

# The arabluatex package

## v1.4.5 – 2016/10/05

Robert Alessi  
[alessi@robertalessi.net](mailto:alessi@robertalessi.net)

## Contents

<b>License and disclaimer</b>	<b>2</b>		
<b>1 Introduction</b>	<b>2</b>		
1.1 arabluatex is for Lua <sup>A</sup> T <sub>E</sub> X . . . . .	4	4.7 Digits . . . . .	21
<b>2 The basics of arabluatex</b>	<b>4</b>	4.7.1 Numerical figures . . . . .	21
2.1 Activating arabluatex . . . . .	4	4.7.2 The <i>abjad</i> . . . . .	21
2.2 Options . . . . .	5	4.8 Additional characters . . . . .	21
2.2.1 Classic contrasted with modern typesetting of Arabic	5	4.9 Arabic emphasis . . . . .	22
2.3 Typing Arabic . . . . .	6	<b>5 Special applications</b>	<b>22</b>
2.3.1 Local options . . . . .	7	<b>6 Transliteration</b>	<b>23</b>
<b>3 Standard Arab<sub>T</sub>E<sub>X</sub> input</b>	<b>8</b>	6.1 Additional note on <code>dmg</code> convention .	25
3.1 Consonants . . . . .	8	6.2 Examples . . . . .	26
3.2 Vowels . . . . .	9	<b>7 Buckwalter input scheme</b>	<b>26</b>
3.2.1 Long vowels . . . . .	9	<b>8 L<sup>A</sup>T<sub>E</sub>X Commands in Arabic environments</b>	<b>29</b>
3.2.2 Short vowels . . . . .	9	8.1 <code>csquotes</code> . . . . .	31
<b>4 arabluatex in action</b>	<b>10</b>	8.2 <code>reledmac</code> . . . . .	32
4.1 The vowels and diphthongs . . . . .	10	<b>9 Future work</b>	<b>32</b>
4.2 Other orthographic signs . . . . .	12	<b>10 Implementation</b>	<b>32</b>
4.3 Special orthographies . . . . .	17	<b>References</b>	<b>37</b>
4.4 Quoting . . . . .	18	<b>Change History</b>	<b>38</b>
4.4.1 Quoting the <i>hamzah</i> . . . . .	19	<b>Index</b>	<b>38</b>
4.5 The “pipe” character ( <code>()</code> ) . . . . .	20		
4.6 Stretching characters: the <i>taṭwīl</i> .	21		

## List of Tables

1 Arab <sub>T</sub> E <sub>X</sub> consonants . . . . .	8	4 “Quoted” <i>hamzah</i> . . . . .	20
2 Arab <sub>T</sub> E <sub>X</sub> long vowels . . . . .	9	5 Additional Arabic codings . . . . .	22
3 Arab <sub>T</sub> E <sub>X</sub> short vowels . . . . .	10	6 Buckwalter scheme . . . . .	27

### Abstract

This package provides for Lua<sup>A</sup>T<sub>E</sub>X an ArabT<sub>E</sub>X-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 ArabT<sub>E</sub>X input notation. Its output can be set in the same modes of vocalization as ArabT<sub>E</sub>X, 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 disclaimer

Copyright © 2016 Robert Alessi

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

- email: <[alessi@robertalessi.net](mailto: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 ArabT<sub>E</sub>X,<sup>1</sup> ArabLuaT<sub>E</sub>X 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 Lua<sup>A</sup>T<sub>E</sub>X users.

---

<sup>1</sup>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 *Nashī* 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”.<sup>2</sup> Even if the same can be said about Unicode, Arab $\TeX$  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 particularly 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).<sup>3</sup> Nonetheless, it remains that inserting all the time these invisible characters in complex documents rapidly becomes confusing and cumbersome.

The great advantage of Arab $\TeX$  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. Arab $\TeX$  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, Arab $\TeX$  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.

---

<sup>2</sup>Lagally (2004, p. 2).

<sup>3</sup>Gáspár Sinai’s Yudit probably has the best Unicode support. See <http://www.yudit.org>.

## 1.1 arablutex is for Lua $\LaTeX$

It goes without saying that arablutex requires Lua $\LaTeX$ .  $\TeX$  and  $\LaTeX$  have arabtex, and X $\LaTeX$  has arabxetex. Both of them are much more advanced than arablutex, as they can process a number of different languages,<sup>4</sup> whereas arablutex can process only Arabic for the time being. More languages will be included in future releases of arablutex.

In comparison to arabxetex, arablutex works in a very different way. The former relies on the `TECkit` engine which converts Arab $\TeX$  input on the fly into Unicode Arabic script, whereas the latter passes Arab $\TeX$  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 from what is not, it is processed by other Lua functions which rely on different sets of correspondence 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 Arab $\TeX$  input yet to be processed—for further processing.

## 2 The basics of arablutex

### 2.1 Activating arablutex

As usual put in your preamble:

```
\usepackage{arablutex}
```

The only requirement of arablutex is Lua $\LaTeX$ ; it will complain if you try to compile your document with another engine. That aside, arablutex 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 arablutex. Assuming that fontspec is loaded, put this in your preamble just above the line that loads arablutex:

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

where  $\langle fontname \rangle$  is the standard name of the Arabic font you wish to use.

By default, if no Arabic font is selected, arablutex will issue a warning message and attempt to load the Amiri font<sup>5</sup> like so:—

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

---

<sup>4</sup>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.

<sup>5</sup>Hosny (2015).

REM. By default Amiri places the *kasrah* in combination with the *tašdīd* below the consonant, like so: َ. That is correct, as at least in the oldest manuscripts َ may stand for َ as well as َ. See Wright (1896, i.14.C–D). The placement of the *kasrah* above the consonant may be obtained by selecting the `ss05` feature of the Amiri font, like so:—<sup>6</sup>

```
\newfontfamily\arabicfont{Amiri}[Script=Arabic,RawFeature={+ss05}]
```

Other Arabic fonts may behave differently.

## 2.2 Options

`arabluatex` may be loaded with four mutually exclusive global options, each of which may be overridden 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: *ḍammah* (◌ُ), *fathah* (◌َ) and *kasrah* (◌ِ). If a vowel is followed by *N*, viz.  $\langle uN, aN, iN \rangle$ , then the corresponding *tanwīn* (◌ِ◌◌, ◌◌◌, ◌◌◌ or ◌◌◌) is generated. Finally,  $\langle u, a, i \rangle$  at the commencement of a word indicate a “connective ‘*alif*” (*‘alifu ‘l-waṣli*), but `voc` mode does not show the *waṣlah* above the ‘*alif*; instead, the accompanying vowel may be expressed at the beginning of a sentence (◌◌◌◌◌).

`fullvoc`

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

`novoc`

None of the diacritics is showed in `novoc` mode, unless otherwise specified (see “quoting” technique below section 4.4 on page 18).

`trans`

This mode transliterates the Arab<sub>TEX</sub> input into one of the accepted standards. At present, two standards are supported (see below section 6 on page 23 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

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

- ‘Classic’ *maddah*: when ‘*alif* and *hamzah* accompanied by a simple vowel or *tanwīn* is preceded by an ‘*alif* of prolongation (◌◌◌), then a mere *hamzah* is written on the line, and a *maddah* is placed over the ‘*alif*, like so:—

samA'uN سَمَاءُ *samā‘un*, ja'a جَاءَ *ǧā‘a*, yatasA'alUna يَتَسَاءَلُونَ *yatasā‘alūna*<sup>7</sup>  
(see on page 13 for further details).

<sup>6</sup>See the documentation of `amiri`, Hosny (2015, p. 5).

<sup>7</sup>Note that in old mss. such forms as سَمَاءُ, جَاءُ are also found; see Wright (1896, i. 24 D).

New feature  
v1.4.4

`\SetArbEasy`  
`\SetArbEasy*`  
`\SetArbDflt`

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

Such refinements may be discarded by the command `\SetArbEasy`, either globally in the preamble or locally at any point of the document. The difference between `\SetArbEasy` and its ‘starred’ version `\SetArbEasy*` is that the former keeps the *sukūn* that is generated by the `fullvoc` mode, while the latter further takes it away. Default ‘classic’ rules may be set back at any point of the document with the command `\SetArbDflt`. Examples follow:—

(a) `\SetArbDflt`:

i. voc وَمَاتَ اسْتَسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَاءِ

ii. fullvoc وَمَاتَ اسْتَسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَاءِ

iii. trans *wa-māta 'stisqā<sup>an</sup> qabla 'ay yutimma kitāba-hu fī nuġūm<sup>i</sup> 's-samā<sup>i</sup>*

(b) `\SetArbEasy`:

i. voc وَمَاتَ اسْتَسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَاءِ

ii. fullvoc وَمَاتَ اسْتَسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَاءِ

iii. trans *wa-māta 'stisqā<sup>an</sup> qabla 'an yutimma kitāba-hu fī nuġūm<sup>i</sup> 's-samā<sup>i</sup>*

(c) `\SetArbEasy*`:

i. voc وَمَاتَ اسْتَسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَاءِ

ii. fullvoc وَمَاتَ اسْتَسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نُجُومِ السَّمَاءِ

iii. trans *wa-māta 'stisqā<sup>an</sup> qabla 'an yutimma kitāba-hu fī nuġūm<sup>i</sup> 's-samā<sup>i</sup>*

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:—

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

```
1 \begin{arab}
2 [...]
3 \end{arab}
```

like so:—

```
1 \begin{arab}
2 'at_A .sadIquN 'il_A ju.hA ya.tlubu min-hu .himAra-hu
3 li-yarkaba-hu fI safraTiN qa.sIraTiN fa-qAla la-hu:
4 \enquote{sawfa 'u`Idu-hu 'ilay-ka fI 'l-masA'-i
5 wa-'adfa`u la-ka 'ujraTaN.} fa-qAla ju.hA:
6 \enquote{'anA 'AsifuN jiddaN 'annI lA 'asta.tI`u 'an
7 'u.haqqiqa la-ka ra.gbata-ka fa-'l-.himAr-u laysa hunA
8 'l-yawm-a.} wa-qabla 'an yutimma ju.hA kalAma-hu bada'a
9 'l-.himAr-u yanhaqu fI 'i.s.tabli-hi. fa-qAla la-hu
10 .sadIqu-hu: \enquote{'innI 'asma`u .himAra-ka yA ju.hA
11 yanhaqu.} fa-qAla la-hu ju.hA: \enquote{.garIbuN
12 'amru-ka yA .sadIQI 'a-tu.saddiqu 'l-.himAr-a
13 wa-tuka_d_diba-nI?}
14 \end{arab}
```

أَتَى صَدِيقٌ إِلَى جَحَّا يَطْلُبُ مِنْهُ حِمَارَهُ لِيُرْكَبَهُ فِي سَفَرَةٍ قَصِيرَةٍ فَقَالَ لَهُ: "سَوْفَ أُعِيدُهُ إِلَيْكَ فِي الْمَسَاءِ  
وَأَدْفَعُ لَكَ أُجْرَةً." فَقَالَ جَحَّا: "أَنَا أَسَفٌ جِدًّا أَيُّ لَا أَسْتَطِيعُ أَنْ أَحَقِّقَ لَكَ رَغْبَتَكَ فَالْحِمَارُ لَيْسَ هُنَا  
الْيَوْمَ." وَقَبِلَ أَنْ يُتَمَّ جَحَّا كَلَامَهُ بَدَأَ الْحِمَارُ يَنْهَقُ فِي إِصْطَبَلِهِ. فَقَالَ لَهُ صَدِيقُهُ: "إِنِّي أَسْمَعُ حِمَارَكَ يَا جَحَّا  
يَنْهَقُ." فَقَالَ لَهُ جَحَّا: "غَرِيبٌ أَمْرُكَ يَا صَدِيقِي أَتَصَدِّقُ الْحِمَارَ وَتُكَلِّمُنِي؟"

### 2.3.1 Local options

As seen above in section 2.2 on page 5, arabuatex 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 overridden 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]{\langle Arabic text \rangle}`;  
`fullvoc`   – `\arb[fullvoc]{\langle Arabic text \rangle}`;  
`novoc`       – `\arb[novoc]{\langle Arabic text \rangle}`;  
`trans`       – `\arb[trans]{\langle Arabic text \rangle}`.

The same optional arguments may be passed to the environment `arab`: one may have `\begin{arab}[\langle mode \rangle] \dots \end{arab}`, where `\langle mode \rangle` 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 <sup>8</sup>		ArabTeX notation
	dmg	loc	
ا <sup>9</sup>	<i>a</i>	<i>a</i>	<code>a</code>
ب	<i>b</i>	<i>b</i>	<code>b</code>
ت	<i>t</i>	<i>t</i>	<code>t</code>
ث	<i>ṭ</i>	<i>th</i>	<code>_t</code>
ج	<i>ǧ</i>	<i>j</i>	<code>˘g</code> or <code>j</code>
ح	<i>ḥ</i>	<i>h</i>	<code>.h</code>
خ	<i>ḫ</i>	<i>kh</i>	<code>_h</code> or <code>x</code>
د	<i>d</i>	<i>d</i>	<code>d</code>
ذ	<i>ḍ</i>	<i>dh</i>	<code>_d</code>
ر	<i>r</i>	<i>r</i>	<code>r</code>
ز	<i>z</i>	<i>z</i>	<code>z</code>
س	<i>s</i>	<i>s</i>	<code>s</code>
ش	<i>š</i>	<i>sh</i>	<code>˘s</code>
ص	<i>ṣ</i>	<i>ṣ</i>	<code>.s</code>
ض	<i>ḍ</i>	<i>ḍ</i>	<code>.d</code>
ط	<i>ṭ</i>	<i>ṭ</i>	<code>.t</code>
ظ	<i>ẓ</i>	<i>ẓ</i>	<code>.z</code>
ع	<i>ʿ</i>	<i>ʿ</i>	<code>`</code>
غ	<i>ǧ</i>	<i>gh</i>	<code>.g</code>
ف	<i>f</i>	<i>f</i>	<code>f</code>
ق	<i>q</i>	<i>q</i>	<code>q</code>
ك	<i>k</i>	<i>k</i>	<code>k</code>
ل	<i>l</i>	<i>l</i>	<code>l</code>
م	<i>m</i>	<i>m</i>	<code>m</code>
ن	<i>n</i>	<i>n</i>	<code>n</code>

<sup>8</sup>See below section 6 on page 23.

<sup>9</sup>For *ʿalif* as a consonant, see Wright (1896, i. 16 D). The *hamzah* itself is encoded `<'>`. See below section 4.2 on page 12.

Letter	Transliteration		ArabTeX notation
	dmg	loc	
ه	<i>h</i>	<i>h</i>	h
و	<i>w</i>	<i>w</i>	w
ي	<i>y</i>	<i>y</i>	y
ة	<i>ah</i>	<i>ah</i>	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 <sup>10</sup>		ArabTeX notation
	dmg	loc	
ا	<i>ā</i>	<i>ā</i>	A
و	<i>ū</i>	<i>ū</i>	U
ي	<i>ī</i>	<i>ī</i>	I
ى <sup>11</sup>	<i>ā</i>	<i>á</i>	_A or Y
أ	<i>ā</i>	<i>ā</i>	_a
و	<i>ū</i>	<i>ū</i>	_u
ي	<i>ī</i>	<i>ī</i>	_i

Table 2: Standard ArabTeX (long vowels)

REM. *a*. The long vowels *ā*, *ū*, *ī*, otherwise called *hurūf*<sup>u</sup> *ʿl-madd*<sup>i</sup>, 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. قَالَا *qāla*, قِيلَا *qīla*, يَقُولُ *yaqūlu*.

REM. *b*. Defective writings, such as ا, *al-ʿalif*<sup>u</sup> *ʿl-maḥḍūfat*<sup>u</sup>, or defective writings of *ū* and *ī* are encoded `_a_u` and `_i` respectively, e.g. `_d_alika` ذَلِكَ, `al-mal_a'ikaT-u` 'l-ra.hm\_an-u الْمَلِكَةُ الرَّحْمَنُ, `.hu_dayfaT-u bn-u` 'l-yamAn\_i الْحَدِيفَةُ بْنُ الْيَمَانِ for *Ḥudayfat*<sup>u</sup> *bn*<sup>u</sup> *ʿl-Yamānī*, etc.

### 3.2.2 Short vowels

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

<sup>10</sup>See below section 6 on page 23.

<sup>11</sup>= *al-ʿalif*<sup>u</sup> *ʿl-maḥḍūrat*<sup>u</sup>.

Letter	Transliteration <sup>12</sup>		ArabTeX notation
	dmg	loc	
َ	<i>a</i>	<i>a</i>	<b>a</b>
ُ	<i>u</i>	<i>u</i>	<b>u</b>
ِ	<i>i</i>	<i>i</i>	<b>i</b>
ْ	<i>an</i>	<i>an</i>	<b>aN</b>
ُن	<i>un</i>	<i>un</i>	<b>uN</b>
ِن	<i>in</i>	<i>in</i>	<b>iN</b>

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.<sup>13</sup>

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 *fathah*, *dammah* or *kasrah* except in the following cases:—

- at the commencement of a word, to indicate that a connective *ʿalif* is needed, with the exception of the article (see below section 4.4 on page 18);
- when `arabluatex` needs to perform a contextual analysis to determine the carrier of the *hamzah*;
- 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  $\langle a, u, i \rangle$ :

\_halaqa (or xalaqa) خَلَقَ *halaqa*, ^samsuN شَمْسُ *šams<sup>un</sup>*, karImuN كَرِيمٌ *Karīm<sup>un</sup>*.

bi-hi بِهٍ *bi-hi*, 'aqi.tuN أَقِيْتُ *'aqiṭ<sup>un</sup>*.

la-hu لَهُ *la-hu*, .hujjaTuN هُجَّةٌ *huḡḡat<sup>un</sup>*.

<sup>12</sup>See below section 6 on page 23.

<sup>13</sup>See below section 4.4 on page 18.

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

qAla قَالَ *qāla*, bI`a بَعَّ *bī`a*, .tUruN طُورُ *ṭūr<sup>un</sup>*, .tInuN طِينٌ *ṭīn<sup>un</sup>*,  
murU'aTuN مَرُوءَةٌ *murū`at<sup>un</sup>*.

**'alif maqṣūrah** It is written ⟨\_A⟩ or ⟨Y⟩:

al-fat\_A الْفَاتَى *al-fatā*, al-maqh\_A الْمَقْهَى *al-maqhā*, 'il\_A إِلَى *'ilā*.

**'alif otiosum** Said *'alif<sup>u</sup> 'l-wiqāyat<sup>i</sup>*, “the guarding *'alif*”, after و at the end of a word, both when preceded by *ḍammah* and by *fathah* is written ⟨UA⟩ or ⟨aW, aWA⟩:

na.sarUA نَصَّرُوا *naṣarū*, katabUA كَتَّبُوا *katabū*, ya.gzUA يَغْزُوا *yaǧzū*, ramaW رَمَّوْا *ramaw*, banaWA بَنَّوْا *banaw*.

**'alif maḥḍūfah and defective ū, ī** They are written ⟨\_a, \_i \_u⟩:

al-l\_ah-u اللهُ *al-lāh<sup>u</sup>*, 'il\_ah\_uN إِلَهٌ *'ilāh<sup>un</sup>*.  
al-ra.hm\_an-u الرَّحْمَنُ *ar-raḥmān<sup>u</sup>*, l\_akin لَكِنْ *lākin*, h\_ahunA هَاهُنَا *hāhunā*,  
.hunayn-u bn-u 'is.h\_aq-a حُنَيْنٌ بْنُ إِسْحَاقَ *Hunayn<sup>u</sup> bn<sup>u</sup> 'Ishāq<sup>a</sup>*, rabb\_i رَبِّ *rabbī*, al-`A.s\_i الْعَاصِ *al-`Āṣī*.

**Silent و/ي** Some words ending with كة are usually written وة or ية instead of كة: see Wright (1896, i. 12 A). arabuatex preserves that particular writing; the same applies to words ending in ية for كة. Long vowels ⟨U, I⟩ shall receive no *sukūn* after a *'alif maḥḍūfah* and are discarded in **trans** mode:

.hay\_aUTuN حَيَاةٌ *ḥayāt<sup>un</sup>*, .sal\_aUTuN صَلَاةٌ *ṣalāt<sup>un</sup>*, mi`sk\_aUTuN مَشْكُوتَةٌ *miškāt<sup>un</sup>*,  
tawr\_aITuN تَوْرِيَةٌ *tawrāt<sup>un</sup>*.

And so also: al-rib\_aIT-u الرِّبِيَّةُ *ar-ribāt<sup>u</sup>*.

**'Amr<sup>un</sup>, and the silent و** To that name a silent و is added to distinguish it from *'Umar<sup>u</sup>*: 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 عَمْرُو *'amr<sup>un</sup>*, `amraNU عَمْرُوا *'amr<sup>an</sup>*, `amriNU عَمْرٍو *'amr<sup>in</sup>*.

When the *tanwīn* falls away (Wright 1896, i. 249 B): `amr-uU bn-u mu.ammadiN عَمْرُو بْنُ مُحَمَّدٍ *'Amr<sup>u</sup> bn<sup>u</sup> Muḥammad<sup>in</sup>*, mu.hammad-u bn-u `amr-iU bn-i\_hAlidiN عَمْرُو بْنُ خَالِدِ *Muḥammad<sup>u</sup> bn<sup>u</sup> 'Amr<sup>i</sup> bn<sup>i</sup> Ḥālid<sup>in</sup>*.

And so also: al-rib\_aUA الرِّبَا *ar-ribā*, ribaNU رِبَا *rib<sup>an</sup>*.

**tanwīn** The marks of doubled short vowels, َ, ِ, ُ, are written  $\langle uN, aN, iN \rangle$  respectively. `arabluatex` deals with special cases, such as َ taking an `l` after all consonants except ڤ, and *tanwīn* preceding ى as in هدى, which is written  $\langle aN\_A \rangle$  or  $\langle aNY \rangle$ :

mAluN مَالٌ *māl<sup>un</sup>*, bAbaN أَبًا *bāb<sup>an</sup>*, madInaTaN مَدِينَةٌ *madīnat<sup>an</sup>*, bintiN بِنْتٌ *bint<sup>in</sup>* maqhaN\_A مَقْهَى *maqha<sup>n</sup>*, fataNY فَتَى *fata<sup>n</sup>*.

`arabluatex` is aware of special orthographies: َشيءٌ *šay<sup>un</sup>*, َشيئاً *šay<sup>an</sup>*, ِشيءٌ *šay<sup>in</sup>*.

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  $\langle aN\_A \rangle$  or  $\langle aNY \rangle$ . Such forms as قَاضٍ may likewise be written  $\langle iNI \rangle$ :—

al-qA.dI الْقَاضِي *al-qāḍī*, qA.diyaN قَاضِيًا *qāḍiy<sup>an</sup>*, qA.diNI قَاضٍ *qāḍi<sup>n</sup>*.

## 4.2 Other orthographic signs

**tā' marbūṭah** It is written  $\langle T \rangle$ :

madInaTuN مَدِينَةٌ *madīnat<sup>un</sup>*, madInaTaN مَدِينَةٌ *madīnat<sup>an</sup>*, madInaTiN مَدِينَةٌ *madīnat<sup>in</sup>*.

**hamzah** It is written  $\langle ' \rangle$ , 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 18.

**Initial hamzah:** 'asaduN أَسَدٌ *'asad<sup>un</sup>*, 'u\_htuN أُخْتُ *'uht<sup>un</sup>*, 'iqlIduN إِقْلِيدٌ *'iqlīd<sup>un</sup>*, 'anna أَنْ *'anna*, 'inna إِنْ *'inna*.

*hamzah* followed by the long vowel و is encoded ' \_U: ' \_U1\_A أُولَى *'ulā*, ' \_U1U أُولُو *'ulū*, ' \_U1A'ika أَوْلَايَكَ *'ulā'ika*.

*hamzah* followed by the long vowel ي is encoded ' \_I: ' \_ImAnuN إِيْمَانٌ *'imān<sup>un</sup>*.

**Middle hamzah:** xA.ti'-Ina خَاطِبِينَ *ḥāṭi'<sup>ina</sup>*, ru'UsuN رُؤُوسٌ *ru'ūs<sup>un</sup>*, xa.tI'aTuN خَاطِبِيَةٌ *ḥaṭi'<sup>at</sup>*, su'ila سُئِلَ *su'ila*, 'as'ilaTuN أَسْئَلَةٌ *'as'ilat<sup>un</sup>*, mas'alaTuN مَسْأَلَةٌ *mas'alat<sup>un</sup>*, 'as'alu أَسْأَلُ *'as'alu*, yataSA'alUna يَتَسَاءَلُونَ *yatasā'alūna*, murU'aTuN مَرُوءَةٌ *murū'at<sup>un</sup>*, ta'xIruN تَأْخِرُ *ta'ḥīr<sup>un</sup>*, ta'axxara تَأَخَّرَ *ta'aḥḥara*, ji'tu-ka جِئْتُكَ *gi'tu-ka*, qA'iluN قَاتِلٌ *qā'il<sup>un</sup>*.

From Wright (1896, i. 14 B):— All consonants, whatsoever, not even *ʿalif hèmzatum* excepted, admit of being doubled and take *tašdīd*. Hence we speak and write ra' 'AsuN رَأْسٌ *raʿās<sup>un</sup>*, sa' 'AluN سَأَلٌ *saʿāl<sup>un</sup>*, na' 'AjuN نَأَجٌ *naʿāġ<sup>un</sup>*.

**Final hamzah:** xa.ta'uN خَطَأٌ *ḥaṭaʿ<sup>un</sup>*, xa.ta'aN خَطَأًا *ḥaṭaʿ<sup>an</sup>*, xa.ta'iN خَطِئًا *ḥaṭaʿ<sup>in</sup>*, 'aqra'u أَقْرَأُ *ʿaqraʿu*, taqra'Inا تَقْرَأُ *taqraʿma*, taqra'Una تَقْرَأُونَ *taqraʿūna*, yaqra'na يَقْرَأُ *yaqraʿna*, yaxba'Anي يَخْبَأُ *yahbaʿāni*, xaba'A خَبَأَ *ḥabaʿā*, xubi'a خُبِئًا *ḥubiʿa*, xubi'UA خُبُوا *ḥubiʿū*, jA'a جَاءَ *ġāʿa*, ridA'uN رَدَأٌ *ridāʿ<sup>un</sup>*, ridA'aN رَدَأًا *ridāʿ<sup>an</sup>*, jI'a جِئًا *ġiʿa*, radI'iN رَدِئًا *radīʿ<sup>in</sup>*, sU'uN سُوءٌ *sūʿ<sup>un</sup>*, .daw'uN دَوُّهُ *ḍawʿ<sup>un</sup>*, qay'iN قَيْءٌ *qayʿ<sup>in</sup>*, ~sifa'I شِفَائِي *šifāʿi*.

~say'uN شَيْءٌ *šayʿ<sup>un</sup>*, ~say'aN شَيْئًا *šayʿ<sup>an</sup>*, ~say'iN شَيْءٍ *šayʿ<sup>in</sup>*, al-~say'-u أَشْيَاءُ *aš-šayʿ<sup>u</sup>*, 'a~syA'-u أَشْيَاءُ *ašyāʿ<sup>u</sup>*, 'a~syA'-a أَشْيَاءُ *ašyāʿ<sup>a</sup>*, .zim'aN ظِمًّا *ẓimʿ<sup>an</sup>*.

**taḥfif<sup>u</sup> ʿl-hamzat<sup>i</sup>:** if the *hamzah* has *ġazmah* and is preceded by *ʿalif hamzatum*, it must be changed into the letter of prolongation that is homogeneous with the preceding vowel; hence: 'a'mana أَمِنَ *āmana*, 'u'minu أَمِنُوا *ūminu*, 'i'mAnuN إِيمَانٌ *īmān<sup>un</sup>*. For other possible ways of encoding such sequences, see on page 12 (*hamzah* followed by و and ي) and the *maddah* on the current page.

**maddah** At the beginning of a syllabe, *ʿalif* with *hamzah* and *fathah* (أ) followed by *ʿalifu ʿl-maddi* (*ʿalif* of prolongation) or *ʿalif* with *hamzah* and *ġazmah* (إ) are both represented in writing *ʿalif* with *maddah*: آ (see Wright 1896, i. 25 A–B).

Hence one should keep to this distinction and encode 'a'kul أَكُلُ *ʿakulu* and 'AkiluN أَكِلُ *ʿakil<sup>un</sup>* respectively.

arabluatex otherwise determines *al-ʿalif<sup>u</sup> ʿl-mamdūdat<sup>u</sup>* by context analysis.

'is'AduN إِسَادٌ *isʿād<sup>un</sup>*, 'AkilUna أَكِلُونُ *ʿakilūna*, 'a'manna أَمَنَّا *āmannā*, al-qur'An-u الْقُرْآنُ *al-qurʿān<sup>u</sup>*.

jA'a جَاءَ *ġāʿa*, yatasA'alUna يَتَسَاءَلُونَ *yatasāʿalūna*, ridA'uN رَدَأٌ *ridāʿ<sup>un</sup>*, xaba'A خَبَأَ *ḥabaʿā*, yaxba'Anي يَخْبَأُ *yahbaʿāni*.

**šaddah** *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:

ʿallaqa عَاقَ ʿallaqa, mAdduN مَادُّ mādd<sup>un</sup>, 'ammara أَمَّرَ ammara, murruN مَرَّرَ murr<sup>un</sup>.

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 arbtex and arabxetex, arbluatex *never requires the solar letter to be written twice*, as it automatically generates the euphonic *tašdīd* above the letter that carries it, whether the article be written in the assimilated form or not, e.g. al-ʿsams-u الشَّمْسُ aš-šams<sup>u</sup>, or aʿs-ʿsams-u الشَّمْسُ aš-šams<sup>u</sup>.

al-tamr-u التَّمْرُ at-tamr<sup>u</sup>, al-ra.hm\_an-u الرَّحْمَنُ ar-rahmān<sup>u</sup>, al-.zulm-u الظُّلْمُ aḡ-ḡulm<sup>u</sup>, al-lu.gaT-u اللَّغَةُ al-luḡat<sup>u</sup>.

(b) With the letters ر, ل, م, و, ي after ن with *ḡazmah*, 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 مِنْ رَبِّيهِ, *mir rabbi-hi*, min layliN مِنْ لَيْلٍ *mil layl<sup>in</sup>*, 'an yaqtula أَنْ يَقْتُلَ *ay yaqtula*.

With *tanwīn*: kitAbuN mubInuN كِتَابٌ مُبِينٌ kitāb<sup>um</sup> mubīn<sup>un</sup>.

(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. لَبِئْتُ for لَبَيْتُ *labīttu*, 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<sup>u</sup> 'l-waṣl<sup>i</sup>** At the beginning of a sentence, اَلْ is never written, as الحمد لله; instead, to indicate that the 'alif is a connective 'alif ('alif<sup>u</sup> 'l-waṣl<sup>i</sup>), the *hamzah* is omitted and only its accompanying vowel is expressed:

al-.hamd-u li-l-l\_ah-i الحمد لله al-ḡamd<sup>u</sup> li-llāh<sup>i</sup>.



**Other cases:** 'awi ismu-hu أو اسمه *'awi 'smu-hu*, zayduN ibn-u `amriNU  
 زيدُ ابنُ عمرو *Zayd<sup>uni</sup> 'bn<sup>u</sup> 'Amr<sup>in</sup>*,<sup>14</sup> `umar-u ibn-u 'l-ha.t.tAb-i عُمرُ  
 ابنُ الخطابِ *Umar<sup>u</sup> 'bn<sup>u</sup> 'l-Haṭṭāb<sup>i</sup>*,<sup>15</sup> imru'-u 'l-qays-i امرؤُ القيسِ *Imru<sup>u</sup>*  
 'l-Qays<sup>i</sup>, la-aymun-u 'l-l\_ah-i لاَ يَمُنُّ اللهُ *la-'ymun<sup>u</sup> 'l-lāh<sup>i</sup>*.

**'alif<sup>u</sup> 'l-waṣl<sup>i</sup> preceded by a long vowel** The long vowel preceding the connective *'alif* is shortened in pronunciation (Wright 1896, i. 21 B–D). This does not appear in the Arabic script, but arablumatex takes it into account in some transliteration standards:—

fI 'l-nAs-i النَّاسِ فِي *fī 'n-nās<sup>i</sup>*, 'abU 'l-wazIr-i أَبُو الْوَزِيرِ *'abu 'l-wazīr<sup>i</sup>*,  
 fI 'l-ibtidA'-i الْإِبْتِدَاءِ فِي *fī 'l-ibtidā'<sup>i</sup>*, \_dU 'l-i`lAl-i ذُو الْأَعْلَالِ *du*  
 'l-i`lāl<sup>i</sup>, maqh\_A 'l-'amIr-i الْأَمِيرِ *maqha 'l-'amīr<sup>i</sup>*.

**'alif<sup>u</sup> 'l-waṣl<sup>i</sup> preceded by a diphthong** The diphthong is resolved into two simple vowels (Wright 1896, i. 21 D–22 A) viz. *ay* → *äi* and *aw* → *äü*. arablumatex detects the cases in which this rule applies:—

fI `aynay 'l-malik-i عَيْنِي الْمَلِكِ فِي *fī 'aynāyi 'l-malik<sup>i</sup>*, ix`say 'l-qawm-a  
 إِيْحَشِي الْقَوْمِ *iḥšāyi 'l-qawm<sup>a</sup>*, mu.s.tafaw 'l-l\_ah-i مُصْطَفَوُ اللهِ *muṣṭafawu*  
 'l-lāh<sup>i</sup>.  
 ramaW 'l-.hijAraT-a رَمَوْا الْحِجَارَةَ *ramawu 'l-ḥiḡārat<sup>a</sup>*, fa-lammA ra'aW 'l-najm-a  
 فَلَمَّا رَأَوْا النَّجْمَ *fa-lammā ra'awu 'n-naġm<sup>a</sup>*.

**'alif<sup>u</sup> 'l-waṣl<sup>i</sup> preceded by a consonant with sukūn** The vowel which the consonant takes is either its original vowel, or that which belongs to the connective *'alif* or the *kasrah*; in most of the cases (Wright 1896, i. 22 A–C), it is encoded explicitly, like so:—

'antumU 'l-kA\_dib-Una أَنْتُمْ الْكَاذِبُونَ *'antumU 'l-kādīb<sup>ūna</sup>*, ra'aytumu  
 رَأَيْتُمْ الرَّجُلَ *ra'aytumu 'r-raġul<sup>a</sup>*, mani 'l-ka\_d\_dAb-u مَنِ  
 الْكَذَّابُ *mani 'l-kaddāb<sup>u</sup>*, qatalati 'l-rUm-u قَتَلَتِ الرُّومُ *qatalati 'r-Rūm<sup>u</sup>*.

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

mu.hammaduN 'l-nabI مُحَمَّدُ النَّبِيُّ *Muḥammad<sup>uni</sup> 'n-nabī*.

<sup>14</sup> “Zayd is the son of ‘Amr”: the second noun is not in apposition to the first, but forms part of the predicate. Hence زيدُ ابنُ عمرو and not زيدُ بنُ عمرو, “Zayd, son of ‘Amr”.

<sup>15</sup> “Umar is the son of al-Ḥaṭṭāb” (see footnote 14).

### 4.3 Special orthographies

**The name of God** The name of God, اللهُ, is compounded of the article اَلْ, and اِلَآه (noted اِله with the defective *alif*) so that it becomes اِلَآه; then the *hamzah* is suppressed, its vowel being transferred to the ل before it, so that there remains اللهُ (I refer to Lane, *Lexicon*, I. 83 col. 1). Finally, the first ل is made quiescent and incorporated into the other, hence the *tasdīd* above it. As arablumatex never requires a solar letter to be written twice (see above, on page 14), the name of God is therefore encoded al-l\_ah-u or 'l-l\_ah-u:—

al-l\_ah-u اللهُ al-lāh<sup>u</sup>, yA|<sup>16</sup> al-l\_ah-u يَا اللهُ yā al-lāh<sup>u</sup>, 'a-fa|<sup>17</sup>-al-l\_ah-i  
 la-ta.g`alanna أَفَاللهُ لَتَنْعَلَنَّ 'a-fa-al-lāh<sup>i</sup> la-taġ'alanna, bi-'l-l\_ah-i  
 بِاللهِ bi-'l-lāh<sup>i</sup>, wa-'l-l\_ah-i وَاللهِ wa-'l-lāh<sup>i</sup>, bi-sm-i 'l-l\_ah-i بِسْمِ  
 اللهُ bi-sm<sup>i</sup> 'l-lāh<sup>i</sup>, al-.hamd-u لِيْلِل\_اللهِ al-ḥamd<sup>u</sup> li-llāh<sup>i</sup>,  
 li-l-l\_ah-i 'l-qA'il-u لِلِ الْقَائِلِ li-llāh<sup>i</sup> 'l-qā'il<sup>u</sup>.

**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 *alif* are encoded 'lla\_dI and 'llatI:—

'a\_hAfu mina 'l-malik-i 'lla\_dI ya.zlimu 'l-nAs-a أَخَافُ مِنَ الْمَلِكِ  
 اَلَّذِي يَظْلِمُ النَّاسَ 'ahāfu mina 'l-malik<sup>i</sup> 'lladī yazlīmu 'n-nās<sup>a</sup>, `udtu  
 'l-~say\_h-a 'lla\_dI huwa marI.duN عُدْتُ الشَّيْخَ الَّذِي هُوَ مَرِيضٌ  
 'š-~sayh<sup>a</sup> 'lladī huwa marīd<sup>un</sup>, mA 'anA bi-'lla\_dI qA'iluN la-ka  
 ~say'aN شَيْئًا لَكَ قَائِلٌ مَا أَنَا بِالَّذِي قَائِلٌ لَكَ شَيْئًا  
 'ari-nA 'lla\_dayni 'a.dallA-nA mina 'l-jinn-i wa-'l-'ins-i  
 اَرِنَا الَّذِيْنَ اَضَلَّانَا مِنَ الْجِنِّ وَالْاِنْسِ  
 wa-'l-'ins<sup>i</sup>.

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 فَانَا نَذْكُرُ الصَّوْتَيْنِ الَّذِيْنَ رَوِيَاهُمَا عَنْ جَهْزَةَ  
 'l-ladayni rawaynā-humā `an Ġahẓat<sup>a</sup>.

<sup>16</sup>Note the “pipe” character ‘|’ here after yA and below after fa before footnote mark 17: it is needed by the dm̄g 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. ‘alā ‘l-ṭarīq<sup>i</sup>. See section 4.5 on page 20 on the “pipe” and section 6 on page 23 on dm̄g mode.

<sup>17</sup>See footnote 16.

And also: al-la\_dAni اللَّذَانِ *al-ladāni*, al-la\_dayni اللَّذَيْنِ *al-ladayni*,  
 al-latAni اللَّاتَانِ *al-latāni*, al-latayni اللَّتَيْنِ *al-latayni*, al-lAtI اللَّاتِي *al-lātī*,  
 al-lA'|Ati<sup>18</sup> اللَّاءَاتِ *al-lā'āti*, al-lA'I اللَّائِي *al-lā'ī*, and so forth.

#### 4.4 Quoting

It is here referred to “quoting” after the package `arabtex`.<sup>19</sup> 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: `عَلَّمَ علم` الهيئة; is it `عَلَّمَ`, *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 *dammah* above the first `ع`, `عَلَّمَ علم الهيئة`: `عَلَّمَ 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`:—

**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 *dammah*, *fathah* or *kasrah* appear above the appropriate consonant:—

``"ullima `ilm-a 'l-hay'aT-i` عَلَّمَ علم الهيئة `'ullima 'ilma 'l-hay'ati`,  
`ya.gz"UA` يَغزُوا `yaǧzū`.

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

`wa-'innA sawfa tudriku-nA 'l-manAyA muqadd"araT"aN` وَإِنَّا سَوْفَ  
 تَدْرِكُنَا الْمَنَاءَ مَقَدَّرَةً, `wa-'innā sawfa tudriku-na 'l-manāyā muqaddaratan`.

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

`qAla isma`" قال اِسْمَعُ` `qāla 'sma', jA'at" hinduN` جَاءَتْ هِنْدُ  
`Hindun, `sabIhuN bi-man q"u.ti`at" qadamA-hu` شَبِيهَ بِنِ قَطَعَتْ  
`sabīhun bi-man quṭī'at qadamā-hu`.

<sup>18</sup>Note here the “pipe” character '|’: as already stated on page 13, the sequence 'A usually encodes *‘alif* with *hamzah* followed by *‘alif* of prolongation, which is represented in writing *‘alif* with *maddah*: `Ā`. The “pipe” character prevents this rule from being applied. See section 4.5 on page 20.

<sup>19</sup>See Lagally (2004, p. 22)

- (d) At the commencement of a word, the straight double quote is interpreted as *'alif<sup>u</sup> 'l-waṣl<sup>i</sup>*:—

wa-"ust"u`mila وأستعمل *wa-'stu'mila*, huwa "inhazama" هو أنهمزم *huwa 'nhazama*, al-"intiqA.du الأنتقاض *al-intiqādu*.

**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 قَائِلٌ قَالَ *qāla qā'il<sup>un</sup>*, ibn-u 'abI 'u.saybi`aT-"a  
 ابنُ أَبِي أُصَيْبَةَ *Ibn<sup>u</sup> 'Abī 'Uṣaybi'at<sup>a</sup>*.

- (b) *tanwīn*:—

madInaT"aN مَدِينَةٌ *madīnat<sup>an</sup>*, bAb"aN بَابًا *bāb<sup>an</sup>*, hud"aN\_A هُدًى *hudā<sup>n</sup>*,  
 شَيْءٌ *ṣay<sup>in</sup>*.

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

- (a) *'alif<sup>u</sup> 'l-waṣl<sup>i</sup>*:—

fI "istisqA'-a أَسْتِسْقَاءٌ فِي *'stisqā'<sup>a</sup>*, wa-"istisqA'-u وَأَسْتِسْقَاءٌ *wa-'stisqā'<sup>u</sup>*,  
 qAla "uhrub fa-lan tuqtala قَالَتْ قَتَلَ أَهْرَبُ فَلَنْ تُقْتَلَ *qāla 'hrub fa-lan*  
*tuqtala*.

- (b) *sukūn*:—

qAla "uqtul" fa-lan tuqtala قَالَتْ قَتَلَ أَهْرَبُ فَلَنْ تُقْتَلَ *qāla 'qtul fa-lan*  
*tuqtala*, mA jA'at" mini imra'aTiN مَا جَاءَتْ مِنْ امْرَأَةٍ *mā ḡā'at mini*  
 'mra'at<sup>in</sup>, kam" qad" ma.dat" min" laylaTiN كَمْ قَدْ مَضَتْ مِنْ لَيْلَةٍ *kam*  
*qad maḡat min laylat<sup>in</sup>*.

**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 الْجَمْرُ الصَّيْفِيُّ  
 الَّذِي كَانَ يَقْرَأُونَ *al-ḡamr<sup>u</sup> 'ṣ-ṣayfiyy<sup>u</sup> 'lladī kāna bi-Qrānnūn<sup>a</sup>*.

#### 4.4.1 Quoting the hamzah

As said above in section 4.2 on page 12, the *hamzah* 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 *hamzah* freely, without any consideration for the context. Table 4 gives the equivalents for all the possible carriers the *hamzah* may take.

Letter	Transliteration <sup>20</sup>		ArabTeX notation
	dmg	loc	
ء	’	’	"'
آ	’ā	’ā	A"'
أ	’	’	a"'
ؤ	’	’	u"'
و	’	’	w"'
إ	’	’	i"'
ي	’	’	y"'

Table 4: “Quoted” *hamzah*

As one can see from table 4, the carrier of the *hamzah* is inferred from the letter that precedes the straight double quote (<">). Of course, any “quoted” *hamzah* may take a short vowel, which is to be written *after* the ArabTeX equivalent for the *hamzah* itself, namely (<'>). For example, ؤ is encoded <w" 'a>, while ؤ 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 أَعْدَاؤُكُمْ 'a`dā'ukum, 'a`dA|" ' 'ukum أَعْدَاءُكُمْ 'a`dā'ukum, 'a`dA'ikum  
 أَعْدَائِكُمْ 'a`dā'ikum, 'a`dA|" ' 'ikum أَعْدَاءُكُمْ 'a`dā'ikum.

## 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 *hamzah* 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 17 and in footnote 18 on page 18. One more example follows:—

bi-qrAn|nUn-a بقرآنونَ *bi-Qrānnūn*<sup>a</sup>, “in Crannon” (Thessaly, Greece).<sup>21</sup>

As one can see, the “pipe” character between the two <n> prevents the necessary *tašdīd* rule (page 14) from being applied.

<sup>20</sup>See below section 6 on page 23.

<sup>21</sup>See more context on page 19.

## 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 and diacritics which stand one above the other:—

.hunayn-u bn-u 'is.h--\_aq-a حُنَيْنُ بْنُ إِسْحَاقَ *Hunayn<sup>u</sup> bn<sup>u</sup> 'Ishāq<sup>a</sup>*

## 4.7 Digits

### 4.7.1 Numerical figures

The *Indian numbers*, *ar-raqam<sup>u</sup> 'l-hindiyy<sup>u</sup>*, 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 ١٠٢٤ فِي سَنَةِ

### 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 *abjad* numbers are usually distinguished from the surrounding words by a stroke placed over them.

`\abjad` *abjad* 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:—

`\abjad{45}` kitAbu-hu fI 'l-`AdAt-i الْعَادَاتِ فِي كِتَابِهِ مَ 45 *kitābu-hu fi 'l-`ādāt<sup>i</sup>*.

REM. *a.* As can be seen in the above given example, `arabluatex` expresses the *abjad* 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 on the next page:—

The `\arb[trans]{'abjad} number` for 1874 is `\abjad{1874}` The *abjad* number for 1874 is غَضَد.

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

## 4.8 Additional characters

In the manuscripts, the unpointed letters, *al-ḥurūfu 'l-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 <sup>22</sup>		ArabTeX notation
	dmg	loc	
ب	<i>b</i>	<i>b</i>	.b
د	<i>ḋ</i>	<i>d</i>	˘d
ف	<i>f</i>	<i>f</i>	.f
ق	<i>q</i>	<i>q</i>	.q
ك	<i>k</i>	<i>k</i>	.k
ن	<i>n</i>	<i>n</i>	.n
﴿	(	(	((
﴾	)	)	))

Table 5: Additional Arabic codings

'afAman.tUs Gal.(M) .fmn.n.ts (sic) Gal.(E1), أفامنطرس Gal.(M) منطس (sic) Gal.(E1), 'afāmantūs Gal.(M) fmnnṫs (sic) Gal.(E1).

## 4.9 Arabic emphasis

As already seen in section 4.7.2 on page 21, the *ʾabjad* 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:—

`\abjad{45}`: kitAbu-hu `\aemph{fI 'l-`AdAt-i}` كِتَابُهُ فِي الْعَادَاتِ 45: *kitābu-hu fi 'l-Ādāt<sup>i</sup>*.

## 5 Special applications

**Linguistics** The same horizontal stroke as the *taṭwīl* (see section 4.6 on page 21) may be encoded  $\langle B \rangle$ ;  $\langle BB \rangle$  will receive the *tašdīd*. This is useful to make linguistic annotations and comments on vowels:—

Bu Ba Bi BuN BaN BiN  $\overset{\text{˘}}{u}$   $\overset{\text{˘}}{a}$   $\overset{\text{˘}}{i}$   $\overset{\text{˘}}{un}$   $\overset{\text{˘}}{an}$   $\overset{\text{˘}}{in}$ , BBu BBa BBi  $\overset{\text{˘}}{u}$   $\overset{\text{˘}}{a}$   $\overset{\text{˘}}{i}$ , B--aN  
 $\overset{\text{˘}}{u}$  -<sup>an</sup>, B<sup>˘</sup> ..

New feature  
v1.4.3

**Brackets** The various bracket symbols are useful in technical documents such as critical editions for indicating that some words or some letters must be added or removed. `arabluatex` will automatically fit those symbols to the direction of the text. For the time being, the following symbols are supported:

- parentheses: `()`
- square brackets: `[]`
- angle brackets: `<>`
- braces: `{}`

`\abraces` Parentheses, square and angle brackets may be input directly at the keyboard; however, words or letters that are to be read between braces must be passed as arguments to the command `\abraces`:—

```
1 \begin{arab}
2   \abraces{wa-qAla} 'inna 'abI kAna mina 'l-muqAtilaTi
3   wa-kAna--<--t> 'ummI min `u.zamA'i buyUti 'l-zamAzimaTi.
4 \end{arab}
```

{وَقَالَ} إِنَّ أَيْ كَانَ مِنَ الْمُقَاتِلَةِ وَكَانَتْ < أُمِّي مِنْ عُظَمَاءِ بُيُوتِ الزَّمَاةِ.

## 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 Arab $\TeX$  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.<sup>23</sup> `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.<sup>24</sup>

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.:—

<sup>22</sup>See below section 6.

<sup>23</sup>See Brockelmann et al. (1935).

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

```

1 \usepackage{arabluatex}
2 \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.

New feature  
v1.4

`\SetTranslitFont` **Font** `\SetTranslitFont{<font selection command>}` allows any specific font to be selected for rendering transliterated text with the font-selecting commands of the `fontspec` or `luaotfload` package. Of course, this font must have been defined properly. To take one example, here is how the *Gentium Plus* font may be used for rendering transliterated text:—

```

1 \newfontfamily\translitfont{Gentium Plus}[Ligatures=TeX]
2 \SetTranslitFont{\translitfont}

```

`\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 initial letter, apart from the article, cannot be uppercased, viz. ' or ' , the letter next to it will be uppercased:—

```

\cap{.hunayn-u} bn-u \cap{'is.h_aq-a} حُنَيْنُ بْنُ إِسْحَاقَ Ḥunaynu bnu
ʾIshāqa, \cap{`u_tm_an-u} عُثْمَانُ ʿUṭmānu, .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} عَبْدُ اللَّهِ بْنِ عَوْفِ بْنِ خَلْدِ سَعْدِ بْنِ زَيْدِ بْنِ ضَرْبِ دَارَابَا Zaydu bnu
Ḥālīdin Sa`da bna ʿAwfi bni ʿAbdi ʿl-Lāhi.

```

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 النَّبِيُّ must be *Muḥammad<sup>umi</sup> ʿn-nabī*, as nouns having the *tanwīn* take a *kasrah* in pronunciation before *ʾalifu ʿl-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 `<'l>`. Fortunately, this can be circumvented in a straightforward way by inserting only part of the noun in the argument of `\cap` viz. up to the first letter that is to be uppercased, like so: `\cap{m}u.hammaduN`.

**Hyphenation** In case transliterated Arabic words break the T<sub>E</sub>X hyphenation algorithm, one may use the command `\-` to insert discretionary hyphens. This command will be discarded in all of the Arabic modes of `arabluatex`, but will be processed by any of the transliteration modes:—

```
\cap{'abU} \cap{bakriN} \cap{\mu\-.ham\-.madu} bnu \cap{za\-.ka
\-.riy\-.yA'a} \cap{'l-rAzI} الرَّابِي الرَّابِي أَبُو بَكْرٍ مُحَمَّدُ بْنُ زَكْرِيَاءَ الزَّكْرِيَّاءُ 'Abū Bakrin Mu-
hammadu bnu Zakariyyāu 'r-Rāzī.
```

## 6.1 Additional note on dmG convention

According to Brockelmann et al. (1935, p. 6), Arabic *'irāb* may be rendered into dmG in three different ways:

- (a) In full: *'Amrun*;
- (b) As superscript text: *'Amr<sup>un</sup>*;
- (c) Discarded: *'Amr*.

`\arbup` By default, `arabluatex` applies rule **b**. Once delimited by a set of Lua functions, *'irāb* 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 *'irāb* is displayed. To take one example, here is how Arabic *'irāb* may be rendered as subscript text:—

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

Arabic dmG transliteration for رَأَيْتُ جَامِعًا مُهْدَمَةً مَثَلْتَهُ: *ra'aytu ḡāmi`<sub>am</sub> muhaddamat<sub>am</sub> mi'danatu-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.

**'irāb boundaries** Every declinable noun (*mu'rab*) may be declined either with or without *tanwīn*, viz. *munṣarīf<sup>un</sup>* or *ḡayr<sup>u</sup> munṣarīf<sup>in</sup>*. The former is automatically parsed by `arabluatex`, whereas the latter has to be delimited with an hyphen, like so:—

```
munṣarīf: mu`allimN مَعْلِمٌ mu'allimun, kA'inuN كَاتِبٌ kā'inun, kA'inAtuN
كَاتِبَاتٌ kā'inānun, \cap{'amraNU} عَمْرُوًا 'Amran, fataN_A فَتَى fatan, qA.diNI
قَاضٍ qādin.
```

ǧayr munṣarīf: al-mu`allim-u <sup>المُعَلِّمُ</sup> *al-mu`allim<sup>u</sup>*, kitAb-Ani كِتَابَانِ *kitāb<sup>āni</sup>*,  
 ra`sa'-Ani رَشَآنِ *rašā`āni*, sAriq-Una سَارِقُونَ *sāriq<sup>āna</sup>*, qA.d-Una قَاضُونَ *qāḍ<sup>āna</sup>*,  
 al-.zulm-Atu الظُّلْمَاتُ *aḏ-ḏulm<sup>ātu</sup>*.

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 *ǧayr munṣarīf* terminations. This can be achieved by simply omitting the hyphen before any *ǧayr munṣarīf* termination:—

kAna .ganiyyaN l\_akinna-hu labisa `gubbaTaN mumazzaqaN 'aydu-hA <sup>كَانَ غَنِيًّا لَكِنَّهُ لَيْسَ جَبَّةً مُمَزَّعًا</sup>  
 كَانُ غَنِِيَّيْ<sup>al</sup> لَآكِينِنَا-hu labisa ḡubbat<sup>am</sup> mumazzaq<sup>am</sup> 'aydu-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 <sup>مَنْعَ النَّاسِ كَافَّةً مِنْ</sup>  
 مَنَا`أ 'l-nās-a kāffat<sup>am</sup> mim muḥāṭabati-hi 'ahad<sup>un</sup> bi-sayyidi-nā.

## 6.2 Examples

Here follows in transliteration the story of Ḡuḥā and his donkey (جُحَا وَحِمَارُهُ). See the code on page 7:—

**‘dmg’ standard:** `atā ṣadīq<sup>un</sup> `ilā Ḡuḥā yaṭlubu min-hu ḥimāra-hu li-yarkaba-hu fī safrat<sup>in</sup> qaṣīrat<sup>in</sup> fa-qāla la-hu: “sawfa `u`īdu-hu `ilay-ka fī `l-masā`i wa-`adfa`u la-ka `uḡrat<sup>an</sup>.” fa-qāla Ḡuḥā: “anā `āsif<sup>un</sup> ḡidd<sup>an</sup> `annī lā `astatī`u an `uḥaqqiqa la-ka raḡbata-ka fa-`l-ḥimār<sup>u</sup> laysa huna `l-yawm<sup>a</sup>.” wa-qabla `ay yutimma Ḡuḥā kalāma-hu bada`a `l-ḥimār<sup>u</sup> yanhaqu fī `iṣṭabli-hi. fa-qāla la-hu ṣadīqu-hu: “`innī `asma`u ḥimāra-ka yā Ḡuḥā yanhaqu.” fa-qāla la-hu Ḡuḥā: “ḡarīb<sup>un</sup> `amru-ka yā ṣadīqī `a-tuṣaddiqu `l-ḥimār<sup>a</sup> wa-tukaddiba-nī?”

**‘loc’ standard:** atā ṣadīqun ilā Juḥā yaṭlubu min-hu ḥimāra-hu li-yarkaba-hu fī safrat<sup>in</sup> qaṣīrat<sup>in</sup> fa-qāla la-hu: “sawfa u`īdu-hu ilay-ka fī al-masā`i wa-`adfa`u la-ka ujratan.” fa-qāla Juḥā: “anā āsifun jiddan annī lā astatī`u an uḥaqqiqa la-ka raḡbata-ka fa-al-ḥimāru laysa hunā al-yawma.” wa-qabla an yutimma Juḥā kalāma-hu bada`a al-ḥimāru yanhaqu fī iṣṭabli-hi. fa-qāla la-hu ṣadīqu-hu: “innī asma`u ḥimāra-ka yā Juḥā yanhaqu.” fa-qāla la-hu Juḥā: “gharībun amru-ka yā ṣadīqī a-tuṣaddiqu al-ḥimāra wa-tukadhḥiba-nī?”

## 7 Buckwalter input scheme

Even though *arabluatex* is primarily designed to process the ArabTeX notation, it can also process the Buckwalter input scheme to a large extent.<sup>25</sup> The Buckwalter scheme is actually processed in two steps, as it is first converted into ArabTeX. Then, once this is accomplished, the ArabTeX scheme is processed through the above described functions. In this way, the Buckwalter input scheme can make the most of the *arabluatex* special features that are presented in section 2.2 on page 5.

The input scheme, which is set to *arabtex* by default, may be changed at

<sup>25</sup>See <http://www.qamus.org/transliteration.htm>

any point of the document by the command `\SetInputScheme{<scheme>}`, where `<scheme>` may be either `arabtex` or `buckwalter`. This command is also accepted in the preamble should one wish to set the input scheme globally, like so:—

```

1 \usepackage{arabluatex}
2 \SetInputScheme{buckwalter}

```

**‘base’, ‘xml’ and ‘safe’ schemes** `arabluatex` can use any of the so-called Buckwalter ‘base’, ‘xml’ or ‘safe’ schemes as they are described in Habash (2010, pp. 25–26).<sup>26</sup> However, the following limitation apply to the ‘base’ and ‘xml’ schemes: the braces { and }, which are used to encode ا and ؤ, must be replaced with square brackets viz. [ and ] respectively.

It is therefore recommended to use the Buckwalter ‘safe’ scheme.

Table 6 gives the Buckwalter equivalents that are currently used by `arabluatex`. The additional characters that are defined in table 5 on page 22 are also available.

Letter	Transliteration <sup>27</sup>		Buckwalter notation	
	dmg	loc	base/xml	safe
ا	<i>a</i>	<i>a</i>	A	A
ب	<i>b</i>	<i>b</i>	b	b
ت	<i>t</i>	<i>t</i>	t	t
ث	<i>ṭ</i>	<i>th</i>	v	v
ج	<i>ǧ</i>	<i>j</i>	j	j
ح	<i>h</i>	<i>h</i>	H	H
خ	<i>ḫ</i>	<i>kh</i>	x	x
د	<i>d</i>	<i>d</i>	d	d
ذ	<i>ḏ</i>	<i>dh</i>	*	V
ر	<i>r</i>	<i>r</i>	r	r
ز	<i>z</i>	<i>z</i>	z	z
س	<i>s</i>	<i>s</i>	s	s
ش	<i>š</i>	<i>sh</i>	\$	c
ص	<i>ṣ</i>	<i>ṣ</i>	S	S
ض	<i>ḏ</i>	<i>ḏ</i>	D	D
ط	<i>ṭ</i>	<i>ṭ</i>	T	T
ظ	<i>ẓ</i>	<i>ẓ</i>	Z	Z
ع	<i>ʿ</i>	<i>ʿ</i>	E	E
غ	<i>ǧ</i>	<i>gh</i>	g	g
ف	<i>f</i>	<i>f</i>	f	f
ق	<i>q</i>	<i>q</i>	q	q
ك	<i>k</i>	<i>k</i>	k	k
ل	<i>l</i>	<i>l</i>	l	l

<sup>26</sup>I am grateful to Graeme Andrews who suggested that the ‘safe’ scheme be included in `arabluatex`.

<sup>27</sup>See section 6 on page 23.

Letter	Transliteration		Buckwalter notation	
	dmg	loc	base/xml	safe
م	<i>m</i>	<i>m</i>	m	m
ن	<i>n</i>	<i>n</i>	n	n
ه	<i>h</i>	<i>h</i>	h	h
و	<i>w</i>	<i>w</i>	w	w
ي	<i>y</i>	<i>y</i>	y	y
ى	<i>ā</i>	<i>á</i>	Y	Y
ة	<i>ah</i>	<i>ah</i>	p	p
ء	’	’	’	C
آ	<i>ā</i>	<i>’ā</i>		M
أ	’	’	>	O
ؤ	’	’	&	W
إ	’	’	<	I
ئ	’	’	]	Q
ـ	—	—	~	~
أ	’	’	[	L
ا	<i>a</i>	<i>a</i>	a	a
و	<i>u</i>	<i>u</i>	u	u
ي	<i>i</i>	<i>i</i>	i	i
ان	<i>an</i>	<i>an</i>	F	F
ان	<i>un</i>	<i>un</i>	N	N
ان	<i>in</i>	<i>in</i>	K	K
و	—	—	o	o
ا	<i>ā</i>	<i>ā</i>	˘	e
- ( <i>taṭwīl</i> )	—	—	-	-

Table 6: Buckwalter scheme

**Transliteration** The Buckwalter notation can also be transliterated into any accepted romanization standard of Arabic. See above section 6 on page 23 for more information. However, it should be pointed out again that only accurate coding produces accurate transliteration. It is therefore at the very least highly advisable to use the hyphen for tying the definite article and the inseparable particles (viz. prepositions, adverbs and conjunctions) to words, like so:—

Al-EaAlamu الْعَالَمُ *al-ālam<sup>u</sup>*, Al-camsu الشَّمْسُ *aš-šams<sup>u</sup>*, bi-SinaAEapi  
Al-T~ib~i, بِصِنَاعَةِ الطِّبِّ *bi-ṣinā‘at<sup>i</sup> ‘t-ṭibb<sup>i</sup>*.

wa-Al-l~ehi وَاللّٰهِ wa-'l-lāh<sup>i</sup>, Al-Hamdu li-l~ehi لِلّٰهِ الْحَمْدُ al-ḥamd<sup>u</sup>  
li-llāh<sup>i</sup>.

Similarly, it is not advisable to use | and [ ('base' and 'xml' schemes) or M and L ('safe' scheme) to encode the 'alif<sup>u</sup> 'l-mamdūdat<sup>i</sup> and the 'alif<sup>u</sup> 'l-waṣl<sup>i</sup> for such signs are supposed to be generated by arabluatex internal functions. Besides, as they do not *per se* convey any morphological information on what they are derived from, they cannot be transliterated accurately. To take one example, <iLY Al-LntiqāADi gives اِلْتِقَاضِي as expected, but only <iLY Al-intiqāADi can be transliterated as ilā 'l-intiqāḍi with the correct vowel ⟨i⟩ in place of the 'alif<sup>u</sup> 'l-waṣl<sup>i</sup>.

## 8 L<sup>A</sup>T<sub>E</sub>X Commands in Arabic environments

**General principle** L<sup>A</sup>T<sub>E</sub>X 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}} العَادَاتِ فِي كِتَابِهِ 45
kitābu-hu fi 'l-`Adāti.28
```

The same applies to footnotes:—

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

إِنَّ أَيْ كَانٍ مِنَ الْمُقَاتِلَةِ<sup>a</sup>, وَكَانَتْ أُمِّي مِنْ عُظَمَاءِ بِيوتِ الزَّمَاذِمَةِ<sup>b</sup>.

<sup>a</sup>المُقَاتِلَةُ: الْمُقَاتِلِينَ.  
<sup>b</sup>الزَّمَاذِمَةُ: طَائِفَةٌ مِنَ الْفُرْسِ.

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{⟨arg⟩}` is designed to typeset its argument from left to right. It may be used

<sup>28</sup>This is odd in Arabic script, but using such features as `\emph` or `\textbf` is a matter of personal taste.

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 L<sup>A</sup>T<sub>E</sub>X 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{<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:—

```

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

```

---

ثُمَّ اتَّخَذَ ذُو الْقَرْنَيْنِ إِلَى أُمَّةٍ أُخْرَى فِي ﴿مَطْلَعِ الشَّمْسِ﴾ وَلَا بِنَاءَ لَهُمْ يَوْمَهُمْ مِنَ الشَّمْسِ.

`\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:—

```

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

```

---

زيد<sup>a</sup> ابن عمرو

---

<sup>a</sup> “*zayd* is the son of ‘*amr*”: 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 L<sup>A</sup>T<sub>E</sub>X 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`.<sup>29</sup>

<sup>29</sup>See <http://ctan.org/pkg/koma-script>; read the documentation of KOMA-script for details about the `\deffootnotemark` and `\deffootnote` commands.



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.

## 8.2 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:—

```
1 \beginnumbering
2 \pstart
3 \begin{arab}
4 wa-ya.sIru ta.hta 'l-jild-i
5 \edtext{\arb{.sadId-uN}}{\Afootnote{M: \arb{.sadId-aN} E1}}
6 \end{arab}
7 \pend
8 \endnumbering
```

## 9 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.

## 10 Implementation

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

```
1 \NeedsTeXFormat{LaTeX2e}
2 \ProvidesPackage{arabluatex}%
3 [2016/10/05 v1.4.5 An ArabTeX-like interface for LuaLaTeX]
4 \RequirePackage{ifluatex}
```

`arabluatex` requires LuaL<sup>A</sup>T<sub>E</sub>X of course. Issue a warning if the document is processed with another engine.

```

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

```

Declare the global options, and define them:

```

12 \DeclareOption{voc}{\def\al@mode{voc}}
13 \DeclareOption{fullvoc}{\def\al@mode{fullvoc}}
14 \DeclareOption{novoc}{\def\al@mode{novoc}}
15 \DeclareOption{trans}{\def\al@mode{trans}}
16 \ExecuteOptions{voc}
17 \ProcessOptions\relax
18 \def\al@mode@voc{voc}
19 \def\al@mode@fullvoc{fullvoc}
20 \def\al@mode@novoc{novoc}
21 \def\al@mode@trans{trans}

```

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

```
22 % \newif\ifal@mode@defined
```

Packages that are required by arabluatex:

```

23 \RequirePackage{fontspec}
24 \RequirePackage{amsmath}
25 \RequirePackage{etoolbox}
26 \RequirePackage{luacode}
27 \RequirePackage{xparse}
28 \RequirePackage{environ}

```

Here begins the real work: load arabluatex.lua:

```
29 \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.

```
30 \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  $\TeX$ live:

```

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

```

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

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

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

```
36 \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:

```
37 \AtBeginDocument{\ifdef{\LR}}%
38   {\RenewDocumentCommand{\LR}{m}{\bgroup\textdir TLT\rmfamily#1\egroup}}
39   {\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.

```
40 \AtBeginDocument{\ifdef{\RL}}%
41   {\RenewDocumentCommand{\RL}{m}{\bgroup\textdir TRT\rmfamily#1\egroup}}
42   {\NewDocumentCommand{\RL}{m}{\bgroup\textdir TRT#1\rmfamily\egroup}}
```

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

```
43 \AtBeginDocument{\ifdef{\aemph}}%
44   {\RenewDocumentCommand{\aemph}{m}{\overline{\text{#1}}}}
45   {\NewDocumentCommand{\aemph}{m}{\overline{\text{#1}}}}
```

`\SetInputScheme` `arabluatex` is designed for processing Arab $\TeX$  input notation. `\SetInputScheme` may be used in the preamble or at any point of the document should the user wish to use a different notation such as the ‘Buckwalter scheme’.

```
46 \def\al@input@scheme{arabtex}
47 \NewDocumentCommand{\SetInputScheme}{m}{\def\al@input@scheme{#1}}
```

`\SetArbEasy` By default, `arabluatex` applies complex rules to generate euphonic *tašdīd*, *ʿalif mamdūdah* and *sukūn* depending on the modes which are selected, either `voc`, `\SetArbEasy*` `fullvoc` or `trans`. Such refinements can be discarded with `\SetArbEasy`, either globally in the preamble or at any point of the document. Note that `\SetArbEasy` keeps the *sukūn* that is generated, while the starred version `\SetArbEasy*` takes it away. Default complex rules can be set back at any point of the document with `\SetArbDflt`.

```
48 \def\al@arb@rules{dflt}
49 \NewDocumentCommand{\SetArbEasy}{s}{%
50   \IfBooleanTF{#1}
51   {\def\al@arb@rules{easynosukun}}
52   {\def\al@arb@rules{easy}}}
53 \NewDocumentCommand{\SetArbDflt}{s}{\def\al@arb@rules{dflt}}
```

`\SetTranslitFont` By default, the font that is used for transliterated text is the main font of the document. Any other font may also be selected with the font-selecting commands of the `fontspec` package.

```
54 \def\al@trans@font{\rmfamily}%
55 \NewDocumentCommand{\SetTranslitFont}{m}{\def\al@trans@font{#1}}
```

`\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:

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

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

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

`\arbup` 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)}`.

```
60 \NewDocumentCommand{\al@arbup@dflt}{m}{\textsuperscript{\thinspace#1}}%
61 \NewDocumentCommand{\al@arbup}{m}{\al@arbup@dflt{#1}}
62 \NewDocumentCommand{\arbup}{m}{\al@arbup{#1}}
63 \NewDocumentCommand{\ArbUpDflt}{}{\let\al@arbup=\al@arbup@dflt}
64 \NewDocumentCommand{\NoArbUp}{}{\RenewDocumentCommand{\al@arbup}{m}{##1}}
65 \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.

```
66 \DeclareDocumentCommand{\cap}{m}%
67 {\luairect{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.

```
68 \DeclareDocumentCommand{\txarb}{+m}{\bgroup\textdir
69 TRT\arabicfont#1\egroup}
70 \DeclareDocumentCommand{\txtrans}{+m}{\bgroup\textdir
71 TLT\al@trans@font#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.

```
72 \DeclareDocumentCommand{\arb}{0{\al@mode} +m}%
73 {\edef\@tempa{#1}%
74 \ifx\@tempa\al@mode@voc%
75 \bgroup\textdir TRT\arabicfont%
76 \luairect{tex.sprint(processvoc(\luastringN{#2},
77 \luastring0{\al@arb@rules}, \luastring0{\al@input@scheme}))}\egroup%
78 \else%
79 \ifx\@tempa\al@mode@fullvoc%
80 \bgroup\textdir TRT\arabicfont%
81 \luairect{tex.sprint(processfullvoc(\luastringN{#2},
```

```

82 \luastring0{\al@arb@rules}, \luastring0{\al@input@scheme}})\egroup%
83 \else%
84 \ifx\@tempa\al@mode@novoc%
85 \bgroup\textdir TRT\arabicfont%
86 \luadirect{tex.sprint(processnovoc(\luastringN{#2},
87 \luastring0{\al@arb@rules}, \luastring0{\al@input@scheme}})\egroup%
88 \else%
89 \ifx\@tempa\al@mode@trans%
90 \bgroup\textdir TLT\al@trans@style%
91 \luadirect{tex.sprint(processtrans(\luastringN{#2},
92 \luastring0{\al@trans@convention},
93 \luastring0{\al@arb@rules},
94 \luastring0{\al@input@scheme}})\egroup%
95 \else%
96 \fi\fi\fi\fi}

```

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

```

97 \NewEnviron{arab}[1][\al@mode]%
98 {\par\edef\@tempa{#1}%
99 \ifx\@tempa\al@mode@voc%
100 \bgroup\pardir TRT\textdir TRT\arabicfont%
101 \luadirect{tex.sprint(processvoc(\luastring0{\BODY},
102 \luastring0{\al@arb@rules}, \luastring0{\al@input@scheme}})\egroup%
103 \else%
104 \ifx\@tempa\al@mode@fullvoc%
105 \bgroup\pardir TRT\textdir TRT\arabicfont%
106 \luadirect{tex.sprint(processfullvoc(\luastring0{\BODY},
107 \luastring0{\al@arb@rules}, \luastring0{\al@input@scheme}})\egroup%
108 \else%
109 \ifx\@tempa\al@mode@novoc%
110 \bgroup\pardir TRT\textdir TRT\arabicfont%
111 \luadirect{tex.sprint(processnovoc(\luastring0{\BODY},
112 \luastring0{\al@arb@rules}, \luastring0{\al@input@scheme}})\egroup%
113 \else \ifx\@tempa\al@mode@trans%
114 \bgroup\pardir TLT\textdir TLT\al@trans@style%
115 \luadirect{tex.sprint(processtrans(\luastring0{\BODY},
116 \luastring0{\al@trans@convention},
117 \luastring0{\al@arb@rules},
118 \luastring0{\al@input@scheme}})\egroup%
119 \else \fi\fi\fi\fi}\par]

```

`\abjad` `\abjad{<number>}` expresses its argument in Arabic letters in accordance with the *abjad* 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.

```

120 \AtBeginDocument{%
121 \ifdefined\abjad%
122 \RenewDocumentCommand{\abjad}{m}%

```

```

123 {\luadirect{tex.sprint(abjadify(#1))}}%
124 \else%
125 \NewDocumentCommand{\abjad}{m}%
126 {\luadirect{tex.sprint(abjadify(#1))}}
127 \fi}

\abraces \abraces{Arabic text} puts its argument between braces. This macro is written
in Lua and is dependent on the current value of tex.textdir.
128 \NewDocumentCommand{\abraces}{+m}{%
129 \luadirect{tex.sprint(abraces(\luastringN{#1}))}}

\LRmarginpar \LRmarginpar is supposed to be inserted in an Arabic environment. It typesets his
argument in a marginal note from left to right.
130 \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.
131 \DeclareDocumentCommand{\LRfootnote}{m}{\bgroup\pdir
132 TLT\LR{\footnote{#1}}\egroup}
133 \DeclareDocumentCommand{\RLfootnote}{m}{\bgroup\pdir
134 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 scrextend if it is not already loaded.
135 \NewDocumentCommand{\FixArbFtnmk}{-}{%
136 \@ifpackageloaded{scrextend}%
137 {\AtBeginDocument{\deffootnote{2em}{1.6em}{\LR{\thefootnotemark}.\enskip}}}%
138 {\RequirePackage{scrextend}}
139 \AtBeginDocument{\deffootnote{2em}{1.6em}{\LR{\thefootnotemark}.\enskip}}}}

That is it. Say goodbye before leaving.
140 \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 collaboration with Ph. S. van Ronkel and Otto Spies. Deutschen Morgenländischen Gesellschaft. Leipzig: Deutschen Morgenländischen Gesellschaft, in Kommission bei F. A. Brockhaus. URL: [http://www.naher-osten.uni-muenchen.de/studium\\_lehre/werkzeugkasten/dmgttransliteration.pdf](http://www.naher-osten.uni-muenchen.de/studium_lehre/werkzeugkasten/dmgttransliteration.pdf).
- Habash, Nizar Y. (2010). *Introduction to Arabic Natural Language Processing*. Synthesis Lectures on Human Language Technologies 10. Toronto: Morgan & Claypool Publishers.

- 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

v1.0	General: Initial release . . . . .	1	v1.4	<code>\SetInputScheme</code> :	
v1.0.1	General: Minor update of the documentation . . . . .	1		<code>\SetInputScheme</code> may be used to process other input schemes such as ‘Buckwalter’ . . . . .	34
v1.1	<code>\abjad</code> : New and more flexible <code>\abjad</code> command. . . . .	36		<code>\SetTranslitFont</code> : For selecting a specific font for transliterated texts . . . . .	34
v1.2	<code>\SetArbEasy</code> : New <code>\SetArbEasy/\SetArbDflt</code> for ‘modern’ or ‘classic’ Arabic styles. . . . .	34	v1.4.3	<code>\abraces</code> : New command <code>\abraces</code> which expresses its argument between braces. . . . .	37
v1.3	<code>\arbup</code> : <i>īrāb</i> is now written as superscript text in <code>dmg</code> mode by default. . . . .	35	v1.4.4	<code>\SetArbEasy*</code> : this starred version discards the <i>sukūn</i> in addition to what is already discarded by <code>\SetArbEasy</code> . . . . .	34

## 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.

<b>Symbols</b>	<b>A</b>	<code>\al@arb@rules</code> . . . . .
<code>\@ifpackageloaded</code> . . . . .		48, 51, 52, 53,
<code>\@tempa</code> . . . . .	<code>\abjad</code> . . . . .	77, 82, 87, 93,
73, 74, 79, 84, 89, 98, 99, 104, 109, 113	<code>\abraces</code> . . . . .	102, 107, 112, 117
	23, <u>128</u>	
<code>\_</code> . . . . .	<code>\aemph</code> . . . . .	<code>\al@arbup</code> . . . . .
32	22, <u>43</u>	. 61, 62, 63, 64, 65



<code>\SetArbEasy*</code> . . . . .	6, <a href="#">48</a>	<code>\textdir</code> . . . . .	<a href="#">35</a> ,	<i>Transliteration der ara-</i>
<code>\SetArbUp</code> . . . . .	25, <a href="#">60</a>		<a href="#">36</a> , <a href="#">38</a> , <a href="#">39</a> , <a href="#">41</a> ,	<i>bischen Schrift,</i>
<code>\SetInputScheme</code> . . . . .	26, <a href="#">46</a>		<a href="#">42</a> , <a href="#">68</a> , <a href="#">70</a> , <a href="#">75</a> ,	<i>Die</i> . . . . .
<code>\setLR</code> . . . . .	31, <a href="#">36</a>		<a href="#">80</a> , <a href="#">85</a> , <a href="#">90</a> , <a href="#">100</a> ,	<a href="#">23</a> , <a href="#">25</a>
<code>\setRL</code> . . . . .	31, <a href="#">35</a>		<a href="#">105</a> , <a href="#">110</a> , <a href="#">114</a> , <a href="#">130</a>	<code>\txarb</code> . . . . .
<code>\SetTranslitConvention</code> . . . . .	23, <a href="#">58</a>	<code>\textsuperscript</code> . . . . .	<a href="#">60</a>	<code>\txtrans</code> . . . . .
<code>\SetTranslitFont</code> 24, <a href="#">54</a>		<code>\thefootnotemark</code> . . . . .	<a href="#">137</a> , <a href="#">139</a>	<b>V</b>
<code>\SetTranslitStyle</code> 24, <a href="#">56</a>				voc (option) . . . . .
		<code>\thinspace</code> . . . . .	<a href="#">60</a>	<b>W</b>
<b>T</b>		<code>trans</code> (option) . . . . .	<a href="#">5</a> , <a href="#">8</a>	Wright, W. LL.D . . . . .
<code>\text</code> . . . . .	<a href="#">44</a> , <a href="#">45</a>			. . . . .
				<a href="#">5</a> , <a href="#">7</a> , <a href="#">8</a> ,
				<a href="#">11</a> , <a href="#">13</a> , <a href="#">14</a> , <a href="#">16</a> , <a href="#">21</a>