

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

Version 1.0/2014-09-06

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 شلاً, 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].

ATTENTION! خطر! Because of bugs in luaT_EX, in some cases appear mysterious blank spaces on the left of the text, word parts are

*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

duplicated, and sometimes T_EX breaks with an error message `This can't happen (sub_disc_widths)`. These errors have been submitted to the luaT_EX team (<http://tracker.luatex.org/view.php?id=912&history=1#history>), which will provide the necessary corrections
إِنْ شَاءَ اللَّهُ.

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 called a “pharyngealized voiced alveolar plosive” (dixit Wikipedia). Here is how [3, 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 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 ‘ḍ’, 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 type-writer), “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 LuaT_EX (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. emphatic consonants (ح = H, ط = T, ص = S, ض = D) as well as long vowels (ا = A, و = U, ي = Y) and *ʾalif maqṣūra* (ى = I) are *uppercased*;
2. some consonants are modified by adding a character h (ذ = dh, ث = th, ش = sh, كه = kh, زه = zh);
3. the stand-alone *hamza* is obtained by a vertical bar | and letter ayn by a grave accent (which, in legacy T_EX produces an inverted curly apostrophe, which is sometimes used to transliterate this letter);
4. 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;
5. more generally, the dash plays the rôle of *zero-width joiner*¹: when writing ب = -b, the letter *bāʾ* will be in final form; ب = b- and ب = -b-

¹Except for the case of letter ذ = dh which is biform and hence is not connected with the following letter. By writing د = d-h one obtains letters *dāl* and *hāʾ*, but the *hāʾ* is not in medial form, as it would be in any other case when preceded by a dash.

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	اُ	i	اُو	u	أَ	aN
اِ	iN	أِ	uN	أِ	+	أِ	+a	أِ	+i	أِ	+u
أِ	+aN	أِ	+iN	أِ	+uN	أِ	a*	أِ	+a*	الله	LLh
پ	p	گ	g	چ	C	ژ	J	ه	e	ف	v
ب	'b	ن	'n	ف	'f	ق	'q	أَ	a**	أِ	+a**

will produce initial and middle letters, provided of course the letter is quadriform (as is letter $b\bar{a}'$ in this example). This is very useful when describing grammar rules, to signify that a letter (or letter group) is an affix;

- the dash can also be used to reestablish contextual forms when combined with $\text{T}_{\text{E}}\text{X}$ commands, for example, to colorize letters. There is only one special case: when we want to colorize a letter of an isolated ligature لآ, instead of a dash, we use digit 4. For the final ligature لآ it will be a digit 5. Example: to colorize the *lāms* of لآلآ, write

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

- finally, there is yet another use of the dash: when doubled, it produces a keshideh stroke: compare لآل = 1Y1 and لآل = 1--Y--1. There

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 لبي وكم يخيب الظن في المستقبل
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 وفي 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 الهويّة 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 صحت 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 بحرها ففي مداها منتهى أمرها
تقارب 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 Ula 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--Al Yns--\kesh--Ab 'am--\kesh--Am--\kesh--Y
zlAl\\
'aUlI 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 Ula qSr mn Al'a'mAr TUl
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\\
Askrrn--\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 الخيام}

\medskip

\begin{minipage}{10cm}

سمعت صوتا هاتفا في السحابة نادت نادى من الغيب رفات البشر
هيو املوا كأس المني قبل ان تملأ كأس العمى
لا تشغل البال بماضي الزمان ولا يأتك آت العيب
وأغصنم من الحاضر لذر لذر فليس فأتك في طبع الليالي
غد يظهر الغيب واليوم لآتكم في وكم يخيب الظن
ولست أتت بالغافل أتت لحتات في أرى جملة نديت في ولا اجتلت في
القلب قد أضناه عشق الجملة والصلوات
يا رب هل يرضيك هذا الظلمة والما تاء ينسأب أم تاء سامت في زلال
أولي بهذا القلب أن يخفق تاء وفي غرام الحات تاء أن يحترق تاء
ما أضيع اليوم الذي مأتت ربت في من غير أن أهوى و أن أعشق تاء
أفق خفيف الظل هذا السحابة نادت نادى دغ النادت نوم ونادت غ
الوت تاء
فمأت تاء أطأت تاء النادت نوم عمأت تاء ولا قصر من الأعمار طول السهات تاء
فكأت تاء تولأت في الليأت تاء لبعأت تاء النهدت تاء وطأت تاء بالأنجأت تاء
المأت تاء
فأمش الهويئة أن هذا الثأت تاء من أعيأت تاء ساحت تاء الاحت تاء
لا توحش النفس بخوف الظن تاء وأغنم من الحاضر أمن اليقيأت تاء
فقد تساوى في الثرى راحأت تاء وماض من الوف السنيأت تاء
اطفي لظى القلب بشهد الرضأت تاء فأنمأت تاء الأيأت تاء مأت تاء السحات تاء
وعيشنأت تاء طيأت تاء خيأت تاء فأنمأت تاء لحتات تاء منأت تاء قبلأت تاء فأت تاء
شوات تاء
لبست ثوب العيش لم استشأت تاء وحات تاء فيات تاء بيأت تاء شتأت في الفكأت تاء
وسوف أنضو الثوب عني ولأت تاء أدرك لمتأت تاء جنتأت تاء أيتأت تاء المغأت تاء
يا من يحار الفهم في قدرأت تاء وتطلأت تاء النفات تاء سجات في طاعتأت تاء
اسكرنأت تاء ولكننأت تاء لحنأت تاء صوت بالأمأت تاء لآت في رحمتأت تاء
إن لم أخلصت في طاعتأت تاء فأنمأت تاء في أطمأت تاء فأت في رحمتأت تاء
وإنمأت تاء يشفأت تاء لآت في بأننأت تاء فأت تاء عشأت تاء لا أشأت تاء في وحدتأت تاء
نخفي عن الناس سني طلعتأت تاء فأنمأت تاء في أطمأت تاء فأت في رحمتأت تاء
فأنمأت تاء مجأت تاء وأنمأت تاء لآت في بأننأت تاء فأت تاء بديأت تاء المصأت تاء فأت في آيتأت تاء
لك
ان تفضل القطرة من بحرمتأت تاء ففأت في مأت تاء مأت تاء منتهأت في أمرمتأت تاء
تقاربأت تاء رب مأت تاء بيننأت تاء مسافأت تاء البعدأت تاء علأت في قدرمتأت تاء
يا عالم الأسرار علم اليقأت تاء من يأت تاء كاشأت تاء الفأت تاء رعت تاء البائسأت تاء
من
يأت تاء قابأت تاء الأعأت تاء دار عدنأت تاء إلات في ظلت تاء فأقبأت تاء تل توبأت تاء
التائبأت تاء

```

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

keshideh) which is exactly equivalent to `\kesh` and has to be placed between Unicode U+0640 ARABIC TATWELL characters.

2.2.1 The ﷲ ligature and Unicode

The ﷲ 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 *lām-lām-hā* (which would normally look like ﷲ) by the complete ligature ﷲ (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.

Nevertheless, for the user’s convenience, we have adopted that behavior also in *خ*, 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 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 *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 *فَلِلَّهِ* the user can choose between one of the following two:

فَ{ }لِلَّهِ

`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.)

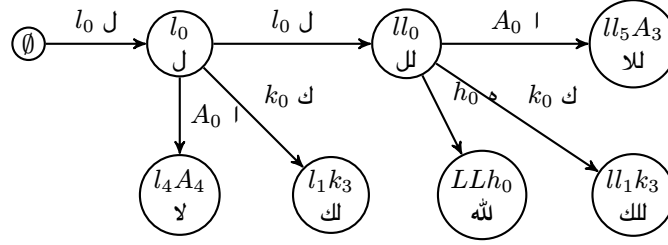


Figure 4: Finite state automaton starting with an isolated *lām* (*ʾalif* ا stands for the set of letter $\mathcal{A} = \{ ا, \text{إ}, \text{ل}, \text{ل}, \text{آ}, \text{آ} \}$; ك stands for any Arabic letter besides ا and set \mathcal{A} .

3 T_EXnicalities

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

References

- [1] Yannis Haralambous. ض (*dād*), a simple arabic typesetting system for mixed latin/arabic alphabet documents. Submitted to *TUGboat*, 2014.
- [2] ام كلثوم كلمات اغاني كوكب الشرق السيده ام كلثوم موسوعه كامله. ساحرالبسمه. <http://alsabbah.yoo7.com/t417-topic>, 2008.
- [3] John R. Smart. *Arabic*. Teach yourself books, 1986.
- [4] Barakat Ahmad Syed. *Introduction to Quranic script*. Curzon Press, 1984.