

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^AT_EX 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_EXnical requirement is LuaT_EX, not because of the Lua language (which is not used, for the moment), but because of features that have survived from LuaT_EX's Ω origins: bidirectionality and use of large fonts (OVF, OFM).

Choose LuaT_EX as your T_EX engine, load the package into your document, and اهلاً شاداً! just start writing in Arabic using command `\arab`.

More information about ض (history, evolution, rationale of technical choices, T_EXnicalities) 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

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 ‘aḏa’, 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;

Table 1: Transliteration of ض system

ء		آ	'A	أ	'a	ؤ	'u	إ	'i	ئ	'I
ا	A	ب	b	ة	t*	ت	t	ث	th	ج	j
ح	H	خ	x	د	d	ذ	dh	ر	r	ز	z
س	s	ش	sh	ص	S	ض	D	ط	T	ظ	Z
ع	'	غ	R	ف	f	ق	q	ك	k	ل	l
م	m	ن	n	ه	h	و	U	ى	I	ي	Y
آ	A*	اَ	o	اِ	a	اِ	i	اُ	u	اَ	aN
اِ	iN	اِ	uN	اِ	+	اِ	+a	اِ	+i	اِ	+u
اِ	+aN	اِ	+iN	اِ	+uN	اِ	a*	اِ	+a*	الله	LLh
پ	p	گ	g	چ	C	ژ	J	ه	e	ف	v
ب	'b	ن	'n	ف	'f	و	'q	اِ	a**	اِ	+a**

6. more generally, the dash plays the rôle of *zero-width joiner*¹: when writing ب = -b, the letter *bā'* will be in final form; ب = b- and ب = -b- will produce initial and middle letters, provided of course the letter is quadriform (as is letter *bā'* 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 ل, instead of a dash, we use digit 4. For the final ligature ل it will be a digit 5. Example: to colorize the *lāms* of تلا, write

```
\arab{t-\textcolor{red}{-15-}-A5%
\textcolor{red}{14-}-A4}
```

8. finally, there is yet another use of the dash: when doubled, it produces a kashida stroke: compare ليل = 1Y1 and ليل = 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:

لل

¹Except for the case of letter ذ = dh which is biform and hence is not connected with the following letter. By writing ده = d-h one obtains letters *dāl* and *hā'*, but the *hā'* 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 \acute{a} = 'a, \acute{i} = 'i, \acute{u} = 'u, \acute{I} = 'I, \acute{A} = 'A but also of undotted letters $b\bar{a}$ ' ب = 'b, $n\bar{a}n$ ن = 'n, $f\bar{a}$ ' ف = 'f and $q\bar{a}f$ ق = 'q;
10. other digraphs end with one or more asterisks: the most frequent one is the $t\bar{a}$ ' *marbuta* 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 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 عَسَقَ (*Koran* 42:2) or كَهَيْتَمَصَّ (*Koran* 19:1)—sometimes it is even combined with a *šadda* (as in الْمَصَّ , *Koran* 7:1 and [5, p. 111] for the *šadda*);
11. there is a special transcription for the ligature لله = LLh used for the *اسم الجلالة* “noun of majesty,” which is the name of God 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 $\text{l\bar{a}m-l\bar{a}m-h\bar{a}}$, for example يُضِلُّهُ (*Koran* 6:39) where we encounter letters له but not with the meaning “God.” Contrarily to other systems, the لله ligature is available also in final form (for 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 وَلَيْلَهُ , *Koran* 2:115 or 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: 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 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) طط (طط are the first two letters of 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 لله ligature and Unicode

The لله ligature is traditionally used for writing the name of God: الله . 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 *inchallah* and in Portuguese as *oxalá*) and in the very common surname 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 $\text{l\bar{a}m-l\bar{a}m-h\bar{a}}$ (which would normally look like له) 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 ‘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 الحب 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 tshRl 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 xYf 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 tUl--\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_EX code of Fig. 1, transliterated input

```

\documentclass{article}
\usepackage{dad}
\begin{document}
\arab{
\begin{center}
\textbf{رباعيات الغيام}
\medskip
\begin{minipage}{10cm}
سمعت صوتا هاتفا في السحابة ترنات عادى من الغيب رفات البشاة ترن
هبوا املاوا كأس المنى قبا تطل أن تملأ كأس العمارة كرف القبا تدر
لا تشعل الببال بماضي الزمات طان ولا بقات العبات ش قبا تطل الأوان
وأغابت من الحاضرات لذات طه فليس فأتطبي طبي تطع الليال تطي الأمات طان
غد يظهر الغيب واليوم لآتطبي وكم يخيب النظرات طن فأتطبي المقبات تطل
ولسأتطبت بالخاف تطل حتاتطى أرى جمات طال دنيا تظاي ولا اجتلت تطى
القلب قد أضناه عشق الجمات طال والصات طدر قاتطد ضاتطاق بما تطل لا يقا تطل
يا رب هل يرضيك هذا الظلمات طال والمات طاء ينسأتطاب أمات طاماتطى زلال
أولي بهذا القلب أن يخفق تطل وفي غرام الحاتطب أن يحترق تطل
ما أضيع اليوم الذي مات طر باتطبي من غير أن أهوى و أن أعشق تطل
أفق خفيف الظل هذا السحابة ترنات عادى دع النات طوم ونات طاغ
الوات ط تر
فمات ط أطل تطل النات طوم عمات طرا ولا قصر من الأعمار طول السهات ط تر
فكأتطم تولد تطى الليات ط تل بعدات طد النهات طار وطات طال بالانجات طم مات طذا 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 طل غدا وماض من الوف السنيات طن
اطفي لظى القلب بشهد الرضات طاب فبانمات طال الأيات طام مات طل السحات طاب
وعيشنات طال طبيات ط عف خبات طال فنات طل حظات طك منات طه قبات طل فتوت الشبات طاب
لبست ثوب العيش لم استشأت طر وجات طرت فيات طه بيدات طن شتات طى الفكات ط تر
وسوف أنضو الثوب عني ولات طم أدرك لمتات ط اذا جنات طت أيدات طن المغات ط تر
يا من يحار الفهم في قدرات طك وتطلات طب النفات طس حما تطل طاعتات طك
أسكرنات طى الإبتات طم ولكننات طى صجات طوت بالأمات طال فأتطى رحمتات طك
إن لم أكن أخلصت في طاعتات طك فإننات طى أطمات طع فأتطى رحمتات طك
وإنمات طال يشفأتطع لآتطى بأننات طى قاتطد عشات طت لا أشات طرك في وحدتات طك
نخفى عن الناس سنى طلعتات طك فإننات طى أطمات طع فأتطى رحمتات طك
فأنات طت مجات طلاه وأنات طت الات طذي تات طرى بيدات طع الصنات طع فأتطى آيات طك
إن تفضل القطرة من بحر مات طال ففات طى مات طد مات طال مننتها تطل أمر مات طال
تقاربات طت يات طال رب مات طال بيننات طال مسافات طة البعات طد علأتطى قدر مات طال
يا عالم الأسرار علم البقات ط يات طن يات طال كاشات طر الضات طر عات طن البانسات طن
يات طال قبات طل الأعات ط دار عدنات طال إلات طى ظلات طك فأقبات طل توبات طة
التائبات طن
\end{minipage}
\end{center}
}
\end{document}

```

Figure 3: T_EX code of Fig. 1, Unicode input

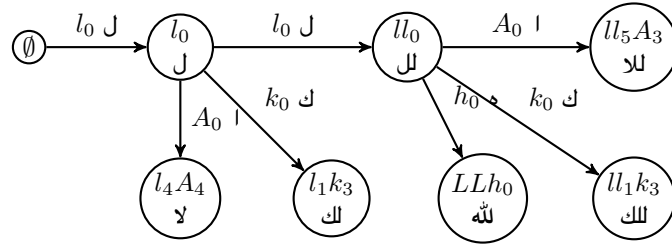


Figure 4: Finite state automaton starting with an isolated *lām* (*ʾalif* *ʾ* stands for the set of letter $\mathcal{A} = \{ ل, \acute{ل}, \grave{ل}, \bar{ل}, \tilde{ل} \}$; *ك* stands for any Arabic letter besides *ه* and set \mathcal{A}).

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-hāʾ* (the first *lām* must not be preceded by a quadriform letter), the system will produce the **الله** 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 **الله** (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 **فَلله** the user can choose between one of the following two:

فَلله{ }
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_EXnicalities

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

References

[1] Yannis Haralambous. A simple arabic typesetting system for mixed latin/arabic alphabet documents: *dād*. *TUGboat*, 35(3):277–283, 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.