution has been found), the author has added “genuine” final and isolated *lam* forms to the typeface and hence the total number of types needed to typeset with ACADEMY becomes 74, instead of 72.

⁴ In the sense that the contextual forms of each letter—with the exception, of course, of letter *heh*—are graphically much closer to each other than before.

‘strange’ isolated *heh*, would make an even stranger isolated *teh marbuta*, and this had to be avoided?

What is also very interesting is the use of the ‘huge’ *kaf*. It is not a secret that the ‘huge’ letters *kaf* and *noon* give a special look to the typeset Quranic text, bringing it closer to calligraphy. The choice of such a glyph for representing the (very frequent in Arabic) letter *kaf* is a very radical one, and contradicts the principles of simplicity and space economy. Of course, one can argue that the ‘huge’ *kaf* has a smaller ascender than the regular *kaf*; but what is gained in height, is lost in width: a text with many *kafs* is considerably wider than the same text with regular ones. A different—and more reasonable—solution would be to bend *kaf*’s ascender and bring it closer to the horizontal, as has been done with the descender of the *meem*.

Why has the ACADEMY project taken such a (apparently illogical) decision? Is it to counterbalance the loss of esthetic quality resulting from the new way of typesetting the final/isolated *beh*, *teh*, *theh* . . . ? Is it to give the new typeface a very specific look? Is it to remind the reader that, although new, this typeface has its roots in traditional Arabic typography?

Many unanswered questions make this a very interesting project to study and to implement. The reader can find a sample of it on fig. 8;⁵ the same text, typeset in a traditional Arabic typeface, can be found on fig. 7.

2.3 Conclusion

This project wants to bring changes without breaking with the esthetics of traditional Arabic typography. It wants to be practical, by re-using types from traditional typefaces, but doesn’t hesitate to draw new glyphs when this is necessary to simplify not only typesetting but also the Arabic writing system in general. If we consider the circumstances under which this project has been developed⁶ it deserves to be called “revolutionary.” It is a pity that it hasn’t be used more widely.

3 The LAKHDAR Project

Ahmed Lakhdar-Ghazal, Director of the Morocco Institute of Research et Studies on Arabization, has been working on the LAKHDAR project for the last forty years. His system has been adapted to all printing methods (traditional, Linotype, phototypesetting, computer). It has been distributed on Letraset foils, IBM electric typewriter disks and balls, and many other media.

In this paper we will present our public domain implementation of the LAKHDAR project, based on a font of the Ω system (Omega Serif Arabic, see [9] [4] and [5]).

⁵ The text of the sample is taken from [1].

⁶ Let us not forget that it is the work of a subcommittee of a committee of an Institution as old and traditional as the Cairo Academy of Arabic Language, and this in the fifties.

3.1 Description

There are two levels of the LAKHDAR system: level 1, the “pure” system, uses 84 types; level 2, the “complete” one uses 107 types. Contrarily to the ACADEMY system, this number also includes types for vowels and other diacritics, digits and punctuation, as well as 5 “foreign” characters (the pseudo-letter *veh*, Farsi *peh* and *gaf*, and Maghrebian dotless *feh* and *hah* with caron accent).

The reader can see the two levels of the LAKHDAR system on fig. 3. The table is actually divided in three parts:

- Part A. The “pure” system.
- Part B. The auxiliary characters which, together with part A, form the “complete system.” The three rows are called: “Priority esthetic letters,” “Secondary esthetic letters,” “Special letters/*shadda*.”
- Part C. Types added by the author for the sake of completeness.

The idea behind the “pure system” was to obtain, as much as possible, a one-to-one correspondence between letters of the alphabet and types. Of course, this is not always possible and two exceptions remain: the *yeh* (and derived letters: *alef maksura* and *hamza* with carrier) and the *heh* (and the derived *teh marbuta*). Nevertheless, the LAKHDAR project gets very close to the “one-to-one” principle. To obtain isolated and final forms, the three “tails” of the third row of table 3 are used.

For the sake of the “one-to-one” principle, many compromises have been taken, and in some cases this is esthetically unacceptable: note, for example, that there is no *lam-alef* ligature, nor an isolated *noon*. Fortunately these types are included in part B of the table. Also are included medial, final and isolated forms of the *ain* and *ghain*, and final/isolated forms of letters *jeem*, *hah*, *khah*, *seen*, *sheen*, *sad* and *dad*. Five “foreign” letters have been added: the Farsi *peh* and *gaf*, the “artificial” *veh* and the Maghrebian letter *hah* and dotless *feh* with caron accent.

The reader will notice that the LAKHDAR system contains letters *alef* with upper and lower *hamza*, as well as with *madda*, but not the corresponding *lam-alef* ligatures. These have been added by the author in part C of the table. Also in the LAKHDAR system there is an initial/medial *hah* with caron but no isolated/final form for this letter: it has been added in part C.

A very interesting aspect of the Lakhdar system is the treatment of vowels and diacritics. The problems with these important ingredients of the Arabic writing system are: (a) it is hard to find where to place them (even for a computer), and (b) depending on the letters they are placed upon, they tend to get too high (or too low) and increase even more the baseline skip.

These problems have been solved in the LAKHDAR system in a very elegant way: vowels/diacritics are placed not above or below letters, but *between them* and their height or depth is never bigger than the one of letters. See, for example, the following vowelized sentence:

نَتَعَلَّمُ الْإِنْكِلِيزِيَّةَ

A.

ابتدئ خذ نرس شصض
طظء غفقكلمنه و يءى قه
أأدئى و - ل ل م
٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩
0987654321 () " ! ? : ; , .

B.

م لا عفن
ج ح خ س ش ص ض ع غ ع غ
٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩

C. (Added by the author)

لله ج لآ لآ

Fig. 3. Table of typographical types of the LAKHDAR project

where, of course, we have omitted the *wasla* on the *alef* since this diacritic is not provided in the LAKHDAR system.

Why is the *shadda* diacritic not part of the “pure system?” The answer is very simple: the LAKHDAR system preconizes complete vocalisation of Arabic [a practice that seems almost utopic in the modern Arabic world]; and since there can never be a double consonant not followed by a vowel, every *shadda* must necessarily be accompanied by a such; the *shadda*+vowel combinations are indeed included in the “pure system.”

The reader can see a sample of our LAKHDAR system implementation, on fig. 10. As our implementation of the LAKHDAR system is based on the Omega Serif Arabic font, we have also typeset the same text in that typeface (fig 9), so that the reader can compare the two approaches.

3.2 Discussion

The LAKHDAR system is first of all an attempt for the ultimate abstraction of Arabic alphabet letters. Ahmed Lakhdar-Ghazal has tried to purify the design of the Arabic letter as much as possible, so that only what is strictly essential for the reader remains.

The greatest achievement of the LAKHDAR project is that this abstraction has been obtained by respecting at the same time typographical esthetics. We can consider the “pure system” as a tribute to abstraction with only intellectual value, while the “complete system” is a real-world Arabic typesetting system, optimized both for readability and for economy in time, effort and paper.

4 The BOUTEMENE Project

Contrarily to ACADEMY and LAKHDAR, BOUTEMENE is a “small” project. It has been developed by a single person (probably in his spare time) and, in all honesty, cannot be considered as “real-world ready.” In other words, the author can hardly imagine somebody reading a whole book written in BOUTEMENE.

Nevertheless, it is a very interesting solution to a certain number of problems, and a very refreshing view of the Arabic writing system: it certainly deserves attention.

4.1 Description

The goal of the BOUTEMENE project is to draw Arabic letters in such a way that they resemble as much as possible to lowercase Latin and Greek alphabet letters. In fact, whenever possible, real Latin and Greek letters have been taken, eventually rotated or inverted. This is by no means a transcription of Arabic; Latin and Greek letters have been used purely graphically.

The reader can find the complete list of BOUTEMENE types on fig. 4. On the first two rows we present the types of the original BOUTEMENE project; the third row has been added by the author, following the same design principles.

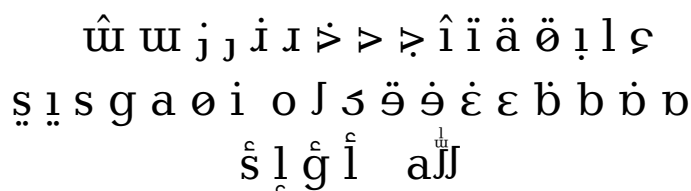


Fig. 4. Table of typographical types of the BOUTEMENE project

Probably the method used by Yahya Boutemène, was to take each one of the Arabic letters, and to try to figure out which Latin lowercase⁷ letter resembles it the most.

Some choices are very convincing: a 180 degrees rotated ‘m’ represents letter *seen*, and the same one with circumflex accent, letter *sheen*⁸; the Greek letter ε is used for *ain*, and a dotted ε for letter *ghain*; a 90 degrees counterclockwise rotated ‘v’ for letter *hah*; letter ‘b’ indeed looks close enough to a *tah*.

Other choices are rather unfortunate: letter ‘o’ has been chosen to represent the *meem*, but evidently the isolated *heh* looks much more like ‘o’ than *meem*; instead, ‘ø’ is used for *heh*: this may be a good choice for the initial form, but is less clear for the medial one. Letter ‘s’ has been chosen for the isolated *alef maksura*; it certainly is the most ‘logical’ choice, but still it is not very convincing graphically. To represent *sad* and *dad*, BOUTEMENE takes a letter ‘b’ without ascender; once again, it is logical: since ‘b’ with ascender denotes *tah*, it is logical that the same letter without ascender denotes *sad*; once again, the resemblance is not breathtaking; letter ‘a’ has been chosen for the final *heh*, here again we do not consider the resemblance strong enough . . .

The remaining choices are simple and logical: letter ‘l’ has been chosen for the *alef*, ‘i’ for the *noon* (it works fine for the initial and medial forms, but not as well for the isolated and final forms), and the different combinations of ‘i’ with dots and circumflex accent for *beh*, *teh*, *theh*, etc. A 180 degrees rotated ‘r’ has been used for the *dal*, a horizontally inverted dotted ‘e’ for the *feh* and derivatives, and finally a special form of letter ‘g’ has been used for the *waw*.

Certain letters had to be specially drawn: a letter ‘l’ with hook (similar to a long dotless ‘j’) for the *lam*, a very low dotless ‘j’ for the *reh*, and a symbol which looks like a headless ‘3’ digit, for the *kaf*. It is very interesting to note that the same symbol, rotated 180 degrees, has been used for the stand-alone *hamza*. On fig. 5 we compare the basic shapes of this system with letters taken from the modern Arabic Ω font.

Undoubtedly, the biggest handicap of BOUTEMENE is that Latin and Greek letters only resemble to *certain forms* of Arabic letters. Letter ‘i’ is a good choice for a medial or

⁷ Only lowercase letters have been used, because in the original BOUTEMENE project, a second table of types uses only uppercase letters. This table being less satisfactory than the lowercase one, we have decided to limit our discussion and implementation to the latter.

⁸ Linguists have observed the similarity between the Arabic س, the Hebrew ש and the Cyrillic ш. Some say that all of these letters derive from the same Egyptian hieroglyph 𓆎 having the form of three papyrus leaves (and the sound ‘sh’). If Boutemène had access to a Cyrillic typeface, he may have taken the Cyrillic ш instead of a rotated Latin ‘m’, identifying hereby letters with very very old common ancestors . . .

ḅ	ḥ	û	j	ṣ	ï	l	ʿ
ب	ح	و	ج	ص	ي	ل	ع
ṣ	g	ā	ö	o	ʃ	ʿ	ğ
ص	غ	ا	و	و	ك	ق	غ

Fig. 5. Comparison between the Ω modern Arabic typeface and BOUTEMENE

! aJəb ʃwi ! aJəb ʃwi
! aJəb ʃwi ! aJəb ʃwi

Fig. 6. The sentence ! اJəb ʃwi نسي طفله! written in four BOUTEMENE typefaces: Times, Palatino, Gill Sans Bold, Optima

initial *noon*, but certainly not for a isolated or final one. Of course, the usual “extreme cases” (the *heh* and the *yeh*) have been solved by adding new letters for different forms: ‘a’ for the final *heh*, ‘s’ for the final/isolated *yeh*. In particular, in this context there can be no notion of ‘tail’ (the graphical element which allows us to obtain a certain number of isolated/final forms, both in *ACADEMY* and *LAKHDAR*), simply because letters are totally disconnected.

4.2 Discussion

After having thoroughly criticized the BOUTEMENE system, let us mention its advantages:

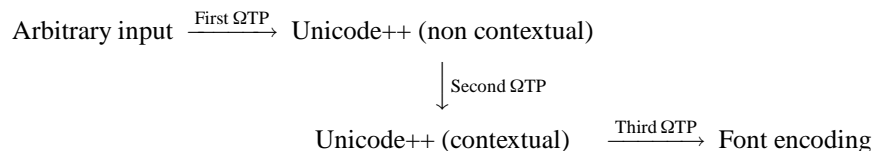
1. Arabic becomes a *member of the LGC (Latin-Greek-Cyrillic) family of alphabets*. All typographical styles can be applied (see fig. 6 for a sample of four well-known typefaces adapted to BOUTEMENE), all typefaces adapted. Arabic and LGC text can be mixed without additional problems.
2. BOUTEMENE is by far the most economic typesetting system: only 39 types are necessary (without counting vowels/diacritics, digits and punctuation), and these have very small widths, compared to other Arabic typefaces.

3. BOUTEMENE letters have the usual Latin descenders and ascenders so that they can share *exactly* the same baseline skip as LGC letters.

Are these advantages sufficient to justify the effort needed to get used to this new writing system? How steep is the learning curve? Which LGC typefaces are the best suited to be BOUTEMENE-ized? We hope that the availability of our public domain BOUTEMENE implementation will allow users from the Arabic reader community to test the system and provide answers to these questions.

5 The implementation

The Ω typesetting system (by John Plaice and the author) is an extension of $\text{T}_{\text{E}}\text{X}$ to 16-bit fonts and data. To typeset Arabic with Ω , one uses internal filters (called Ω TPs = Ω Translation Processes); these filters, written in Lex-like syntax, are dynamically loaded during compilation of the document by Ω .



The possibility of combining these filters has been used in the following way: a first filter reads the data input (for example, the text document on disk) and converts it to Unicode++ (a superset for the Unicode encoding, intended for typography). A second filter performs contextual analysis, sending Unicode++ Arabic characters to a certain number of tables in the higher range, taking into account the contextual form of each character. Finally, a third filter sends contextually analyzed Arabic to the font table in use.

Dividing the processing of Arabic into three steps proves to be extremely useful in cases such as the ones described in this paper: to adapt Ω to ACADEMY, LAKHDAR and BOUTEMENE we only had to rewrite the third filter: the mapping of already contextually analyzed characters to positions in the font table; this required only a minimal effort. This method also makes font encoding totally independent of input encoding: whether the input is in 7-bit transcribed Arabic, or ASMO, or ISO 8859, or any other encoding, only the first filter needs to change: once we are inside Unicode++, the input encoding is already forgotten by the procedure, and the third filter never finds out about it.

Ω TPs serve also for other transformations depending on the script or language, for example upper/lowercasing [which is irrelevant for Arabic, but can seriously corrupt 7-bit transcribed Arabic], applying punctuation conventions, etc. In the case of traditional Arabic, and additional Ω TP is applied (between the second and third filter) to introduce esthetic ligatures: having a separate filter for doing this allows dynamic activation/deactivation.

6 Conclusion

We have presented three systems of simplification of the Arabic writing system. There are other proposals which deserve to be studied and implemented⁹ and the author hopes to do this in a future publication. For the moment, the readers are invited to use the proposed implementations and give their feedback [weither positive or negative] so that these systems can be perfectionated (or completely abandoned, if considered useless).

All three typefaces are available as Type 1 PostScript fonts, together with OVF and OCP files, to be used with the Ω system by John Plaice and the author. These files can be retrieved by anonymous ftp from the following URL:

`ftp://ftp.ens.fr/pub/tex/yannis/arabic-simpli`

References

1. ١٩٧٩، القاهرة، المعارف، دار المعارف، المكتبة الخضراء للأطفال، أطفال الغابة، محمد عطية الإبراشي،
2. Y. Haralambous, *Typesetting the Holy Qur'an with TeX*, Proceedings of the 2nd International Conference on Multilingual Computing (Latin and Arabic script), Durham, 1992
3. Y. Haralambous, *The Traditional Arabic Typecase Extended to the Unicode Set of Glyphs*, Electronic Publishing **8** (2 & 3) 111-123 (1995)
4. Y. Haralambous and J. Plaice, *Multilingual Typesetting with Ω , a Case Study: Arabic*, Proceedings of the International Symposium on Multilingual Information Processing, 26-28 March 1997, Tsukuba, pp. 137-154
5. Y. Haralambous and J. Plaice, *The Design and Use of a Multiple-Alphabet Font with Ω* , submitted to RIDT'98.
6. A. Lakhdar-Ghazal, *Arabe Standard Voyellé — Code Arabe*, Institut d'Études et de Recherches pour l'Arabisation, Rabat, 1988
7. R. Meynet, *L'écriture arabe en question. Les projets de l'Académie de Langue Arabe du Caire de 1938 à 1968*, Dar el-Machreq Éditeurs, Beyrouth, 1971
8. N. Richert, *Arabisation et Technologie*, Institut d'Études et de Recherches pour l'Arabisation, Rabat, 1987
9. <http://www.ens.fr/omega>

⁹ In particular the proposal by Nasri Hattar, which combines the advantages of LAKHDAR and BOUTEMENE by using non-connected simplified Arabic letter shapes.

كان لاحد الملوك القدماء اخت تعيش معه فى قصره، بعد ان ماتت زوجته، وتركت له من الاولاد ثلاثة: اميرين واميرة. وقد ازداد حب الملك لاولاده، بعد وفاة والدتهم الملكة، واحبهم حبا كثيرا؛ ليعوضهم ما فقدوه من عطف امهم وحبها لهم، وتفكيرها فيهم؛ فكان يسال عنهم كلما حضر، ويفكر فيهم كلما دخل، ويوصى بهم كلما خرج، ويطلبهم كلما جلس لتناول طعام الافطار او الغداء او الشاى او العشاء.

محمة اخيها لاولاده، وصممت فيما بينها وبين نفسها ان تعمل سرا كل وسيلة ممكنة لابعادهم عن ابيهم والتخلص منهم.

وفى يوم من الايام كان الاميران يلعبان مع اختهما الاميرة فى حدائق القصر بعد خروج الملك، فشوقتهم عمتهم وحببت اليهم الذهاب معها الى الغابة للعب فيها، ووعدتهم ان تريحهم اشياء جميلة والعبا لذيدة سارة تحت الاشجار هناك.

فصدق الاميران والاميرة ما قالته عمتهم، ولم يعرفوا ما تخفيه عنهم من الشر، وذهبوا معها للعب والرياضة فى الغابة، ومشاهدة الاشياء الجميلة فيها، ورؤية الالعاب الغريبة تحت اشجارها.

Fig. 7. Text sample typeset in traditional Arabic typeface

كان لأحد الملوك القدماء أخت تعيش معه في قصره، بعد أن ماتت زوجته، وتركت له من الأولاد ثلاثة: أميرين وأميرة. وقد ازداد حب الملك لأولاده، بعد وفاة والدتهم الملكة، وأحبهم حبا كثيرا؛ ليعوضهم ما فقدوه من عطف أمهم وحبها لهم، وتفكيرها فيهم؛ فكان يسأل عنهم كلما حضر، ويفكر فيهم كلما دخل، ويوصي بهم كلما خرج، ويطلبهم كلما جلس لتناول طعام الإفطار أو الغداء أو الشاي أو العشاء. محبة أخيها لأولاده، وصممت فيما بينها وبين نفسها أن تعمل سرا كل وسيلة ممكنة لإبعادهم عن أبيهم والتخلص منهم.

وفي يوم من الأيام كان الأميران يلعبان مع أختهما الأميرة في حدائق القصر بعد خروج الملك، فشوقتهم عمتهم وحببت إليهم الذهاب معها إلى الغابة للعب فيها، ووعدتهم أن تريحهم أشياء جميلة وألعابا لذيذة سارة تحت الأشجار هناك.

فصدق الأميران والأميرة ما قالته عمتهم، ولم يعرفوا ما تخفيه عنهم من الشر، وذهبوا معها للعب والرياضة في الغابة، ومشاهدة الأشياء الجميلة فيها، ورؤية الألعاب الغريبة تحت أشجارها.

Fig. 8. Text sample typeset in ACADEMY

كان لأحد الملوك القدماء أخت تعيش معه في قصره،
بعد أن ماتت زوجته، وتركت له من الأولاد ثلاثة:
أميرين وأميرة. وقد ازداد حبّ الملك لأولاده، بعد وفاة
والدتهم الملكة، وأحبّهم حبّاً كثيراً؛ ليعوّضهم ما فقدوه
من عطف أمّهم وحبّها لهم، وتفكيرها فيهم؛ فكان
يسأل عنهم كلّما حضر، ويفكّر فيهم كلّما دخل، ويوصي
بهم كلّما خرج، ويطلبهم كلّما جلس لتناول طعام الإفطار
أو الغداء أو الشّاي أو العشاء.
محّمة أخيها لأولاده، وصمّمت فيما بينها وبين نفسها
أن تعمل سرّاً كلّ وسيلة ممكنة لإبعادهم عن أبيهم
والتخلّص منهم.
وفي يوم من الأيام كان الأميران يلعبان مع أختهما
الأميرة في حدائق القصر بعد خروج الملك، فشوّقتهم
عمّتهم وحبّبت إليهم الذهاب معها إلى الغابة للعب
فيها، ووعدتهم أن تريحهم أشياء جميلة وألعاباً لذيذة
سارّة تحت الأشجار هناك.
فصدّق الأميران والأميرة ما قالت عمّتهم، ولم يعرفوا ما
تخفيه عنهم من الشرّ، وذهبوا معها للعب والريّاضة في
الغابة، ومشاهدة الأشياء الجميلة فيها، ورؤية الألعاب
الغريبة تحت أشجارها.

Fig. 9. Text sample typeset in modern Arabic (Ω system)

كان لأحد الملوك القدماء أخت تعيش معه فى قصره، بعد أن ماتت زوجته، وتركت له من الأولاد ثلاثة: أميرين وأميرة. وقد ازداد حب الملك لأولاده، بعد وفاة والدتهم الملكة، وأحبهم حبا كثيرا؛ ليعوضهم ما فقدوه من عطف أمهم وحبها لهم، وتفكيرها فيهم؛ فكان يسأل عنهم كلما حضر، ويفكر فيهم كلما دخل، ويوصى بهم كلما خرج، ويطلبهم كلما جلس لتناول طعام الإفطار أو الغداء أو الشاي أو العشاء.

محنة أخيها لأولاده، وصممت فيما بينها وبين نفسها أن تعمل سرا كل وسيلة ممكنة لإبعادهم عن أبيهم والتخلص منهم.

وفى يوم من الأيام كان الأميران يلعبان مع أختهما الأميرة فى حدائق القصر بعد خروج الملك، فشوقتهم عمتهم وحببت إليهم الذهاب معها إلى الغابة للعب فيها، ووعدتهم أن تريحهم أشياء جميلة وألعابا لذيذة سارة تحت الأشجار هناك.

فصدق الأميران والأميرة ما قالتهم عمتهم، ولم يعرفوا ما تخفيه عنهم من الشر، وذهبوا معها للعب والرياضة فى الغابة، ومشاهدة الأشياء الجميلة فيها، وروية الألعاب الغريبة تحت أشجارها.

Fig. 10. Text sample typeset in LAKHDAR

ïïlo îl̄ ɹɛɪ̄ ,øɹp̄ə s̄ə aεo ũɹ̄ɛɪ̄ i>̄l̄ ɛloɹ̄ə̄l̄ ɹ̄glōl̄ ɹ>̄l̄ il̄s̄
 ɹ̄> ɹl̄ɹ̄l̄ ɹ̄ə̄ḡ .ə̄ɹ̄iōl̄ḡ iɹ̄ɹ̄iōl̄ :ǟil̄l̄̄ il̄lḡl̄l̄ io al̄ īɹ̄j̄iḡ ,āī>ḡj̄
 !l̄ɹ̄īs̄ l̄ɹ̄> oøɹ̄>l̄ḡ ,ǟs̄lōl̄ oø̄īl̄ḡ əl̄ə̄ḡ ɹɛɪ̄ ,ø̄ɹl̄lḡl̄ s̄lōl̄
 :oøɹ̄ə̄ l̄ə ɹ̄ɹ̄s̄ə̄iḡ ,oø̄l̄ l̄əɹ̄>ḡ oøōl̄ ə̄bε io ø̄ḡɹ̄ə̄ə̄ lo oø̄b̄ḡɛɹ̄l̄
 oøɹ̄ s̄ōḡɹ̄iḡ ,ɹ̄>ɹ̄ lōl̄s̄ oø̄ɹ̄ə̄ ɹ̄s̄ə̄iḡ ,ɹ̄b̄> lōl̄s̄ oø̄īε l̄̄ɹ̄ɹ̄ il̄s̄ə̄
 ḡl̄ ɛl̄ɹ̄ḡl̄ ḡl̄ ɹ̄l̄b̄ə̄l̄l̄ ol̄εb̄ l̄ḡlīil̄ ɹ̄l̄> lōl̄s̄ oø̄ɹ̄l̄b̄ɹ̄iḡ ,>ɹ̄> lōl̄s̄
 .ɛl̄ɹ̄ɹ̄ɹ̄l̄ ḡl̄ sl̄ɹ̄l̄l̄
 l̄ɹ̄ɹ̄ lōɛɪ̄ il̄ l̄əɹ̄ə̄ī iɹ̄iḡ l̄əīɹ̄ l̄oɹ̄ə̄ īoōōḡ ,ø̄ɹl̄lḡl̄ l̄əɹ̄>̄l̄ äo>o
 .oø̄io ɹ̄l̄>̄il̄lḡ oø̄īl̄̄ iε oø̄ɹ̄l̄ɛɹ̄l̄l̄ äīs̄oō ǟl̄ɹ̄ɹ̄ḡ l̄s̄
 ə̄s̄l̄ɹ̄> s̄ə ə̄ɹ̄iōl̄l̄ loø̄ī>̄l̄ εo il̄ɹ̄ɹ̄ɹ̄ il̄ɹ̄iōl̄l̄ il̄s̄ ol̄īl̄l̄ io ōḡɹ̄ s̄ə̄ḡ
 ɹ̄l̄əīl̄ oø̄ɹ̄l̄l̄ īɹ̄>ḡ oø̄īoε oø̄īə̄ḡɹ̄ɹ̄ə̄ ,s̄lōl̄ >ḡɹ̄> ɹɛɪ̄ ɹ̄ə̄l̄
 ǟl̄io> ɛl̄īɹ̄l̄̄ oø̄ɹ̄ɹ̄ il̄ oø̄īɹ̄εḡḡ ,l̄əɹ̄ə̄ ɹ̄l̄ɹ̄l̄ ǟl̄ɹ̄ḡl̄ sl̄̄ l̄əεo
 .s̄l̄io ɹ̄l̄>̄ɹ̄l̄l̄ i>̄ī ə̄ɹ̄l̄ɹ̄ ə̄ɹ̄īl̄ l̄l̄ɹ̄l̄ḡ
 aɹ̄ə̄>̄ī lo l̄ḡə̄ɹ̄ɹ̄ ol̄ḡ ,oø̄īoε āīl̄ə̄ lo ə̄ɹ̄iōl̄l̄ḡ il̄ɹ̄iōl̄l̄ ə̄ɹ̄ə̄
 ə̄ɹ̄ə̄l̄ɹ̄ōḡ ,ǟl̄ɹ̄ḡl̄̄ s̄ə ǟb̄l̄ɹ̄l̄ḡ ɹ̄l̄ɹ̄l̄ l̄əεo l̄ḡɹ̄ə̄iḡ ,ɹ̄ɹ̄l̄l̄ io oø̄īε
 .l̄əɹ̄l̄>̄ɹ̄l̄̄ i>̄ī ǟīɹ̄ḡl̄ ɹ̄l̄ɹ̄l̄l̄ ǟīḡɹ̄ḡ ,l̄əɹ̄ə̄ ǟl̄io>̄l̄ ɛl̄ɹ̄l̄l̄l̄

Fig. 11. Text sample typeset in BOUTEMENE