The arabluatex package
v1.3 – 2016/05/28

Robert Alessi
alessi@robertalessi.net

Contents

List of Tables

Abstract

This package provides for LuaLaTeX an ArabTeX-like interface to generate Arabic writing from an ASCII transliteration. It is particularly well-suited for complex documents such as technical documents or critical editions where a lot of left-to-right commands intertwine with Arabic writing. arabluatex is able to process any ArabTeX input notation. Its output can be set in the same modes of vocalization as ArabTeX, or in different roman transliterations. It further allows many typographical refinements. It will eventually interact with some other packages yet to come to produce from `.tex` source files, in addition to printed books, `TEI xml` compliant critical editions and/or lexicons that can be searched, analyzed and correlated in various ways.

License and disclamer

Copyright © 2016 Robert Alessi

Please send error reports and suggestions for improvements to Robert Alessi:

- email: <alessi@robertalessi.net>
- website: http://www.robertalessi.net/arabluatex
- development: http://git.robertalessi.net/arabluatex
- comments, feature requests, bug reports: http://issues.robertalessi.net

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

This release of arabluatex consists of the following source files:

- `arabluatex.ins`
- `arabluatex.dtx`
- `arabluatex.lua`
- `arabluatex_voc.lua`
- `arabluatex_fullvoc.lua`
- `arabluatex_novoc.lua`
- `arabluatex_trans.lua`

1 Introduction

In comparison to Prof. Lagally's outstanding ArabTeX,[1] ArabLuaTeX is at present nothing more than a modest piece of software. Hopefully—if I may say so—it will eventually provide all of its valuable qualities to the LuaLaTeX users.

[1] See http://ctan.org/pkg/arabtex

arabtex dates back to 1992. As far as I know, it was then the first and only way to typeset Arabic texts with TEX and LATEX. To achieve that, arabtex provided—and still does—an Arabic font in *Nasḫī* style and a macro package that defined its own input notation which was, as the author stated, "both machine, and human, readable, and suited for electronic transmission and e-mail communication".[2] Even if the same can be said about Unicode, ArabTEX ASCII input notation still surpasses Unicode input, in my opinion, when it comes to typesetting complex documents, such as scientific documents or critical editions where footnotes and other kind of annotations can be particulary abundant. It must also be said that most text editors have trouble in displaying Arabic script connected with preceding or following LATEX commands: it often happens that commands seem misplaced, not to mention punctuation marks, or opening or closing braces, brackets or parentheses that are unexpectedly displayed in the wrong direction. Of course, some text editors provide ways to get around such difficulties by inserting invisible Unicode characters, such as LEFT-TO-RIGHT or RIGHT-TO-LEFT MARKS (U+200E, U+200F), RTL/LTR "embed" characters (U+202B, U+202A) and RLO/LRO "bidi-override" characters (U+202E, U+202D).[3] Nonetheless, it remains that inserting all the time these invisible characters in complex documents rapidly becomes confusing and cumbersome.

The great advantage of ArabTEX notation is that it is immune from all these difficulties, let alone its being clear and straightforward. One also must remember that computers are designed to process code. ArabTEX notation is a way of encoding Arabic language, just as TEX "mathematics mode" is a way of processing code to display mathematics. As such, not only does it allow greater control over typographical features, but it also can be processed in several different ways: so without going into details, depending on one's wishes, ArabTEX input can be full vocalized Arabic (*scriptio plena*), vocalized Arabic or non-vocalized Arabic (*scriptio defectiva*); it further can be transliterated into whichever romanization standard the user may choose.

But there may be more to be said on that point, as encoding Arabic also naturally encourages the coder to vocalize the texts—without compelling him to do so, of course. Accurate coding may even have other virtuous effects. For instance, hyphens may be used for tying particles or prefixes to words, or to mark inflectional endings, and so forth. In other words, accurate coding produces accurate texts that can stand to close grammatical scrutiny and to complex textual searches as well.

Having that in mind, I started arabluatex. With the help of Lua, it will eventually interact with some other packages yet to come to produce from `.tex` source files, in addition to printed books, `TEI xml` compliant critical editions and/or lexicons that can be searched, analyzed and correlated in various ways.

1.1 arabluatex is for LuaLATEX

It goes without saying that arabluatex requires LuaLATEX. TEX and LATEX have arabtex, and XELATEX has arabxetex. Both of them are much more advanced than arabluatex,

[2] Lagally (2004, p. 2).

[3] Gáspár Sinai's Yudit probably has the best Unicode support. See http://www.yudit.org.

as they can process a number of different languages,[4] whereas arabluatex can process only Arabic for the time being. More languages will be included in future releases of arabluatex.

In comparison to arabxetex, arabluatex works in a very different way. The former relies on the TECkit engine which converts ArabTEX input on the fly into Unicode Arabic script, whereas the latter passes ArabTEX input on to a set of Lua functions. At first, LATEX commands are taken care of in different ways: some, as \emph, \textbf and the like are expected to have Arabic text as arguments, while others, as \LR, for "left-to-right text", are not. Then, once what is Arabic is carefully separated form what is not, it is processed by other Lua functions which rely on different sets of correpondence tables to do the actual conversion in accordance with one's wishes. Finally, Lua returns to TEX the converted strings—which may in turn contain some other ArabTEX input yet to be processed—for further processing.

2 The basics of arabluatex

2.1 Activating arabluatex

As usual put in your preamble:

```
\usepackage{arabluatex}
```

The only requirement of arabluatex is LuaLATEX; it will complain if you try to compile your document with another engine. That aside, arabluatex does not load packages such as polyglossia or luabidi. It can work with polyglossia though, but does not require it.

Font setup If you wish to use your own Arabic font, you can define it before loading arabluatex. Assuming that fontspec is loaded, put this in your preamble just above the line that loads arabluatex:

```
\newfontfamily\arabicfont[Script=Arabic]{⟨fontname⟩}
```

where ⟨*fontname*⟩ is the standard name of the Arabic font you wish to use.

By default, if no Arabic font is selected, arabluatex will issue a warning message and attempt to load the Amiri font[5] like so:—

```
\newfontfamily\arabicfont[Script=Arabic]{Amiri}
```

> Rem. By default Amiri places the *kasra* in combination with the *tašdīd* below the consonant, like so: ـِّـ. That is correct, as at least in the oldest manuscripts ـٍّـ may stand for ـٍّـ as well as ـِّـ. See Wright (1896, i.14.C–D). The placement of the *kasra* above the consonant may be obtained by selecting the ss05 feature of the Amiri font, like so:—[6]
>
> ```
> \newfontfamily\arabicfont[Script=Arabic,RawFeature={+ss05}]{Amiri}
> ```
>
> Other Arabic fonts may behave differently.

[4] To date, both packages support Arabic, Maghribi, Urdu, Pashto, Sindhi, Kashmiri, Uighuric and Old Malay; in addition to these, arabtex also has a Hebrew mode, including Judeo-Arabic and Yiddish.

[5] Hosny (2015).

[6] See the documentation of amiri, Hosny (2015, p. 5).

2.2 Options

arabluatex may be loaded with four mutually exclusive global options, each of which may be overriden at any point of the document (see below section 2.3.1 on page 7):

voc `default`

In this mode, which is the one selected by default, every short vowel written generates its corresponding diacritical mark: *ḍamma* (ـُ), *fatḥa* (ـَ) and *kasra* (ـِ). If a vowel is followed by ⟨*uN, aN, iN*⟩, then the corresponding *tanwīn* (ـٌ, ـً, ـٍ, ـًى or ـً) is generated. Finally, ⟨*u, a, i*⟩ at the commencement of a word indicate a "connective *ʾalif*" (*ʾalifu ʾl-waṣli*), but `voc` mode does not show the *waṣla* above the *ʾalif*; instead, the accompanying vowel may be expressed at the beginning of a sentence (اُ اَ اِ).

`fullvoc`

In addition to what the `voc` mode does, `fullvoc` expresses the *sukūn* and the *waṣla*.

`novoc`

None of the diacritics is showed in `novoc` mode, unless otherwise specified (see "quoting" technique below section 4.4 on page 17).

`trans`

This mode transliterates the ArabTEX input into one of the accepted standards. At present, two standards are supported (see below section 6 on page 22 for more details):
dmg *Deutsche Morgenländische Gesellschaft*, which is selected by default;
loc *Library of Congress*.
More standards will be included in future releases of arabluatex.

2.2.1 Classic contrasted with modern typesetting of Arabic

New feature v.1.2

By default, arabluatex typesets Arabic in a classic, traditional style the most prominent features of which are the following:

- 'Classic' *madda*: when *ʾalif* and *hamza* accompanied by a simple vowel or *tanwīn* is preceded by an *ʾalif* of prolongation (ـَا), then a mere *hamza* is written on the line, and a *madda* is placed over the *ʾalif*, like so:—

  `samA'uN` سَمَآءٌ *samāʾun*, `jA'a` جَآءَ *ǧāʾa*, `yatasA'alUna` يَتَسَآءَلُونَ *yatasāʾalūna*[7]
  (see on page 13 for further details).

- The euphonic *tašdīd* is generated (see on page 13).
- Assimilation rules laid on item b on page 14 are applied.
- In `fullvoc` mode, the *sukūn* is expressed.

`\SetArbEasy` `\SetArbDflt`

Such refinements may be discarded by the command `\SetArbEasy`, either globally in the preamble or locally at any point of the document. Default 'classic' rules may be set back at any point of the document with the command `\SetArbDflt`. Examples follow:—

(a) `\SetArbDflt`:

i. `voc` وَمَاتَ استِسقَآءً قَبلَ أَن يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَآءِ

ii. `fullvoc` وَمَاتَ ٱسْتِسْقَآءً قَبْلَ أَن يُتِمَّ كِتَابَهُ فِي نُجُومِ ٱلسَّمَآءِ

[7] Note that in old mss. such forms as سَمَآأٌ, جَآأَ are also found; see Wright (1896, i. 24 D).

iii. `trans` *wa-māta ’stisqāʾan qabla ’ay yutimma kitāba-hu fī nuǧūmi ’s-samāʾi*

(b) `\SetArbEasy`:

i. `voc` وَمَاتَ استِسقَاءً قَبلَ أَن يُتِمَّ كِتَابَهُ فِي نُجُومِ السَمَاءِ

ii. `fullvoc` وَمَاتَ ٱستِسقَاءً قَبلَ أَن يُتِمَّ كِتَابَهُ فِي نُجُومِ ٱلسَّمَاءِ

iii. `trans` *wa-māta ’stisqāʾan qabla ’an yutimma kitāba-hu fī nuǧūmi ’s-samāʾi*

> Please note that this document is typeset with `\SetArbDflt` throughout.

2.3 Typing Arabic

`\arb` Once arabluatex is loaded, a command `\arb{`⟨*Arabic text*⟩`}` is available for inserting Arabic text in paragraphs, like so:—

```
From \textcite[i. 1 A]{Wright}:--- Arabic, like Hebrew and
Syriac, is written and read from right to left. The letters
of the alphabet (\arb{.hurUf-u 'l-hijA'-i}, \arb{.hurUf-u
'l-tahajjI}, \arb{al-.hurUf-u 'l-hijA'iyyaT-u}, or
\arb{.hurUf-u 'l-mu`jam-i}) are twenty-eight in number and
are all consonants, though three of them are also used as
vowels (see §~3).
```

From Wright (1896, i. 1 A):— Arabic, like Hebrew and Syriac, is written and read from right to left. The letters of the alphabet (حُرُوفُ الهِجَآءِ, حُرُوفُ التَّهَجِّي, اَلحُرُوفُ الهِجَآئِيَّةُ, or حُرُوفُ المُعجَمِ) are twenty-eight in number and are all consonants, though three of them are also used as vowels (see § 3).

`arab` Running paragraphs of Arabic text should rather be placed inside an *Arabic environment*

```
  \begin{arab}
  [...]
  \end{arab}
```

like so:—

```
\begin{arab}
  'at_A .sadIquN 'il_A ju.hA ya.tlubu min-hu .himAra-hu
  li-yarkaba-hu fI safraTiN qa.sIraTiN. wa-qAla la-hu:
  \enquote{sawfa 'u`Idu-hu 'ilay-ka fI 'l-masA'-i
  wa-'adfa`u la-ka 'ujraTaN.} fa-qAla ju.hA:
  \enquote{'anA 'AsifuN jiddaN 'annI lA 'asta.tI`u 'an
  'u.haqqiqa la-ka .garbata-ka fa-'l-.himAr-u laysa hunA
  'l-yawm-a.}  wa-qabla 'an yutimma ju.hA kalAma-hu bada'a
```

```
    'l-.himAr-u yanhaqu fI 'i.s.tabili-hi. fa-qAla la-hu
    .sadIqu-hu: \enquote{'innI 'asma`u .himAra-ka yA ju.hA
    yanhaqu.} fa-qAla la-hu ju.hA: \enquote{.garIbuN
    'amru-ka yA .sadIqI 'a-tu.saddiqu 'l-.himAr-a
    wa-tuka_d_dibu-nI?}
  \end{arab}
```

أَتَى صَدِيقٌ إِلَى جُحَا يَطلُبُ مِنهُ حِمَارَهُ لِيَركَبَهُ فِي سَفرَةٍ قَصِيرَةٍ. وَقَالَ لَهُ: "سَوفَ أُعِيدُهُ إِلَيكَ فِي المَسَاءِ وَأَدفَعُ لَكَ أُجرَةً." فَقَالَ جُحَا: "أَنَا آسِفٌ جِدًّا أَنِّي لَا أَستَطِيعُ أَن أُحَقِّقَ لَكَ غَرَبَتَكَ فَالحِمَارُ لَيسَ هُنَا اليَومَ." وَقَبلَ أَن يُتِمَّ جُحَا كَلَامَهُ بَدَأَ الحِمَارُ يَنهَقُ فِي إِصطَبلِهِ. فَقَالَ لَهُ صَدِيقُهُ: "إِنِّي أَسمَعُ حِمَارَكَ يَا جُحَا يَنهَقُ." فَقَالَ لَهُ جُحَا: "غَرِيبٌ أَمرُكَ يَا صَدِيقِي أَتُصَدِّقُ الحِمَارَ وَتُكَذِّبُنِي؟"

2.3.1 Local options

As seen above in section 2.2 on page 5, arabluatex may be loaded with four mutually exclusive global options: `voc` (which is the default option), `fullvoc`, `novoc` and `trans`. Whatever choice has been made globally, it may be overriden at any point of the document, as the `\arb` command may take any of the `voc`, `fullvoc`, `novoc` or `trans` modes as optional arguments, like so:—

`voc` – `\arb[voc]{`⟨*Arabic text*⟩`}`;
`fullvoc` – `\arb[fullvoc]{`⟨*Arabic text*⟩`}`;
`novoc` – `\arb[novoc]{`⟨*Arabic text*⟩`}`;
`trans` – `\arb[trans]{`⟨*Arabic text*⟩`}`.

The same optional arguments may be passed to the environment `arab`: one may have `\begin{arab}[`⟨*mode*⟩`]...\end{arab}`, where ⟨*mode*⟩ may be any of `voc`, `fullvoc`, `novoc` or `trans`.

3 Standard ArabTEX input

3.1 Consonants

Table 1 gives the ArabTEX equivalents for all of the Arabic consonants.

Letter	Transliteration[8]		ArabTEX notation
	`dmg`	`loc`	
ا[9]	*a*	*a*	a
ب	*b*	*b*	b
ت	*t*	*t*	t

[8] See below section 6 on page 22.
[9] For *'alif* as a consonant, see Wright (1896, i. 16 D).

Letter	Transliteration dmg	loc	ArabTEX notation
ث	*ṯ*	*th*	_t
ج	*ǧ*	*j*	^g or j
ح	*ḥ*	*ḥ*	.h
خ	*ḫ*	*kh*	_h or x
د	*d*	*d*	d
ذ	*ḏ*	*dh*	_d
ر	*r*	*r*	r
ز	*z*	*z*	z
س	*s*	*s*	s
ش	*š*	*sh*	^s
ص	*ṣ*	*ṣ*	.s
ض	*ḍ*	*ḍ*	.d
ط	*ṭ*	*ṭ*	.t
ظ	*ẓ*	*ẓ*	.z
ع	ʿ	ʿ	`
غ	*ġ*	*gh*	.g
ف	*f*	*f*	f
ق	*q*	*q*	q
ك	*k*	*k*	k
ل	*l*	*l*	l
م	*m*	*m*	m
ن	*n*	*n*	n
ه	*h*	*h*	h
و	*w*	*w*	w
ي	*y*	*y*	y
ة	*a*	*ah*	T

Table 1: Standard ArabTEX (consonants)

3.2 Vowels

3.2.1 Long vowels

Table 2 gives the ArabTEX equivalents for the Arabic long vowels.

Letter	Transliteration[10] dmg	loc	ArabTEX notation
ا	*ā*	*ā*	A
و	*ū*	*ū*	U
ي	*ī*	*ī*	I

[10] See below section 6 on page 22.

Letter	Transliteration dmg	loc	ArabTEX notation
ى[11]	*ā́*	*á*	_A or Y
ـٰ	*ā*	*ā*	_a
ـُ	*ū*	*ū*	_u
ـِ	*ī*	*ī*	_i

Table 2: Standard ArabTEX (long vowels)

Rem. *a.* The long vowels *ā, ū, ī*, otherwise called *ḥurūf*^{u} *'l-madd*^{i}, *the letters of prolongation*, involve the placing of the short vowels *a, u, i* before the letters ا, و, ي respectively. arabluatex does that automatically in case any from `voc`, `fullvoc` or `trans` modes is selected e.g. قَالَ *qāla*, قِيلَ *qīla*, يَقُولُ *yaqūlu*.

Rem. *b.* Defective writings, such as ـٰ, *al-'alif*^{u} *'l-maḥḏūfat*^{u}, or defective writings of *ū* and *ī* are encoded `_a` `_u` and `_i` respectively, e.g. `_d_alika` ذٰلِكَ, `al-mal_a'ikaT-u 'l-ra.hm_an-u` اَلْمَلٰٓئِكَةُ الرَّحْمٰنُ, `.hu_dayfaT-u bn-u 'l-yamAn_i` حُذَيْفَةُ بْنُ ٱلْيَمَانِ for *Ḥuḏayfat*^{u} *bn*^{u} *'l-Yamānī*, etc.

3.2.2 Short vowels

Table 3 gives the ArabTEX equivalents for the Arabic short vowels.

Letter	Transliteration[12] dmg	loc	ArabTEX notation
ـَ	*a*	*a*	a
ـُ	*u*	*u*	u
ـِ	*i*	*i*	i
ـً	*an*	*an*	aN
ـٌ	*un*	*un*	uN
ـٍ	*in*	*in*	iN

Table 3: Standard ArabTEX (short vowels)

Whether Arabic texts be vocalized or not is essentially a matter of personal choice. So one may use `voc` mode and decide not to write vowels except at some particular places for disambiguation purposes, or use `novoc` mode, not write vowels—as `novoc` normally does not show them—except, again, where disambiguation is needed.[13]

[11] = *al-'alif*^{u} *'l-maqṣūrat*^{u}.

[12] See below section 6 on page 22.

[13] See below section 4.4 on page 17.

However, it may be wise to always write the vowels, leaving to the various modes provided by arabluatex to take care of showing or not showing the vowels.

That said, there is no need to write the short vowels *fatḥa*, *ḍamma* or *kasra* except in the following cases:

- at the commencement of a word, to indicate that a connective *ʾalif* is needed, with the exception of the article (see below section 4.4 on page 17);
- when arabluatex needs to perform a contextual analysis to determine the carrier of the *hamza*;
- in the various transliteration modes, as vowels are always expressed in romanized Arabic.

4 arabluatex in action

4.1 The vowels and diphthongs

Short vowels As said above, they are written ⟨*a, u, i*⟩:

`_halaqa` (or `xalaqa`) خَلَقَ *ḫalaqa*, `^samsuN` شَمْسٌ *šams*^{un}, `karImuN` كَرِيمٌ *Karīm*^{un}.

`bi-hi` بِهِ *bi-hi*, `'aqi.tuN` أَقِطٌ *ʾaqi.*^{tun}.

`la-hu` لَهُ *la-hu*, `.hujjaTuN` حُجَّةٌ *ḥuǧǧat*^{un}.

Long vowels They are written ⟨*U, A, I*⟩:

`qAla` قَالَ *qāla*, `bI`a` بِيعَ *bīʿa*, `.tUruN` طُورٌ *ṭūr*$^{un}$, `.tInuN` طِينٌ *ṭīn*$^{un}$, `murU'aTuN` مُرُوءَةٌ *murūʾat*^{un}.

ʾalif maqṣūra It is written ⟨*_A*⟩ or ⟨*Y*⟩:

`al-fat_A` اَلْفَتَى *al-fatā*, `al-maqh_A` اَلْمَقْهَى *al-maqhā*, `'il_A` إِلَى *ʾilā*.

ʾalif otiosum Said *ʾalif*^{u} *ʾl-wiqāyat*^{i}, "the guarding *ʾalif*", after و at the end of a word, both when preceded by *ḍamma* and by *fatḥa* is written ⟨*UA*⟩ or ⟨*aW, aWA*⟩:

`na.sarUA` نَصَرُوا *naṣarū*, `katabUA` كَتَبُوا *katabū*, `ya.gzUA` يَغْزُوا *yaġzū*, `ramaW` رَمَوْا *ramaw*, `banaWA` بَنَوْا, *banaw*.

ʾalif maḥḍūfa* and defective *ū, ī They are written ⟨*_a, _i _u*⟩:

`al-l_ah-u` اَللّٰهُ *al-lāh*^{u}, `'il_ahuN` إِلٰهٌ *ʾilāh*^{un}.

`al-ra.hm_an-u` اَلرَّحمٰنُ *ar-raḥmān*^{u}, `l_akin` لٰكِن *lākin*, `h_ahunA` هٰهُنَا *hāhunā*, `.hunayn-u bn-u 'is.h_aq-a` حُنَيْنُ بْنُ إِسْحٰقَ *Ḥunayn*^{u} *bn*^{u} *ʾIsḥāq*^{a}, `rabb_i` رَبِّ *rabbī*, `al-`A.s_i` اَلعَاصِ *al-ʿĀṣī*.

Silent و/ي Some words ending with ـاة are usually written ـوة or ـوة instead of ـاة: see Wright (1896, i. 12 A). arabluatex preserves that particular writing; the same applies to words ending in ـية for ـاة. Long vowels ⟨*U, I*⟩ shall receive no *sukūn* after a *ʾalif maḥḍūfa* and are discarded in `trans` mode:

`.hay_aUTuN` حَيٰوةٌ *ḥayāt*^{un}, `.sal_aUTuN` صَلٰوةٌ *ṣalāt*^{un}, `mi^sk_aUTuN` مِشْكٰوةٌ *miškāt*^{un}, `tawr_aITuN` تَوْرِيةٌ *tawrāt*^{un}.

And so also: `al-rib_aIT-u` اَلرِّبٰوِيةُ *ar-ribāt*^{u}.

***ʿAmr*^{un}, and the silent و** To that name a silent و is added to distinguish it from *ʿUmar*^{u}: see Wright (1896, i. 12 C). In no way this affects the sound of the *tanwīn*, so it has to be discarded in `trans` mode:

``amruNU` عَمْرٌو *ʿamr*$^{un}$, ``amraNU` عَمْرًا *ʿamr*^{an}, ``amriNU` عَمْرٍو *ʿamr*^{in}.

When the *tanwīn* falls away (Wright 1896, i. 249 B): `amr-uU bn-u mu.hammadiN` عَمْرُو بْنُ مُحَمَّدٍ *ʿAmr*^{u} *bn*^{u} *Muḥammad*^{in}, `mu.hammad-u bn-u `amr-iU bn-i _hAlidiN` مُحَمَّدُ بْنُ عَمْرِو بْنِ خَالِدٍ *Muḥammad*^{u} *bn*^{u} *ʿAmr*^{i} *bn*^{i} *Ḫālid*^{in}.

And so also: `al-rib_aUA` اَلرِّبٰوا *ar-ribā*, `ribaNU` رِبًوا *rib*^{an}.

tanwīn The marks of doubled short vowels, ـٌ ـً ـٍ, are written ⟨*uN, aN, iN*⟩ respectively. arabluatex deals with special cases, such as ـً taking an ا after all consonants except ة, and *tanwīn* preceding ى as in هُدًى, which is written ⟨*aN_A*⟩ or ⟨*aNY*⟩:

`mAluN` مَالٌ *māl*^{un}, `bAbaN` بَابًا *bāb*^{an}, `madInaTaN` مَدِينَةً *madīnat*^{an}, `bintiN` بِنتٍ *bin*^{tin} `maqhaN_A` مَقهًى *maqhạ*^{n}, `fataNY` فَتًى *fatạ*^{n}.

arabluatex is aware of special orthographies: `^say'uN` شَيْءٌ *šayʾ*^{un}, `^say'aN` شَيْئًا *šayʾ*^{an}, `^say'iN` شَيْءٍ *šayʾ*^{in}.

In some cases, it may be useful to mark the root form of defective words so as to produce a more accurate transliteration of ending *tanwīn*. As seen above, *tanwīn* preceding ى is written ⟨*aN_A*⟩ or ⟨*aNY*⟩. Such forms as قَاضٍ may likewise be written ⟨*iNI*⟩:—

`al-qA.dI` اَلقَاضِي *al-qāḍī*, `qA.diyaN` قَاضِيًا *qāḍiy*^{an}, `qA.diNI` قَاضٍ *qāḍi*^{n}.

4.2 Other orthographic signs

tāʾ marbūṭa It is written ⟨*T*⟩:

> madInaTuN مَدِينَةٌ *madīnat*^{un}, madInaTaN مَدِينَةً *madīnat*^{an}, madInaTiN مَدِينَةٍ *madīnat*^{in}.

hamza It is written ⟨'⟩, its carrier being determined by contextual analysis. In case one wishes to bypass this mechanism, he can use the "quoting" feature that is described below in section 4.4 on page 17.

> **Initial *hamza***: 'asaduN أَسَدٌ *ʾasad*^{un}, 'u_htuN أُختٌ *ʾuḫt*^{tun}, 'iqlIduN إِقلِيدٌ *ʾiqlīd*^{un}, 'anna أَنَّ *ʾanna*, 'inna إِنَّ *ʾinna*.
>
> *hamza* followed by the long vowel و is encoded '_U: '_Ul_A أُولَى *ʾūlā*, '_UlU أُولُو *ʾūlū*, '_UlA'ika أُولَآئِكَ *ʾūlāʾika*.
>
> *hamza* followed by the long vowel ي is encoded '_I: '_ImAnuN إِيمَانٌ *ʾīmān*^{un}.
>
> **Middle *hamza***: xA.ti'-Ina خَاطِئِينَ *ḫāṭiʾ*^{ina}, ru'UsuN رُؤُوسٌ, *ruʾūs*^{un}, xa.tI'aTuN خَطِيئَةٌ *ḫaṭīʾat*^{un}, su'ila سُئِلَ *suʾila*, 'as'ilaTuN أَسئِلَةٌ *ʾasʾilat*^{un}, mas'alaTuN مَسأَلَةٌ *masʾalat*^{un}, 'as'alu أَسأَلُ *ʾasʾalu*, yatasA'alUna يَتَسَآءَلُونَ *yatasāʾalūna*, murU'aTuN مُرُوءَةٌ *murūʾat*^{un}, ta'xIruN تَأْخِيرٌ *taʾḫīr*^{un}, ta'axxara تَأَخَّرَ *taʾaḫḫara*, ji'tu-ka جِئتُكَ *ǧiʾtu-ka*, qA'iluN قَآئِلٌ *qāʾil*^{un}.
>
> From Wright (1896, i. 14 B):— All consonants, whatsoever, not even *ʾalif hèmzatum* excepted, admit of being doubled and take *tašdīd*. Hence we speak and write ra''AsuN رَأَّاسٌ *raʾʾās*^{un}, sa''AluN سَأَّالٌ *saʾʾāl*^{un}, na''AjuN نَأَّاجٌ *naʾʾāǧ*^{un}.
>
> **Final *hamza***: xa.ta'uN خَطَأٌ *ḫaṭaʾ*^{un}, xa.ta'aN خَطَأً *ḫaṭaʾ*^{an}, xa.ta'iN خَطَإٍ *ḫaṭaʾ*^{in}, 'aqra'u أَقرَأُ *ʾaqraʾu*, taqra'Ina تَقرَئِينَ *taqraʾīna*, taqra'Una تَقرَؤُونَ *taqraʾūna*, yaqra'na يَقرَأْنَ *yaqraʾna*, yaxba'Ani يَخبَآنِ *yaḫbaʾāni*, xaba'A خَبَآ *ḫabaʾā*, xubi'a خُبِئَ *ḫubiʾa*, xubi'UA خُبِئُوا *ḫubiʾū*, jA'a جَآءَ *ǧāʾa*, ridA'uN رِدَآءٌ *ridāʾ*^{un}, ridA'aN رِدَآءً *ridāʾ*^{an}, jI'a جِيءَ *ǧīʾa*, radI'iN رَدِيءٍ *radīʾ*^{in}, sU'uN سُوءٌ *sūʾ*^{un}, .daw'uN ضَوءٌ *ḍawʾ*^{un}, qay'iN قَيءٍ *qayʾ*^{in}.

`^say'uN` شَيْءٌ *šayʾun*, `^say'aN` شَيْئًا *šayʾan*, `^say'iN` شَيْءٍ *šayʾin*, `al-^say'-u` اَلشَّيْءُ *aš-šayʾu*, `'a^syA'-u` أَشْيَآءُ *ʾašyāʾu*, `'a^syA'-a` أَشْيَآءَ *ʾašyāʾa*, `.zim'aN` ظِمْئًا *ẓimʾan*.

taḫfīfu ʾl-hamzati: if the *hamza* has *ǧazma* and is preceded by *ʾalif hamzatum*, it must be changed into the letter of prolongation that is homogeneous with the preceding vowel; hence: `'a'mana` آمَنَ *ʾāmana*, `'u'minu` أُومِنُ *ʾūminu*, `'i'mAnuN` إِيمَانٌ *ʾīmānun*. For other possible ways of encoding such sequences, see on page 12 (*hamza* followed by و and ي) and the *madda* on this page.

madda At the beginning of a syllabe, *ʾalif* with *hamza* and *fatḥa* (أَ) followed by *ʾalifu ʾl-maddi* (*ʾalif* of prolongation) or *ʾalif* with *hamza* and *ǧazma* (أْ) are both represented in writing *ʾalif* with *madda*: آ (see Wright 1896, i. 25 A–B).

> Hence one should keep to this distinction and encode `'a'kulu` آكُلُ *ʾākulu* and `'AkiluN` آكِلٌ *ʾākilun* respectively.

arabluatex otherwise determines *al-ʾalifu ʾl-mamdūdatu* by context analysis.

`'is'AduN` إِسْآدٌ *ʾisʾādun*, `'AkilUna` آكِلُونَ *ʾākilūna*, `'a'mannA` آمَنَّا *ʾāmannā*, `al-qur'An-u` اَلْقُرْآنُ *al-qurʾānu*.

`jA'a` جَآءَ *ǧāʾa*, `yatasA'alUna` يَتَسَآءَلُونَ *yatasāʾalūna*, `ridA'uN` رِدَآءٌ *ridāʾun*, `xaba'A` خَبَآ *ḫabaʾā*, `yaxba'Ani` يَخْبَآنِ *yaḫbaʾāni*.

šadda *tašdīd* is either *necessary* or *euphonic*.

The necessary *tašdīd* always follows a vowel, whether short or long (see Wright 1896, i. 15 A–B). It is encoded in writing the consonant that carries it twice:

`` `allaqa `` عَلَّقَ *ʿallaqa*, `mAdduN` مَادٌّ *māddun*, `'ammara` أَمَّرَ *ammara*, `murruN` مُرٌّ *murrun*.

The euphonic *tašdīd* always follows a vowelless consonant which is passed over in pronunciation and assimilated to a following consonant. It may be found (Wright 1896, i. 15 B–16 C):—

(a) With the *solar* letters ت, ث, د, ذ, ر, ز, س, ش, ص, ض, ط, ظ, ل, ن, after the article اَلْ:—

> Unlike arabtex and arabxetex, arabluatex *never requires the solar letter to be written twice*, as it automatically generates the euphonic *tašdīd* above the let-

ter that carries it, whether the article be written in the assimilated form or not, e.g. `al-^sams-u` اَلشَّمسُ *aš-šams*^{u}, or `a^s-^sams-u` اَلشَّمسُ *aš-šams*^{u}.

`al-tamr-u` اَلتَّمرُ *at-tamr*^{u}, `al-ra.hm_an-u` اَلرَّحمٰنُ *ar-raḥmān*^{u}, `al-.zulm-u` اَلظُّلمُ *aẓ-ẓulm*^{u}, `al-lu.gaT-u` اَللُّغَةُ *al-luġat*^{u}.

(b) With the letters ر, ل, م, و, ي after ن with *ğazma*, and also after the *tanwīn*:—

Note the absence of *sukūn* above the passed over ن in the following examples, each of which is accompanied by a consistent transliteration: `min rabbi-hi` مِن رَبِّهِ *mir rabbi-hi*, `min layliN` مِن لَّيلٍ *mil layl*^{in}, `'an yaqtula` أَن يَقتُلَ *'ay yaqtula*.

With *tanwīn*: `kitAbuN mubInuN` كِتَابٌ مُّبِينٌ *kitāb*^{um} *mubīn*^{un}.

(c) With the letter ت after the dentals ث, د, ذ, ض, ط, ظ in certain parts of the verb: this kind of assimilation, e.g. لَبِثُّ for لَبِثتُ *labiṯtu*, will be discarded here, as it is largely condemned by the grammarians (see Wright 1896, i. 16 B–C).

The definite article and the *'alif*^{u} *'l-waṣl*^{i} At the beginning of a sentence, اٱ is never written, as اَلحَمدُ لِلّٰهِ; instead, to indicate that the *'alif* is a connective *'alif* (*'alif*^{u} *'l-waṣl*^{i}), the *hamza* is omitted and only its accompanying vowel is expressed:

`al-.hamd-u li-l-l_ah-i` اَلحَمدُ لِلّٰهِ *al-ḥamd*^{u} *li-llāh*^{i}.

As said above on page 5, `fullvoc` is the mode in which arabluatex expresses the *sukūn* and the *waṣla*. arabluatex will take care of doing that automatically provided that the vowel which is to be absorbed by the final vowel of the preceding word is properly encoded, like so:—

(a) Definite article at the beginning of a sentence is encoded `al-`, or `a<solar letter>-` if one wishes to mark the assimilation—which is in no way required, as arabulatex will detect all cases of assimilation.

(b) Definite article inside sentences is encoded `'l-` or `'<solar letter>-`.

(c) In all remaining cases of elision, the *'alifu 'l-waṣli* is expressed by the vowel that accompanies the omitted *hamza*: ⟨*u, a, i*⟩.

Article: `bAb-u 'l-madrasaT-i` بَابُ اَلمَدرَسَةِ *bāb*^{u} *'l-madrasat*^{i}, `al-maqAlaT-u 'l-'_Ul_A` اَلمَقَالَةُ اَلأُولَى *al-maqālat*^{u} *'l-'ūlā*, `al-lu.gaT-u 'l-`arabiyyaT-u` اَللُّغَةُ اَلعَرَبِيَّةُ *al-luġat*$^{u}$ *'l-ʿarabiyyat*$^{u}$, `fI .sinA`aT-i 'l-.tibb-i` فِي صِنَاعَةِ اَلطِّبِّ *fī ṣināʿat*^{i} *'ṭ-ṭibb*^{i}, `'il_A 'l-intiqA.d-i` إِلَى اَلاِنتِقَاضِ *'ilạ 'l-intiqāḍ*^{i}, `fI 'l-ibtidA'-i` فِي اَلاِبتِدَاءِ *fi 'l-ibtidā'*^{i}, `'abU 'l-wazIr-i` أَبُو اَلوَزِيرِ *'abu*

ʾl-wazīr^{i}, fa-lammA ra'aW 'l-najm-a فَلَمَّا رَأَوُا ٱلنَّجْمَ *fa-lammā raʾawu ʾn-nağm*^{a}.

Particles:—

(a) *li-*: *ʾalif*^{u} *ʾl-waṣl*^{i} is omitted in the article اَلْ when it is preceded by the preposition لِ: li-l-rajul-i لِلرَّجُلِ *li-r-rağul*^{i}.
If the first letter of the noun be لِ, then the ل of the article also falls away, but arabluatex is aware of that: li-l-laylaT-i لِلَّيْلَةِ *li-llaylat*^{i}.

(b) *la-*: the same applies to the affirmative particle لَ: la-l-.haqq-u لَلْحَقُّ *la-l-ḥaqq*^{u}.

(c) With the other particles, *ʾalif*^{u} *ʾl-waṣl*^{i} is expressed: fI 'l-madInaT-i فِي ٱلْمَدِينَةِ *fi ʾl-madīnat*^{i}, wa-'l-rajul-u وَٱلرَّجُلُ *wa-ʾr-rağul*^{u}, bi-'l-qalam-i بِٱلْقَلَمِ *bi-ʾl-qalam*^{i}, bi-'l-ru`b-i بِٱلرُّعْبِ *bi-ʾr-ruʿb*^{i}.

Perfect active, imperative, nomen actionis: qAla isma` قَالَ ٱسْمَعْ *qāla ʾsmaʿ*, qAla uqtul قَالَ ٱقْتُلْ *qāla ʾqtul*, huwa inhazama هُوَ ٱنْهَزَمَ *huwa ʾnhazama*, wa-ustu`mila وَٱسْتُعْمِلَ *wa-ʾštuʿmila*, qad-i in.sarafa قَدِ ٱنْصَرَفَ *qadi ʾnṣarafa*, al-iqtidAr-u اَلِاقْتِدَارُ *al-iqtidār*^{u}, 'il_A 'l-intiqA.d-i إِلَى ٱلِانْتِقَاضِ *ʾilạ ʾl-intiqāḍ*^{i}, lawi istaqbala لَوِ ٱسْتَقْبَلَ *lawi ʾštaqbala*.

Other cases: 'awi ismu-hu أَوِ ٱسْمُهُ *ʾawi ʾsmu-hu*, zayduN ibn-u `amriNU زَيْدٌ ٱبْنُ عَمْرٍو *Zayd*$^{uni}$ *ʾbn*$^{u}$ *ʿAmr*$^{in}$,[14] `umar-u ibn-u 'l-_ha.t.tAb-i عُمَرُ ٱبْنُ ٱلْخَطَّابِ *ʿUmar*^{u} *ʾbn*^{u} *ʾl-Ḫaṭṭāb*^{i},[15] imru'-u 'l-qays-i اِمْرُؤُ ٱلْقَيْسِ *Imruʾ*^{u} *ʾl-Qays*^{i}, la-aymun-u 'l-l_ah-i لَأَيْمُنُ ٱللهِ *la-ʾymun*^{u} *ʾl-lāh*^{i}.

***ʾalif*^{u} *ʾl-waṣl*^{i} preceded by a long vowel** The long vowel preceding the connective *ʾalif* is shortened in pronunciation (Wright 1896, i. 21 B–D). This does not appear in the Arabic script, but arabluatex takes it into account in some transliteration standards:—

fI 'l-nAs-i فِي ٱلنَّاسِ *fi ʾn-nās*^{i}, 'abU 'l-wazIr-i أَبُو ٱلْوَزِيرِ *ʾabu ʾl-wazīr*^{i}, fI 'l-ibtidA'-i فِي ٱلِابْتِدَآءِ *fi ʾl-ibtidāʾ*^{i}, _dU 'l-i`lAl-i ذُو ٱلِاعْلَالِ *ḏu ʾl-iʿlāl*^{i}, maqh_A 'l-'amIr-i مَقْهَى الأَمِيرِ *maqhạ ʾl-ʾamīr*^{i}.

***ʾalif*^{u} *ʾl-waṣl*^{i} preceded by a diphthong** The diphthong is resolved into two simple vowels (Wright 1896, i. 21 D–22 A) viz. *ay* → *ăĭ* and *aw* → *ăŭ*. arabluatex detects the cases in which this rule applies:—

[14] "*Zayd* is the son of *ʿAmr*": the second noun is not in apposition to the first, but forms part of the predicate. Hence زَيْدٌ ابْنُ عَمْرٍو and not زَيْدُ بْنُ عَمْرٍو, "Zayd, son of ʿAmr".

[15] "*ʿUmar* is the son of *al-Ḫaṭṭāb*" (see footnote 14).

`fI `aynay 'l-malik-i` فِي عَيْنَيِ ٱلْمَلِكِ *fī ʿaynayi ’l-malik*$^{i}$, `ix^say 'l-qawm-a` اِخْشَيِ ٱلْقَوْمَ *iḫšayi ’l-qawm*$^{a}$, `mu.s.tafaw 'l-l_ah-i` مُصْطَفَوُ ٱللهِ *muṣṭafawu ’l-lāh*^{i}.

`ramaW 'l-.hijAraT-a` رَمَوُا ٱلْحِجَارَةَ *ramawu ’l-ḥiǧārat*^{a}, `fa-lammA ra'aW 'l-najm-a` فَلَمَّا رَأَوُا ٱلنَّجْمَ *fa-lammā raʾawu ’n-naǧm*^{a}.

ʾalif*^{u} *’l-waṣl*^{i} preceded by a consonant with *sukūn The vowel which the consonant takes is either its original vowel, or that which belongs to the connective *ʾalif* or the *kasra*; in most of the cases (Wright 1896, i. 22 A–C), it is encoded explicitly, like so:—

`'antumu 'l-kA_dib-Una` أَنْتُمُ ٱلْكَاذِبُونَ *ʾantumu ’l-kāḏib*^{ūna}, `ra'aytumu 'l-rajul-a` رَأَيْتُمُ ٱلرَّجُلَ *raʾaytumu ’r-raǧul*^{a}, `mani 'l-ka_d_dAb-u` مَنِ ٱلْكَذَّابُ *mani ’l-kaḏḏāb*^{u}, `qatalati 'l-rUm-u` قَتَلَتِ ٱلرُّومُ *qatalati ’r-Rūm*^{u}.

However, the Arabic script does not shows the *kasra* which is taken by the nouns having *tanwīn* although it is explicit in pronunciation and must appear in some transliteration standards. arabluatex takes care of that automatically:—

`mu.hammaduN 'l-nabI` مُحَمَّدٌ ٱلنَّبِي *Muḥammad*^{uni} *’n-nabī*.

4.3 Special orthographies

The name of God The name of God, اَللهُ, is compounded of the article اَلْ, and إِلَاهُ (noted إِلٰهُ with the defective *ʾalif*) so that it becomes اَلْإِلَاهُ; then the *hamza* is suppressed, its vowel being transferred to the ل before it, so that there remains اَلْلٰهُ (I refer to Lane, *Lexicon*, I. 83 col. 1). Finally, the first ل is made quiescent and incorporated into the other, hence the *tašdīd* above it. As arabluatex never requires a solar letter to be written twice (see above, on page 13), the name of God is therefore encoded `al-l_ah-u` or `'l-l_ah-u`:—

`al-l_ah-u` اَللهُ *al-lāh*^{u}, `yA` | [16] `al-l_ah-u` يَا اَللهُ *yā al-lāh*^{u}, `'a-fa` | [17] `-al-l_ah-i la-ta.g`alanna` أَفَاَللهِ لَتَغْعَلَنَّ *ʾa-fa-al-lāh*$^{i}$ *la-taġʿalanna*, `bi-'l-l_ah-i` بِٱللهِ *bi-’l-lāh*$^{i}$, `wa-'l-l_ah-i` وَٱللهِ *wa-’l-lāh*$^{i}$, `bi-sm-i 'l-l_ah-i` بِسْمِ ٱللهِ *bi-sm*$^{i}$ *’l-lāh*$^{i}$, `al-.hamd-u li-l-l_ah-i` اَلْحَمْدُ لِلهِ *al-ḥamd*$^{u}$ *li-llāh*$^{i}$, `li-l-l_ah-i 'l-qA'il-u` لِلهِ ٱلْقَائِلُ *li-llāh*^{i} *’l-qāʾil*^{u}.

[16] Note the “pipe” character ‘|’ here after `yA` and below after `fa` before footnote mark 17: it is needed by the `dmg` transliteration mode as in this mode any vowel at the commencement of a word preceded by a word that ends with a vowel, either short or long, is absorbed by this vowel viz. *ʿalạ ’ṭ-ṭarīq*^{i}. See section 4.5 on page 20 on the “pipe” and section 6 on page 22 on `dmg` mode.

[17] See footnote 16.

The conjunctive اَلَّذِي Although it is compounded of the article اَلْ, the demonstrative letter ل and the demonstrative pronoun ذَا, both masculine and feminine forms that are written defectively are encoded `alla_dI` and `allatI` respectively. Forms starting with the connective *ʾalif* are encoded `'lla_dI` and `'llatI`:—

`'a_hAfu mina 'l-malik-i 'lla_dI ya.zlimu 'l-nAs-a` أَخَافُ مِنَ ٱلْمَلِكِ ٱلَّذِي يَظْلِمُ ٱلنَّاسَ *ʾaḫāfu mina ʾl-malik^{i} ʾllaḍī yaẓlimu ʾn-nās^{a}*, `` `udtu 'l-^say_h-a 'lla_dI huwa marI.duN `` عُدْتُ ٱلشَّيْخَ ٱلَّذِي هُوَ مَرِيضٌ *ʿudtu ʾš-šayḫ^{a} ʾllaḍī huwa marīḍ^{un}*, `mA 'anA bi-'lla_dI qA'iluN la-ka ^say'aN` مَا أَنَا بِٱلَّذِي قَآئِلٌ لَّكَ شَيْئًا *mā ʾanā bi-ʾllaḍī qāʾil^{ul} la-ka šayʾ^{an}*.

`'ari-nA 'lla_dayni 'a.dallA-nA mina 'l-jinn-i wa-'l-'ins-i` أَرِنَا ٱلَّذَيْنِ أَضَلَّانَا مِنَ ٱلْجِنِّ وَٱلْإِنْسِ *ʾari-nā ʾllaḍayni ʾaḍallā-nā mina ʾl-ǧinn^{i} wa-ʾl-ʾins^{i}*.

The other forms are encoded regularly as `al-l` or `'l-l`:—

`fa-'innA na_dkuru 'l-.sawt-ayni 'l-la_dayni rawaynA-humA `an ja.h.zaT-a` فَإِنَّا نَذْكُرُ ٱلصَّوْتَيْنِ ٱللَّذَيْنِ رَوَيْنَاهُمَا عَنْ جَحْظَةَ *fa-ʾinnā naḍkuru ʾṣ-ṣawt^{ayni} ʾl-laḍayni rawaynā-humā ʿan Ǧaḥẓat^{a}*.

And also: `al-la_dAni` اَللَّذَانِ *al-laḍāni*, `al-la_dayni` اَللَّذَيْنِ *al-laḍayni*, `al-latAni` اَللَّتَانِ *al-latāni*, `al-latayni` اَللَّتَيْنِ *al-latayni*, `al-lAtI` اَللَّاتِي *al-lātī*, `al-lA'|Ati`[18] اَللَّآءَاتِ *al-lāʾāti*, `al-lA'I` اَللَّآءِي *al-lāʾī*, and so forth.

4.4 Quoting

It is here referred to "quoting" after the package arabtex.[19] The "quoting" mechanism of arabluatex is designed to be very similar in effect to the one of arabtex.

To start with an example, suppose one types the following in `novoc` mode: علّم علم الهيئة; is it عُلِّمَ, *he was taught the science of astronomy*, or عَلَّمَ, *he taught the science of astronomy*? In order to disambiguate this clause, it may be sensible to put a *ḍamma* above the first ع: عُلّم علم الهيئة, which is achieved by "quoting" the vowel u, like so: `` `"ullima ``, or, with no other vowel than the required u: `` `"ullm ``.

This is how the "quoting" mechanism works: metaphorically speaking, it acts as a *toggle switch*. If something, in a given mode, is supposed to be visible, "quoting" hides it; conversely, if it is supposed not to, it makes it visible.

As shown above, "quoting" means inserting one straight double quote (`"`) *before* the letter that is to be acted upon. Its effects depend on the mode which is currently selected, either `novoc`, `voc` or `fullvoc`:—

[18] Note here the "pipe" character '|': as already stated on page 13, the sequence `'A` usually encodes *ʾalif* with *hamza* followed by *ʾalif* of prolongation, which is represented in writing *ʾalif* with *madda*: آ. The "pipe" character prevents this rule from being applied. See section 4.5 on page 20.

[19] See Lagally (2004, p. 22)

novoc In this mode, “quoting” essentially means make visible something that ought not to be so.

(a) Quoting a vowel, either short or long, makes the *ḍamma*, *fatḥa* or *kasra* appear above the appropriate consonant:—

`"ullima `ilm-a 'l-hay'aT-i عُلِّم علم الهيئة *ʿullima ʿilm*^{a} *ʾl-hayʾat*^{i}, ya.gz"UA يغزُوا *yaġzū*.

(b) The same applies when “quoting” the *tanwīn*:—

wa-'innA sawfa tudriku-nA 'l-manAyA muqadd"araT"aN وإنّا سوف تدركنا المنايا مقدَّرةً, *wa-ʾinnā sawfa tudriku-na ʾl-manāyā muqaddarat*^{an}.

(c) If no vowel follows the straight double quote, then a *sukūn* is put above the preceding consonant:—

qAla isma`" قال اسمعْ *qāla ʾsmaʿ*, jA'at" hinduN جآءَتْ هِندٌ *ǧāʾat Hind*$^{un}$, ^sabIhuN bi-man q"u.ti`at" qadamA-hu شبيه بمن قُطعتْ قدماه *šabīh*^{un} *bi-man quṭiʿat qadamā-hu*.

(d) At the commencement of a word, the straight double quote is interpreted as *ʾalif*^{u} *ʾl-waṣl*^{i}:—

wa-"ust"u`mila وٱستُعمل *wa-ʾstuʿmila*, huwa "inhazama هو ٱنهزم *huwa ʾnhazama*, al-"intiqA.du الٱنتقاض *al-intiqāḍu*.

voc In accordance with the general rule, in this mode, “quoting” makes the vowels and the *tanwīn* disappear, should this feature be required for some reason:—

(a) Short and long vowels:—

q"Ala q"A'iluN قالَ قآئِلٌ *qāla qāʾil*^{un}, ibn-u 'abI 'u.saybi`aT-"a اِبنُ أَبِي أُصَيبِعَة *Ibn*^{u} *ʾAbī ʾUṣaybiʿat*^{a}.

(b) *tanwīn*:—

madInaT"aN مَدِينَة *madīnat*^{an}, bAb"aN بَابا *bāb*^{an}, hud"aN_A هُدى *huḍa*^{n}, ^say'"iN شَيئ *šay*^{ʾin}.

One may more usefully “quote” the initial vowels to write the *waṣla* above the *ʾalif* or insert a straight double quote after a consonant not followed by a vowel to make the *sukūn* appear:—

(a) *ʾalif*^{u} *ʾl-waṣl*^{i}:—

fI "istisqA'-a فِي ٱستِسقَآءَ *fi ʾstisqāʾa*, wa-"istisqA'-u وَٱستِسقَآءُ *wa-ʾstisqāʾu*, qAla "uhrub fa-lan tuqtala قَالَ ٱهرُب فَلَن تُقتَلَ *qāla ʾhrub fa-lan tuqtala*.

(b) *sukūn*:—

qAla "uqtul" fa-lan tuqtala قَالَ ٱقتُلْ فَلَن تُقتَلَ *qāla ʾqtul fa-lan tuqtala*, mA jA'at" mini imra'aTiN مَا جَآءَتْ مِنِ امرَأَةٍ *mā ǧāʾat mini ʾmraʾat*^{in}, kam" qad" ma.dat" min" laylaTiN كَمْ قَدْ مَضَتْ مِنْ لَيلَةٍ *kam qad maḍat min laylat*^{in}.

`fullvoc` In this mode, "quoting" may be used to take away any short vowel (or *tanwīn*, as seen above) or any *sukūn*:—

al-jamr-u 'l-.sayfiyy-u 'lla_dI kAna bi-q"rAn"|nUn-a اَلْجَمْرُ ٱلصَّيْفِيُّ ٱلَّذِي كَانَ بِقرَانُنُونَ *al-ǧamr^{u} ʾṣ-ṣayfiyy^{u} ʾllaḍī kāna bi-Qrānnūn^{a}*.

4.4.1 Quoting the *hamza*

As said above in section 4.2 on page 12, the *hamza* is always written ⟨'⟩, its carrier being determined by contextual analysis. "Quoting" that straight single quote character like so: ⟨"'⟩ allows to determine the carrier of the *hamza* freely, without any consideration for the context. Table 4 gives the equivalents for all the possible carriers the *hamza* may take:

Letter	Transliteration[20]		ArabTEX notation
	dmg	loc	
ء	ʾ	’	|"'
آ	ʾā	’ā	A"'
أ	ʾ	’	a"'
أ	ʾ	’	u"'
ؤ	ʾ	’	w"'
إ	ʾ	’	i"'
ئ	ʾ	’	y"'

Table 4: "Quoted" *hamza*

As one can see from table 4, the carrier of the *hamza* is inferred from the letter that precedes the straight double quote ⟨"⟩. Of course, any "quoted" *hamza* may take a short vowel, which is to be written *after* the ArabTEX equivalent for the *hamza* itself, namely ⟨'⟩. For example, وَ is encoded ⟨w"'a⟩, while وْ is encoded ⟨w"'"⟩. In the latter example, the second straight double quote encodes the *sukūn* in `voc` mode in accordance with the rule laid above on pages 18–19.

'a`dA'ukum أَعْدَآؤُكُمْ *ʾaʿdāʾukum*, 'a`dA|"'ukum أَعْدَآءُ كُمْ *ʾaʿdāʾukum*, 'a`dA'ikum أَعْدَآئِكُمْ *ʾaʿdāʾikum*, 'a`dA|"'ikum أَعْدَآءِ كُمْ *ʾaʿdāʾikum*.

[20] See below section 6 on page 22.

4.5 The “pipe” character (|)

In the terminology of ArabTEX, the “pipe” character ‘|’ is referred to as the “invisible consonant”. Hence, as already seen above in section 4.4.1 on page 19, its usage to encode the *hamza* alone, with no carrier: `|"'` ء.

Aside from that usage, the “pipe” character is used to prevent almost any of the contextual analysis rules that are described above from being applied. Two examples have already been given to demonstrate how that particular mechanism works in footnote 16 on page 16 and in footnote 18 on page 17. One more example follows:—

`bi-qrAn|nUn-a` بِقرَانُنونَ *bi-Qrānnūn*^{a}, “in Crannon” (Thessaly, Greece).[21]

As one can see, the “pipe” character between the two ⟨*n*⟩ prevents the necessary *tašdīd* rule (page 13) from being applied.

4.6 Stretching characters: the *taṭwīl*

A double hyphen ⟨--⟩ stretches the ligature in which one letter is bound to another. Although it is always better to rely on automatic stretching, this technique may be used to a modest extent, especially to increase legibility of letters an diacritics which stand one above the other:–

`.hunayn-u bn-u 'is.h--_aq-a` حُنَيْنُ بنُ إِسحٰقَ *Ḥunayn*^{u} *bn*^{u} *ʾIsḥāq*^{a}

4.7 Digits

4.7.1 Numerical figures

The *Indian numbers*, *ar-raqam*^{u} *ʾl-hindiyy*^{u}, are ten in number, and they are compounded in exactly the same way as our numerals:—

`1874` ١٨٧٤, `123-456,789` ١٢٣-٤٥٦,٧٨٩, `fI sanaT-i 1024` فِي سَنَةِ ١٠٢٤

4.7.2 The *abjad*

The numbers may also be expressed with letters from right to left arranged in accordance with the order of the Hebrew and Aramaic alphabets (see Wright 1896, i. 28 B–C). The *ʾabǧad* numbers are usually distinguished from the surrounding words by a stroke placed over them.

`\abjad` *ʾabǧad* numbers are inserted with the command `\abjad{`⟨*number*⟩`}` in any of the `voc`, `fullvoc` and `novoc` modes, where ⟨*number*⟩ may be any number between 1 and 1999, like so:—

New feature v.1.1

`\abjad{45} kitAbu-hu fI 'l-`AdAt-i` مَة كِتَابُهُ فِي العَادَاتِ *45 kitābu-hu fi ʾl-ʿādāt*^{i}.

[21] See more context on page 19.

Rem. *a.* As can be seen in the above given example, arabluatex expresses the *ʾabǧad* numbers in Roman numerals if it finds the command `\abjad` in any of the transliteration modes.

Rem. *b.* `\abjad` may also be found outside Arabic environments. In that case, arabluatex does not print the stroke as a distinctive mark over the number for it is not surrounded by other Arabic words. In case one nonetheless wishes to print the stroke, he can use the `\aemph` command that is described below in section 4.9:—

```
The \arb[trans]{'abjad} number for 1874 is \abjad{1874} The ʾabǧad number for 1874 is غضعد.
The \arb[trans]{'abjad} number for 1874 is \aemph{\abjad{1874}} The ʾabǧad number for 1874 is غضعد.
```

4.8 Additional characters

In the manuscripts, the unpointed letters, *al-ḥurūfu ʾl-muhmalatu*, are sometimes further distinguished from the pointed by various contrivances, as explained in Wright (1896, i. 4 B–C). One may find these letters written in a smaller size below the line, or with a dot or another mark below. As representing all the possible contrivances leads to much complexity and also needs to be agreed among scholars, new ways of encoding them will be proposed and gradually included as arabluatex will mature.

For the time being, the following is included:—

Letter	Transliteration[22] dmg	loc	ArabTEX notation
ب	*ḅ*	*b*	.b
د	*ḍ*	*d*	^d
ف	*f̣*	*f*	.f
ق	*q̣*	*q*	.q
ک	*k*	*k*	.k
ں	*ṇ*	*n*	.n
﴾	*(*	*(*	((
﴿	*)*	*)*	))

Table 5: Additional Arabic codings

`'afAman.tUs` Gal.(M) `.fmn.n.ts` (sic) Gal.(E1), أفامنطوس Gal.(M) ڤمنطس (sic) Gal.(E1), *ʾafāmanṭūs* Gal.(M) *f̣mnṇṭs* (sic) Gal.(E1).

4.9 Arabic emphasis

As already seen in section 4.7.2 on page 20, the *ʾabǧad* numbers are distinguished from the surrounding words by a stroke placed over them. This technique is used to distinguish further words that are proper names or book titles.

`\aemph` One may use the command `\aemph{`⟨*Arabic text*⟩`}` to use the same technique to emphasize words, like so:—

[22] See below section 6 on the next page.

\abjad{45}: kitAbu-hu \aemph{fI 'l-`AdAt-i} كِتَابُهُ فِي العَادَاتِ :مَة 45:
kitābu-hu fi ’l-ʿĀdāt^{i}.

5 Special applications

Linguistics The same horizontal stroke as the *taṭwīl* (see section 4.6 on page 20) may be encoded ⟨*B*⟩; ⟨*BB*⟩ will receive the *tašdīd*. This is useful to make linguistic annotations and comments on vowels:—

Bu Ba Bi BuN BaN BiN ـٌـٍـً *u a i* $^{un\ an\ in}$, BBu BBa BBi ـُّـَّـِّ *u a i*, B--aN ـًا $^{-an}$, B" ـْ.

6 Transliteration

It may be more appropriate to speak of “romanization” than “transliteration” of Arabic. As seen above in section 2.2 on pages 5–7, the “transliteration mode” may be selected globally or locally.

This mode transliterates the ArabTEX input into one of the accepted standards. As said above on page 5, two standards are supported at present:

dmg *Deutsche Morgenländische Gesellschaft*, which was adopted by the International Convention of Orientalist Scholars in Rome in 1935.[23] dmg transliteration convention is selected by default;

loc *Library of Congress*: this standard is part of a large set of standards for romanization of non-roman scripts adopted by the American Library Association and the Library of Congress.[24]

More standards will be included in future releases of arabluatex.

\SetTranslitConvention

Convention The transliteration mode, which is set to dmg by default, may be changed at any point of the document by the command \SetTranslitConvention{⟨*mode*⟩}, where ⟨*mode*⟩ may be either dmg or loc. This command is also accepted in the preamble should one wish to set the transliteration mode globally, eg.:—

```
\usepackage{arabluatex}
\SetTranslitConvention{loc}
```

\SetTranslitStyle

Style Any transliterated Arabic text is printed in italics by default. This also can be changed either globally in the preamble or locally at any point of the document by the command \SetTranslitStyle{⟨*style*⟩}, where ⟨*style*⟩ may be any font shape selection command, eg. \upshape, \itshape, \slshape, and so forth. Any specific font may also be selected using the font-selecting commands of the fontspec package.

[23] See Brockelmann et al. (1935).

[24] See http://www.loc.gov/catdir/cpso/roman.html for the source document concerning Arabic language.

\cap **Proper names** Proper names or book titles that must have their first letters uppercased may be passed as arguments to the command `\cap{`⟨*word*⟩`}`. `\cap` is a clever command, for it will give the definite article *al-* in lower case in all positions. Moreover, if the inital letter, apart from the article, cannot be uppercased, viz. ʾ or ʿ, the letter next to it will be uppercased:—

`\cap{.hunayn-u} bn-u \cap{'is.h_aq-a}` حُنَيْنُ بْنُ إِسْحٰقَ *Ḥunayn*^{u} *bn*^{u} *ʾIsḥāq*^{a}, `\cap{`u_tm_an-u`}` عُثْمٰنُ *ʿUṭmān*^{u}, `.daraba \cap{zayd-u} bn-u \cap{_h_alidiN} \cap{sa`d-a} bn-a \cap{`awf-i} bn-i \cap{`abd-i} \cap{'l-l_ah-i}` ضَرَبَ زَيْدُ بْنُ خٰلِدِ سَعْدَ بْنَ عَوْفِ بْنِ عَبْدِ ٱللّٰهِ *ḍaraba Zayd*^{u} *bn*^{u} *Ḫālid*^{in} *Saʿd*^{a} *bn*^{a} *ʿAwf*^{i} *bn*^{i} *ʿAbd*^{i} *ʾl-Lāh*^{i}.

However, `\cap` must be used cautiously in some very particular cases, for the closing brace of its argument may prevent a rule from being applied. To take an example, as seen above on page 16, the transliteration of مُحَمَّدٌ ٱلنَّبِيِّ must be *Muḥammad*^{uni} *ʾn-nabī*, as nouns having the *tanwīn* take a *kasra* in pronunciation before *ʾalifu ʾl-waṣli*. In that case, encoding مُحَمَّدٌ like so: `\cap{mu.hammaduN}` is wrong, because the closing brace would prevent arabluatex from detecting the sequence ⟨*-uN*⟩ immediately followed by ⟨*ʾ-l*⟩. Fortunately, this can be circumvented in a straightforward way by inserting only part of the noun in the argument of `\cap` vz. up to the first letter that is to be uppercased, like so: `\cap{m}u.hammaduN`.

6.1 Additional note on dmg convention

New feature v1.3

According to Brockelmann et al. (1935, p. 6), Arabic *ʾiʿrāb* may be rendered into dmg in three different ways:
(a) In full: *ʿAmrun*;
(b) As superscript text: *ʿAmr*^{un};
(c) Discarded: *ʿAmr*.

\arbup By default, arabluatex applies rule b. Once delimited by a set of Lua functions, *ʾiʿrāb* is passed as an argument on to a `\arbup` command which is set to `\textsuperscript`.

\NoArbUp \ArbUpDflt `\NoArbUp` may be used either in the preamble or at any point of the document in case one wishes to apply rule a. The default rule b can be set back with `\ArbUpDflt` at any point of the document.

\SetArbUp Finally, `\SetArbUp{`⟨*formatting directives*⟩`}` may be used to customize the way *ʾiʿrāb* is displayed. To take one example, here is how Arabic *ʾiʿrāb* may be rendered as subscript text:—

```
1 \SetArbUp{\textsubscript{#1}}
2 Arabic |dmg| transliteration for \arb{ra'aytu ^gAmi`aN
3 muhaddamaTaN mi'_danatu-hu}: \arb[trans]{ra'aytu
4 ^gAmi`aN muhaddamaTaN mi'_danatu-hu.}
```

Arabic dmg transliteration for رَأَيتُ جَامِعًا مُهَدَّمَةً مِّئذَنتُهُ: *raʾaytu ǧāmiʿ$_{am}$ muhaddamat$_{am}$ miʾḏanatu-hu.*

As shown in the above example, `#1` is the token that is replaced with the actual *tanwīn* in the formatting directives of the `\SetArbUp` command.

***ʾiʿrāb* boundaries** Every declinable noun (*muʿrab*) may be declined either with or without *tanwīn*, viz. *munṣarif^{un}* or *ġayr^{u} munṣarif^{in}*. The former is automatically parsed by arabluatex, whereas the latter has to be delimited with an hyphen, like so:—

munṣarif: `mu`allimuN` مُعَلِّمٌ *muʿallim$^{un}$*, `kA'inuN` كَآئِنٌ *kāʾin$^{un}$*, `kA'inAtuN` كَآئِنَاتٌ *kāʾin$^{ātun}$*, `\cap{`amraNU}` عَمرًوا *ʿAmr^{an}*, `fataN_A` فَتًى *fataⁿ*.

ġayr munṣarif: `al-mu`allim-u` اَلْمُعَلِّمُ *al-muʿallim$^{u}$*, `kitAb-Ani` كِتَابَانِ *kitāb$^{āni}$*, `ra^sa'-Ani` رَشَآنِ *raša$^{x̄ni}$*, `sAriq-Una` سَارِقُونَ *sāriq$^{ūna}$*, `qA.d-Una` قَاضُونَ *qāḍ$^{ūna}$*, `al-.zulm-Atu` اَلظُّلمَاتُ *aẓ-ẓulm^{ātu}*.

Rem. *a.* As the *tanwīn* is passed over in pronunciation when it is followed by the letters ر, ل, م, و, ي (see item b on page 14), it may be desirable to further distinguish it by putting it above the line, but not to do the same for *ġayr munṣarif* terminations. This can be achieved by simply omitting the hyphen before any *ġayr munṣarif* termination:—

`kAna .ganiyyaN l_akinna-hu labisa ^gubbaTaN mumazzaqaN 'aydu-hA` كَانَ غَنِيًّا لٰكِنَّهُ لَبِسَ جُبَّةً مُّمَزَّقًا أَيدُهَا *kāna ġaniyy^{al} lākinna-hu labisa ǧubbat^{am} mumazzaq^{an} ʾaydu-hā.*

Rem. *b.* Although the hyphen before the *tanwīn* is optional as arabluatex always parses nouns with such termination, it may also be used to mark better the inflectional endings:—

`mana`a 'l-nAs-a kAffaT-aN min mu_hA.tabati-hi 'a.had-uN bi-sayyidi-nA` مَنَعَ النَّاسَ كَافَّةً مِن مُخَاطَبَتِهِ أَحَدٌ بِسَيِّدِنَا *manaʿa ʾn-nās^{a} kāffat^{am} mim muḫāṭabati-hi ʾaḥad^{un} bi-sayyidi-nā.*

6.2 Examples

Here follows in transliteration the story of *Ǧuḥā* and his donkey (جُحَا وَحِمَارُهُ). See the code on page 6:—

***'dmg' standard**: ʾatā ṣadīq^{un} ʾilā Ǧuḥā yaṭlubu min-hu ḥimāra-hu li-yarkaba-hu fī safrat^{in} qaṣīrat^{in}. wa-qāla la-hu: "sawfa ʾuʿīdu-hu ʾilay-ka fi 'l-masāʾi wa-ʾadfaʿu la-ka ʾuǧrat^{an}." fa-qāla Ǧuḥā: "ʾanā ʾāsif^{un} ǧidd^{an} ʾannī lā ʾaṣṭaṭīʿu ʾan ʾuḥaqqiqa la-ka ġarbata-ka fa-'l-ḥimār^{u} laysa huna 'l-yawm^{a}." wa-qabla ʾay yutimma Ǧuḥā kalāma-hu badaʾa 'l-ḥimār^{u} yanhaqu fī ʾiṣṭabili-hi. fa-qāla la-hu ṣadīqu-hu: "ʾinnī ʾasmaʿu ḥimāra-ka yā Ǧuḥā yanhaqu." fa-qāla la-hu Ǧuḥā: "ġarīb^{un} ʾamru-ka yā ṣadīqī ʾa-tuṣaddiqu 'l-ḥimār^{a} wa-tukaḏḏibu-nī?"*

***'loc' standard**: atá ṣadīqun ilá Ǧuḥā yaṭlubu min-hu ḥimāra-hu li-yarkaba-hu fī safratin qaṣīratin. wa-qāla la-hu: "sawfa u'īdu-hu ilay-ka fī al-masā'i wa-adfaʿu la-ka ujratan." fa-qāla Ǧuḥā: "anā āsifun jiddan annī lā aṣṭaṭīʿu an uḥaqqiqa la-ka gharbata-ka fa-al-ḥimāru laysa hunā al-yawma." wa-qabla an yutimma Ǧuḥā kalāma-hu bada'a*

al-ḥimāru yanhaqu fī iṣṭabili-hi. fa-qāla la-hu ṣadīqu-hu: "innī asmaʿu ḥimāra-ka yā Ǯuḥā yanhaqu." fa-qāla la-hu Ǯuḥā: "gharībun amru-ka yā ṣadīqī a-tuṣaddiqu al-ḥimāra wa-tukadhdhibu-nī?"

7 LaTeX Commands in Arabic environments

General principle LaTeX commands are accepted in Arabic environments. The general principle which applies is that single-argument commands (`\command{`⟨*arg*⟩`}`) such as `\emph{`⟨*text*⟩`}`, `\textbf{`⟨*text*⟩`}` and the like, are assumed to have Arabic text as their arguments:—

```
\abjad{45} kitAbu-hu \emph{fI 'l-\cap{`AdAt-i}}
```

مَهَ كِتَابُهُ فِي لعَادَاتِ *45 kitābu-hu fi 'l-ʿĀdāt*[i].[25]

The same applies to footnotes:—

```
\renewcommand{\footnoterule}%
   {\hfill\noindent\rule[1mm]{.4\textwidth}{.15mm}}
\begin{arab}
'inna 'abI kAna mina 'l-muqAtilaT-i\footnote{al-muqAtilaT-i:
al-muqAtil-Ina.}, wa-kAnat 'ummI min `u.zamA'-i buyUt-i
'l-zamAzimaT-i\footnote{al-zamAzimaT-u: .tA'ifaT-u mina
'l-furs-i.}.
\end{arab}
```

إِنَّ أَبِي كَانَ مِنَ المُقَاتِلَةِ[a], وَكَانَت أُمِّي مِن عُظَمَآءِ بُيُوتِ الزَّمَازِمَةِ[b]ـ

[a] اَلمُقَاتِلَة: اَلمُقَاتِلِينَ.
[b] اَلزَّمَازِمَة: طَآئِفَةٌ مِنَ الفُرسِ.

Some commands, however, do not expect running text in their arguments, or one may wish to insert English text eg. in footnotes or in marginal notes. arabluatex provides a set of commands to handle such cases.

`\LR` `\LR{`⟨*arg*⟩`}` is designed to typeset its argument from left to right. It may be used in an Arabic environment, either `\arb{`⟨*Arabic text*⟩`}` or `\begin{arab}` ⟨*Arabic text*⟩ `\end{arab}`, for short insertions of left-to-right text, or to insert any LaTeX command that would otherwise be rejected by arabluatex, such as commands the argument of which is expected to be a dimension or a unit of measurement.

`\RL` `\RL{`⟨*arg*⟩`}` does the same as `\LR{`⟨*arg*⟩`}`, but typesets its argument from right to left. Even in an Arabic environment, this command may be useful.

For example, to distinguish words with a different color, one may proceed like so:—

[25] This is odd in Arabic script, but using such features as `\emph` or `\textbf` is a matter of personal taste.

```
\begin{arab}
_tumma "intalaqa _dU 'l-qarn-ayni 'il_A 'ummaT-iN 'u_hr_A fI
\LR{\textcolor{red}{\arb[fullvoc]{((ma.tli`-i 'l-^sams-i))}}}
wa-lA binA'-a la-hum yu'amminu-hum mina 'l-^sams-i.
\end{arab}
```

ثُمَّ ٱنطَلَقَ ذُو القَرنَينِ إِلَى أُمَّةٍ أُخرَى فِي ﴿مَطْلِعِ ٱلشَّمْسِ﴾ وَلَا بِنَآءَ لَهُم يُؤَمِّنُهُم مِنَ الشَّمسِ۔

\LRfootnote
\RLfootnote

\LRfootnote{⟨*text*⟩} and \RLfootnote{⟨*text*⟩} typeset left-to-right and right-to-left footnotes respectively in Arabic environments. Unlike \footnote{⟨*text*⟩}, the arguments of both \LRfootnote and \RLfootnote are not expected to be Arabic text. For example, \LRfootnote may be used to insert English footnotes in running Arabic text:—

```
\arb[fullvoc]{\cap{z}ayd-uN\LRfootnote{%
\enquote{\arb[trans]{\cap{z}ayd} is the son of
\arb[trans]{\cap{`a}mr}}: the second
noun is not in apposition to the first, but forms
part of the predicate\ldots} "ibn-u \cap{`a}mr-iNU}
```

زَيْدٌ[a] ٱبْنُ عَمْرٍو

[a] "*zayd* is the son of *ʿamr*": the second noun is not in apposition to the first, but forms part of the predicate…

When footnotes are typeset from right to left, it may happen that the numbers of the footnotes that are at the bottom of the page be typeset in the wrong direction. For example, instead of an expected number 18, one may get 81. arabluatex is not responsible for that, but should it happen, it may be necessary to redefine in the preamble the LaTeX macro \thefootnote like so:—

```
\renewcommand*{\thefootnote}{\textsuperscript{\LR{\arabic{footnote}}}}
```

\FixArbFtnmk

Another solution is to put in the preamble, below the line that loads arabluatex, the command \FixArbFtnmk. However, for more control over the layout of footnotes marks, it is advisable to use the package scrextend.[26]

\LRmarginpar

The command \LRmarginpar does for marginal notes the same as \LRfootnote does for footnotes. Of course, it is supposed to be used in Arabic environments. Note that \marginpar also works in Arabic environments, but it acts as any other single-argument command inserted in Arabic environments. The general principle laid on page 25 applies.

\setRL
\setLR

\setRL and \setLR may be used to change the direction of paragraphs, either form left to right or from right to left. As an example, an easy way to typeset a right-to-left sectional title follows:—

[26] See http://ctan.org/pkg/koma-script; read the documentation of KOMA-script for details about the \deffootnotemark and \deffootnote commands.

```
\setRL
\section*{\arb{barzawayhi li-buzurjumihra bn-i 'l-buxtikAni}}
\setLR
\begin{arab}
qAla barzawayhi bn-u 'azhar-a, ra's-u 'a.tibbA'-i fAris-a...
\end{arab}
```

بَرزَوَيهِ لِبُزُرجُمهرَ بنِ البُختِكَانِ

قَالَ بَرزَوَيهِ بنُ أَزهَرَ، رَأسُ أَطِبَّآءِ فَارِسَ---

7.1 csquotes

The recommended way of inserting quotation marks in running Arabic text is to use csquotes. With the help of the `\DeclareQuoteStyle` command, one can define an Arabic style, like so:—

```
\usepackage{csquotes}
\DeclareQuoteStyle{arabic}
{\rmfamily\textquotedblright}{\rmfamily\textquotedblleft}
{\rmfamily\textquoteright}{\rmfamily\textquoteleft}
```

Then, use this newly defined style with `\setquotestyle`, like so:—

```
\setquotestyle{arabic}
\begin{arab}
  fa-qAla la-hu ju.hA: \enquote{.garIb-uN 'amru-ka yA .sadIqI
    'a-tu.saddiqu 'l-.himAr-a wa-tuka_d_dibu-nI?}
\end{arab}
\setquotestyle{english}
```

فَقَالَ لَهُ جُحَا: "غَرِيبٌ أَمرُكَ يَا صَدِيقِي أَتُصَدِّقُ الحِمَارَ وَتُكَذِّبُنِي؟"

Rem. Do not forget to set back the quoting style to its initial state once the Arabic environment is closed. See the last line in the code above.

7.2 reledmac

The two-arguments command `\edtext{`⟨*lemma*⟩`}{`⟨*commands*⟩`}` is supported inside `\begin{arab}` ... `\end{arab}`. As an example, one may get arabluatex and reledmac to work together like so:—

```
\beginnumbering
\pstart
\begin{arab}
wa-ya.sIru ta.hta 'l-jild-i
\edtext{\arb{.sadId-uN}}{\Afootnote{M: \arb{.sadId-aN} E1}}
\end{arab}
\pend
\endnumbering
```

8 Future work

A short, uncommented, list of what is planned in the versions of arabluatex to come follows:

(a) Short-term:
   i. Support for typesetting Arabic poetry.
   ii. The *Qurʾān*: support for typesetting the *Qurʾān*.
   iii. `TEI xml` support: arabluatex will interoperate with `TEI xml` through new global and local options that will output Arabic in a `TEI xml` compliant file in addition to the usual PDF output: see on page 3.

(b) Medium-term:
   i. More languages: the list of supported languages will eventually be the same as arabtex: see footnote 4 on page 4.
   ii. Formulate propositions for extending the ArabTEX notation and the transliteration tables. Include them in arabluatex. See section 4.8 on page 21.

9 Implementation

The most important part of arabluatex relies on Lua functions and tables. Read the `.lua` files that accompany arabluatex for more information.

```
\NeedsTeXFormat{LaTeX2e}
\ProvidesPackage{arabluatex}%
[2016/05/28 v1.3 ArabTeX-like interface for LuaLaTeX]
\RequirePackage{ifluatex}
```

arabluatex requires LuaLaTeX of course. Issue a warning if the document is processed with another engine.

```
\ifluatex\else
 \PackageError{arabluatex}{lualatex needed}{%
  Package `arabluatex' needs LuaTeX.\MessageBreak
  So you should use `lualatex' to process your document.\MessageBreak
  See documentation of `arabluatex' for further information.}%
  \expandafter\expandafter\expandafter\csname endinput\endcsname
\fi
```

Declare the global options, and define them:

```
\DeclareOption{voc}{\def\al@mode{voc}}
\DeclareOption{fullvoc}{\def\al@mode{fullvoc}}
\DeclareOption{novoc}{\def\al@mode{novoc}}
\DeclareOption{trans}{\def\al@mode{trans}}
\ExecuteOptions{voc}
\ProcessOptions\relax
\def\al@mode@voc{voc}
\def\al@mode@fullvoc{fullvoc}
\def\al@mode@novoc{novoc}
\def\al@mode@trans{trans}
```

The following line will be used in the next release of arabluatex:

```
% \newif\ifal@mode@defined
```

Packages that are required by arabluatex:

```
\RequirePackage{fontspec}
\RequirePackage{amsmath}
\RequirePackage{etoolbox}
\RequirePackage{luacode}
\RequirePackage{xparse}
\RequirePackage{environ}
```

Here begins the real work: load `arabluatex.lua`:

```
\luadirect{dofile(kpse.find_file("arabluatex.lua"))}
```

This is needed by the current versions of polyglossia and luabidi. luabidi provides a `\Footnote` command. Use it as well if it is loaded.

```
\luadirect{tex.enableprimitives("luatex",tex.extraprimitives("omega"))}
```

Font setup. If no Arabic font is selected, issue a warning message and attempt to load the Amiri font which is included in TeXlive:

```
\AtBeginDocument{\ifdefined\arabicfont\relax\else
\PackageWarning{arabluatex}{\string\arabicfont\ is not defined.^^JI
  will try to load Amiri}%
\newfontfamily\arabicfont[Script=Arabic]{Amiri}\fi}%
```

`\setRL` This neutralizes what is defined by the same command in luabidi:

```
\AtBeginDocument{\def\setRL{\pardir TRT\textdir TRT}}
```

`\setLR` The same applies to `\setLR`:

```
\AtBeginDocument{\def\setLR{\pardir TLT\textdir TLT}}
```

`\LR` This command typesets its argument from left to right. As `\LR` may be already defined, we need to redefine for it to suit our purpose:

```
\AtBeginDocument{\ifdef{\LR}%
  {\RenewDocumentCommand{\LR}{m}{\bgroup\textdir TLT\rmfamily#1\egroup}}
  {\NewDocumentCommand{\LR}{m}{\bgroup\textdir TLT\rmfamily#1\egroup}}}
```

`\RL` This one typesets its argument from right to left. Same remark as above regarding the need of redefinition.

```
\AtBeginDocument{\ifdef{\RL}%
```

```
  {\RenewDocumentCommand{\RL}{m}{\bgroup\textdir TRT\rmfamily#1\egroup}}
  {\NewDocumentCommand{\RL}{m}{\bgroup\textdir TRT#1\rmfamily\egroup}}}
```

\aemph Arabic emphasis. Needs to be redefined as well.

```
\AtBeginDocument{\ifdef{\aemph}%
  {\RenewDocumentCommand{\aemph}{m}{$\overline{\text{#1}}$}}
  {\NewDocumentCommand{\aemph}{m}{$\overline{\text{#1}}$}}}
```

\SetArbEasy
\SetArbDflt
By default, arabluatex applies complex rules to generate euphonic *tašdīd*, *ʾalif mamdūda* and *sukūn* depending on the modes which are selected, either `voc`, `fullvoc` or `trans`. Such refinements can be discarded with `\SetArbEasy`, either globally in the preamble or at any point of the document. Default complex rules can be set back at any point of the document with `\SetArbDflt`.

```
\def\al@arb@rules{dflt}
\NewDocumentCommand{\SetArbEasy}{}{\def\al@arb@rules{easy}}
\NewDocumentCommand{\SetArbDflt}{}{\def\al@arb@rules{dflt}}
```

\SetTranslitStyle By default any transliterated Arabic text is printed in italics. This can be changed either globally in the preamble or at any point of the document:

```
\def\al@trans@style{\itshape}%
\NewDocumentCommand{\SetTranslitStyle}{m}{\def\al@trans@style{#1}}
```

\SetTranslitConvention `\SetTranslitConvention{⟨convention⟩}` may be used to change the transliteration convention, which is `dmg` by default:

```
\def\al@trans@convention{dmg}
\NewDocumentCommand{\SetTranslitConvention}{m}{\def\al@trans@convention{#1}}
```

\arbup
\NoArbUp
\ArbUpDflt
\SetArbUp
By default, `\arbup` is set to `\textsuperscript`. This is how the *tanwīn* that takes place at the end of a word should be displayed in `dmg` mode. `\NoArbUp` may be used either in the preamble or at any point of the document in case one wishes to have the *tanwīn* on the line. The default rule can be set back with `\ArbUpDflt` at any point of the document. Finally `\SetArbUp` may be used to customize the way *tanwīn* is displayed: this command takes the formatting directives as argument, like so: `\SetArbUp{⟨code⟩}`.

```
\NewDocumentCommand{\al@arbup@dflt}{m}{\textsuperscript{#1}}%
\NewDocumentCommand{\al@arbup}{m}{\al@arbup@dflt{#1}}
\NewDocumentCommand{\arbup}{m}{\al@arbup{#1}}
\NewDocumentCommand{\ArbUpDflt}{}{\let\al@arbup=\al@arbup@dflt}
\NewDocumentCommand{\NoArbUp}{}{\RenewDocumentCommand{\al@arbup}{m}{##1}}
\NewDocumentCommand{\SetArbUp}{m}{\RenewDocumentCommand{\al@arbup}{m}{#1}}
```

\cap Proper Arabic names or book titles should be passed to the command `\cap` so that they have their first letters uppercased. `\cap` is actually coded in Lua.

```
\DeclareDocumentCommand{\cap}{m}%
  {\luadirect{tex.sprint(cap(\luastringN{#1}))}}
```

\txarb `\txarb` sets the direction to right-to-left and selects the Arabic font. As it is supposed to be used internally by several Lua functions, this command is not documented, but available to the user should he wish to insert `utf8` Arabic text in his document.

\txtrans `\txtrans` is used internally by several Lua functions to insert transliterated Arabic text.

```
\DeclareDocumentCommand{\txarb}{+m}{\bgroup\textdir
  TRT\arabicfont#1\egroup}
\DeclareDocumentCommand{\txtrans}{+m}{\bgroup\textdir
  TLT\rmfamily#1\egroup}
```

\arb The `\arb` command detects which Arabic mode is to be used, either globally if no option is set, or locally, then passes its argument to the appropriate Lua function.

```
\DeclareDocumentCommand{\arb}{O{\al@mode} +m}%
{\edef\@tempa{#1}%
  \ifx\@tempa\al@mode@voc%
  \bgroup\textdir TRT\arabicfont%
  \luadirect{tex.sprint(processvoc(\luastringN{#2},
    \luastringO{\al@arb@rules}))}\egroup%
  \else%
  \ifx\@tempa\al@mode@fullvoc%
  \bgroup\textdir TRT\arabicfont%
  \luadirect{tex.sprint(processfullvoc(\luastringN{#2},
    \luastringO{\al@arb@rules}))}\egroup%
  \else%
  \ifx\@tempa\al@mode@novoc%
  \bgroup\textdir TRT\arabicfont%
  \luadirect{tex.sprint(processnovoc(\luastringN{#2},
    \luastringO{\al@arb@rules}))}\egroup%
  \else%
  \ifx\@tempa\al@mode@trans%
  \bgroup\textdir TLT\al@trans@style%
  \luadirect{tex.sprint(processtrans(\luastringN{#2},
    \luastringO{\al@trans@convention},
    \luastringO{\al@arb@rules}))}\egroup%
  \else%
  \fi\fi\fi\fi}
```

arab The `arab` environment does for paragraphs the same as `\arb` does for short insertions of Arabic text.

```
\NewEnviron{arab}[1][\al@mode]%
{\par\edef\@tempa{#1}%
  \ifx\@tempa\al@mode@voc%
  \bgroup\pardir TRT\textdir TRT\arabicfont%
  \luadirect{tex.sprint(processvoc(\luastringO{\BODY},
    \luastringO{\al@arb@rules}))}\egroup%
  \else%
  \ifx\@tempa\al@mode@fullvoc%
  \bgroup\pardir TRT\textdir TRT\arabicfont%
  \luadirect{tex.sprint(processfullvoc(\luastringO{\BODY},
    \luastringO{\al@arb@rules}))}\egroup%
  \else%
  \ifx\@tempa\al@mode@novoc%
```

```
  \bgroup\pardir TRT\textdir TRT\arabicfont%
  \luadirect{tex.sprint(processnovoc(\luastringO{\BODY},
    \luastringO{\al@arb@rules}))}\egroup%
  \else \ifx\@tempa\al@mode@trans%
  \bgroup\pardir TLT\textdir TLT\al@trans@style%
  \luadirect{tex.sprint(processtrans(\luastringO{\BODY},
    \luastringO{\al@trans@convention},
    \luastringO{\al@arb@rules}))}\egroup%
  \else \fi\fi\fi\fi}[\par]
```

`\abjad` `\abjad{⟨number⟩}` expresses its argument in Arabic letters in accordance with the *ʾabǧad* arrangement of the alphabet. ⟨*number*⟩ must be between 1 and 1999. It is now coded in Lua so that polyglossia is no longer needed. See `arabluatex.lua` for more information.

```
\AtBeginDocument{%
  \ifdefined\abjad%
  \RenewDocumentCommand{\abjad}{m}%
  {\luadirect{tex.sprint(abjadify(#1))}}%
  \else%
  \NewDocumentCommand{\abjad}{m}%
  {\luadirect{tex.sprint(abjadify(#1))}}
\fi}
```

`\LRmarginpar` `\LRmarginpar` is supposed to be inserted in an Arabic environment. It typsets his argument in a marginal note from left to right.

```
\DeclareDocumentCommand{\LRmarginpar}{m}{\marginpar{\textdir TLT #1}}
```

`\LRfootnote` `\LRfootnote` and `\RLfootnote` are supposed to be used in Arabic environments for insertions of non Arabic text. `\LRfootnote` typesets its argument left-to-right...

`\RLfootnote` while `\RLfootnote` typesets its argument left-to-right.

```
\DeclareDocumentCommand{\LRfootnote}{m}{\bgroup\pardir
  TLT\LR{\footnote{#1}}\egroup}
\DeclareDocumentCommand{\RLfootnote}{m}{\bgroup\pardir
  TRT\LR{\footnote{#1}}\egroup}
```

`\FixArbFtnmk` In the preamble, just below `\usepackage{arabluatex}`, `\FixArbFtnmk` may be of some help in case the footnote numbers at the bottom of the page are printed in the wrong direction. This quick fix uses and loads screxteend if it is not already loaded.

```
\NewDocumentCommand{\FixArbFtnmk}{}{%
  \@ifpackageloaded{screxteend}%
  {\AtBeginDocument{\deffootnote{2em}{1.6em}{\LR{\thefootnotemark}.\enskip}}}%
  {\RequirePackage{screxteend}
  \AtBeginDocument{\deffootnote{2em}{1.6em}{\LR{\thefootnotemark}.\enskip}}}}
```

That is it. Say goodbye before leaving.

```
\endinput
```

References

Brockelmann, Carl et al. (1935). "Die Transliteration der arabischen Schrift in ihrer Anwendung auf die Hauptliteratursprachen der islamischen Welt". In: *Denkschrift dem 19. internationalen Orientalistenkongreß in Rom*. In collab. with Ph. S. van Ronkel and Otto Spies. Deutschen Morgenländischen Gesellschaft. Leipzig: Deutschen Morgenländischen Gesellschaft, in Kommission bei F. A. Brockaus. URL: http://www.naher-osten.uni-muenchen.de/studium_lehre/werkzeugkasten/dmgtransliteration.pdf.

Hosny, Khaled (2015). *Amiri*. URL: http://www.amirifont.org/.

Lagally, Klaus (2004). *ArabTEX. Typesetting Arabic and Hebrew*. User Manual Version 4.00. Version 4.00. URL: http://mirrors.ctan.org/language/arabic/arabtex/doc/html/arabtex.htm.

Lane, Edward William (1863–1893). *An Arabic-English lexicon*. 8 vols. London – Edinburgh: Williams and Norgate.

Wright, W. LL.D (1896). *A Grammar of the Arabic Language*. Rev. by W. Robertson Smith and M. J. de Goeje. With a forew. by Pierre Cachia. 3rd ed. 2 vols. Beirut: Librairie du Liban.

Change History

Index

Numbers written in italic refer to the page where the corresponding entry is described; numbers underlined refer to the code line of the definition; numbers in roman refer to the code lines where the entry is used.

V

W