

# 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 L<sup>A</sup>T<sub>E</sub>X 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 T<sub>E</sub>Xnical requirement is LuaT<sub>E</sub>X, not because of the Lua language (which is not used, for the moment), but because of features that have survived from LuaT<sub>E</sub>X's Ω origins: bidirectionality and use of large fonts (OVF, OFM).

Choose LuaT<sub>E</sub>X as your T<sub>E</sub>X engine, load the package into your document, and اهلاً شلاً! just start writing in Arabic using command `\arab`.

More information about ض (history, evolution, rationale of technical choices, T<sub>E</sub>Xnicalities) 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 Télécom, Télécom Bretagne, UMR CNRS 6285 Lab-STICC, Technopôle Brest Iroise CS 83818, 29238 Brest Cedex 3, France, [yannis.haralambous@telecom-bretagne.eu](mailto: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 ض

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 LuaTeX (for change of direction and OVF/OFM compliance).

To typeset in Arabic, one uses the command `\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 `الكتاب` one would write in transliteration `\arab{AlkitAb}` or in Unicode `\arab{الكتاب}`. By writing `\arabtt{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 ح = H, emphatic ص = S, ض = D, ط = T, ظ = Z and velar غ = R are *uppercased*—do not confuse them with glottal ه = h, non-emphatic س = s, د = d, ت = t, ز = z, and alveolar ر = r;
2. long vowels (ا = A, و = U, ي = Y) and *ʾalif maqṣūra* (ى = I) are also *uppercased*;
3. some consonants are modified by adding a character h (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 سه = s-h and ش = sh, or ته = t-h and ث = th;



9. some digraphs start with an apostrophe: it is the case of *hamza*-carriers  $\acute{a}$  = 'a,  $\acute{i}$  = 'i,  $\acute{u}$  = 'u,  $\acute{I}$  = 'I,  $\acute{A}$  = 'A but also of undotted letters  $b\bar{a}$ '  $\text{ب}$  = 'b,  $n\bar{a}n$   $\text{ن}$  = 'n,  $f\bar{a}$ '  $\text{ف}$  = 'f and  $q\bar{a}f$   $\text{ق}$  = 'q;
10. other digraphs end with one or more asterisks: the most frequent one is the  $t\bar{a}$ ' *marbuta*  $\text{ة}$  =  $t^*$  (which can be used also in initial and medial, and then becomes a regular  $t\bar{a}$ '). The asterisk is also used for the *waṣla* (which is only placed on the *ʾalif*)  $\acute{a}$  =  $A^*$  as well as for the vertical *fatha* (as in  $\text{هَذَا}$  =  $ha^*dhA$ ) and the *madda*. The latter is normally used only on the *ʾalif* ( $\acute{a}$  = 'A) but can be found also in the notorious *muqattaʿat* in the Koran, as in  $\text{عَسَقَ}$  (*Koran* 42:2) or  $\text{كَهَيْتَصَرَ}$  (*Koran* 19:1)—sometimes it is even combined with a *šadda* (as in  $\text{الْمَصَّر}$ , *Koran* 7:1 and [5, p. 111] for the *šadda*);
11. there is a special transcription for the ligature  $\text{لله}$  = LLh used for the *اسم الجلالة* “noun of majesty,” which is the name of God  $\text{الله}$ : in this case—and in this case only—an uppercase L is used. The reason is that we wish to avoid ambiguity with other uses of the trigram  $\text{lām-lām-hā}$ ʾ, for example  $\text{يُضِلُّهُ}$  (*Koran* 6:39) where we encounter letters  $\text{له}$  but not with the meaning “God.” Contrarily to other systems, the  $\text{لله}$  ligature is available also in final form (for  $\text{فَلَيْلَهُ}$  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  $\text{وَلَيْلَهُ}$ , *Koran* 2:115 or  $\text{لَيْلَهُ}$ , *Koran* 2:165).

## 2.2 Unicode input

Input can be transliterated or provided directly in Unicode Arabic: `\arab{YAnis}` or `\arab{يانيس}` or even `\arab{يا nis}` or `\arab{YA نيس}` will produce the same result:  $\text{يانيس}$ .

All cells of Table 1 can be obtained by the corresponding Unicode characters (mostly via a single character, except for *šadda* + vowel combinations which require two characters). There is a special case, though: the  $\text{لله}$  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)  $\text{طط}$  (طط are the first two letters of  $\text{تطويل}$  = *tatwyl*, the Arabic name of kashida) which is exactly equivalent to `\kesh` and has to be placed between Unicode U+0640 ARABIC TATWELL characters.

### 2.2.1 The $\text{لله}$ ligature and Unicode

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

The problem with this ligature is that it contains a rather rare diacritic (a *šadda* combined with a vertical *fatha*—the latter is available on Apple Arabic keyboard layout but not on the Microsoft one) and, as a convenience, most standard fonts will replace the character string  $\text{lām-lām-hā}$ ʾ (which would normally look like  $\text{له}$ ) by the complete ligature  $\text{لله}$  (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 ‘fi’ or ‘لا’) 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 الخيام

سمعت صوتا هاتفا في السحر نادى من الغيب رفات البشر  
هبوا املأوا كأس المنى قبل أن تملأ كأس العمر كف القدر  
لا تشغل البال بماضي الزمان ولا بآت العيش قبل الأوان  
وأغنم من الحاضر لذاته فليس في طبع الليالي الأمان  
غد بظهر الغيب واليوم لسي وكم يخيب الظن في المقبل  
ولست بالغافل حتى أرى جمال دنياي ولا اجتلي  
القلب قد أضناه عشق الجمال والصدر قد ضاق بما لا يقال  
يا رب هل يرضيك هذا الظلما والماء ينساب أمامي زلال  
أولى بهذا القلب أن يخفقا و في غرام الحسب أن يحترقا  
ما أضيع اليوم الذي مر بي من غير أن أهوى و أن أعشقا  
أفق خفيف الظل هذا السحر نادى دع النوم وناغ الوتر  
فما أطال النوم عمرا ولا قصر من الأعمار طول السهر  
فكم تولى الليل بعد النهار وطال بالأنجم هذا المدار  
فأمش الهويبة أن هذا الثرى من أعين ساحرة الاحرار  
لا توحش النفس بخوف الظنون وأغنم من الحاضر أمن اليقين  
فقد تساوى في الثرى راحل غدا وماض من الوف السنين  
اطفى لظى القلب بشهد الرضاب فإنما الأيام مثل السحاب  
وعيشنا طيف خيال فنل حظك منه قبل فوت الشباب  
لبست ثوب العيش لم استشر وحررت فيه بين شتى الفكر  
وسوف أنضو الثوب عني ولم أدرك لماذا جئت أين المغر  
يا من يحار الفهم في قدرتك وتطلب النفس حمى طاعتك  
اسكرني الإثم و لكنني صحوت بالأمال في رحمتك  
إن لم أكن أخلصت في طاعتك فإنني أطمع في رحمتك  
وإنما يشفع لي بأنني قد عشت لا أشرك في وحدتك  
نخفي عن الناس سنى طلعتك فإنني أطمع في رحمتك  
فأنت مجلاه وأنت الذي ترى بديع الصنع في آيتك  
ان تفضل القطرة من بحرهما ففي مداها منتهى أمرها  
تقاربت يا رب ما بيننا مسافة البعد على قدرها  
يا عالم الأسرار علم اليقين يا كاشف الضر عن البائسين  
يا قابل الأعذار عدنا إلى ظلك فأقبل توبة التائبين

Figure 1: The lyrics of the song رباعيات الخيام (Oum Kalthoum, 1950) [2]

```

\documentclass{article}
\usepackage{dad}
\begin{document}
\arab{
\begin{center}
\textbf{rba'YAt AlxYAm}

\medskip

\begin{minipage}{10cm}
sm't SUTa hAtfA fY Alsh--\kesh--r n--\kesh--AdI mn AlRYb rfAt Albsh--\kesh--r\
hbUA Aml'aUA k'as AlmnI qb--\kesh--l 'an tml'a k--\kesh--'as Al'm--\kesh--r kf
Alq--\kesh--dr\
lA tshRI AlbAl bmADY Alzm--\kesh--An UIA b--\kesh--'At Al'Y--\kesh--sh qb--\kesh--l
Al'aUAn\
U'aR--\kesh--nm mn AlHAD--\kesh--r ldhAt--\kesh--h flYs f--\kesh--Y Tb--\kesh--'
AllyAl--\kesh--Y Al'am--\kesh--An\
Rd bZhr AlRYb UAlYUm l--\kesh--Y Ukm YxYb AlZ--\kesh--n f--\kesh--Y Almqb--\kesh--l\
Uls--\kesh--t bAlRAF--\kesh--l Ht--\kesh--I 'arI jm--\kesh--Al dnY--\kesh--AY U
lA Ajtl--\kesh--I\
Alqlb qd 'aDnAh 'shq Aljm--\kesh--Al UAlS--\kesh--dr q--\kesh--d D--\kesh--Aq bm--\kesh--A
lA Yq--\kesh--Al\
YA rb hl YrDYk hdhA AlZlm--\kesh--A UAlm--\kesh--A| Yns--\kesh--Ab 'am--\kesh--Am--\kesh--Y
zAl\
'aUII bhdhA Alqlb 'an Yxfq--\kesh--A U fY RrAm AlH--\kesh--b 'an YHtrq--\kesh--A\
mA 'aDY' AlYUm AldhY m--\kesh--r b--\kesh--Y mn RYr 'an 'ahUI U 'an 'a'shq--\kesh--A\
'afq xfyf AlZl hdhA Alsh--\kesh--r n--\kesh--AdI d' Aln--\kesh--Um Un--\kesh--AR
AlUt--\kesh--r\
fm--\kesh--A 'aT--\kesh--Al Aln--\kesh--Um 'm--\kesh--rA UIA qSr mn Al'a'mAr TUI
Als-h--\kesh--r\
fk--\kesh--m tUI--\kesh--I Ally--\kesh--l b'--\kesh--d Alnh--\kesh--Ar UT--\kesh--Al
bAl'anj--\kesh--m h--\kesh--dhA Alm--\kesh--dAr\
f'amsh AlhUYnt* 'an hdhA Alc--\kesh--rI m--\kesh--n 'a'Y--\kesh--n sAH--\kesh--rt*
AlAH--\kesh--UrAr\
lA tUHsh Alnfs bxUf AlZn--\kesh--Un U'aRnm mn AlHADr 'amn AlYqY--\kesh--n\
fqd tsAUI fY AlcrI rAH--\kesh--l RdA UmAD mn AlUf AlsnY--\kesh--n\
ATf'I lZI Alqlb bshhd AlrD--\kesh--Ab f'inm--\kesh--A Al'aY--\kesh--Am mc--\kesh--l
Alsh--\kesh--Ab\
U'Yshn--\kesh--A TY--\kesh--f xY--\kesh--Al fn--\kesh--l HZ--\kesh--k mn--\kesh--h
qb--\kesh--l f--\kesh--Ut Alshb--\kesh--Ab\
lbst cUb Al'Ysh lm Astsh--\kesh--r UH--\kesh--rt fY--\kesh--h bY--\kesh--n sht--\kesh--I
Alfk--\kesh--r\
UsUf 'anDU AlcUb 'nY Ul--\kesh--m 'adrk lm--\kesh--AdhA j'I--\kesh--t 'aY--\kesh--n
AlmR--\kesh--r\
YA mn YHAr Alfhm fY qdrt--\kesh--k UtTl--\kesh--b Alnf--\kesh--s Hm--\kesh--I TA't--\kesh--k\
AskRn--\kesh--Y Al'ic--\kesh--m U lknn--\kesh--Y SH--\kesh--Ut bAl'am--\kesh--Al
f--\kesh--Y rHmt--\kesh--k\
'in lm 'akn 'axlSt fY TA't--\kesh--k f'inn--\kesh--Y 'aTm--\kesh--' f--\kesh--Y
rHmt--\kesh--k\
U'inm--\kesh--A Yshf--\kesh--' l--\kesh--Y b'ann--\kesh--Y q--\kesh--d 'sh--\kesh--t
lA 'ash--\kesh--rk fY UHdt--\kesh--k\
nxfY 'n AlnAs snI Tl't--\kesh--k f'inn--\kesh--Y 'aTm--\kesh--' f--\kesh--Y rHmt--\kesh--k\
f'an--\kesh--t mj--\kesh--lAh U'an--\kesh--t Al--\kesh--dhY t--\kesh--rI bdY--\kesh--'
AlSn--\kesh--' f--\kesh--Y 'AYt--\kesh--k\
An tfDl AlqTrt* mn bHrh--\kesh--A ff--\kesh--Y m--\kesh--dAh--\kesh--A mnt-h--\kesh--I
'amrh--\kesh--A\
tqArb--\kesh--t Y--\kesh--A rb m--\kesh--A bYnn--\kesh--A msAf--\kesh--t* Alb'--\kesh--d
'l--\kesh--I qdrh--\kesh--A\
YA 'Alm Al'asrAr 'lm AlYq--\kesh--Y--\kesh--n Y--\kesh--A kAsh--\kesh--f AlD--\kesh--r
'--\kesh--n Alba'IsY--\kesh--n\
Y--\kesh--A qAb--\kesh--l Al'a--\kesh--dhAr 'dn--\kesh--A 'il--\kesh--I Zl--\kesh--k
f'aqb--\kesh--l tUb--\kesh--t* AltA'IbY--\kesh--n
\end{minipage}
\end{center}
}
\end{document}

```

Figure 2: T<sub>E</sub>X code of Fig. 1, transliterated input

```

\documentclass{article}
\usepackage{dad}
\begin{document}
\arab{
\begin{center}
\textbf{رباعيات الخيام}

\medskip

\begin{minipage}{10cm}

سمعت صوتا هاتفا في السحابة ترنات عادى من الغيب رفات البشاة ترن
هبوا املاوا كأس المنى قبات تل أن تملأ كأس العمارة كرف القبات در
لا تشغل البال بماضي الزمان ولا بتاتات العبات تش قبائل الأوان
وأغابت عن من الحاضر تر لذات طه فليس فأتى طبيطع الليال تفي الأماتان
غد يظهر الغيب واليوم لآتى وكم يخيب الظل تن فأتى المقبات تل
ولسأت بت بالخاف تطل حتاتى أرى جمات طال دنيا تاي ولا اجتلت تى
القلب قد أضناه عشق الجمات طال والصات ط در قات طه ضات طاق بمات ط لا يقا
يا رب هل يرضيك هذا الظلمات ط والمات طاء ينسأت طاب أمات طامات طي زلال
أولى بهذا القلب أن يخفق تى وفي غرام الحات ط ب أن يحترق تى
ما أضيع اليوم الذي مات طر بات طي من غير أن أهوى و أن أعشق تى
أفق خفيف الظل هذا السحابة ترنات عادى دع الناطوم ونات طاغ
الوات ط تر

فمات ط أطلت طال الناطوم عمات طرا ولا قصر من الأعمار طول السهات ط تر
فكأت طم تولت طى الليات ط تل بعات طه النهات طار وطات طال بالانجات طم مات ط هذا المات ط دار
فأمش الهوينة أن هذا الثبات طرى مات طن أعبات طن ساحات طرة الاحات ط ورا
لا توحش النفس بخوف الظنات طون وأغنم من الحاضر أمن اليقيات طن
فقد تساوى في الثرى راحت ط تل غدا وماض من الوف السنيات طن
اطفى لظى القلب بشهد الرضات طاب فبانمات طال الأيات طام مات ط تل السحات طاب
وعيشنات ط طيات ط عف خبات طال فنات ط تل حظات طك منات طه قبات ط تل فبات طوت الشبات طاب
لبست ثوب العيش لم استشأت طر وحات طرت فيات طه بيدات طن شتات طى الفكأت ط تر
وسوف أنضو الثوب عني ولت طم أدرك لمت ط اذا جنات طت أبات طن المغات ط تر
يا من يحار الفهم في قدرات طك وتطلات ط ب النفات طس حماطى طاعت طك
أسكرنات طى الإبتات طم ولكننات طى صجات طوت بالأمات طال فأتى رحمت طك
إن لم أكن أخلصت في طاعت طك فإننات طى أطمات طع فأتى رحمت طك
وإنمات ط يشفأت طع لآتى باننات طى قات طه عشات ط تل لا أشات طرك في وحدت طك
نخفى عن الناس سنى طلعت طك فإننات طى أطمات طع فأتى رحمت طك
فأنات ط تل مجات طلاه وأنات ط تل الات ط لذي تات طرى بيدات طع الصنات طع فأتى آيات طك
إن تفضل القطرة من بحر مات ط ففات طى مات طه مات ط مننت طى أمر مات ط
تقاربات ط تل مات ط رب مات ط بيننات ط مسافات طة البعات طه علآت طى قدر مات ط
يا عالم الأسرار علم اليقيات ط تل ينات ط كاشات طه الضات طر عات طن البانسات طن
يات ط قابات ط تل الأعات ط دار عدنات طال إلت طى ظلات طك فأقبات ط تل توبات طة
التائبات طن

```

Figure 3: T<sub>E</sub>X code of Fig. 1, Unicode input

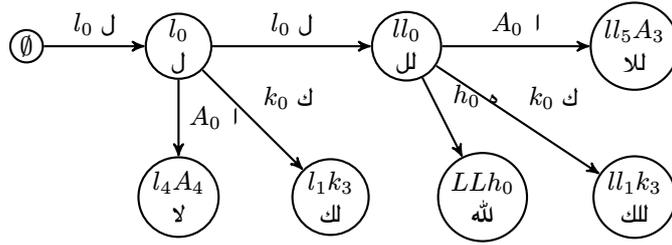


Figure 4: Finite state automaton starting with an isolated *lām* (*ʾalif* ا stands for the set of letter  $\mathcal{A} = \{ ا, \acute{ا}, \grave{ا}, \bar{ا}, \tilde{ا} \}$ ; ك stands for any Arabic letter besides ه and set  $\mathcal{A}$ ).

Nevertheless, for the user’s convenience, we have adopted that behavior also in ض, but only in the case of Unicode input. Therefore when the user types Unicode *lām-lām-hā’* (the first *lām* must not be preceded by a quadriform letter), the system will produce the لله ligature.

This method will not work if a diacritic is inserted between the two *lāms*, or if the first *lām* follows a quadriform letter and hence will be medial. For that case, we have defined a macro لله (the macro name is in Arabic script so that right-to-left direction is not disrupted) which takes an argument: the vowel between the two *lāms*. Hence, to obtain لله the user can choose between one of the following two:

فَ/لله{ }  
faLiLhi

(The dotted circle, used to show the combining nature of short vowels and other diacritics, can be obtained by the macros `\arabdottedcircle` or دائرة / with the macro name in Arabic script.)

### 3 T<sub>E</sub>Xnicalities

More information about ض (history, evolution, rationale of technical choices, T<sub>E</sub>Xnicalities) can be found in [1].

### References

[1] Yannis Haralambous. A simple arabic typesetting system for mixed latin/arabic alphabet documents: *dād*. *TUGboat*, 35(3), 2014.

[2] ام كلثوم كلمات اغاني كوكب الشرق السيده ام كلثوم موسوعه كامله. ساحرالبسمه. <http://alsabbah.yoo7.com/t417-topic>, 2008.

[3] Karin C. Ryding. *Arabic. A linguistic introduction*. Cambridge University Press, Cambridge, 2014.

[4] John R. Smart. *Arabic*. Teach yourself books, 1986.

[5] Barakat Ahmad Syed. *Introduction to Quranic script*. Curzon Press, 1984.