

The arabluatex package v1.4.5 – 2016/10/05

Robert Alessi
alessi@robertalessi.net

Contents

| | | | |
|--|-----------|--|-----------|
| License and disclaimer | 2 | | |
| 1 Introduction | 2 | | |
| 1.1 arabluatex is for Lua [®] TeX | 4 | 4.7 Digits | 21 |
| 2 The basics of arabluatex | 4 | 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 | 5 | 4.9 Arabic emphasis | 22 |
| modern typesetting of Arabic | 5 | 5 Special applications | 22 |
| 2.3 Typing Arabic | 6 | 6 Transliteration | 23 |
| 2.3.1 Local options | 7 | 6.1 Additional note on <code>dmg</code> convention | 25 |
| 3 Standard ArabTeX input | 8 | 6.2 Examples | 26 |
| 3.1 Consonants | 8 | 7 Buckwalter input scheme | 26 |
| 3.2 Vowels | 9 | 8 L^AT_EX Commands in Arabic environments | 29 |
| 3.2.1 Long vowels | 9 | 8.1 <code>csquotes</code> | 31 |
| 3.2.2 Short vowels | 9 | 8.2 <code>reledmac</code> | 32 |
| 4 arabluatex in action | 10 | 9 Future work | 32 |
| 4.1 The vowels and diphthongs | 10 | 10 Implementation | 32 |
| 4.2 Other orthographic signs | 12 | References | 37 |
| 4.3 Special orthographies | 17 | Change History | 38 |
| 4.4 Quoting | 18 | Index | 38 |
| 4.4.1 Quoting the <i>hamzah</i> | 19 | | |
| 4.5 The “pipe” character (<code> </code>) | 20 | | |
| 4.6 Stretching characters: the <i>taṭwīl</i> | 21 | | |

List of Tables

| | | | |
|------------------------|----|-----------------------------|----|
| 1 ArabTeX consonants | 8 | 4 “Quoted” <i>hamzah</i> | 20 |
| 2 ArabTeX long vowels | 9 | 5 Additional Arabic codings | 22 |
| 3 ArabTeX short vowels | 10 | 6 Buckwalter scheme | 27 |

Abstract

This package provides for Lua^AT_EX an ArabT_EX-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_EX input notation. Its output can be set in the same modes of vocalization as ArabT_EX, 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>
- 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_EX,¹ ArabLuaT_EX 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^AT_EX users.

¹See <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”.² 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).³ 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.

²Lagally (2004, p. 2).

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

1.1 arablualatex is for Lua⁴LaTeX

It goes without saying that arablualatex requires Lua⁴LaTeX. TeX and LaTeX have arabtex, and Xe⁵LaTeX has arabxetex. Both of them are much more advanced than arablualatex, as they can process a number of different languages,⁴ whereas arablualatex can process only Arabic for the time being. More languages will be included in future releases of arablualatex.

In comparison to arabxetex, arablualatex 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 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 ArabTeX input yet to be processed—for further processing.

2 The basics of arablualatex

2.1 Activating arablualatex

As usual put in your preamble:

```
\usepackage{arablualatex}
```

The only requirement of arablualatex is Lua⁴LaTeX; it will complain if you try to compile your document with another engine. That aside, arablualatex 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 arablualatex. Assuming that fontspec is loaded, put this in your preamble just above the line that loads arablualatex:

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

where *<fontname>* is the standard name of the Arabic font you wish to use.

By default, if no Arabic font is selected, arablualatex will issue a warning message and attempt to load the Amiri font⁵ like so:—

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

⁴To 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.

⁵Hosny (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:—⁶

```
\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: *dammah* (ُ), *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 (ٱٱ).

`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 `ArabTeX` 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 (ٱ), 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*⁷
(see on page 13 for further details).

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

⁷Note 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 اسْتِسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نَجُومِ السَّمَاءِ
- ii. `fullvoc` وَمَاتَ اسْتِسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نَجُومِ السَّمَاءِ
- iii. `trans` *wa-māta 'stisqā^{an} qabla 'ay yutimma kitāba-hu fī nuġūmⁱ 's-samāⁱ*

(b) `\SetArbEasy`:

- i. `voc` وَمَاتَ اسْتِسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نَجُومِ السَّمَاءِ
- ii. `fullvoc` وَمَاتَ اسْتِسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نَجُومِ السَّمَاءِ
- iii. `trans` *wa-māta 'stisqā^{an} qabla 'an yutimma kitāba-hu fī nuġūmⁱ 's-samāⁱ*

(c) `\SetArbEasy*`:

- i. `voc` وَمَاتَ اسْتِسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نَجُومِ السَّمَاءِ
- ii. `fullvoc` وَمَاتَ اسْتِسْقَاءَ قَبْلَ أَنْ يُتِمَّ كِتَابَهُ فِي نَجُومِ السَّمَاءِ
- iii. `trans` *wa-māta 'stisqā^{an} qabla 'an yutimma kitāba-hu fī nuġūmⁱ 's-samāⁱ*

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 الْمَجَاءِ) 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 جَحَا يَطْلُبُ مِنْهُ حِمَارَهُ لِيَرْكَبَهُ فِي سَفَرَةٍ قَصِيرَةٍ فَقَالَ لَهُ: "سَوْفَ أُعِيدُهُ إِلَيْكَ فِي الْمَسَاءِ
وَأَدْفَعُ لَكَ أُجْرَةً." فَقَالَ جَحَا: "أَنَا آسِفٌ جِدًّا أَنِّي لَا أَسْتَطِيعُ أَنْ أَحَقِّقَ لَكَ رَغْبَتَكَ فَالْحِمَارُ لَيْسَ هُنَا
الْيَوْمَ." وَقَبْلَ أَنْ يَتِمَّ جَحَا كَلَامَهُ بَدَأَ الْحِمَارُ يَنْهَقُ فِي إِصْطَبَلِهِ. فَقَالَ لَهُ صَدِيقُهُ: "إِنِّي أَسْمَعُ حِمَارَكَ يَا جَحَا
يَنْهَقُ." فَقَالَ لَهُ جَحَا: "غَرِيبُ أَمْرِكَ يَا صَدِيقِي أَتَصَدِّقُ الْحِمَارَ وَتَكْذِبُنِي؟"

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 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:—

| | |
|----------------------|--|
| <code>voc</code> | – <code>\arb[voc]{\langle Arabic text \rangle};</code> |
| <code>fullvoc</code> | – <code>\arb[fullvoc]{\langle Arabic text \rangle};</code> |
| <code>novoc</code> | – <code>\arb[novoc]{\langle Arabic text \rangle};</code> |
| <code>trans</code> | – <code>\arb[trans]{\langle Arabic text \rangle}.</code> |

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 ⁸ | | ArabTeX notation |
|----------------|------------------------------|-----------|-----------------------------------|
| | dmg | loc | |
| ا ⁹ | <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>ḏ</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>‘</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> |

⁸See below section 6 on page 23.

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

| Letter | Transliteration | | ArabT _E X 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 ArabT_EX (consonants)

3.2 Vowels

3.2.1 Long vowels

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

| Letter | Transliteration ¹⁰ | | ArabT _E X 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 ¹¹ | <i>ā</i> | <i>á</i> | _A or Y |
| أ | <i>ā</i> | <i>ā</i> | _a |
| ـ | <i>ū</i> | <i>ū</i> | _u |
| ـ | <i>ī</i> | <i>ī</i> | _i |

Table 2: Standard ArabT_EX (long vowels)

REM. *a*. The long vowels *ā*, *ū*, *ī*, otherwise called *hurūf^u 'l-maddⁱ*, 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*, قَالَ يَقُولُ *yaqūlu*.

REM. *b*. Defective writings, such as ا, *al-'alif^u 'l-mahḍūfat^u*, 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 حَضِيْفَةُ بَنِ الْيَمَانِ for *Huḍayfat^u bn^u 'l-Yamānī*, etc.

3.2.2 Short vowels

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

¹⁰See below section 6 on page 23.

¹¹= *al-'alif^u 'l-maqṣūrat^u*.

| Letter | Transliteration ¹² | | ArabTeX notation |
|--------|-------------------------------|-----------|------------------|
| | dmg | loc | |
| َ | <i>a</i> | <i>a</i> | a |
| ُ | <i>u</i> | <i>u</i> | u |
| ِ | <i>i</i> | <i>i</i> | i |
| ْ | <i>an</i> | <i>an</i> | aN |
| ُن | <i>un</i> | <i>un</i> | uN |
| ِنْ | <i>in</i> | <i>in</i> | 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

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*, *damma* 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 شَمْسٌ *šams^{un}*, karImuN كَرِيمٌ *Karīm^{un}*.

bi-hi بِهِ *bi-hi*, 'aqi.tuN أَقِطُ *'aqiṭ^{un}*.

la-hu لَهُ *la-hu*, .hujjaTuN هُجَّةٌ *huġġat^{un}*.

¹²See below section 6 on page 23.

¹³See below section 4.4 on page 18.

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

qAla قَالَ *qāla*, bI`a بَعَّ *bī`a*, .tUruN طُور *ṭūr^{un}*, .tInuN طِين *ṭīn^{un}*,
murU'aTuN مُرُوَّةٌ *murū`at^{un}*.

‘alif maqṣūrah It is written ⟨_A⟩ or ⟨Y⟩:

al-fat_A أَلْفَتَى *al-fatā*, al-maqh_A أَلْمَقْهَى *al-maqhā*, 'il_A إِلَى *ilā*.

‘alif otiosum Said ‘*alif^u* ‘*l-wiqāyatⁱ*, “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 نَصَرُوا *naṣarū*, katabUA كَتَبُوا *katabū*, ya.gzUA يَغْزُوا *yağzū*, ramaw
رَمَوْا *ramaw*, banaWA بَنَوْا *banaw*.

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

al-l_ah-u أَللهُ *al-lāh^u*, 'il_ahuN إِلَهٌ *ilāh^{un}*.
al-ra.hm_an-u الرَّحْمَنُ *ar-raḥmān^u*, l_akin لَكِنْ *lākin*, h_ahunA هَاهُنَا *hāhunā*,
.hunayn-u bn-u 'is.h_aq-a حُنَيْنٌ بْنُ إِسْحَقَ *Hunayn^u bn^u 'Ishāq^a*, rabb_i
رَبِّ *rabbī*, al-`A.s_i الْعَاصِ *al-Āṣī*.

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

.hay_aUTuN حَيَوةٌ *hayāt^{un}*, .sal_aUTuN صَلَوةٌ *ṣalāt^{un}*, mi`sk_aUTuN مَشْكُوةٌ *miškāt^{un}*,
tawr_aITuN تَوْرِيَّةٌ *tawrāt^{un}*.

And so also: al-rib_aIT-u أَلرِبِيَّةُ *ar-ribāt^u*.

‘Amr^{un}, and the silent و To that name a silent و is added to distinguish it from ‘*Umar^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 عَمَرُوا *amr^{un}*, `amraNU عَمَرُوا *amr^{an}*, `amriNU عَمَرُوا *amrⁱⁿ*.

When the *tanwīn* falls away (Wright 1896, i. 249 B): `amr-uU bn-u
mu.hammadIN عَمَرُوا بْنُ مُحَمَّدٍ *Amr^u bn^u Muḥammadⁱⁿ*, mu.hammad-u bn-u
`amr-iU bn-i_hAlidiN عَمَرُوا بْنُ خَالِدٍ *Muḥammad^u bn^u Amrⁱ bnⁱ*
Hālidⁱⁿ.

And so also: al-rib_aUA أَلرِبَا *ar-ribā*, ribaNU رِبَا *rib^{an}*.

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 ة, and *tanwīn* preceding ى as in هدى, which is written $\langle aN_A \rangle$ or $\langle aNY \rangle$:

mAluN مَال *māl^{un}*, bAbaN أَبَا *bāb^{an}*, madInaTaN مَدِينَة *madīnat^{an}*, bintiN بنت *bintⁱⁿ* maqhaN_A مَقْهَى *maqhaⁿ*, fataNY فَتَى *fataⁿ*.

`arabluatex` is aware of special orthographies: َ شَيْءٌ *šay^{un}*, َ شَيْئًا *šay^{an}*, ِ شَيْءٍ *šayⁱⁿ*.

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^{an}*, qA.diNI قَاضٍ *qāḍiⁿ*.

4.2 Other orthographic signs

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

madInaTuN مَدِينَة *madīnat^{un}*, madInaTaN مَدِينَة *madīnat^{an}*, madInaTiN مَدِينَة *madīnatⁱⁿ*.

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^{un}*, 'u_htuN أُخْتُ *uht^{un}*, 'iqlIduN إِقْلِيدٌ *iqlīd^{un}*, 'anna أَنْ *anna*, 'inna إِنْ *inna*.

hamzah followed by the long vowel ُ is encoded ' _U: ' _U1_A أُولَى *ulā*, ' _U1U أُولُو *ulū*, ' _U1A'ika أُولَآئِكَ *ulā'ika*.

hamzah followed by the long vowel ِ is encoded ' _I: ' _ImAnuN إِيْمَانٌ *īmān^{un}*.

Middle hamzah: xA.ti'-Ina خَاطِنٌ *hāṭi^{ina}*, ru'UsuN رُوُوسٌ *ru'ūs^{un}*, xa.tI'aTuN خَاطِنَةٌ *hāṭi'at^{un}*, su'ila سُئِلَ *su'ila*, 'as'ilaTuN أَسْئِلَ *as'ilat^{un}*, mas'alaTuN مَسْأَلَةٌ *mas'alat^{un}*, 'as'alu أَسْأَلَ *as'alu*, yataSA'alUna يَتَسَاءَلُونَ *yatasā'alūna*, murU'aTuN مُرُوءَةٌ *murū'at^{un}*, ta'xIruN تَأْخِرُ *ta'hīr^{un}*, ta'axxara تَأْخَرُ *ta'ahhara*, ji'tu-ka جِئْتُكَ *gi'tu-ka*, qA'iluN قَاتِلٌ *qā'il^{un}*.

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 رَأْسُ *ra'ās^{un}*, sa' 'AluN سَأَلَ *sa'āl^{un}*, na' 'AjuN نَأَجَّ *na'āğ^{un}*.

Final hamzah: xa.ta'uN خَطَأُ *ḥaṭa^{un}*, xa.ta'aN خَطَأٌ *ḥaṭa^{an}*, xa.ta'iN خَطِئَ *ḥaṭaⁱⁿ*, 'aqra'u أَقْرَأُ *'aqra'u*, taqra'Inا تَقْرَأُ *taqra'īna*, taqra'Una تَقْرَأُونَ *taqra'ūna*, yaqra'na يَقْرَأُ *yaqra'na*, yaxba'Anي يَخْبَأُ *yaxba'āni*, xaba'A خَبَأَ *ḥaba'ā*, xubi'a خُبِئَ *ḥubi'a*, xubi'UA خُبِئُوا *ḥubi'ū*, jA'a جَاءَ *ğā'a*, ridA'uN رَدَأَ *ridā^{un}*, ridA'aN رَدَأَ *ridā^{an}*, jI'a جِئَ *ğī'a*, radI'iN رَدِئَ *radīⁱⁿ*, sU'uN سُوءُ *sū^{un}*, .daw'uN دَوُّ *daw^{un}*, qay'iN قِيَّ *qayⁱⁿ*, ~sifA'I شِفَائِي *šifā'ī*.

~say'uN شَيْءٌ *šay^{un}*, ~say'aN شَيْئاً *šay^{an}*, ~say'iN شَيْءٍ *šayⁱⁿ*, al-~say'-u أَشْيَاءُ *aš-šay^u*, 'a~syA'-u أَشْيَاءُ *ašyā^u*, 'a~syA'-a أَشْيَاءُ *ašyā^a*, .zim'aN ظَمْنَا *ẓim^{an}*.

taḥfīf^u 'l-hamzatⁱ: if the *hamzah* has *ğazmah* 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 آمَنَ *āmana*, 'u'minu أُمِنَ *ūminu*, 'i'mAnuN إِيمَانٌ *īmān^{un}*. 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 'l-maddi* (*ʿalif* of prolongation) or *ʿalif* with *hamzah* and *ğazmah* (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'kulu أَكُلُ *ākulu* and 'AkiluN أَكَلُ *ākil^{un}* respectively.

arabluatex otherwise determines *al-ʿalif^u 'l-mamdūdāt^u* by context analysis.

'is'AduN إِسَادٌ *is'ād^{un}*, 'AkilUna أَكِلُونُ *ākilūna*, 'a'manna آمَنَّا *āmannā*, al-qur'An-u الْقُرْآنُ *al-qur'ān^u*.

jA'a جَاءَ *ğā'a*, yatasA'alUna يَتَسَاءَلُونَ *yatasā'alūna*, ridA'uN رَدَأَ *ridā^{un}*, xaba'A خَبَأَ *ḥaba'ā*, yaxba'Anي يَخْبَأُ *yaxba'āni*.

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 عَلَقَ *allaqa*, mAdduN مَادَّ *mādd^{un}*, 'ammara أَمَّرَ *ammara*, murruN مَرَّرَ *murr^{un}*.

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 letter that carries it, whether the article be written in the assimilated form or not, e.g. al-^ˆsams-u الشَّمْسُ *aš-šams^u*, or a^ˆs-^ˆsams-u الشَّمْسُ *aš-šams^u*.

al-tamr-u التَّمَرُ *at-tamr^u*, al-ra.hm_an-u الرَّحْمَنُ *ar-raḥmān^u*, al-.zulm-u الظُّلْمُ *aḏ-ḏulm^u*, al-lu.gaT-u اللُّغَةُ *al-luḡat^u*.

- (b) With the letters ر, ل, م, 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 رَبِّي *mir rabbi-hi*, min layliN مِنْ لَيْلٍ *mil laylⁱⁿ*, 'an yaqtula أَنْ يَقْتُلَ *ay yaqtula*.

With *tanwīn*: kitAbuN mubInuN كِتَابٌ مُبِينٌ *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. لَبِثْتُ for لَبِثْتُ *labittu*, 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-waṣlⁱ At the beginning of a sentence, اَ is never written, as اَلْحَمْدُ لِلّٰهِ; instead, to indicate that the 'alif is a connective 'alif ('alif^u 'l-waṣlⁱ), the *hamzah* is omitted and only its accompanying vowel is expressed:

al-.hamd-u li-l-l_ah-i اَلْحَمْدُ لِلّٰهِ *al-ḥamd^u li-llāhⁱ*.

As said above on page 5, `fullvoc` is the mode in which `arabluatex` expresses the *sukūn* and the *waṣlah*. `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 `arabluatex` 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 *hamzah*: $\langle u, a, i \rangle$.

Article: `bAb-u 'l-madrasaT-i` *بَابُ الْمَدْرَسَةِ* *bāb^u 'l-madrasatⁱ*, `al-maqAlaT-u 'l-'_U1_A` *الْمَقَالَةُ الْأُولَى* *al-maqālat^u 'l-'ūlā*, `al-lu.gaT-u 'l-'_arabiyyaT-u` *اللُّغَةُ الْعَرَبِيَّةُ* *al-luġat^u 'l-'arabiyyat^u*, `fI .sinA`aT-i 'l-.tibb-i` *فِي صِنَاعَةِ الطِّبِّ* *fī ṣinā'atⁱ 'l-ṭibbⁱ*, `'il_A 'l-intiqA.d-i` *إِلَى الْإِنْتِقَاضِ* *'ila 'l-intiqādⁱ*, `fI 'l-ibtidA'-i` *فِي الْإِبْتِدَاءِ* *fī 'l-ibtidā'ⁱ*, `'abU 'l-wazIr-i` *أَبُو الْوَزِيرِ* *'abu 'l-wazīrⁱ*, `fa-lammA ra'aW 'l-najm-a` *فَلَمَّا رَأَوْا النَّجْمَ* *fa-lammā ra'awu 'n-naġm^a*.

Particles:—

- (a) *li-*: *'alif^u 'l-waṣlⁱ* is omitted in the article *أَلْ* when it is preceded by the preposition *لِ*: `li-l-rajul-i` *لِلرَّجُلِ* *li-r-raġulⁱ*.
If the first letter of the noun be *لِ*, then the *لِ* of the article also falls away, but `arabluatex` is aware of that: `li-l-laylaT-i` *لِلَّيْلَةِ* *li-l-laylatⁱ*.
- (b) *la-*: the same applies to the affirmative particle *لَ*: `la-l-.haqq-u` *لِلْحَقِّ* *la-l-ḥaqq^u*.
- (c) With the other particles, *'alif^u 'l-waṣlⁱ* is expressed: `fI 'l-madInaT-i` *فِي الْمَدِينَةِ* *fī 'l-madīnatⁱ*, `wa-'l-rajul-u` *وَالرَّجُلُ* *wa-'r-raġul^u*, `bi-'l-qalam-i` *بِالْقَلَمِ* *bi-'l-qalamⁱ*, `bi-'l-ru`b-i` *بِالرُّعْبِ* *bi-'r-ru`bⁱ*.

Perfect active, imperative, nomen actionis: `qAla isma`` *قَالَ أَسْمَعُ* *qāla 'sma'*, `qAla uqtul` *قَالَ أَقْتُلْ* *qāla 'qtul*, `huwa inhazama` *هُوَ أَنْهَزَمَ* *huwa 'nhazama*, `wa-ustu`mila` *وَأَسْتَعْمَلَ* *wa-'stu'mila*, `qad-i in.sarafa` *قَدْ أَنْصَرَفَ* *qadi 'nṣarafa*, `al-iqtidAr-u` *الْإِقْتِدَارُ* *al-iqtidār^u*, `'il_A 'l-intiqA.d-i` *إِلَى الْإِنْتِقَاضِ* *'ila 'l-intiqādⁱ*, `lawi istaqbala` *لَوْ اسْتَقْبَلَ* *lawi 'staqbala*.

Other cases: 'awi ismu-hu ^{أَوْ اسْمُهُ} 'awi 'smu-hu, zayduN ibn-u 'amriNU
 زَيْدُ بْنُ عَمْرٍو Zayd^{uni} 'bn^u 'Amrⁱⁿ,¹⁴ 'umar-u ibn-u 'l-ha.t.tAb-i عُمَرُ
 ابْنُ الْخَطَّابِ Umar^u 'bn^u 'l-Haṭṭābⁱ,¹⁵ imru'-u 'l-qays-i اِمْرُؤُ الْقَيْسِ Imru^u
 'l-Qaysⁱ, la-aymun-u 'l-l-ah-i لَا يُؤْمِنُ اللَّهُ la-'ymun^u 'l-lāhⁱ.

'alif^u 'l-waṣlⁱ 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 arabluatex takes it into account in some transliteration standards:—

fI 'l-nAs-i النَّاسِ فِي fi 'n-nāsⁱ, 'abU 'l-wazIr-i أَبُو الْوَزِيرِ 'abu 'l-wazīrⁱ,
 fI 'l-ibtidA'-i الْإِبْتِدَاءِ فِي fi 'l-ibtidāⁱ, _dU 'l-i`lAl-i ذُو الْأَعْلَالِ du
 'l-i'lālⁱ, maqh_A 'l-'amIr-i الْأَمِيرِ maqha 'l-'amīrⁱ.

'alif^u 'l-waṣlⁱ 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*. arabluatex detects the cases in which this rule applies:—

fI `aynay 'l-malik-i عَيْنِي الْمَلِكِ fi 'aynayi 'l-malikⁱ, ix`say 'l-qawm-a
 اخْتَبَيْ الْقَوْمَ ihšayi 'l-qawm^a, mu.s.tafaw 'l-l-ah-i مُصْطَفَوُ اللَّهِ muṣṭafawu
 'l-lāhⁱ.
 ramaW 'l-.hijAraT-a رَمَوْا الْحِجَارَةَ ramawu 'l-ḥiğārat^a, fa-lammA ra'aW 'l-najm-a
 فَلَمَّا رَأَوْا النَّجْمَ fa-lammā ra'awu 'n-nağm^a.

'alif^u 'l-waṣlⁱ 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-kādib^{ūna}, ra'aytumU
 رَأَيْتُمُ الرَّجُلَ ra'aytumU 'r-rağul^a, mani 'l-ka_d_dAb-u مَنِ
 الْكَذَّابُ mani 'l-kaddāb^u, qatalati 'l-rUm-u قَتَلَتِ الرُّومُ qatalati 'r-Rūm^u.

However, the Arabic script does not shows the *kasrah* 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 مُحَمَّدُ النَّبِيّ Muḥammad^{uni} 'n-nabī.

¹⁴ “Zayd is the son of ‘Amr”: the second noun is not in apposition to the first, but forms part of the predicate. Hence زَيْدُ بْنُ عَمْرٍو and not زَيْدُ بْنُ عَمْرٍو, “Zayd, son of ‘Amr”.

¹⁵ “Umar is the son of al-Haṭṭāb” (see footnote 14).

4.3 Special orthographies

The name of God The name of God, ﷲ, is compounded of the article ﷲ, and ﷲ (noted ﷲ with the defective ‘*alif*’) so that it becomes ﷲ; then the *hamzah* is suppressed, its vowel being transferred to the ﷲ before it, so that there remains ﷲ (I refer to Lane, *Lexicon*, I. 83 col. 1). Finally, the first ﷲ 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 14), the name of God is therefore encoded *al-l_ah-u* or ‘*l-l_ah-u*:—

al-l_ah-u ﷲ *al-lāh*^u, yA¹⁶ *al-l_ah-u* ﷲ *yā al-lāh*^u, ‘a-fa¹⁷-*al-l_ah-i*
la-ta.g`alanna ﷲ ﷲ *‘a-fa-al-lāhⁱ la-taḡ‘alanna*, bi-‘*l-l_ah-i*
ﷲ *bi-‘l-lāhⁱ*, wa-‘*l-l_ah-i* ﷲ *wa-‘l-lāhⁱ*, bi-sm-i ‘*l-l_ah-i* ﷲ
ﷲ *bi-smⁱ ‘l-lāhⁱ*, al-.hamd-u li-l-l_ah-i ﷲ *al-ḥamd^u li-llāhⁱ*,
li-l-l_ah-i ‘*l-qA‘il-u* ﷲ *li-llāhⁱ ‘l-qā‘il^u*.

The conjunctive ﷲ Although it is compounded of the article ﷲ, the demonstrative letter ﷲ and the demonstrative pronoun ﷲ, 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* ﷲ
ﷲ *‘ahāfu mina ‘l-malikⁱ ‘lladī yazlimu ‘n-nās^a*, ‘*udtu*
‘*l-ˆsay_h-a* ‘*lla_dI* huwa marI.duN ﷲ ﷲ *‘udtu*
‘*š-ˆsayh^a ‘lladī huwa marīd^{un}*, mA ‘anA bi-‘*lla_dI* qA‘iluN la-ka
ˆsay‘an ﷲ ﷲ *‘anā bi-‘lladī qā‘il^{ul} la-ka šay‘an*.
‘ari-nA ‘*lla_dayni* ‘a.dallA-nA mina ‘*l-jinn-i* wa-‘*l-‘ins-i*
ﷲ ﷲ *‘ari-nā ‘lladayni ‘aḡallā-nā mina ‘l-ḡinnⁱ*
wa-‘l-‘insⁱ.

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 ﷲ ﷲ *fa-‘innā nadkuru ‘š-ṣawt^{ayni}*
‘l-ladayni rawaynā-humā ‘an Ḡaḡṣat^a.

¹⁶Note the “pipe” character ‘|’ here after *yA* and below after *fa* before footnote mark 17: it is needed by the *dmḡ* 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-ṭarīqⁱ*. See section 4.5 on page 20 on the “pipe” and section 6 on page 23 on *dmḡ* mode.

¹⁷See footnote 16.

And also: al-la_dAni اللّٰذَانِ *al-ladāni*, al-la_dayni اللّٰذَيْنِ *al-ladayni*,
al-latAni اللّٰتَانِ *al-latāni*, al-latayni اللّٰتَيْنِ *al-latayni*, al-lAtI اللّٰتِي *al-lātī*,
al-lA'Ati¹⁸ اللّٰءَاتِ *al-lā'āti*, al-lA'I اللّٰئِي *al-lā'ī*, 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: `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`, 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 علم الهيئة` *'ullima 'ilm^a 'l-hay'atⁱ*,
`ya.gz"UA يغزوا` *yaǧzū*.

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

`wa-'innA sawfa tudriku-nA 'l-manAyA muqadd"araT"aN وإنا سوف`
`تدركنا المنايا مقدرة`, *wa-'innā sawfa tudriku-na 'l-manāyā muqaddarat^{an}*.

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

`qAla isma`" قال اسمع` *qāla 'sma', jA'at" hinduN جاءت هند* *ǧā'at*
`Hindun, `sabIhuN bi-man q"u.ti`at" qadamA-hu شبيه بن قطعت`
`قدماه` *ṣabīh^{un} bi-man quṭi'at qadamā-hu*.

¹⁸Note 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.

¹⁹See Lagally (2004, p. 22)

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

wa-"ust"u`mila وأَسْتَعْمَل *wa-'stu'mila*, huwa "inhazama هو أَنَهَزَم *huwa 'nhazama*, al-"intiqA.ḏu الِاتِّقَاضُ *al-intiqāḏu*.

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 قَالَتْ قَالَتْ *qāla qā'il^{un}*, ibn-u 'abI 'u.saybi`aT-"a
ابْنُ أَبِي أُصَيْبَةَ *Ibn^u 'Abī 'Uṣaybi'at^a*.

- (b) *tanwīn*:—

madInaT"aN مَدِينَةٌ *madīnat^{an}*, bAb"aN أَبَا *bāb^{an}*, hud"aN_A هُدًى *hudāⁿ*,
شَيْءٌ *say[']"iN šayⁱⁿ*.

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^u 'l-waṣl'*:—

fI "istisqA'-a فِي اسْتِسْقَاءٍ *fi 'stisqā'^a*, wa-"istisqA'-u وَاسْتِسْقَاءٌ *wa-'stisqā'^u*,
qAla "uhrub fa-lan tuqtala قَالَهُ هَرُبْ فَلَنْ تُقْتَلَ *qāla 'hrub fa-lan tuqtala*.

- (b) *sukūn*:—

qAla "uqtul" fa-lan tuqtala قَاتِلُ فَلَنْ تُقْتَلَ *qāla 'qtul fa-lan tuqtala*, mā jA'at" mini imra'aTIN مَا جَاءَتْ مِنْ امْرَأَةٍ *mā ḡā'at mini 'mra'atⁱⁿ*, kam" qad" ma.dat" min" laylaTiN كَمْ قَدْ مَضَتْ مِنْ لَيْلَةٍ *kam qad maḏat min laylatⁱⁿ*.

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 *al-ḡamr^u 'ṣ-ṣayfiyy^u 'llaḏī kāna bi-Qrānnūn^a*.

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 ²⁰ | | ArabTeX notation |
|--------|-------------------------------|-----|------------------|
| | dmg | loc | |
| ء | ’ | ’ | "' |
| آ | ’ā | ’ā | A"' |
| أ | ’ | ’ | a"' |
| إ | ’ | ’ | u"' |
| ؤ | ’ | ’ | w"' |
| إ | ’ | ’ | i"' |
| ئ | ’ | ’ | 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, ءُ is encoded <w" 'a>, while ءُ 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'ikُم أَعْدَائِكُم 'a`dā'ikُم, 'a`dA|" "'ikُم أَعْدَاءُكُمْ 'a`dā'ikُم.

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 بِقْرَانُونْ *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 14) from being applied.

²⁰See below section 6 on page 23.

²¹See 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^u bn^u 'Ishāq^a*

4.7 Digits

4.7.1 Numerical figures

The *Indian numbers*, *ar-raqam^u 'l-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 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 *ʿabjad* numbers are usually distinguished from the surrounding words by a stroke placed over them.

`\abjad` *ʿabjad* 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 45 *kitābu-hu fi 'l-ādātⁱ*.

REM. *a.* As can be seen in the above given example, `arabluatex` expresses the *ʿabjad* 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]{ʿabjad}` number for 1874 is `\abjad{1874}` The *ʿabjad* number for 1874 is غَضَد.

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

For the time being, the following is included:—

Table 5: Additional Arabic codings

4.9 Arabic emphasis

`\aemph` One may use the command `\aemph{\i{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*.

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 *tašdid*. This is useful to make linguistic annotations and comments on vowels:—

Bu Ba Bi BuN BaN BiN $\overset{\textcircled{a}}{\text{---}} \overset{\textcircled{b}}{u} \overset{\textcircled{c}}{a} \overset{\textcircled{d}}{i} \overset{\textcircled{e}}{un} \overset{\textcircled{f}}{an} \overset{\textcircled{g}}{in}$, BBu BBa BBi $\overset{\textcircled{h}}{\text{---}} \overset{\textcircled{i}}{u} \overset{\textcircled{j}}{a} \overset{\textcircled{k}}{i}$, B $\overset{\textcircled{l}}{\text{---}}$ aN
 $\overset{\textcircled{m}}{\text{C}}_{-an}$, B $\overset{\textcircled{n}}{\text{---}}$.

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: `{}`

`\abracas` 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 `\abracas`:—

```
1 \begin{arab}
2   \abracas{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 الزَّمَازِمَةِ.

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.:—

²²See below section 6.

²³See Brockelmann et al. (1935).

²⁴See <http://www.loc.gov/catdir/cpsd/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{}` 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 إِسْحَقَ *Hunayn^u bn^u*
Ishāq^a, `\cap{`u_tm-an-u}` عُثْمَانُ *Utmān^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}` عَبْدُ اللَّهِ بْنُ عَوْفِ بْنِ خَلْدٍ سَعْدُ بْنُ زَيْدُ بْنُ ضَرَبَ *daraba Zayd^u bn^u*
Hālidⁱⁿ Sa`d^a bn^a `Awfⁱ bnⁱ `Abdⁱ `l-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^{umi} 'n-nabī*, as nouns having the *tanwīn* take a *kasrah* in pronunciation before *'alifu 'l-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_EX 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} أَبُو بَكْرٍ مُحَمَّدُ بْنُ زَكْرِيَّا الرَّازِي Abū Bakrin Mu-
hammadu bnu Zakariyyāa 'r-Rāzī.
```

6.1 Additional note on dm̄g convention

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

- (a) In full: *Amrun*;
- (b) As superscript text: *Amr^{un}*;
- (c) Discarded: *Amr*.

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

`\NoArbUp` `\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 *ʾiṣrāb* is displayed. To take one example, here is how Arabic *ʾiṣrāb* may be rendered as subscript text:—

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

Arabic dm̄g transliteration for رَأَيْتُ جَامِعًا مُهْدَمَةً مِثْلَتَهُ: *ra'aytu ḡāmi^c_{am} muhaddamat_{am} 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.

ʾiṣrāb boundaries Every declinable noun (*muṣrab*) may be declined either with or without *tanwīn*, viz. *munṣarīf^{un}* or *ḡayr^u munṣarīfⁱⁿ*. 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'inātun, \cap{'amraNU} عَمْرُو Amran, fataN_A فَتَى fatan, qA.diNI
قَاضٍ qāḍin.
```

ġayr munṣarif: al-mu`allim-u ^{المُعَلِّمُ} *al-mu'allim^u*, kitAb-Ani كِتَابَانِ *kitāb^{āni}*,
 ra`sa'-Ani رَسَائِنِ *raša'āni*, sAriq-Una سَارِقُونَ *sāriq^{āna}*, qA.d-Una قَاضُونَ *qāḍ^{āna}*,
 al-.zulm-Atu الظُّلْمَاتُ *aẓ-ẓulm^{ātu}*.

REM. *a.* As the *tanwīn* is passed over in pronunciation when it is followed by the letters ر, ل, م, 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 ġubbaTaN mumazzaqaN 'aydu-hA ^{كَانَ غَنِيًّا لَكِنَّهُ لَبَسَ جُبَةً مُمَزَّجًا}
^{أَيْدُهَا} *kāna ġaniyy^{al} lākinna-hu labisa ġubbat^{am} mumazzaq^{am} '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 ^{مَنَّعَ النَّاسَ كُلَّهُ مِنْ}
^{بِسَيِّدِنَا} *mana'a 'n-nās^a kāffat^{am} mim muḥāṭabati-hi 'ahad^{un} bi-sayyidi-nā.*

6.2 Examples

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

‘dmg’ standard: ‘atā ṣadīq^{un} ‘ilā Ġuḥā yaṭlubu min-hu ḥimāra-hu li-yarkaba-hu fī safratⁱⁿ qaṣīratⁱⁿ fa-qāla la-hu: “sawfa u’idu-hu ‘ilay-ka fī ‘l-masā’i wa-’adfa’u la-ka ‘uġrat^{an}.” fa-qāla Ġuḥā: “anā ‘āsif^{un} ġidd^{an} ‘annī lā ‘astatī’u an ‘uḥaqqiqa la-ka raġbata-ka fa-‘l-ḥimār^u laysa huna ‘l-yawm^a.” wa-qabla ‘ay yutimma Ġuḥā kalāma-hu bada’a ‘l-ḥimār^u yanhaqu fī ‘iṣṭabli-hi. fa-qāla la-hu ṣadīq^u-hu: “innī ‘asma’u ḥimāra-ka yā Ġuḥā yanhaqu.” fa-qāla la-hu Ġuḥā: “ġarīb^{un} ‘amru-ka yā ṣadīqī ‘a-tuṣaddiqu ‘l-ḥimār^a wa-tukaddiba-nī?”

‘loc’ standard: atā ṣadīqun ilā Juḥā yaṭlubu min-hu ḥimāra-hu li-yarkaba-hu fī safratⁱⁿ qaṣīratⁱⁿ fa-qāla la-hu: “sawfa u’idu-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 raġbata-ka fa-al-ḥimāru laysa huna al-yawma.” wa-qabla an yutimma Juḥā kalāma-hu bada’a al-ḥimāru yanhaqu fī iṣṭabli-hi. fa-qāla la-hu ṣadīq^u-hu: “innī asma’u ḥimāra-ka yā Juḥā yanhaqu.” fa-qāla la-hu Juḥā: “gharībun amru-ka yā ṣadīqī a-tuṣaddiqu al-ḥimāra wa-tukadhdhiba-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.²⁵ 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

²⁵See <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).²⁶ However, the following limitation apply to the ‘base’ and ‘xml’ schemes: the braces { and }, which are used to encode ا and ئ, 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 ²⁷ | | 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>ḥ</i> | <i>ḥ</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>ḍ</i> | <i>ḍ</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 |

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

²⁷See 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 | — | — | ~ | ~ |
| ـ | ’ | ’ | [| 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 |
| ون | <i>un</i> | <i>un</i> | N | N |
| ين | <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^u*, Al-camsu الشَّمْسُ *aš-šams^u*, bi-SinaAEapi
Al-T~ib~i, بِصِنَاعَةِ الطِّبِّ *bi-ṣinā‘atⁱ ṭ-ṭibbⁱ*.

wa-Al-l~ehi وَاللهِ *wa-l-lāhⁱ*, Al-Hamdu li-l~ehi اَلْحَمْدُ لِلّٰهِ *al-ḥamd^u*
li-llāhⁱ.

Similarly, it is not advisable to use | and [(‘base’ and ‘xml’ schemes) or M and L (‘safe’ scheme) to encode the ‘*alif^u*’ ‘*l-mamdūdātⁱ*’ and the ‘*alif^u*’ ‘*l-waṣlⁱ*’ 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 اِلَى الْاِتِّقَاضِ as expected, but only `<iLY Al-intiqāADi` can be transliterated as *ila l-intiqāḍi* with the correct vowel ⟨i⟩ in place of the ‘*alif^u*’ ‘*l-waṣlⁱ*’.

8 L^AT_EX Commands in Arabic environments

General principle L^AT_EX 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}}` مَهْ كِتَابُهُ فِي الْعَادَاتِ
kitābu-hu fī l-‘Ādātⁱ.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 buyŪt-i
6 'l-zamAzimaT-i\footnote{al-zamAzimaT-u: .tA'ifaT-u mina
7 'l-furs-i.}.
8 \end{arab}
```

إِنَّ أَيْ كَانَ مِنَ الْمُقَاتِلَةِ^a, وَكَانَتْ أُمِّي مِنْ عُظَمَاءِ بُيُوتِ الزَّمَاذِمَةِ^b.

^aالْمُقَاتِلَةُ: الْمُقَاتِلِينَ.
^bالزَّمَاذِمَةُ: 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

²⁸This 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^AT_EX 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:—

```
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 أُخْرَىٰ فِي **مَطْلَعِ الشَّمْسِ** وَلَا بِنَاءَ لَهُمْ يَوْمَهُمْ مِنَ الشَّمْسِ.

`\LRfootnote` `\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}
```

زَيْدٌ ^aابْنُ عَمْرٍو

^a “*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^AT_EX 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`.²⁹

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

`\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 29 applies.

`\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:—

```

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

```

بَرْزَوِيهِ لِبَرْجُمِهِرِ بْنِ الْبُخْتِكَانِ
قَالَ بَرْزَوِيهِ بْنُ أَزْهَرَ، رَأْسُ أَطْبَاءِ فَارَسَ...

8.1 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:—

```

1 \usepackage{csquotes}
2 \DeclareQuoteStyle{arabic}
3 {\rmfamily\textquotedblright}{\rmfamily\textquotedblleft}
4 {\rmfamily\textquoteright}{\rmfamily\textquoteleft}

```

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

```

1 \setquotestyle{arabic}
2 \begin{arab}
3     fa-qAla la-hu ju.hA: \enquote{.garIb-uN 'amru-ka yA .sadIqI
4         'a-tu.saddiqu 'l-.himAr-a wa-tuka_d_diba-nI?}
5 \end{arab}
6 \setquotestyle{english}

```

فَقَالَ لَهُ جَحَا: "غَرِيبُ أَمْرِكَ يَا صَدِيقِي أَتُصَدِّقُ الْحِمَارَ 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.

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 Arab_{TEX} 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 Lua_{AT_EX} 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 T_EXlive:

```

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_T_E_X 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*
`\SetArbEasy*` *mamdūdah* and *sukūn* depending on the modes which are selected, either `voc`,
`\SetArbDflt` `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}{}{\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(processstrans(\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(processstrans(\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}

\abracess \abracess{(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{\abracess}{+m}{%
129 \luadirect{tex.sprint(abracess(\luastringN{#1}))}}

\LRmarginpar \LRmarginpar is supposed to be inserted in an Arabic environment. It typsets 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 col- lab. with Ph. S. van Ronkel and Otto Spies. Deutschen Morgenländischen Gesellschaft. Leipzig: Deutschen Morgenländischen Gesellschaft, in Kommis- sion bei F. A. Brockhaus. URL: http://www.naher-osten.uni-muenchen.de/studium_lehre/werkzeugkasten/dmgtransliteration.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 | \SetInputScheme: | |
| v1.0.1 | General: Minor update of the documentation | 1 | | \SetInputScheme may be used to process other input schemes such as ‘Buckwalter’ | 34 |
| v1.1 | \abjad: New and more flexible \abjad command. | 36 | | \SetTranslitFont: For selecting a specific font for transliterated texts | 34 |
| v1.2 | \SetArbEasy: New \SetArbEasy/\SetArbDflt for ‘modern’ or ‘classic’ Arabic styles. | 34 | v1.4.3 | \abraces: New command \abraces which expresses its argument between braces. | 37 |
| v1.3 | \arbup: <i>īrāb</i> is now written as superscript text in <code>dmg</code> mode by default. | 35 | v1.4.4 | \SetArbEasy*: this starred version discards the <i>sukūn</i> in addition to what is already discarded by \SetArbEasy. | 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.

| Symbols | A | \al@arb@rules |
|-----------------------------|-----------------------------------|-------------------------|
| \@ifpackageloaded | | 48, 51, 52, 53, |
| \@tempa 73, 74, | \abjad 21, <u>120</u> | 77, 82, 87, 93, |
| 79, 84, 89, 98, | \abraces 23, <u>128</u> | 102, 107, 112, 117 |
| 99, 104, 109, 113 | | |
| _ 32 | \aemph 22, <u>43</u> | \al@arbup |
| | | . 61, 62, 63, 64, 65 |

| | | |
|---|---|--|
| <code>\al@arbup@dflt</code> | D | <code>luaotfload</code> (package) . . . 24 |
| 60, 61, 63 | <code>\DeclareOption</code> | <code>\luastringN</code> 67, |
| <code>\al@input@scheme</code> | 12, 13, 14, 15 | 76, 81, 86, 91, 129 |
| 46, 47, | <code>\deffootnote</code> . . 137, 139 | <code>\luastringO</code> . 77, 82, |
| 77, 82, 87, 94, | E | 87, 92, 93, 94, |
| 102, 107, 112, 118 | environments: | 101, 102, 106, |
| <code>\al@mode</code> 12, | arab 7, 97 | 107, 111, 112, |
| 13, 14, 15, 72, 97 | <code>\ExecuteOptions</code> 16 | 115, 116, 117, 118 |
| <code>\al@mode@fullvoc</code> | F | M |
| 19, 79, 104 | <code>\FixArbFtnmk</code> . . . 30, 135 | <code>\marginpar</code> 130 |
| <code>\al@mode@novoc</code> | fontspec (package) | N |
| 20, 84, 109 | 4, 24, 34 | <code>\NewEnviron</code> 97 |
| <code>\al@mode@trans</code> | <code>\footnote</code> 132, 134 | <code>\newfontfamily</code> 34 |
| 21, 89, 113 | fullvoc (option) 5, 8 | <code>\NoArbUp</code> 25, 60 |
| <code>\al@mode@voc</code> . 18, 74, 99 | G | novoc (option) 5, 8 |
| <code>\al@trans@convention</code> | <i>Grammar of the Arabic</i> | O |
| . . . 58, 59, 92, 116 | <i>Language, A</i> | options: |
| <code>\al@trans@font</code> | 5, 7, 8, | fullvoc 5, 8 |
| 54, 55, 71 | 11, 13, 14, 16, 21 | novoc 5, 8 |
| <code>\al@trans@style</code> | H | trans 5, 8 |
| . . . 56, 57, 90, 114 | Habash, Nizar Y. 27 | voc 5, 8 |
| <i>Amiri</i> 4, 5 | Hosny, Khaled 4, 5 | <code>\overline</code> 44, 45 |
| amiri (package) 5 | I | P |
| arab (environment) . 7, 97 | <code>\ifal@mode@defined</code> . 22 | <code>\par</code> 98, 119 |
| <i>ArabTeX</i> 3, 18 | <code>\IfBooleanTF</code> 50 | <code>\pardir</code> 35, |
| <i>Arabic-English Lexicon,</i> | <code>\ifluatex</code> 5 | 36, 100, 105, |
| <i>An</i> 17 | <i>Introduction to Arabic</i> | 110, 114, 131, 133 |
| <code>\arabicfont</code> . 31, 32, | <i>Natural Lan-</i> | polyglossia (package) . |
| 34, 69, 75, 80, | <i>guage Process-</i> | 4, 33, 36 |
| 85, 100, 105, 110 | <i>ing</i> 27 | <code>\ProcessOptions</code> 17 |
| arabluatex (package) . | <code>\itshape</code> 56 | <code>\ProvidesPackage</code> 2 |
| 2–7, | K | R |
| 9–18, 21–27, | KOMA-script (package) 30 | reledmac (package) . . . 32 |
| 29, 30, 32–34 | L | <code>\RequirePackage</code> |
| arabtex (package) | Lagally, Klaus 3, 18 | 4, 23, 24, |
| . . . 3, 4, 14, 18, 32 | Lane, Edward William . 17 | 25, 26, 27, 28, 138 |
| arabulatem (package) . . 15 | <code>\LR</code> 29, 37, | <code>\RL</code> 30, 40 |
| arabxetex (package) . 4, 14 | 132, 134, 137, 139 | <code>\RLfootnote</code> . . . 30, 131 |
| <code>\arb</code> 6, 72 | <code>\LRfootnote</code> . . . 30, 131 | <code>\rmfamily</code> |
| <code>\arbup</code> 25, 60 | <code>\LRmarginpar</code> . . . 31, 130 | . 38, 39, 41, 42, 54 |
| <code>\ArbUpDflt</code> 25, 60 | luabidi (package) . . . 4, 33 | S |
| B | <code>\luadirect</code> 29, 30, 67, | scrxextend (package) 30, 37 |
| <code>\BODY</code> 101, 106, 111, 115 | 76, 81, 86, 91, | <code>\SetArbDflt</code> 6, 48 |
| Brockelmann, Carl . 23, 25 | 101, 106, 111, | <code>\SetArbEasy</code> 6, 48 |
| C | 115, 123, 126, 129 | |
| <code>\cap</code> 24, 66 | | |
| csquotes (package) . . . 31 | | |

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