



HAL
open science

The Traditional Arabic Typecase Extended to the Unicode Set of Glyphs

Yannis Haralambous

► **To cite this version:**

Yannis Haralambous. The Traditional Arabic Typecase Extended to the Unicode Set of Glyphs. Electronic Publishing, 1994, 8 (2/3), pp.125-138. hal-02100480

HAL Id: hal-02100480

<https://hal.science/hal-02100480>

Submitted on 25 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Traditional Arabic Typecase Extended to the Unicode Set of Glyphs

YANNIS HARALAMBOUS

Atelier Fluxus Virus
187, rue Nationale, 59800 Lille, France
Fax (33) 20.40.28.64, Internet yannis@pobox.com

SUMMARY

Unicode, and other “universal” encodings have shown the need for adapting resources until now available only for major scripts, to the “ethnic” extensions of these scripts. In this paper we describe such an example: the adaptation of the traditional Arabic typecase to the needs of other languages using the Arabic script. We present an implementation of this extension : Al-Amal, based on T_EX, METAFONT, and Flex/Bison filters.

1 Introduction

The first Arabic book, a 5 × 11 cm volume titled *كتاب صلوات السواعي* (*Book of the prayer of hours*), was printed in 1514 by Grégoire de Grégoire in Venice and Fano, under the protection of pope Leo the 10th ([1, p. 18–19]). It took about two centuries for Arabic book printing to move to the East: in 1727 the Ottoman printing agency is founded in Constantinople and starts printing using Dutch types and technology ([8, p. 156]). A similar institution was founded in Cairo in 1821.

Undoubtedly a script like the Arabic one, having deep roots in calligraphy, was rather difficult to adapt to typography, a technique where strict standardization and repetition of forms is necessary. When Aldus Manutius created the first *italic* font in 1501, out of manuscript calligraphic forms, he did a certain number of choices—and these choices became a standard of occidental typography. Similar choices had to be made for Arabic: calligraphy had to be “tamed”, so that the result be homogeneous, reproducible, and flexible enough to be pleasant to the eye.

This standardization took place in 1906, in Cairo, when the *المطابع الأماري* (*Almaṭāb al’amārya*) typecase is defined. This typecase (see fig. 1 and 2), divided in four parts (as opposed to “upper” and “lower” case of the Occident), uses a total of 470 characters. Astonishing as it may seem, this typesetting system has been kept in use until today: books typeset in a traditional way, all around the Arabic world, are still using the same set of characters, and the same conventions and rules.¹ In fig. 1 and 2, the reader can see the four parts of this typecase.

The reader knowing the technical limitations of computerized typesetting can already imagine the effect of computers on the Arabic script: not being able to cope with the com-

¹ In [4, p. 102–103], a book published in 1880 (!) the reader can find 30 rules for typesetting Arabic, which are still *strictly* applied today by traditional typographers.

س	ش	س	ش	ب	ت	ث	س	ش	س	ش	ب	ت	ث
ص	ض	ص	ض	ب	ت	ث	ص	ض	ص	ض	ب	ت	ث
ع	غ	ع	غ	ج	ح	خ	ع	غ	ع	غ	ج	ح	خ
ف	ق	ف	ق	ح	ج	خ	ف	ق	ف	ق	ح	ج	خ
ل	ل	ل	ل	ل	ل	ل	ل	ل	ل	ل	ل	ل	ل
م	م	م	م	م	م	م	م	م	م	م	م	م	م
لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا
لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا

ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا
ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب
ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف
ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع
لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا
لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا
ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا	ا
ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب
ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف
ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع	ع
لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا
لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا	لا

Figure 1. The Cairo typecase—cases 1 and 2

منذ ظهور طبعة معجم بروكلان الثانية في عام ١٩٢٨ ، أصبح لدى دارسي اللغة السريانية اداة عمل ممتازة ، مقتبسة من النصوص ذاتها وعلى نحو يكاد يكون تاماً في كل ما له صلة بالنصوص المنشورة لغاية ذلك التاريخ . فضلاً عن ان صاحب المعجم يعتبر مرجعاً وحجة في جميع الأبواب المختصة بمصدر الكلمات واشتقاقها ، كما ان سعة اطلاعه وحصافته معترف بهما في مجالي التحريك وضبط الكتابة (الاملاء) . غير انه ليس من الميسور للمبتدئين او لتلامذة الاكليزيكيات ان يقفوا على كنوز مصنف ضخم ، مكتوب باللاتينية ومعتمد على الاشارات الاصطلاحية ومستند احياناً في ترتيبه الابددي الى مصدر خفي . فهذا ما حدانا على وضع معجم مختصر ، من النوع المدرسي ، آملين ان يجدوا فيه ما هم بحاجة اليه ، مع مراعاة الامانة قدر المستطاع .

ورق معدّ إعداداً خاصاً تطبع عليه الصور في غشاء يمكن نقله إلى سطح آخر لم يكن مستطاعاً إدخاله في ماكينات الطبع ، مثل الخشب أو الزجاج أو الخنزف أو الآلات . لتحضيره ، يغمس ورق ذو مسام في محلول من النشا والزلال والجلسرين ، وتطبع عليه الصور بعد جفافه ، ثم يطبع عدسة مرات بالحبر الأبيض غير الشفاف ، ثم يطلّى الوجه المطبوع بطبقة من الغراء القابل للذوبان في الماء . وعندما يبيل الورق ويلصق على السطح المطلوب النقل اليه ، تنزع الورقة المبللة ويثبت الطبع على ذلك السطح . وتطبع الصور المنقولة إلى الأواني الخزفية بأحبار معدنية ، ثم تدخل في الأفران لكي تصمد بعد ذلك للغسل بالماء الساخن .

حلايب مع التوتر الآخر الذي تشيره اتهامات مصر للسودان بدعم الجماعات المتطرفة. وقد ذكرت الدوائر الرسمية في القاهرة امس أن سلطات الامن المصرية احكمت قبضتها على كافة المنافذ الحدودية تحسباً لتسرب عناصر ارهابية او متطرفة من الدول المجاورة بعد ضبط شبكة تضم عناصر سودانية وايرانية وارمنية. ومن المتوقع ان تعلن النيابة العامة في مصر خلال ايام تفاصيل التحقيقات التي تجريها مع اعضاء الشبكة البالغ عددهم ٦٢ متهما. وقد اعتقلت اجهزة الامن المصرية مساء امس الاول ٦٤ من اعضاء الجماعات المتطرفة في مدينة قنا بعد صدامهم مع قوات الامن بأرمنت يوم الاربعاء

Figure 3. Samples of printed Arabic: Beirut 1963, Leipzig 1981, London 1992.

plexity of the Cairo typecase, the computer industry has tried (and was finally able) to impose new standards of simplified typesetting,² covering most of the time only the fundamental properties of Arabic script, without any typographical enhancement. Was it the computers, which have simplified Arabic printed script, or was it a deeper change in Arabic society and mentality? This is hard to say; nevertheless, even today, commercial computer typesetting systems are—a few isolated exceptions apart—unable to reach the typographic quality of 'Almaṭāb' al'amārya. In fig. 3 one can see different samples of printed Arabic material, showing the evolution and simplification of Arabic script; these examples are extreme cases: the first one is taken from a scholarly book printed in Lebanon (it contains almost all ligatures of the 'Almaṭāb' al'amārya typecase), the second from a technical book printed in East Germany (a fewer number of ligatures), and the third from a daily newspaper printed in the U.K. (almost no ligatures).

This paper describes the author's solution to this problem: الأمل (Al-Amal), a typesetting system based on T_EX (actually T_EX--X_EL^AT_E), emulating the 'Almaṭāb' al'amārya typecase.

² For more information on the Arabic script and the computer see also [3] and [10].

منذ طهور طبعة معجم بر و كلمان الثانية في عام ١٩٢٨، أصبح لدى دارسى اللغة السريانية اداة عمل ممتازة، مقتبسة من النصوص ذاتها وعلى نحو يكاد يكون تاماً في كل ما له صلة بالنصوص المنشورة لغاية ذلك التاريخ. فضلاً عن ان صاحب المعجم يعتبر مرجعاً وحجة في جميع الأبواب المختصة بمصدر الكلمات واشتقاقها، كما ان سعة اطلاعة وحصافته معترف بهما في مجالي التحريك وضبط الكتابة (الاملاء).

غير انه ليس من الميسور للمبتدئين او لتلامذة الاكلير يكتيات ان يقفوا على كوز مصنف ضخ، مكتوب باللاتينية ومعتمد على الاشارات الاصطلاحية ومستند احياناً في ترتيبه الابجدي الى مصدر خفي. فهذا ما حدانا على وضع معجم مختصر، من النوع المدرسى، املين ان يجدوا فيه ما هم بحاجة اليه، مع مراعاة الامانة قدر المستطاع.

Figure 4. Samples of text typeset with Al-Amal, with Cairo typecase ligatures

This system (already presented in [6] and [7]), has been recently extended to the complete set of Unicode Arabic alphabet characters; problems and open questions arising from this extension are discussed at the end of the paper.

2 The Cairo typecase

Arabic letters have contextual forms, depending on surrounding letters in the same word: a typical three letter word will start with a letter in initial form, followed by a letter in medial form and, finally, by a letter in final form (the hypothetical word consisting of three times the letter ‘ghayn’ is written غنغ). A fourth form is used for isolated letters (this is also the form used in crosswords or Scrabble-like games, where letters have to be placed in boxes, independently of their context). Some letters appear only in isolated and final form (and sometimes even only in isolated form), so that the letters immediately following them must be written in initial (or isolated) form, although they are located inside the word.

These are the basic contextual rules of the Arabic script: they are independent of style and medium, and are applied in all cases, without exception; they are as basic as the dot on the Latin lowercase ‘i’, or the horizontal bar of the ‘t’.

But besides these contextual forms, ‘Almaṭāb’ al’amārya also combines letters into *ligatures*, not unlike the ‘f’ + ‘i’ → ‘fi’ phenomenon in Latin alphabet typesetting. The rules are described in [4, p. 102–103] and have been confirmed by careful examination of various printed texts of different origin. In fig. 4 and 5 the reader can compare the same text (actually the text of the first example of fig. 3) typeset via the Al-Amal system, with and without ‘Almaṭāb’ al’amārya ligatures.

In the following we start by giving the mandatory ligatures (those that are part of every font), then we give the second level ligatures (those that are characteristic of the Cairo typecase), and finally we give the variant characters (form changes applied to single characters). By “foo-like”, where “foo” is some common Arabic letter, we mean all characters having

منذ ظهور طبعة معجم بر وكلمان الثانية في عام ١٩٢٨، أصبح لدى دارسي اللغة السريانية اداة عمل ممتازة، مقتبسة من النصوص ذاتها وعلى نحو يكاد يكون تاماً في كل ما له صلة بالنصوص المنشورة لغاية ذلك التاريخ. فضلاً عن ان صاحب المعجم يعتبر مرجعاً وحجة في جميع الأبواب المختصة بمصدر الكلمات واشتقاقها، كما ان سعة اطلاعه وحصافته معترف بهما في مجالي التحريك وضبط الكتابة (الاملاء).
غير انه ليس من الميسور للمبتدئين او لتلامذة الا كبير يكتيات ان يقفوا على كوز مصنف ضخم، مكتوب باللاتينية ومعتمد على الاشارات الاصطلاحية ومستند احياناً في ترتيبه الابجدي الى مصدر خفي. فهذا ما حدانا على وضع معجم مختصر، من النوع المدرسي، أملين ان يجدوا فيه ما هم بحاجة له، مع مراعاة الامانة قدر المستطاع.

Figure 5. Samples of text typeset with Al-Amal, without Cairo typecase ligatures

the same base form as letter “foo” but different dots and other diacritics (for example, the ba-like family is the set of characters ب, ت, ث, ط, etc.).

2.1 Mandatory ligatures

2.1.1 L1.

A lam-like letter followed by an alif-like letter: لا, جميلا, etc.

2.1.2 L2.

the second part of the word *Allah* (God): الله.

2.2 Typographical ligatures

2.2.1 L3.

A lam-like letter followed by meem: لم, علم, سامى, للاح, etc.

2.2.2 L4.

A ba-like letter followed by a ra-like one: ربيز, فتر, ثبر, etc.

2.2.3 L5.

A ba-like letter followed by a final noon-like one: فنن, فلسطين, فنن, etc.

2.2.4 L6.

A lam-like letter followed by a ha-like one:

2.2.5 L7.

A ba-like letter followed by a ya-like one: أسخواتى, سسكى, سسى, etc.

2.2.6 L8.

A gim-like letter followed by meem: سس, سس, etc.

2.2.7 L9.

A lam-like letter followed by a gim-like, and eventually a meem: لسس, لسس, لسس, etc.

2.2.8 L10.

A ba-like letter followed by a ha-like: سسلس, سسلس, etc.

2.2.9 L11.

A ba-like letter followed by meem: سسلسس, سسلسس, سسلسس, etc.

2.2.10 L12.

A ba-like letter followed by a gim-like one, and eventually a meem: سسلسلس, سسلسلس, سسلسلس, etc.

2.2.11 L13.

A lam-like letter followed by a ya-like one: سسلسلس, etc.

2.2.12 L14.

A kaf-like letter followed by an alif-like, or a lam-like, or a final kaf-like: سسلس, سسلس, سسلس, etc. Such a two-letter ligature can be extended to a three-letter or even four-letter ligature, by adding a ya-like letter, or a ha-like letter, or lam-like letter, or lam-alif-like ligature, etc.: سسلسلس, سسلسلس, etc.

2.2.13 L15.

A ba-like letter, followed by a meem and a gim-like letter:

2.2.14 L16.

A lam-like letter followed by a lam-like letter and eventually by a meem or a gim-like letter:

2.2.15 L17.

A kaf-like letter followed by a meem and eventually other letters: لكمة, لكم, كم, لكان, كك, etc.

2.2.16 L18.

A meem followed by a gim-like letter and eventually a meem: مجد, مع, محل, etc.

2.2.17 L19.

A šad-like, ha-like, fa-like or kaf-like letter followed by a gim-like one: فخر, فحل, فصح, فصحى, فخر, etc.

2.2.18 L20.

A ba-like, or lam-like, followed by meem, or a meem followed by a ba-like, followed by meem, or a lam-like followed by two meems:

2.2.19 L21.

A sin-like, or šad-like, or fa-like, or ayn-like, or gim-like, followed by a gim-like and eventually by a meem: جحفل, جحجج, ججلان, ججج, شخ, صحفى, صبح, عجل, عج, عجر, عجمى, عجمة, etc.

2.2.20 L22.

A lam-like letter or lam-meem-like ligature, followed by a gim-like letter, and eventually a meem: لجم, لجم, لجم, لجم, etc.

2.2.21 L23.

The name “Mohammad” محمد

2.3 Variant forms

2.3.1 V1.

An initial ba-like letter in front of a sin-like, šad-like, ayin-like, waw-like or ha-like one grows higher: بسملة, تعليق, ثورة, نهمرة, يونان, etc.

2.3.2 V2.

a medial ba-like letter between two ba-like letters, or in front of a sin-like letter grows higher: متيسر, تقييش, تستيت, بوييس, ثبت, تتبع

2.3.3 V3.

An initial or medial gim-like letter in front of an alif-like or lam-like letter takes a rounder closed form: جاي, حاجت, جليل, etc.

2.3.4 V4.

An initial meem in front of a ra-like letter, a ha-like letter or a ya-like letter gets smaller and non-hollow: امي, مهل, مربع, etc.

2.3.5 V5.

A ra-like letter following a gim-like, ta-like, ayn-like, fa-like, kaf-like, ha-like letter or a meem, takes a more calligraphic form: گريه, كريم, قزمية, غزل, عربي, ظرف, طرطر, حرفي, هرع, etc.

3 Porting the Cairo case to Unicode

The first plane of ISO 10646-1, also known as Unicode, provides characters for the following languages: Arabic (modern and classical), Farsi, Urdu, Pashto, Sindhi, Ottoman Turkish, Baluchi, Kashmiri, Kazakh, Lahnda, Dargwa, Uighur, Turkic, Berber, Hausa, Malay, Adighe, Ingush, Kirghiz ([12]).³

Similarly to European languages which have diacritized letters of the Latin alphabet to adapt them to their phonetic needs, the languages stated in the previous paragraph have added diacritics to the letters of the basic Arabic alphabet. There is a slight difference though: historically, Arabic alphabet was first written without dots;⁴ so in a sense, dots are already “diacritics”. It is only natural that these languages have first tried to use new combinations of dots and letter forms: almost every combination of basic form and sets of one, two, three, or even four dots, over or under the word has been used to obtain new characters.

The author has expanded the Al-Amal system to cover all these characters derived from the basic Arabic script; in Fig. 6 the reader can see an example of Sindhi text (kindly provided to the author by) typeset in Al-Amal. In most of the cases, the extension to Unicode has been a straightforward task. Nevertheless, in some cases the fact of applying a ligature or even just a contextual form similar to those of the basic Arabic alphabet brought up ambiguities. These will be discussed below.

³ This set of characters is quite complete; nevertheless, the author encountered characters not provided in Unicode, in four cases: for typesetting the Qur’ān, a ba-like letter without dot is needed [2, p. 102–103] (one new character), for typesetting old manuscripts, all characters are needed without dots (2 new characters, in ba-like and qaf-like forms), Salem Chaker’s proposal for the transcription of Berber into Arabic script [5] (one new character), and Ahmed Lakhdar’s proposal for the writing of African languages [9] (7 new characters and 6 new diacritics).

⁴ Take for example letters ب (‘b’), ت (‘t’), ث (‘th’ like in ‘thought’); they differ only by the number and position of dots. Originally, these letters were all written without dots, and the reader had to guess their pronunciation from the context(!).

رڳو ڳالهيون ڪندي ۽ نمرن هٽندي اسان جي قوم چاليهه سال پيڙائون ۽ عذاب پوڳيا آهن ۽ انهن نمرن اسان جي قوم لاءِ وڌيڪ پيڙائون ۽ عذاب نازل ڪيا آهن. جيڪڏهن اسان ۾ اڄ قوم جي اميد پيدا ٿي آهي ته اها اسان جي عمل ۽ اسان جي بي لوٽ جدوجهد جي ڪري پيدا ٿي آهي ۽ ماڻهو اسان ڏانهن واجهائي رهيا آهن. ته اسان ئي آهيون جيڪي ڪجهه نه ڪجهه ڪندا سون.

پر اسان ڪا ڏسڻو آهي ته دنيا جي اندر ڇا ٿي رهيو آهي ۽ اسان جو دشمن ڪيئن حالتن ڪا پنهنجن مفادن ۾ ڪتب آڻڻ جي ڪوشش ڪري رهيو آهي، انهيءَ جي لاءِ ضروري آهي ته اسان پاڻ ۾ ڏاهپ پيدا ڪريون ۽ پاڻ ۾ ڄاڻ جو هڪ وسيع خزانو پيدا ڪريون.

۽ اسان موجوده صورت حال ڪا سمجهڻ لاءِ روزمره جي ميڊيا ۽ دنيا جي اندر ٿيندڙ ڪاروائين تي گهري نظر رکون ته دشمن ۽ جارحيت پسند قوتون ۽ اسان تي قابض قوتون دنيا جي اندر ٿيندڙ تبديلين ڪا سندن حق ۾ ۽ سندن مفادن جي حق ۾، سنڌ تي جارحيت قائم رکڻ جي حق ۾، سنڌ ڪا مستقل قبضي ۾ ڪرڻ جي حق ۾ ڪيئن ڪتب آڻي رهيون آهن.

Figure 6. Sindhi text typeset in Al-Amal

3.1 Cases where contextuality leads to confusion between characters

3.1.1 Letters fa and qaf.

In basic Arabic, letters fa ف (and its artificial derivative va و) and qaf ق have different forms: the former is longer and flatter, while the latter is rounder and deeper. This difference is visible only in the isolated and final forms: compare ف ففف and ق ققق. Since these letters differ mainly in the number of dots (one for fa, two for qaf, three for va), the shape difference is of minor importance, and in some modern Arabic typefaces it is totally ignored.

The problems arise with Unicode characters 06A7 (ARABIC LETTER QAF WITH DOT ABOVE) and 06A8 (ARABIC LETTER QAF WITH THREE DOTS ABOVE), which use the basic shape of letter qaf, and have the same number of dots as fa and va. These characters are used in Maghribi Arabic. In initial and medial forms, as well as in ligatures involving these forms, they are indistinguishable from the basic Arabic letters fa and va.

3.1.2 Letters ta, noon and ya.

In basic Arabic, letters ta ت and noon ن have different forms: the former is longer and flatter and the latter is rounder and deeper. Once again, the difference can only be seen in isolated and initial forms: compare ت تت and ن نن. Since these letters differ mainly in the number of dots (one above for noon, two above for ta, etc.) the shape difference is of minor importance.

Unicode characters 06BB (ARABIC LETTER RNOON), 06BD (ARABIC LETTER NOON WITH THREE DOTS ABOVE) use the letter form of the Arabic noon and the dots of the Urdu letter tteh and the Arabic letter tha. These letters are used in Sindhi and Malay. Their initial and medial forms, as well as all ligatures involving initial and medial forms are indistinguishable from the Urdu and Arabic counterparts.

The situation is even more complicated since the Arabic letter ya ي shares the same initial and medial forms as ba, noon and friends: ي ييي. Nevertheless, the isolated and final forms of this letter are significantly different from those of the ba and noon letter shapes. Once again, in basic Arabic the number and position of dots is sufficient for determining the letter (ya carries two horizontally aligned dots below).

Unicode character 067B (ARABIC LETTER BEEH) has the form of ba and carries two vertically aligned dots below; this is also the case of 06D0 (ARABIC LETTER E) which carries the same set of dots, but has the form of an Arabic ya. Furthermore, 06D1 (ARABIC LETTER YEH WITH THREE DOTS BELOW) carries three dots below, exactly as does Arabic letter tha: the former has the letter form of a ya, while the latter the one of a ba.

3.1.3 Arabic and Sindhi letters kaf.

In Arabic, the letter kaf is written with an oblique ascender stroke in initial and medial form, and with a hamza-like diacritic in final and isolated form. Sindhi uses a kaf-like letter 06A9 (ARABIC LETTER KEHEH) which has oblique ascender strokes in all forms and no hamza-like diacritic. This letter is indistinguishable from the Arabic kaf, in initial and medial forms, as well as in all ligatures involving these forms.

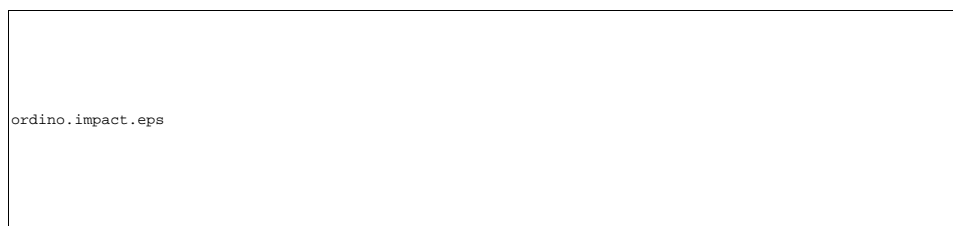


Figure 7. The Al-Amal internal structure

3.2 Cases where ligatures obstruct proper diacritization of characters

1. The Pashto ring (as in ټ) is incompatible with the ba-like + gim-like ligature (for example ټځ) and the initial/isolated ba-like + meem ligature (for example ټم). Either the ring must be designed like “a hanging drop”—a dubious esthetic result—or the ligature must be broken. The author has tried to design a ligature of isolated form ټم , but the result is not entirely satisfying.

2. The Uighur character ۋ (ARABIC LETTER HIGH HAMZA ALEF) can hardly take part in a lam-alef-like ligature: the hamza would be too far to the right.⁵

3. The fact that ‘*Almaṭāb*’ *al’amārya* ligatures have been designed without taking into account Indic characters, makes many ligatures with non-standard dots ambiguous: is ٲٳ the combination of ٲ and ٳ (06A5 ARABIC LETTER FEH WITH THREE DOTS BELOW) and the standard Arabic ḥah or of ٲ and ٳ (06A1 ARABIC LETTER DOTLESS FEH and 0686 ARABIC LETTER TCHEH)? Theoretically, one can distinguish them by slightly moving the dots to the right in the former case (ٲٳ vs. ٲٳ); but still the two forms are very close graphically, and it may be difficult to the reader to distinguish them at first sight.

4 Technical details

4.1 Preprocessing

The extended Al-Amal system is build of four modules, as shown in diagram 7:

1. re-encoding to the (extended) Unicode encoding;
2. standard contextual analysis and mandatory ligatures;
3. (optional) Cairo typecase ligatures processing;
4. preparation to output (conversion into $\text{T}_{\text{E}}\text{X}$ code).

The first three modules are independent of $\text{T}_{\text{E}}\text{X}$. To avoid ligatures one simply removes module 3 from the processing chain. Preprocessors have been written in C, using Flex and Bison tools: writing a grammar for Arabic ligatures avoids tedious pattern matching.

4.2 The fonts

The Al-Amal have been designed in the METAFONT language, to obtain maximum possibilities of optical scaling. Many ligatures have been splitted in several parts and are re-

⁵ Not to mention the fact that in the Qur’ān one finds a lam-alef ligature with a central hamza, not included in Unicode.

combined by TEX (this is also the task of preprocessor module 4). One can consider these fonts as *glyph banks*, providing glyphs which TEX combines into characters and ligatures. This approach has allowed minimization of storage space and time/energy needed for font design. The author was able to produce all possible Cairo typecase ligatures on the Unicode Arabic character set, using only six 8-bit (partially filled) font tables, that is less than 1500 glyphs.

REFERENCES

1. Josée Balagna, *L'imprimerie arabe en occident*, Maisonneuve & Larose, Paris 1984.
2. Syed Barakat Ahmad, *Introduction to Qur'anic Script*, Curzon Press, London, 1985.
3. Joseph D. Becker, *Arabic Word Processing*, Communications of the ACM, **30** (7), 1987.
4. Claus Faulmann, *Das Buch der Schrift, enthaltend die Schriftzeichen und Alphabete aller Zeiten und aller Völker des Erdkreises*, Vienna, 1880 (reprint by Franz Greno, Nördlingen, 1985).
5. Yannis Haralambous, *Un système TEX berbère*, Études et documents berbères, **11** 43–54, 1991.
6. Yannis Haralambous, *Towards the revival of traditional Arabic typography... through TEX* , Proceedings of the Euro TEX 92 conference, Prague, 1992
7. Yannis Haralambous, *Typesetting the Holy Qur'ān with TEX* , Proceedings of the 2nd International Conference on Multilingual Computing (Latin and Arabic script), Durham, 1992.
8. Klaus Lagally, *Arab TEX — Typesetting Arabic with Vowels and Ligatures*, Proceedings of the Euro TEX 92 conference, Prague, 1992
9. Ahmed Lakhdar-Ghazal, *Caractères arabes diacritiques selon l'ASV-CODAR (pour imprimer les langues arabes)*, Institut d'Études et de Recherches pour l'Arabisation, Rabat, 1993.
10. Pierre MacKay, *Typesetting problem scripts*, BYTE 11, **2** 1986, 201–218.
11. Roland Meynet, *L'écriture arabe en question*, Dar el-Maghreb Éditeurs, Beyrouth, 1971.
12. The Unicode Consortium, *The Unicode Standard*, Version 1.0, Vol. 1, Addison-Wesley, 1991.