User’s Guide of ض (dād),
a Simple Arabic Typesetting System
for Mixed Latin/Arabic Alphabet Documents

Version 1.1/2014-10-11

Yannis Haralambous∗

ض is a package for typesetting Arabic in the simplest possible way. It is particularly well suited for mixed Arabic/Latin documents. “Simplest possible” means:

- it is compatible with all LATEX style files, since the code is minimal and all the complexity is in the font;
- input can be done in Unicode or in transliteration, the latter being often the best choice when mixing left-to-right and right-to-left scripts;
- the only TExnical requirement is LuaTEx, not because of the Lua language (which is not used, for the moment), but because of features that have survived from LuaTEx’s Ω origins: bidirectionality and use of large fonts (OVF, OFM).

Choose LuaTEx as your TEx engine, load the package into your document, and اهلاً شا، just start writing in Arabic using command \arab.

More information about ض (history, evolution, rationale of technical choices, TExnicalities) can be found in [1].

1 The name

Thanks to the Internet, search engines, social media, and the like, people are becoming more and more aware of other languages and writing systems. Why not give this package an Arabic name, be it a single letter?

The author has chosen letter ض, called dād, because Arabic is traditionally called the “language of the dād,” since this sound was considered as being unique to Arabic.

The reader is probably wondering how to pronounce this letter, technically a “voiced velarized alveolar stop” [3, p. 16]. Here is how [4, p. 10] describes its pronunciation:

Pronounce the regular sound ‘d’ and you will find that the tip of your tongue will touch in the region of the upper front teeth/gum. Now pronounce the sound

∗Institut Mines Téléc, Télécom Bretagne, UMR CNRS 6285 Lab-STICCC, Technopôle Brest Iroise CS 83818, 29238 Brest Cedex 3, France, yannis.haralambous@telecom-bretagne.eu
again and at the same time depress the middle of the tongue. This has the effect of creating a larger space between the tongue and the roof of the mouth and gives the sound produced a distinctive ‘hollow’ characteristic, which also effects the surrounding vowels. It is difficult to find a parallel in English, but the difference between ‘Sam’ and ‘psalm’ (standard English pronunciation) gives a clue. Tense the tongue muscles in pronouncing ‘psalm’ and you are nearly there. Now pronounce the a-vowel of ‘psalm’ before and after ‘d’, saying ‘aḍa’, keeping the tongue tense, and that’s as near as we can get to describing it in print.

2 How to use

The package provides three PostScript Type 1 fonts (plain, bold and typewriter), “real” fonts (regular TFM) and large virtual fonts (OVF and OFM files). There are also rudimentary FD and STY files, a MAP file, Perl scripts for conversion to (and from) UTF-8, the Perl script which builds the font and finally adjustment files, in case the user wants to change kerning and diacritic placement.

Once the package is installed, to use it just call

\usepackage{dad}

Notice however that it requires Lua\TeX{} (for change of direction and OVF/OFM compliance).

To typeset in Arabic, one uses the command \textbackslash arab (which is “long”: paragraph changes are allowed in its argument).

Arabic text can be input in transliteration, as described in Table 1 or in Unicode UTF-8 (§2.2).

For example, to obtain الكُتاب one would write in transliteration \textbackslash arab\{AlkitAb\} or in Unicode \textbackslash arab\{الكتاب\}. By writing \textbackslash arab\texttt{tt\{AlkitAb\}} one obtains the typewriter version الكتاب (which is less appealing, but fits quite nicely with the Computer Modern Typewriter font).

2.1 Rationale of the transliteration

Here are the rules of the proposed transliteration:

1. pharyngeal خ = H, emphatic ص = S, ض = D, ط = T, ظ = Z and velar غ = R are uppercased—do not confuse them with glottal ح = h, non-emphatic س = s, د = d, ت = t, ز = z, and alveolar ر = r;
2. long vowels (أ = A, و = U, ي = Y) and ُلِف maṣūra (ى = I) are also uppercased;
3. some consonants are modified by adding a character ح (ذ = dh, ث = th, ش = sh);
4. the stand-alone hamza is obtained by a vertical bar | and letter ayn by a grave accent (which, in legacy \TeX{} produces an inverted curly apostrophe, which is sometimes used to transliterate this letter);
5. to avoid confusion between pairs of letters and letters obtained by digraphs, one has to use a dash to separate characters: compare س-ح and ش = sh, or ت-ح and ث = th;
| \( \text{ض} \) | \( \text{ش} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ط} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) | \( \text{ق} \) |
| \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) | \( \text{ض} \) |

6. more generally, the dash plays the rôle of zero-width joiner\(^1\): when writing \( \text{ض} = -b \), the letter \( \text{ض} \) will be in final form; \( \text{ض} = b- \) and \( \text{ض} = -b- \) will produce initial and middle letters, provided of course the letter is quadriform (as is letter \( \text{ض} \) in this example). This is very useful when describing grammar rules, to signify that a letter (or letter group) is an affix;  

7. the dash can also be used to reestablish contextual forms when combined with \TeX\ commands, for example, to colorize letters. There is only one special case: when we want to colorize a letter of an isolated ligature \( \text{ض} \), instead of a dash, we use digit 4. For the final ligature \( \text{ض} \) it will be a digit 5. Example: to colorize the \( \text{ض} \) of \( \text{ض} \), write

\begin{verbatim}
\arab{t-\textcolor{red}{-l5-}-A5% \textcolor{red}{l4-}-A4}
\end{verbatim}

8. finally, there is yet another use of the dash: when doubled, it produces a kashida stroke: compare \( \text{ض} = 1Y1 \) and \( \text{ض} = 1--Y--1 \). There is also a \kesh\ command for extensible kashida (it is equivalent to a \hrulefill\ using the default rule thickness font dimension \fontdimen8\); \( 1--\kesh--Y--\kesh--1 \). will produce:

\( \text{ض} \)

---

\(^1\)Except for the case of letter \( \text{ض} = \text{ض} \) which is biform and hence is not connected with the following letter. By writing \( \text{ض} = d-h \) one obtains letters \( \text{ض} \) and \( \text{ض} \), but the \( \text{ض} \) is not in medial form, as it would be in any other case when preceded by a dash.
9. some digraphs start with an apostrophe: it is the case of hamza-carriers ی = 'a, ی = 'i, ی = 'u, ی = 'I, ی = 'A but also of undotted letters ی = 'b, ی = 'n, ی = 'f and ی = 'q;

10. other digraphs end with one or more asterisks: the most frequent one is the ی marbuṭa ی = t* (which can be used also in initial and medial, and then becomes a regular ی).

11. there is a special transcription for the ligature لله لـ used for the noun of majesty,” which is the name of God لله in this case—and in this case only—an uppercase ل is used. The reason is that we wish to avoid ambiguity with other uses of the trigram لـ لـ، for example (Koran 6:39) where we encounter letters لله but not with the meaning “God.” Contrarily to other systems, the لله ligature is available also in final form (for م which occurs six times in the Koran, for example Koran 6:149), and it is possible to add diacritics to its first glyph (as in ﯾ، Koran 2:115 or لله، Koran 2:165).

2.2 Unicode input

Input can be transliterated or provided directly in Unicode Arabic: \arab{YAnis} or \arab{بانيس} or even \arab{nis} or \arab{YA} will produce the same result: ﯾ.

All cells of Table 1 can be obtained by the corresponding Unicode characters (mostly via a single character, except for ﯾ + vowel combinations which require two characters). There is a special case, though: the لله ligature (see next section).

For the convenience of the user who wants to write kashida (so that Arabic input is not disrupted) we have defined a command (in Arabic characters) ﯾ ﯾ are the first two letters of تط، the Arabic name of kashida) which is exactly equivalent to ﯾ and has to be placed between Unicode U+0640 arabic tatwell characters.

2.2.1 The لله ligature and Unicode

The لله ligature is traditionally used for writing the name of God: لله. It can be found in religious texts, but also in expressions (for example, إن شاء الله which means “hopefully” appears even in French language as incheilla and in Portuguese as oxalá) and in the very common surname ﯾ. Abdallah.

The problem with this ligature is that it contains a rather rare diacritic (a ﯾ combined with a vertical ﯾ—the latter us available on Apple Arabic keyboard layout but not on the Microsoft one) and, as a convenience, most standard fonts will replace the character string لـ لــ (which would normally look like لله) by the complete ligature لله (in other words: the font not only changes the glyphs but, at the same time, also adds the diacritics). This behavior is barely legitimate: a ligature (as in ‘f’ or ‘z’) is normally limited to a change of glyphs, and should not add new characters (in this case, characters U+0651 ARABIC SHADDA and U+0671 ARABIC LETTER SUPERSCRIPT ALEF) since this means that what is rendered does not correspond anymore to the underlying Unicode character string.
رباعيات الخيم

سمعت صوتًا هائلاً في السحّر ننادي إلى الغيب، رفاته السفيرة
هيوا أملاً كأس السلك، قبل أن تملأ كأس العمر كف القلّ.
لا تشتعل النيل بمضات الرزمان ولا يبق في البصر قبّل الأوان.
وأغنى من الخامير لذائه، فليس في مطهو السيف الآمان.
وعند رفع الغب واليوم ليس كتم مخبأ الفضّن فسأ القبّل،
وليسه بالغافل، حتى أرى جمالي دنيام وسأجتدي
القلب قد أضراه عشق الجمال والصدر قد ضاق بما لا يقال
يا رب هل بريضك هذا الفضّن، ولصعى بالنسبة لأعمال زوال.
أولى بهذا القلب أن يخفقنا وفريق الحبّ أن يحترقنا
ما يضيق اليوم الذي مسر بي من غبر أن أهوى وان أعلقنا
أرف خفي الفضّن هذا السحّر ننادي إلى النوم واليومن الوتر،
فما أطلال النوم عصرا ولا قصر من الأعماز طول السهر.
فكم تولي الليل بعد النهار وطال بالانجح هذا المدار.
فمَّش الهوئية أن هذا الشهّر مين عيّن ساحرة الآحوار.
لا تهوي النفس يخفى الفضّن وأغنى من الحامير أمر القيّم.
فقد تضاوّل في الأثر راحب الله عماض من الفوقيّين.
إطعنا لقي القلب بشهد الرضوان فإيانا الأبام. مثل السحاب،
وعيشنا طيفي هبل قيل حظك منه قبل فوت الشبيبة.
ليست ثوب العيش لم استشر وحرى فيه بين شتى القكر.
وبسوف أنضو الثوب عن وسم أدرك لميّاً جهّتي أيسن المغر.
يا من ضيّق الفهم في قناتك وتطسب النفس بمتي طاعتك.
أشركتي الأزهار وكتبني قصصي بالفساد في نوحين.
إن لم أكن أُعلمن في طاعتك فانّلي أطمّع في رحبنك.
وإنما يشفع لي باتني قد علمت لا أشرك في وحيدك.
تخفى عن الناس سنت طاعتك وانستي أطمّع فسي رحماك.
فانت bí jحلا وانت الذي تريد يطيب الصقر في آرك
أن تقبل القطرة من يردها ففي مماها منهني أمراً
تقرّبت يا رب ما بينا مسافة البعيد على قدرها
يا عالم الأسرار علم البلقيين يا كأشف الضمر عن الناسبيين.
يا قابل الأذار عدننا إلى ظلل فاقتب توبة الثانيين.

Figure 1: The lyrics of the song (Oum Kalthoum, 1950) [2]
Figure 2: TeX code of Fig. 1, transliterated input
Figure 3: TeX code of Fig. 1, Unicode input
Nevertheless, for the user’s convenience, we have adopted that behavior also in ضٓ ضٓ ضٓ كٓ كٓ كٓ ٔ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ كٓ K