ArabTeX
a System for Typesetting Arabic
User Manual Version 3.00

Klaus Lagally

November 22, 1993

\footnote{Report Nr. 1993/11, Universität Stuttgart, Fakultät Informatik, Breitwiesenstraße 20–22, 70565 Stuttgart, Germany}
\footnote{This Report supersedes Report Nr. 1992/06}
Overview

Arab\TeX is a package extending the capabilities of \TeX to generate the Arabic writing from an ASCII transliteration for texts in several languages using the Arabic script. It consists of a \TeX macro package and an Arabic font in several sizes, presently only available in the Naskhi style. Arab\TeX will run with Plain \TeX and also with \La\TeX. It is compatible with NFSS, NFSS2 and the EDMAC package; other additions to \TeX have not been tried.

Arab\TeX is primarily intended for generating the Arabic writing, but the standard scientific transliteration can also be easily produced. For languages other than Arabic that are customarily written in the Arabic script some limited support is available.

Arab\TeX defines its own input notation which is both machine, and human, readable, and suited for electronic transmission and Email communication. However, texts in some of the Arabic standard encodings can also be processed.

Arab\TeX is copyrighted, but free use for scientific, experimental and other strictly private, noncommercial purposes is granted. Offprints of publications using Arab\TeX are welcome. Using Arab\TeX otherwise requires a license agreement. There is no warranty of any kind, either expressed or implied. The entire risk as to the quality and performance rests with the user.

Please send error reports, suggestions and inquiries to the author:

Prof. Klaus Lagally
Institut für Informatik
Universität Stuttgart
Breitwiesenstraße 20-22
70565 Stuttgart
GERMANY
lagally@informatik.uni-stuttgart.de

Copyright © 1992, 1993, Klaus Lagally
Contents

1 Activating ArabTeX 5

2 Input to ArabTeX 6
  2.1 Arabic text elements ........................................ 6
  2.2 Commands in an Arabic context ............................ 7

3 Language selection 10

4 Font selection 11

5 Input coding conventions 12
  5.1 Standard Arabic and Persian characters ................ 12
  5.2 Quoting ...................................................... 15
  5.3 Ligatures ..................................................... 16
  5.4 Vowelization ................................................ 16
  5.5 Verbatim input ............................................. 17
  5.6 Alternate input codings ................................... 17

6 Transliteration 19
  6.1 ZDMG transliteration style ................................. 19
  6.2 Encyclopedia of Islam style ............................... 20

7 Support for other languages besides Arabic 21
  7.1 Persian (Farsi, Dari), also Ottoman, Kurdish .......... 21
CONTENTS

7.2 Urdu ................................. 22
7.3 Pashto (Afghanic) ...................... 22
7.4 Maghribi .............................. 23
7.5 Other languages ....................... 24

8 Miscellaneous features ..... 25
  8.1 Automatic stretching ............... 25
  8.2 Dots on yū’ .......................... 25
  8.3 Additional codings ................... 25
  8.4 Progress report ....................... 26
  8.5 Verbatim copy of the input .......... 27
  8.6 Using ArabTeX with EDMAC .......... 27

9 Acknowledgments ..... 28

10 References ..... 29

A Obtaining ArabTeX ............. 32

B Installing ArabTeX ......... 33

C Release history .............. 34

D Sample ArabTeX input ........... 36

E Sample ArabTeX output ...... 37

F Coding examples for Arabic ...... 38

G Coding examples for Persian ...... 45

H Alternate input encodings ...... 48
  H.1 ASMO 449 = ISO 9036 ............... 48
  H.2 ASMO 449E = ISO 8859-6 ............ 50
I Miscellaneous utilities 52
  1.1 twobkls.sty .......................................................... 52
  1.2 abjad.sty ............................................................ 53
  1.3 MLS2ARAB ............................................................ 53

Index 54
List of Tables

5.1 Standard codings for Arabic and Persian. ................. 13
5.2 Additional codings generally available. .................. 14
5.3 Verbatim codings for the carrier of hamza ............... 17

7.1 Additional codings for Urdu. ............................ 23
7.2 Additional codings for Pashto. ........................... 24

8.1 Additional codings for special purposes. .................. 26

H.1 ASMO 449 code table .................................. 49
H.2 ISO 8859-6 code table .................................. 51
Chapter 1

Activating ArabTeX

With Plain TeX, load the ArabTeX macros by \input arabetex.tex. With \TeX, include the option "arabetex" in the document header. In both cases some additional files will be loaded automatically.

ArabTeX defines several user commands as indicated below. There is also a large number of (hidden) internal commands which could lead to storage (hash table\(^1\)) overflow in a small \TeX implementation. All internal commands contain an “at” sign (@) in their names and thus should not interfere with any user defined commands (but could possibly with other \TeX extensions we do not know about).

With Plain TeX, the Arabic font by default is only available at the normal 14 point size which ought to cooperate well with the "cm" fonts at 10 points. A bold variant is also provided. For other sizes, the user has to change the \magnification or to define additional font identifiers himself. To change the default, inspect the file "arabetex.tex" and redefine the \texttt{\pnash} and/or \texttt{\pnashbf} command accordingly. With \TeX, the usual size changing commands will also operate on the Arabic font.

\(^1\)A \TeX hash table size of 3000 to 3500 is recommended
Chapter 2

Input to ArabTEX

After activating ArabTEX, select one of the Arabic writing styles, e.g., \setarab (see Section 3). Your modified TeX/\LaTeX system will recognize the following items:

- normal TeX/\LaTeX text and commands,
- short Arabic quotations bracketed by < and >. These must normally fit onto one line of output, except if explicitly broken up by \ or \ commands (see below). A quotation may also be started with \ except inside a \LaTeX \{tabbing\} environment.
- longer Arabic texts which are bracketed by \begin{arabtext} and \end{arabtext}, (even when using Plain TeX!), called Arabic Environments in the sequel. An Arabic Environment consists of one or more paragraphs separated by blank lines or \par commands.

Arabic quotations and Arabic environments are called Arabic contexts in the sequel.

2.1 Arabic text elements

Every Arabic paragraph and every Arabic quotation is a sequence of the following kinds of Arabic items, separated by blank spaces or newlines:

- isolated punctuation marks, interpreted as the corresponding Arabic punctuation mark;
• “numbers”, i.e. character sequences starting with a digit. A “number” will be processed using the normal writing sequence from left to right even if it contains letters and/or special characters; however, if the final character is a punctuation mark, it will be split off and processed separately.

• “Arabic quotes” coded as two left quotes or two right quotes each; they may also be written directly adjacent to a word.

• “words”, i.e. character sequences starting with a letter or a special (non-digit) character followed by a letter. A final punctuation mark will be split off and processed separately. The (coded) characters of a word will in the output be arranged from right to left.

• a sequence of words, numbers, and special characters enclosed in curly braces { and }. This introduces a new level of \TeX grouping; otherwise the constituents are processed normally. This feature may be nested.

Output from all items will be arranged from right to left, lines will be broken as necessary.

Inside an \textit{Arabic Environment}, or in an \textit{Arabic quotation}, you may also have:

• \LaTeX\ commands with or without parameters. These will be executed immediately.

• Some, but not all, \TeX\ commands (see below). These will be executed immediately.

• Short mathematical insertions, bracketed by \textit{single} $ signs. They must fit on one output line and are processed as usual. \TeX\ Display mode within an \textit{Arabic environment} is not provided; if it is required, the user has to leave the \textit{Arabic environment} temporarily.

• short \textit{non-Arabic (“Roman”)} quotations, containing text and possibly also \TeX\ commands, bracketed by < and >. These must fit on one output line and introduce a new level of grouping, so if they contain any \TeX\ assignments the effects of these will be local by default. This feature is not available within an \textit{Arabic quotation}. The alternate notation \texttt{\langle} is also not provided.

2.2 Commands in an \textit{Arabic context}

A control sequence inside an \textit{Arabic context} must be separated from the preceding text item by at least one blank space, newline, or another control sequence, and may be of the following kinds:
Arab/\TeX\ option changing commands. These may also be used outside an
Arabic Context, and usually follow the \TeX\ grouping rules.

\\ for a line break; the last line will be padded on the left with spaces.

\| for a line break; the last line will be aligned. If it comes out very badly
spaced, automatic stretching might help (see Section 8).

\indent or \par (or a blank line) for a new paragraph, \noindent for a
new paragraph without indentation; (not inside Arabic quotations).

\texttt{\textbackslash emphasize \texttt{Arabic item}} will put a bar over the Arabic item.

\texttt{\textbackslash emphasize \{group of Arabic items\}} will put a bar over the indicated
group of Arabic items.

\texttt{\textbackslash setnash, \textbackslash setnashbf, \textbackslash setnastaliq} font selection commands, see
Section 4.

size changing \LaTeX\ commands like \texttt{\textbackslash large} etc., only if \LaTeX\ is used!

the following commands: \texttt{\textbackslash footnote} (observe that the syntax for Plain
\TeX\ and \LaTeX\ is different!), \texttt{\textbackslash marginpar} (also with Plain \TeX, analogous
to the \LaTeX\ usage).

the \TeX/\LaTeX\ commands \texttt{\textbackslash smallskip, \textbackslash medskip, \textbackslash bigskip, \textbackslash input,
\textbackslash hfill, \textbackslash (for a space), \textbackslash space} with their usual meaning.

\texttt{\textbackslash nospace} will place the adjacent items in the output in contact, without
any intervening space.

\texttt{\textbackslash hspace \{width\}} will introduce the indicated amount of spacing in the
output.

\texttt{\textbackslash mbox \{text\}} puts the text into a box that will not be split across a line
break.

\texttt{\textbackslash spreadbox \{width\}\{text\}} spreads out the text to the indicated width.
This may be useful e.g., when typesetting poetry.
\texttt{\textbackslash spreadbox \{width\}\{text\}\textbackslash hfill} will inhibit the spreading,
\texttt{\textbackslash spreadbox \{width\}\{\textbackslash hfill text\textbackslash hfill\}} will center the text inside the
box.
\texttt{\textbackslash spreadbox \{width\}\{\textbackslash hfill\}} or \texttt{\textbackslash spreadbox \{width\}\{\textbackslash hfill\}} just
introduces the indicated amount of horizontal space, as will \texttt{\textbackslash hspace \{width\}}.

If two boxing commands follow each other without any intervening blank
space in the input, there will also be no resulting space between the boxes
in the output.
\centerline{text} will start a new line whose contents are centered (not inside Arabic quotations).

\spreadline{text} will start a new line whose contents are spread out over the whole width of the page (not inside Arabic quotations). It is approximately equivalent to \spreadbox{\hspace{text}}.

User defined commands whose expansion produces legal Arab\TeX{} input may be called by \docommand{command and parameters}. The command is expanded exactly once,\footnote{This is no strong restriction as the expansion may contain \docommand{} calls again.} and the result is processed by Arab\TeX{} again. Any side effects of the expansion will be local.

Parameter assignments inside an Arabic context may be performed by \doassign{parameter}{value}. The effect is normally local except if the form \doassign{\global parameter}{value} is used.

Any non-recognized command will generate an error message and will be echoed verbatim in the output. Even though Arab\TeX{} tries hard to get into synchronization again, additional spurious errors may occur.

inside an Arabic Context no further \LaTeX{} or Arab\TeX{} environment may be nested (with the possible future exception of list environments; these are not yet implemented.)

For a list of all available commands, consult the Index to this report. As a reminder, a list of all commands that are valid inside Arabic text will appear in the log file.
Chapter 3

Language selection

The processing of input text to be written in the Arabic script is somewhat language dependent. Thus before the first Arabic quotation or Arabic environment you have to indicate the desired processing mode by one of the commands \setarab, \setfarsi, \seturdu, \setpashto, \setmaghribi, or \setverb (no special processing; see however Section 5.5). The processing mode may be changed at any time, even inside an Arabic environment or an Arabic quotation.

After selecting a language, the symbols < and > serve to bracket short insertions in the chosen language. Whereas this is usually convenient, observe that they can thus no more be used for other purposes, except in mathematical mode where they retain their normal meaning as relational operators. To temporarily return them to their normal mode of operation, deselect the language by \setnone. Arabic insertions may also be started by \<.\footnote{Note for advanced \TeX{} users: All language selecting commands except \setnone set the character < active. If Arabic insertions are not needed, or are always started with \<, the user may re-use the command < for other purposes, or deactivate it by \texttt{\catcode{12}} to return it to its normal meaning.}

For further details on supported languages, see Section 7.
Chapter 4

Font selection

For space economy, only the Naskh font is available by default. With \LaTeX, additional fonts can be loaded by the document style options "\texttt{nashbf}" (for bold-face) and/or "\texttt{nastaliq}" (when available). Users of Plain \TeX are considered specialists and have to define and load suitable fonts at the required sizes themselves.

The following font selection commands are available:

- \texttt{\setnash} (default) selects the Naskh font.
- \texttt{\setnashbf} selects a bold-face version of Naskh.
- \texttt{\setnastaliq} selects the Nasta\'liq font.

If a font is not available or has not been loaded, the corresponding command will select the default font.

With \LaTeX, the size changing commands will also operate on the additional fonts.
Chapter 5

Input coding conventions

The ASCII input notation for Arabic text has been modelled closely after the transliteration standards ISO/R 233 and DIN 31 635. As these standards do not guarantee unique re-transliteration and are also not 7-bit ASCII compatible, some modifications were necessary. These follow the general rules:

- whenever the transliteration uses a single letter, code that letter;
- whenever the transliteration uses a letter with a diacritical mark, put the punctuation character most closely resembling the diacritical mark before the letter (and not behind it as in some other coding proposals, as otherwise the readability of the input would suffer).
- use capital letters for writing variants

5.1 Standard Arabic and Persian characters

The standard codings for Arabic and Persian are given in Table 5.1 and Table 5.2.

- For long vowels, use the capital letters <A>, <I>, <U> or <aa>, <iy>, <uw>.

- To get the defective writing of long vowels, use <_a>, <_i>, <_u>.
- 'Alif maqṣūra is <_a> or <Y>.
- The short vowels fatha, kasra, damma are coded <a>, <i>, <u> and need not normally be written except in the following cases:
Table 5.1: Standard codings for Arabic and Persian.

- at the beginning of a word where they generate 'alif,
- adjacent to hamza where they will influence its carrier,
- when the transliteration is required,
- in the \fullvocalize mode.

- Tanwīn is coded <šš>, <šh>, or <šš>. A silent 'alif, if required, is supplied automatically; it may also be explicitly written: <ššA>. Likewise, a silent wāw may be written <wU> as in <ššamrudU>.

- hamza is denoted by a single right quote '. After selecting a language by \setarab etc., the hamza carrier will be determined from the context according to the rules for writing Arabic words; if that is not wanted, “quote” the hamza (see Section 5.2 below). In the \setverb mode, the hamza carrier is determined by the following letter; see Section 5.5.

- madda on 'alif is generated by a right quote (hamza) before <A>: <'A>.
CHAPTER 5. INPUT CODING CONVENTIONS

<table>
<thead>
<tr>
<th>Code</th>
<th>Arabic Letter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>خ</td>
<td>&quot;hā'&quot; with hamza</td>
</tr>
<tr>
<td>^c</td>
<td>ج</td>
<td>&quot;ghīm&quot; with three dots (below)</td>
</tr>
<tr>
<td>,c</td>
<td>خ</td>
<td>&quot;hā'&quot; with three dots (above)</td>
</tr>
<tr>
<td>^z</td>
<td>ز</td>
<td>&quot;zāy&quot; with three dots (above)</td>
</tr>
<tr>
<td>^n</td>
<td>ن</td>
<td>&quot;kāf&quot; with three dots (Ottoman)</td>
</tr>
<tr>
<td>&quot;l</td>
<td>ل</td>
<td>&quot;lām&quot; with a bow accent (Kurdish)</td>
</tr>
<tr>
<td>.r</td>
<td>ر</td>
<td>&quot;rā'&quot; with two bows (Kurdish)</td>
</tr>
</tbody>
</table>

Table 5.2: Additional codings generally available.

It may also be written "<$>"; likewise, "<Γ>" and "<Τ>" will produce madda on "gā" and on "wāw", as required in some older writing conventions.

- The coding <`>` for "ayn is a single left quote, beware of confusing it with hamza!
- The “invisible consonant” <|> may be inserted in order to break unwanted ligatures and to influence the hamza writing. It will not show in the Arabic output or in the transliteration. At the beginning of a word it will suppress a following short vowel; otherwise it acts like a consonant.
- The sequence <|> will insert a small space, as does <|> (see Section 5.2 below). The adjacent characters will not be connected.
- Ṣadāda is indicated by doubling the appropriate letter coding.
- The definite article is separated from the following word by a hyphen. It may be written in the assimilated form (if it exists): <as-salaam>, or always as <al->; in that case a subsequent “sun letter” must be doubled: <al-ssalaamu>, to receive a ṣadāda, and to prevent a sukūn on the lām. The transliteration in both cases is identical.
- Hyphens <-> are used for tying words together, or for indicating a connecting vowel in Arabic, or an izāfe connection in Persian. They may be used freely, and generally do not change the writing, but will show up in the transliteration. Additionally, at the beginning and the end of an
otherwise isolated word they enforce the use of the connecting form of the adjacent letter (if it exists), like e.g. in the date <1400 ه>.

- A double hyphen -- between two otherwise joining letters will break any ligature and will insert a horizontal stroke (تاء، كاِدا) without appearing in the transliteration. It may be used repeatedly. See also Section 8: automatic stretching.

For special applications, it can also be coded <B>; and <[^B]> will behave like an ordinary consonant and may carry vowel indicators, تاء، سكَّن، and, in the combination[^B]: سَدَد.

### 5.2 Quoting

In \texttt{novocalize} mode (see Section 5.4), a double quote "" will modify the meaning of the following character as follows:

- if a short vowel follows, the appropriate diacritical mark فَظَا، كَسْرَا، َذَمْمَا will be put on the preceding character.
  - If ئ follows the short vowel, the appropriate form of تاء will be generated instead.
  - At the beginning of a word, 'اَلِيف is assumed as the first character.

- if the following character is a single right quote, a حَمْزَا mark will be put on the preceding character even if in conflict with the حَمْزَا rules.

At the beginning of a word, an isolated حَمْزَا will be generated.

- if the following character is the “invisible consonant” ئ, the connection between the adjacent letters will be broken and a small space inserted. This can also be denoted "|" instead of "|

At the beginning of a word, 'اَلِيف with وَاَشْلَا will be generated.

- otherwise: a سكَّن will be put on the preceding character. The following character will be processed again.

The double quote will not show up in the transliteration.

In \texttt{vocalize} mode, (see Section 5.4), quoting will turn a short vowel off; likewise, in \texttt{fullvocalize} mode, quoting will also turn a سكَّن off. Put differently: quoting will toggle the generation of short vowel indicators and سكَّن on and off.
5.3 Ligatures

There is no way to explicitly enforce ligatures as a large number of them are generated automatically. The results will not always look satisfactory, so we recommend inspecting the output after the first run. Any unwanted ligature can be suppressed by interposing the invisible character $<$l$>$ between the two letters otherwise combined into a ligature. After \texttt{ligsfalse}, in the middle of a word fewer ligatures will be produced; for some texts this looks better. You can return to the normal strategy by \texttt{ligstrue}.

5.4 Vowelization

There are three modes of rendering short vowels:

- \texttt{\textbackslash fullvocalize}: 
  - Every short vowel written will generate the corresponding diacritical mark $fath\,a$, $kasra$, $damma$, except if quoted.
  - If \texttt{\textbackslash A} follows a short vowel, the corresponding form of $taun\,\nu$ is generated instead.
  - Defective writing: The coding \texttt{\_a} will produce a Qur'an 'alif accent (also called dagger 'alif) instead of an explicit 'alif character which would be coded \texttt{\textbackslash A} or \texttt{aa}. Likewise, \texttt{\_i} will produce a small 'alif below the preceding consonant in place of \texttt{\textbackslash I} (\texttt{iy}), and \texttt{\_u} will produce an inverted $damma$ in place of \texttt{\textbackslash U} (\texttt{uw}).
  - If a long vowel follows a consonant, the corresponding short vowel is implied. The long vowel itself carries no diacritical mark.
  - If no vowel is given after a consonant, $suk\,\nu$ will be generated except if a double quote precedes the next consonant. The $\texttt{\textbackslash L}$ of the definite article receives no $suk\,\nu$ if a double “sun letter” follows.
  - 'alif at the beginning of a word carries $wasla$ instead of the vowel indicator if the preceding word ended with a vowel.

- \texttt{\textbackslash vocalize}: As above, but $suk\,\nu$ and $wasla$ will not be generated except if explicitly indicated by “quoting”.

- \texttt{\textbackslash novocalize}: No diacritics will be generated except if explicitly asked for by “quoting”.

In all modes, a double consonant will generate $sadda$, and \texttt{\textbackslash A} always generates $madda$ on 'alif.
After `<ام>` the silent 'ال 'character is generated if necessary. The silent 'ال 'may also be explicitly indicated by `<اNA>` or coded literally as `<A>` in `\novocalize` mode. If a silent 'ال 'maqsūra is wanted instead, write `<ام_A>`, `<aNY>`, `<_A>` or `<_Y>`.

The *tanūn fatha* is normally put on the last consonant of the word, even if a silent 'ال 'follows. If it is instead supposed to go onto the 'ال 'as in some modern Arabic conventions, or in Persian, this behaviour can be achieved by the option `\newtanwin`. The option `\oldtanwin` will restore the classical behaviour.

A silent 'ال 'after وَ in indicated by `<UA>` or `<WA>` (with a capital `<و>`!).

### 5.5 Verbatim input

<table>
<thead>
<tr>
<th>'ا</th>
<th>ٌ</th>
<th>hamza on 'ال</th>
<th>'ي</th>
<th>ٌ</th>
<th>hamza below 'ال</th>
</tr>
</thead>
<tbody>
<tr>
<td>'و</td>
<td>ٌ</td>
<td>hamza on وَ</td>
<td>'ي</td>
<td>ُ</td>
<td>hamza on a tooth</td>
</tr>
<tr>
<td>'ح</td>
<td>ُ</td>
<td>hamza on حَا’</td>
<td>'ب</td>
<td>ُ</td>
<td>hamza on the line</td>
</tr>
<tr>
<td>'ل</td>
<td>ِ</td>
<td>isolated hamza</td>
<td>'ا</td>
<td>ِ</td>
<td>madda on 'ال</td>
</tr>
</tbody>
</table>

Table 5.3: Verbatim codings for the carrier of *hamza*

After disabling language specific processing by `\setverb` or `\setnone`, *ArabTeX* will not use any context information to determine the carrier of *hamza*. Instead the user has to supply this information himself by the next character typed after `<`>. Generally this character will be used as the carrier; for examples and some exceptions see Table 5.3. A short vowel indicator may follow.

To ease automatic conversion, an initial 'ال may also be coded `<A>`.

### 5.6 Alternate input codings

The *ArabTeX* input notation has been very carefully designed for flexibility, readability, and ease of use for linguists confined to standard 7-bit ASCII equipment for processing and transmitting data. However, it does not make much sense recoding existing machine-readable text files coded according to other standards. Thus, some alternate reading modules have been written (as there
are more than 10 different codings in current use, this is an open-ended activity), and a general code switching procedure has been provided.

An alternate reading module, e.g. `asmo449.sty` for the ASMO 449 code, is installed by adding its name (`asmo449`) as a \TeX\ style option, or by `
input asmo449.sty`. Afterwards, a `code_name` (in this case `asmo449`) is defined.

Input coding is switched by the command `\setcode{code_name}` that changes the coding for Arabic text globally, or by the environment `\begin{setcode}{code_name} \cdots \end{setcode}` which follows the normal \TeX\ grouping rules.

Coding may be switched several times in the same document, provided the appropriate reading modules are installed; `\setcode{arabtex}` reverts to the standard Arab\TeX\ notation.

Please observe that only Arabic text is affected by `\setcode{code_name}`; text outside of Arabic contexts, and control sequence names, are still assumed to be in 7-bit ASCII. As existing text files presumably do not contain any control sequences or non-Arabic text anyway, we suggest using a small ASCII \TeX\/\LaTeX\ driver file setting all relevant options and containing any non-Arabic text, and calling the Arabic text files by `\input {file_name}` from within an Arabic environment.

For details on available additional reading modules, see Appendix II.
Chapter 6

Transliteration

6.1 ZDMG transliteration style

In addition to the arabic writing, the standard scientific transliteration may also be obtained from a fully vowelized input text. This mode is activated by \texttt{\textbackslash transtrue} and may be switched off again by \texttt{\textbackslash transfalse}. If only the transliteration is wanted, you can deactivate the arabic writing by \texttt{\textbackslash arabfalse}; it can be reactivated by \texttt{\textbackslash arabtrue}. If both modes are active their output will be interleaved line by line.

The transliteration mode assumes that the input text is in the Arabic or Persian language and has been coded according to the rules given above. For words from other languages the transliteration might be in error. For Arabic text, the following special cases are handled:

- after the definite article, a double consonant will be assimilated;
- an initial vowel will be replaced by an apostrophe whenever the preceding word ended with a vowel (in this case a \\
  waşla appears in the Arabic writing). If that is not wanted, start with hamza.
- a silent 'alif or 'alif maqṣūra after \\
  <ا> (tanvin) and <א> is omitted in the transliteration. The same happens after "waw" if it is written as a capital <א>.
- To correctly reproduce some historical writings, a silent long vowel after \\
  <א> is omitted in the transliteration. For examples, see the Appendix.

For economy of space, the transliteration module is not loaded by default. If
you want to use it, add the style option "atrans" with \LaTeXX; and with Plain \TeX, say \texttt{input atrans.sty} after loading Arab\TeXX.

6.2 Encyclopedia of Islam style

For special purposes, the standard transliteration output may be modified by including the \LaTeXX option "etrans", or by loading the file "etrans.sty" when working with Plain \TeX. After this modification, the transliteration will follow the style of the Encyclopedia of Islam.
Chapter 7

Support for other languages besides Arabic

Arab\TeX is primarily intended for typesetting texts in classical and modern Arabic, but it also provides some support for several other languages that are customarily written in the Arabic alphabet.

In order to switch to the conventions for one of these languages, say `\setfarsi`, `\seturdu`, `\setpashto`, `\setmaghribi`; `\setverb` will switch off any language specific processing. `\setarab` can be used to switch back to the Arabic conventions. After selecting the language, `<` and `>` serve as delimiters for quotations; `\setnone` will, like `\setverb`, deselect any language, and will also return `<` and `>` to their normal \TeX meaning.

This part of Arab\TeX relies heavily on contributions from the user community; we want to especially mention Ivan Dershanski who completely reimplemented the routines for processing Persian. As we extensively modified these contributions while integrating the system, we are solely responsible for any remaining, or newly introduced, errors.

7.1 Persian (Farsi, Dari), also Ottoman, Kurdish

- All characters needed for writing Farsi are available by default. The short vowels `<e>` and `<o>` are mapped to `<i>` and `<u>`, the long vowels `<E>` and `<O>` to `<I>` and `<U>` without a vowel indicator. `<E>` denotes final silent ـhāʾ. This ـhāʾ receives no sukūn even in fully vowelized mode.
CHAPTER 7. SUPPORT FOR OTHER LANGUAGES BESIDES ARABIC

- For fatḥa or kasra followed by a final silent hāʾ you can also write <,a> or <,e> in place of <ah> and <eh>.
- The izāfet connection may always be written <-i> or <-e> (with hyphen); then the correct spelling will be determined from the context. Likewise the yāʾ-i-wahdat can always be written <-I> or <-E>.
- The present tense forms of the copula are coded <-am>, <-ı>, <-ast>, <-ım>, <-ıd>, <-and>. In the output they are written as separate words after a little space.
- The final yāʾ carries no dots. Farsi uses the Nastaʿliq font if available, otherwise Naskh.

For further details see Appendix G.

7.2 Urdu

- For Urdu, additional codings are available, see Table 7.1. Some of the given codings also occur in Pashto but with a different meaning, see Section 7.3.
- The short vowels <e> and <o> are mapped to <i> and <u>. <H>, <a> and <,e> are used as in Persian.
- Even in fully voweled mode, an aspirated consonant before ʾах receives no sukun since the two are technically a single letter.
- Urdu uses the Nastaʿliq font if available, otherwise Naskh.

7.3 Pashto (Afghanic)

- For Pashto, additional codings are available, see Table 7.2. Some of the given codings also occur in Urdu but with a different meaning, see Section 7.2.
- The short vowel <e> is indicated by a zwarekay, <o> by an inverted damma.

Observe also the following codings:

<z> " hamza on wā w
<ח> " hamza on hāʾ, if not generated by izāfet
CHAPTER 7. SUPPORT FOR OTHER LANGUAGES BESIDES ARABIC

<table>
<thead>
<tr>
<th>letter</th>
<th>Arabic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>⌅</td>
<td>⌅</td>
<td>always denotes the “two-eyed” hā’</td>
</tr>
<tr>
<td>⌉</td>
<td>⌉</td>
<td>the “wavy” hā’ letter</td>
</tr>
<tr>
<td>⌄</td>
<td>⌄</td>
<td>ẓā’ with a small fā’ accent</td>
</tr>
<tr>
<td>⌃</td>
<td>⌃</td>
<td>dāl with a small fā’ accent</td>
</tr>
<tr>
<td>⌂</td>
<td>⌂</td>
<td>rā’ with a small fā’ accent</td>
</tr>
<tr>
<td>⌁</td>
<td>⌁</td>
<td>nān without a dot</td>
</tr>
<tr>
<td>⌀</td>
<td>⌀</td>
<td>ẓā’, yā’ barī’ in the final position</td>
</tr>
<tr>
<td>⌉</td>
<td>⌉</td>
<td>ae the diphthong ae</td>
</tr>
<tr>
<td>⌈</td>
<td>⌈</td>
<td>ao the diphthong ao</td>
</tr>
<tr>
<td>⌇</td>
<td>⌇</td>
<td>the long vowel ā</td>
</tr>
<tr>
<td>⌆</td>
<td>⌆</td>
<td>the long vowel ā</td>
</tr>
</tbody>
</table>

Table 7.1: Additional codings for Urdu.

- The codings <ɐ>, <ɐ> and <ɐ> are used as in Persian. The rules for izāfet and yā’-i-wahdat apply.
- For writing some Pashto words in the Urdu style, write the command \seturdu and afterwards switch back to the Pashto conventions by \setpashto.

7.4 Maghribi

Nearly like Arabic but using a different writing convention. fā’ is written with one dot below the letter, qāf with one dot above the normal letter form of fā’. The three dots of vā’ are put below the letter.
Table 7.2: Additional codings for Pashto.

### 7.5 Other languages

This is up to experimentation by the user. If `\setarab` or `\setfarsi` will not produce the desired result, try `\setverb` for verbatim mode.

The vowelization and the transliteration cannot generally be expected to be correct, but might work by accident.

In case some character variants not yet provided are needed, feel free to ask the author for help. There is no simple way for the user to modify the script.
Chapter 8

Miscellaneous features

8.1 Automatic stretching

For special purposes, e.g. for headlines and for Arabic paragraphs containing long mathematical or non-Arabic insertions, the connection between adjacent Arabic letters may be made “elastic”, if they form no ligature. Thus a kašīda is inserted whose length will be adjusted automatically to uniformly fill the output line.

This feature very easily leads to storage overflow during the processing, and should only be used whenever necessary. It is switched on with \spreadtrue and switched off again with \spreadfalse. Inside an Arabic Environment, it will also be switched off automatically at the end of every paragraph.

8.2 Dots on yā’

Whether yā’ in the final position carries dots or not is controlled by the chosen language convention. You can override this, after selecting the language, by \yahdots and \yahnodots.

8.3 Additional codings

To reproduce exotic, erroneous or archaic texts exactly as they are written, some additional codings are available, see Table 8.1.
Table 8.1: Additional codings for special purposes.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\k</td>
<td>kāf in the final position without a mark</td>
</tr>
<tr>
<td>\d</td>
<td>dāl with a dot below</td>
</tr>
<tr>
<td>\f</td>
<td>fā’ without a dot</td>
</tr>
<tr>
<td>\b</td>
<td>bā’ without a dot</td>
</tr>
<tr>
<td>\n</td>
<td>nūn without a dot (not available in Pashto mode)</td>
</tr>
<tr>
<td>\y</td>
<td>'ālif maqṣūra; yā’ without dots in all positions</td>
</tr>
</tbody>
</table>

If further variants are needed, write to the author and indicate:

- the required shape,
- the assumed transliteration,
- a suggestion for the input coding,
- some information on the intended use.

We are willing to consider any suggestion. Adding a new character might be easy, or else it might be impossible. Arab\TeX is flexible, but there are some technical limitations.

### 8.4 Progress report

As Arab\TeX is slow, it will produce some terminal output while running to indicate it is still alive. If that is not wanted, e.g., on a very fast system, or while running a batch job, say \texttt{\quiet} or \texttt{\tracingarab = 0} (outside an Arabic Environment; otherwise say \texttt{\doassign \{\tracingarab \{0\} \}}). \texttt{\tracingarab = 1} will only report Arabic paragraphs, a value of 2: Arabic lines and insertions, a value of 3 or more: individual Arabic items.
8.5 Verbatim copy of the input

For test purposes, the Arabic input may be reproduced verbatim after `\showtrue` in addition to the normal output; `\showfalse` switches this feature off again. Commands will not usually be shown. The output will generally not look pleasant, and this feature is only provided in order to trace down errors, or to demonstrate the operation of Arab\TeX as in the appendix.

8.6 Using Arab\TeX with EDMAC

Arab\TeX will cooperate with EDMAC, a Plain \TeX macro package for critical editions, written by John Lavagnino and Dominik Wujastyk. If EDMAC is already present when Arab\TeX is loaded, the EDMAC commands will, after suitable modifications, be available inside an Arabic environment. Their arguments are considered Roman text but may contain Arabic quotations.

For further details, see the EDMAC documentation.
Chapter 9

Acknowledgments

The development of Arab\TeX would not have been possible without the assistance of many people, and it is impossible to acknowledge every individual contribution. Besides our local team, i.e. Udo Merkel and Heribert Schlebbé, helpful advice came, among others, from Chahriar Assad, Benno van Dalen, Ivan Derzhanski, Wolfdietrich Fischer, Ahmed El-Hadi, Yannis Haralambous, Abdelsalam Heddaya, Nicholas Heer, Iqbal Khan, Tom Koornwinder, Eberhard Krüger, Asif Lakehsar, Jan Lodder, Richard Lorch, Pierre MacKay, Eberhard Mattes, Fathy Neamat-Allah, Bernd Raichle, Ulrich Relstock, Mohamed Saba, Waheed Samy, Annemarie Schimmel, Nariman Shehab, Dominik Wujastyk, and Michio Yano. We also have to thank all users who sent error reports, comments, and suggestions.
Chapter 10

References


C. Faulmann: *Das Buch der Schrift, enthaltend die Schriften und Alphabete aller Zeiten und aller Völker des gesammtten (sic!) Erdkreises.*

W. D. Fischer: *Grammatik des Klassischen Arabisch.*

A. Grohmann: *Arabische Paläographie (Teil I und II).*

E. Harder, A. Schimmel: *Arabische Sprachlehre.*


ISO 9036: *Information processing — Arabic 7-bit coded character set for information interchange.*

D. E. Knuth: *The METAFONT Book.*


D. E. Knuth and P. MacKay: *Mixing right-to-left texts with left-to-right texts.*
TUGboat 8/1, 14-25 (1987).

Ann K. S. Lambton: *Persian Grammar.*
Cambridge University Press, 1953.

L. Lamport: *\LaTeX, A Document Preparation System.*

M. Lorenz: *Lehrbuch des Pashto (Afghanisch).*

P. A. MacKay: *Typesetting Problem Scripts.*
BYTE 11/2, 201-216 (1986).


Appendix A

Obtaining ArabTEX

The ArabTEX system is available from the author’s institution (by anonymous FTP from ftp.informatik.uni-stuttgart.de (129.69.211.2), in the directory pub/arabtex) and from many other common servers, e.g. the CTAN network (Aston, Niord, Stuttgart). The files may be transferred individually or as a package: arabtex.zip for PC systems, arabtex.tar.Z for U*IX systems; we recommend to get and inspect the README file first. Successful operation on the Apple Macintosh in conjunction with OzTeX has also been reported.

At the time of this writing, version 3.00 is current. The Nasta'liq font is still under development; Naskh will be substituted automatically. Version 2 is downward compatible; the old version 1 is obsolete and should no more be used.

ArabTEX is copyrighted, but free use for scientific, experimental and other strictly private, noncommercial purposes is granted. Offprints of any publications using ArabTEX are welcome. Using ArabTEX otherwise requires a license agreement.
Appendix B

Installing ArabTEX

The installation procedure is strongly system dependent, and we recommend securing the assistance of a local \TeX{}pert. You have to install the "nashid" font with its "*.pk" and "*.tfm" files on the font search path of your \TeX{} system, and the "*.sty" files and "arabtex.tex" on the source search path (usually TEXINPUT) of your system. Possibly you will also have to rename the "*.pk" files according to local conventions, and as a last resort you can try to recreate the fonts from the "*.mf" METAFONT sources. Additional fonts, whenever available, are installed analogously.

Arab\TeX{} has been found to cooperate well with \TeX{} versions 3.xxx, \LaTeX{} versions 2.09 of 1991 or later, NFSS and NFSS2 (not required), and previewers that can handle fonts of more than 128 characters. \TeX{}\–\\TT{} or \TeX{}\–\\TT{\sf} are not required, and their additional features are presently not exploited. The \TeX{} "hash size" should be at least 3000 to 3500, especially when using Arab\TeX{} in conjunction with \LaTeX{}, and if the transliteration module is used. Use of a BIG \TeX{} may be necessary when using the NFSS due to the latter's high demand on string storage. Space and time requirements are not negligible, and have increased during development; however, Arab\TeX{} currently still runs, albeit slowly, even on a PC XT standard configuration.
Appendix C

Release history

There was a Version 1 which is no more supported.

Version 2 was not fully compatible with Version 1; however, moving to the new version usually caused little problems. Apart from some extensions, most changes were introduced in order to better conform to the transliteration standards, and to have less compatibility problems with TeX and ET
tex. Further versions are expected to be upward compatible if no serious problems will turn up.

The main differences between versions 1 and 2 are:

- The font size has increased, so the document layout may change. The old font "nash10" can no more be used as the character locations have been assigned differently.

- Some Arabic characters are now coded differently: ‘ain is denoted by a left quote, and <c>, <z>, <t>, and <n> have been assigned new meanings in order to better conform to the standard transliteration.

- There are many more ligatures than before. This normally need not concern the user.

- \vocalize will no more generate sukān and wasla except if explicitly indicated by quoting. See \fullvocalize.

- Arabic Environments are now always bracketed by the new control sequences \begin{arabtext} and \end{arabtext} even if only the transliteration is wanted.

We strongly recommend converting any still existing version 1 input files to the new notation. To assist in this migrating procedure, the \LaTeX option
"oldarabtex" and/or the command \oldarabtex will switch to a mode where virtually all places where the old conventions are used, will either produce a \TeX error message or will be flagged in the output.

The changes introduced since the release of Version 2.00 up to now (Version 3.00) fall into one of two categories: error corrections, and upward compatible extensions. Details are not given here, but are documented in the text file CHANGES that is part of the distribution package of Arab\TeX.

Version 3 is upwards compatible with version 2. All supported features are documented in this manual.
Appendix D

Sample ArabTEX input

\documentstyle[12pt,arabtex]{article}
\begin{document}

\setarab % choose the language conventions
\vocalize % diacritics for short vowels on
\trantrue % additionally switch on the transliteration
\arabtrue % print arabic text ... is on anyway
\spreadtrue % spread out caption

\centerline{"gu.hA wa-.himArulu"}

\begin{arabtext}
'at_A .sadiqu.N 'il_A -gu.hA ya.tlubu minhu .himArahu li-yrkabahu
fi safraTiN qa.sIraTiN wa-qAla lahu:

sura 'u'Iduhu 'ilayka fi al-masA'i, wa-'adfa'u laka 'u^graTaN. \|
fa-qAla ^gu.hA:

'anA 'AsifuN ^giddaN 'annI lA 'asta.tI'u 'an 'u.haqqiqa
laka ra.gbataka, fa-al.himAru laysa hunA al-yawma. \|
wa-qabla 'an yutimmu ^gu.hA kalAmahu
bada'a al-.himAru yanaqahu fi I.s.tablihi. \|
fa-qAla lahu .sadiquhu:

'inN 'asma'u .himAraka yA ^gu.hA yanaqhu. \|
fa-qAla lahu ^gu.hA:

,garIbuN 'amruka yA .sadIqI!
'a-tu.saddiqu al-.himArA wa-tuka_d_dibunI?
\end{arabtext}

\end{document}
Appendix E

Sample ArabTEX output


guhā wa-himāruhu
ratā sabīqun ilā guhā yallubu mahu himāruhu l-yarkabahu fi safrāt qasīratān wa-qāla laku:

أَنِّي صديقُ إِلَيْهِ حَمَّارُ جَمَارتُ لِلْيَكْبِيَةِ فِي سَرِّهِ قَصِيرَةٍ وَقَالَ لَهُ:
saefu urdhu shayka fi 'masā`, wa-adfaru laka qīratān.

سُوْفُ أَيَّدُهُ إِلَيْهِ فِي الْمَنِيَّةِ، وَأَدْفَعُ لَهُ أَجْرَهُ.

fa-qāla guhā:

فَقَالَ لَهُ:

anā 'āsifun giddan annī li 'astaffru an 'uhaqqīqa laka ragbataka, fa-'himāru layṣa kunā 'nyauma.

أَنَا أَيَّد حَمَّارُ جَمَارَتُ لِلْيَكْبِيَةِ فَأَحْتفَقُ لَهُ رَغَبَاتُكَ، فَأَحْمَارُ لَيْسُ هَذَا الْيَوْمُ

wa-qāla 'an yuṭimmu guhā kāfimahu hadarā 'himāru yankaqu fi ṭabluhi.

وَقَبِلَ أَنْ يَمُّهُ جَمَارَتُهُ بَدَا الْمَحْرَازُ يَنْبَغُي فِي اسْتِفْتِهِ.

fa-qāla laku sadqaku:

فَقَالَ لَهُ صَدِيقُهُ:

innī 'asmaru himāraka gā guhā yankaqu.

لَيْسِي أَحْفَقَ جَمَارتُ يَا حَمَّارُ يَنْبَغِي:

fa-qāla laku guhā:

فَقَالَ لَهُ حَمَّارُ:

jarīban yamruka yā sadqī! 'a-tusaddiqu 'himāru wa-tukaddibānī?

عَرَبِي أَمَّلُ يَا صَدِيقُهُ، أَنْصِدِقُ الْمَحْرَازُ وَنَكَّذِبَيْنِ! 

37
Appendix F

Coding examples for Arabic

The short vowels *fatḥa, kasra, dammā* are denoted, as in the transliteration, by the small letters a, i, u:

- *mana’ā* مَنَّعْ, *dahābaََ ذُهِبَ, sarībaََ غَرَبَ*.

The long vowels ā, ī, ū are denoted by capitals Ā, Ī, Ū or by aa, iy, uw:

- *qātala* قَاتِلَ, *nūzra* نُؤُزْرَ, *lūmī* لَوْمِيَ,

*Alif maqṣūm* is coded as _ā_ or _y_.

- *ramā* رَمَّ, *dikrī* ذِكْرِي, *al_ā* عَلِيَ, *bal_ā* بَالِ,

Silent *ʿalif*: The plural suffixes -ā, -aw of the verb are denoted UA, āw or awA:

- *katabUA* كَتَبُوا, *yaktubUA* يَكْتُبُوا,

---

The defective notation of ā, ī, ā can be indicated by _a, _i, _u and leads to the appropriate spelling:

dāru-h_u ʌl_dāru-hā, ri₉ɡli-h_i ʌl_rigli-hā,
however: ramā-h_u ʌl_ramā-ḥu, yarmā-h_i ʌl_yarmā-ḥi;

_dīh_i ʌl_dīh, h_a-dīh_i ʌl_hādīḥ, tīh_i ʌl_tīḥ, hātīh_i ʌl_hāṭīḥ,
rabb_i ʌl_rabb, sAl_i ʌl_sāl, hum_u ʌl_humū;

qiy'amatun ʌl_qiy'amatun, 'il_ahun ʌl_ahūn,
sam_awanun ʌl_samāwānun, _tal_a_tun ʌl_talātun,
lakin ʌl_lākān, h_a-da ʌl_dādā, 'al-lāhu ʌl_lāhū,
'al-raḥmānu ʌl-ar-rahmānu, _d_alika ʌl_dālīka.

To reproduce the historical writing correctly, a silent long vowel or 'alif maqsūra after _a receives no sukūn and is ignored in the transliteration:

sal_aTun ʌl_salātun, hay_aTun ʌl_hayātun,
zak_aTun ʌl_zakātun, mi₉sk_aTun ʌl_miskātun,
ar-rib_aU ʌl_ar-ribā, tawr_aTun ʌl_tawrātun,
ram_aYhu ʌl_rāmah, SAM_aYhum ʌl_samāhum.

The short vowel _u can be written as a long vowel by _U:

_UYO ʌlā, _ULA'i ʌlā', _ULUO ʌlā, _ULAKA ʌlāka, _ULA'ika ʌlā', _ULUO ʌlā', _ULAKA ʌlāka.

Tanwin: The plural suffixes -un, -in, -an are written -uN, -iN, -aN or aIaN. Silent 'alif in -an may be indicated by A or omitted; if necessary it is supplied from the context.

ra'guluN ʌl_ragulum, ra'guliN ʌl_ragulīm, ra'gulāN ʌl_ragulām, madinaTAN ʌl_madīnatān, gamilTaTAN ʌl_gamīlatān,
i_dāN ʌl_idān, samA'iN ʌl_samān.

There is a special case:

ribAN ʌl_ribān, 'amruN ʌl_amru, 'amruN ʌl_amru, 'amruN ʌl_amru, however: 'amru ʌl_'amru.
**Tanwīn fatha** is traditionally put on the last consonant even if a silent 'alif follows. Some modern conventions, and also Persian practice, require to put it on the 'alif in this case. This behaviour may be switched on by `\newtanwin`, and off by `\oldtanwin`. `\newtanwin` mode is the default for Persian.

raːgulaʔ, 'iː_dan, ʔiːdan.

A silent 'alif maṣūra after tanwīn is written alNy or all_A:

hudaNY هُدَى kudan, fataN_A fatan;

compare:

al-hudY al-hudā, 'al-fat_A al-fatā.

**Tāʾ marbūṭa** is denoted by T:

kalimaTun كِلَمَتُني kalimatun, kalimaTin كِلَمَتْني kalimatın,

kalimaTn كِلَمَتْنِ kalimatın; fatATun فَتَأْتُنَّ fatatun,

fatATin فَتَأْتْنِ fatatın, fatATn فَتَآْتُنَّ fatatan.

**Hamza** is indicated by '؛ the appropriate carrier is determined by the context:

'amruN ʾamrun, 'ibiluN إِبِيلَ ʾيَبِيلَ ʾيَبِيلَ viehun, ʾu_htuN ʾعُهْتُنَّ ʿuhtun;

raːsuN rasun, ʿarʿasu, ʾarʿasun, saʿala, qaraʾa ʿarʿasun, buʾsuN ʿبُوُسُنَّ ḫusun, ʿabʿusuN ʿعَبْعُسُنَّ ḫusun,

ruːfa raufa, ruʿasʾuN ʿرُوُسُعَنَّ ḫusun, biʾruN بِيُرُنَّ ḫirun,

'asʾilaTuN ʾاِسْيَلَتُنَّ, masʿalatun, kaʾiba كِبْيَةَ qāʾimun, riʾAsaTuN رَيْسَأَتُنَّ, suʿila سَعِيَّةَ samāʾun,

barIʾuN بَرِيُّنَّ barun, suʿuN سَوُعَنَّ, saʿala سَعَالَةَ samāʾun,

sayʾuN ʿسَأُيْنَّ, ʿsayʾiN ʿسَأِيْنَّ ʿsayun, ʿsayʾaN ʿسَأِيْنَّ ʿsayun;

saʿala سَعَالَةَ masʿalatun, sawatun, ha.tIʾaTuN حَطِيْتُنَّ haṭfatun.
Old *Hamza* convention: In an older writing style that is used, e.g., in some Qur’an editions, the *hamza* is sometimes put below its carrier or on the connecting line. This style may be switched on by \textbackslash oldhamza (and off again by \textbackslash newhamza):

\[
\text{'asilaTun} \quad \text{asilatun}, \quad \text{ka'iba} \quad \text{ka'ibat}, \quad \text{qA'imun} \quad \text{qA'imun,}
\]
\[
\text{su ila} \quad \text{su ila}, \quad \text{say 'al} \quad \text{sayr}, \quad \text{ha.Tu} \quad \text{ha tarun}.
\]

*Madda* in the context ‘\(\text{a}\) is generated automatically:

\[
\text{'Akilun} \quad \text{'Aki}un, \quad \text{qur} \quad \text{qur\dagger nun,} \quad \text{ra'Ah} \quad \text{ra'Ah}.
\]

To reproduce the historic writing correctly, it can also be explicitly written in other contexts:

\[
\text{'a.sdiq'A'uh}\text{.u} \quad \text{'asdiq\dagger uh};
\]
\[
\text{ya}^{'\dagger r} \quad \text{yagru,} \quad \text{s'U'ila} \quad \text{s\dagger ila}.
\]

*Sadha*: A double consonant must be written twice, even if it is coded by more than one character:

\[
\text{nazzala} \quad \text{nazzala}, \quad \text{ba's}s\text{Aru}\text{n} \quad \text{ba's}arun, \quad \text{nawara} \quad \text{nawara},
\]
\[
\text{sayyidun} \quad \text{sayyidun}, \quad \text{sa'}\text{Alun} \quad \text{sa'}\text{alun},
\]
\[
\text{sabiyyu} \quad \text{sabiyyun}, \quad \text{a'du} \quad \text{a'dun}.
\]

Instead of \text{iy}, \text{uw} one can also write \text{i\dagger y}, \text{u\dagger w}:

\[
\text{sabiyyu} \quad \text{sabiyyun}, \quad \text{a'du} \quad \text{a'dun}.
\]

*Assimilation*: the definite article may be always written \text{al}-; a following “sun letter” must be written twice like in the Arabic spelling. The transliteration and the use of suk\dagger n are adjusted accordingly:

\[
\text{al-ddAru} \quad \text{al-d\dagger aru}, \quad \text{al-rra}\text{gulu} \quad \text{ar-ra\dagger gulu},
\]
\[
\text{al-ssanaTu} \quad \text{as-sanatu}, \quad \text{al-nnAru} \quad \text{an-n\dagger ru};
\]
\[
\text{al-\dagger gAru} \quad \text{al-g\dagger ru}, \quad \text{al-bAgu} \quad \text{al-b\dagger hru};
\]
\[
\text{al-llaylaTu} \quad \text{al-laylatu}, \quad \text{al-llisAnu} \quad \text{al-lis\dagger nu},
\]
\[
\text{al-llahu} \quad \text{al-l\dagger hru}.
\]
The article may also be written in the assimilated form, with identical result:


In some special cases the literal spelling must be used:


However:


Wasla: an auxiliary vowel at the beginning of a word is always written, but in the middle of a sentence generally without hamza. If a vowel precedes the word, the auxiliary vowel will be omitted in the transliteration, and the wasla sign will be used in the spelling:

wa-ismu hu wa-smahu, f-a-in sara fa fa-nṣara fa.

This also works across word boundaries:

ya ibnI yā, h_a_dā ibnuh_u, hādā ‘bnuhā, qāla ‘hrug.

An auxiliary vowel at the end of the preceding word may be separated by a hyphen:

qad-i in sara fa, qad-i nṣara fa, ra'aw-u al-bāba, ra'aw-u 'l-bāba, min-i ibnih_i, min-i 'bnihī.

This also works for the article preceding 'alif al-wašt:


And even if the auxiliary vowel is omitted in the spelling:

ra'guluN-i ibnātuh_u, gamIlaTuN, ra'guluN-i 'bnatuhā, gamīlatun, muhammaduN-i al-qura's lively, muhammadun-i 'l-qura'sīyu.
The particles *li-* and *la-* must be combined with the article except before *lām*:
*li*-rā'gulī *la*-rā'gulī, *la*-māl-*gdu* ḍalāl-*māgdu*;
however:
*li*-llaylaTim lāl-*llaylati*, *li*-ll_ahi lāl-*lliḥi*.

The Name of God is written with a special ligature if it is recognized from the input sequence *ll_ahi*:
'āl-*ll_ahi* *āl-lāhun* *ta*-al-*ll_ahi* lāl-*llāh*.  

Increased spacing (*Tatwīl*) between adjoining characters may be produced by a double hyphen --; note the position of the vowel marks:
*qabila* ṭīb *qabila*, qa--bi---la ṭīb *qabila*, q--ab--ila *qabila*.
q--a--b--i--la ṭīb *qabila*, qa---bi----la ṭīb *qabila*.

Ties between words are indicated by a single hyphen:
*bi*-baladīn *bi*-baladīn, *ta*-al-*ll_ahi* lāl-*llāhi*.
*sa*-ya'tī *sa*-ya'tī, *li*-yafra*h*a ḍalāl-*yafra*h*a.
*wa*-iswadda *wa*-iswadda, *ba'da-mā* li-*ba'da*mā*.
*`āla*-mā *`āla*mā*, *fi*-ma ḍalāl-*`āla*mā* *al-*ma.  

A single hyphen at the beginning or end of a word will enforce the use of the joining form of the first resp. the last character, if that form exists (for special uses only):

<table>
<thead>
<tr>
<th>s</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td></td>
</tr>
</tbody>
</table>

Digit sequences are written in the natural order:
1234567890 1234567890 1234567890
Ligatures are generated automatically; they can be suppressed by \:

'al-islamu, 'al-islamu;
'al-garu, 'al-garu, 'al-garu, 'al-garu;
_tumma, _tumma, _tumma;
muhammadun, muhammadun, muhammadun.

Abbreviations and emphasis are indicated by \emph:

\emph{slm} صام
\emph{abj} آبج

If necessary, use grouping by curly braces:
\emph{\'alayhi as-salam} عليه السلام
Appendix G

Coding examples for Persian

The short vowels a (ä), e (ē), o (ū) are denoted by the lowercase letters a, e or i, o or u:
bar ʿbɑr, beh ʿbɛh, bon ʿbɔn.

The long vowels a (ā), i (ī), u (ū, ū) are denoted by the capital letters A, I or E, U or O. ʿAlef maʿdde is automatically generated for word-initial a:
Ab ʿɑb, bād ʿbɑd, bīd ʿbɪd, būd ʿbʊd, bād ʿbɑd.

Note that I yields a ya-ye maʿrafi(with zīr), whilst E yields a ya-ye maʿjūl (without zīr). Similarly, U yields a waaw-e maʿrafi(with pis), whilst O yields a waaw-e maʿjūl(without pis):
tīr ʿtɪr, tīg ʿtɪɡ, dūr ʿdʊr, zūr ʿzʊr, zūr ʿzʊr.

The diphthongs ēi and ōu are written ay and aw:
pay ʿpɛy, naw ʿnɔw.

Intervocalic hæmze is written ¹:
pA′Iz ʿpɛɑɬezi, miyA′I ʿmɪɛɑɬ, miyGU′I ʿmɪɡuɑɬ.

¹We gratefully acknowledge the voluntary help by Ivan Derzhanski who wrote this chapter, and implemented the language-specific processing. As we extensively modified his routines during system integration, all responsibility for any remaining, or new, errors rests with us.
tawānā'ī, tawānā-.tā, zanā-sū'ī, zanāsū-.

Silent word-final waw is generated by _U or ِ:
\[ t\_U \to \text{tu}, d\_U \to \text{du}; t\_U \to \text{tu}, d\_U \to \text{du} \text{ dū}. \]

Waw-e ma'dul is written \( w \); it is omitted in the transliteration and is not joined to the preceding xe receives no jazm:
\[ _\text{hwāb} \text{ kāhe \text{hwāb}, _\text{hwā}'} \text{ s āh}, \_\text{hwā} \text{ dār, } \_\text{hwā} \text{ dār.} \]

Ha-ye ha-vawaz-e ma'xi is generated by ِ, or optionally by , , a or , ِ. It does not receive a jazm even in fully vocalised mode and is not joined to a following letter:
\[ _\text{hānēh} \text{ kāhe } \text{hānēh}, _\text{hānēh-hā} \text{ hānēh-hā.} \]

Short edafe is written -e or -i:
\[ \text{ketāb-e } \overline{\text{a}}, \text{rāh-e } t\_U \to \text{rāh-e tu}, \]
\[ \text{nāmeh-i } \overline{\text{a}}, \text{bānē-i } \overline{\text{a}}, \text{pā-i } \overline{\text{a}}, \text{bāzū-i } \overline{\text{a}}. \]

Long edafe is written -e or -i:
\[ \text{dār-e } \overline{\text{a}}, \text{hū-i } \overline{\text{a}}. \]

Hāmase as ya-ye wahda'et/nesbāt/zetāb is likewise written -e:
\[ \text{nāmeh-e } \overline{\text{a}}, \text{sormehe-e } \overline{\text{a}}, \text{goftehe-e } \overline{\text{a}}. \]

Ye-ye wahda'et is written -e or -e:
\[ \text{ketāb-e } \overline{\text{a}}, \text{rāh-e } \overline{\text{a}}, \text{nāmeh-e } \overline{\text{a}}, \text{dānē-e } \overline{\text{a}}, \text{pārū-e } \overline{\text{a}}, \text{dānē-e } \overline{\text{a}}, \text{pārū-e } \overline{\text{a}}. \]
The present tense forms of the verb \textit{budan} and the pronominal clitics are written as they are spoken:

- rafte\textit{h}-\textit{am} ṭafte\textit{h}-\textit{am}, rafte\textit{h}-\textit{Im} ṭafte\textit{h}-\textit{Im},
- rafte\textit{h}-\textit{I} ṭafte\textit{h}-\textit{I}, rafte\textit{h}-\textit{Id} ṭafte\textit{h}-\textit{Id},
- rafte\textit{h}-\textit{ast} ṭafte\textit{h}-\textit{ast}, rafte\textit{h}-\textit{and} ṭafte\textit{h}-\textit{and},
- mard-\textit{Id} ṭ\textit{mard}-\textit{Id}, asb-\textit{et\textit{An}} ṭ\textit{asb}-\textit{et\textit{An}};
- An\textit{\textacuten}-\textit{st} ṭ\textit{An\textacuten}-\textit{st}, U-\textit{st} ṭ\textit{U-}\textit{st}, t-\textit{U-}\textit{st} ṭ\textit{t-U-}\textit{st};
- ketAb-\textit{I-}\textit{st} ṭ\textit{ketAb-I-}\textit{st}, nAmeH-\textit{I-}\textit{st} ṭ\textit{nAmeH-I-}\textit{st}.

The preposition \textit{be-} can be written with or without a hyphen:

- be-\textit{man} ṭ\textit{be-}\textit{man}, be-\textit{tu} ṭ\textit{be-}\textit{tu},
- be-\textit{An} ṭ\textit{be-}\textit{An}, be-\textit{In} ṭ\textit{be-}\textit{In}, be\textit{U} ṭ\textit{be-}\textit{U}.

The components of compounds can be separated by | or ~:

- \textit{\textacuten-s\textacuten-heb}|\textit{h\textacuten-}\textit{aneh} ṭ\textit{\textacuten-s\textacuten-heb}|\textit{h\textacuten-}\textit{aneh},
- ta\textit{-ht-e-}|\textit{h\textacuten-ab} ṭ\textit{taht-e-h\textacuten-ab};
- pas\textit{|and\textacuten-Az} ṭ\textit{pasand\textacuten-Az}, naw\textit{|and\textacuten-Uz} ṭ\textit{nawam\textacuten-Uz},
- b\textit{|h\textacuten-d} ṭ\textit{b\textacuten-d}. 

The position \textit{be-} can be written with or without a hyphen:
Appendix H

Alternate input encodings

H.1 ASMO 449 = ISO 9036

The file \texttt{asmo449.sty} contains a reading module for the ASMO 449 code (identical to ISO 9036). It is installed by the \LaTeX option \texttt{asmo449} or by \texttt{\textbackslash input asmo449.sty}. The module is activated by \texttt{\textbackslash setcode \{asmo449\}} or \texttt{\textbackslash setcode \{iso9036\}}; all following Arabic text will be considered to be coded according to the ASMO 449 standard. The Arab\TeX notation may be reactivated by \texttt{\textbackslash setcode \{arabtex\}}.

ASMO 449 (see Table H.1) is a 7-bit code, differing from ASCII (ISO 646) mainly by replacing the letters by the Arabic letter characters and diacritical marks; the Arabic digits share their positions with the ASCII digits. The positions of special and control characters in both codes are identical.

A minimal driver file for processing, e.g. a file \texttt{asmotext.dat}, could be structured as follows:

\begin{verbatim}
\documentstyle [arabtex,asmo449]{article}
\begin {document}
\setcode {asmo449}
\begin {arabtext}
\input asmotext.dat

\% the preceding blank line is required if "asmotext.dat" did not
\% end with a blank line itself; this is strange and embarrassing
\end {arabtext}
\end {document}
\end{verbatim}
<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>NUL</td>
<td>DLE</td>
<td>SP</td>
<td>@</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
</tr>
<tr>
<td>01</td>
<td>SOH</td>
<td>DC1</td>
<td>!</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>٤</td>
</tr>
<tr>
<td>02</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>٤</td>
</tr>
<tr>
<td>03</td>
<td>ETX</td>
<td>DC3</td>
<td>#</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ل</td>
</tr>
<tr>
<td>04</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>05</td>
<td>ENQ</td>
<td>NAK</td>
<td>%</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷳ</td>
</tr>
<tr>
<td>06</td>
<td>ACK</td>
<td>SYN</td>
<td>&amp;</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>07</td>
<td>BEL</td>
<td>ETB</td>
<td>'</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>08</td>
<td>BS</td>
<td>CAN</td>
<td>)</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>09</td>
<td>HT</td>
<td>EM</td>
<td>(</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>10</td>
<td>LF</td>
<td>SUB</td>
<td>*</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>11</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>12</td>
<td>FF</td>
<td>IS4</td>
<td>,</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>13</td>
<td>CR</td>
<td>IS3</td>
<td>-</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>14</td>
<td>SO</td>
<td>IS2</td>
<td>&gt;</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
<tr>
<td>15</td>
<td>SI</td>
<td>IS1</td>
<td>/</td>
<td>٠</td>
<td>١</td>
<td>٢</td>
<td>٣</td>
<td>ﷲ</td>
</tr>
</tbody>
</table>

Table H.1: ASMO 449 code table
As texts coded in ASMO 449 are always rendered verbatim the commands \novocalize, \vocalize, \fullyvocalize and the language selection commands \setarab etc. make no sense and are temporarily disabled.

Texts in ASMO 449 are usually not fully vowelized. Thus the transliteration cannot be expected to be correct. This is especially true for Egyptian texts which commonly do not differentiate between َā and َ'alif maqṣūra.

**H.2 ASMO 449E = ISO 8859-6**

The file iso88596.sty contains a reading module for the ISO 8859-6 code (extended ASMO 449 = ASMO 449E). It is installed by the \input iso88596.sty option or by \setcode {iso8859-6}; all following Arabic text will be considered to be coded according to the ISO 8859-6 standard. The Arab\TeX notation may be reactivated by \setcode {arabtex}.

ISO 8859-6 (see Table H.2) is an 8-bit code closely related both to 7-bit ASCII and to ASMO 449; whereas the lower 128 positions are identical to ASCII (ISO 646), the upper 128 positions contain the Arabic characters of ASMO 449 in the analogous places, plus a few additional graphic and control characters.

We exploit the close relationship of these codes by reusing the ASMO 449 reading routines, after suitable modification of the input. This only works correctly if the input text does not contain genuine ASCII letters, as we project the Arabic characters onto their locations in ASMO 449. Some of the code switching messages in the log file are spurious; do not worry.

The notes on vowelization and transliteration of ASMO 449 apply also.

The driver file indicated for ASMO 449 will be usable after the obvious modifications; however, your \TeX installation must be capable of processing 8-bit data input. This is nowadays usually the case; otherwise you can try to locally find some utility program that will strip the highest order bit off the characters in your file, and process the result via ASMO 449.
|    | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 00 | NUL | LDE | SP |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 01 | SOH | DC1 | ! | 1 | A | Q | a | q |  |  |  |  |  |  |  |  |
| 02 | STX | DC2 | " | 2 | B | R | b | r |  |  |  |  |  |  |  |  |
| 03 | ETX | DC3 | # | 3 | C | S | c | s |  |  |  |  |  |  |  |  |
| 04 | EOT | DC4 | $ | 4 | D | T | d | t |  |  |  |  |  |  |  |  |
| 05 | ENQ | NAK | % | 5 | E | U | e | u |  |  |  |  |  |  |  |  |
| 06 | ACK | SYN & | 6 | F | V | f | v |  |  |  |  |  |  |  |  |  |
| 07 | BEL | ETB | ' | 7 | G | W | g | w |  |  |  |  |  |  |  |  |
| 08 | BS  | CAN ) | 8 | H | X | h | x |  |  |  |  |  |  |  |  |  |
| 09 | HT  | EM ( | 9 | I | Y | i | y |  |  |  |  |  |  |  |  |  |
| 10 | LF  | SUB * | : | J | Z | j | z |  |  |  |  |  |  |  |  |  |
| 11 | VT  | ESC + | ; | K | | l | k | ) |  |  |  |  |  |  |  |  |
| 12 | FF  | IS4 . | > | L \ | I | |  |  |  |  |  |  |  |  |  |  |
| 13 | CR  | IS3 - | = | M | E | m | { | SHY |  |  |  |  |  |  |  |  |
| 14 | SO  | IS2 . | < | N - | n | - |  |  |  |  |  |  |  |  |  |  |
| 15 | SI  | IS1 / | ? | O . | o |  |  |  |  |  |  |  |  |  |  |  |

Table II.2: ISO 8859-6 code table
Appendix I

Miscellaneous utilities

The following packages are not part of Arab\TeX{} proper, and are not supported in any way, but are distributed along with Arab\TeX{} as possibly a convenience to the users. There is no warranty whatsoever.

I.1 \texttt{twoblks.sty}

This \LaTeX{} option will define a command \texttt{\textbackslash twoblocks \{#1\}\{#2\}} which will place the two parameters \#1 and \#2, usually two paragraphs, into two boxes side by side, separated by space of length \texttt{\textbackslash colsep}. If necessary, the resulting boxes will be split across a page boundary.

This feature is useful if two versions of a text are to be compared. They may be in different languages, and one of them might be in Arabic (if enclosed in \texttt{\textbackslash begin \{arabtext\} \ldots \textbackslash end \{arabtext\}}).

This sentence has been written twice: in the English language and in the Arabic language.

Otherwise this command does not depend on Arab\TeX{} in any way, and indeed originated in a completely different context.

Beware that the two “blocks” should each not contain much more than one, not too long, paragraph of text, otherwise \TeX{}’s main storage might overflow. There must be no \texttt{\textbackslash verbatim} text inside the parameters of \texttt{\textbackslash twoblocks}, nor any \texttt{\textbackslash catcode} changes; and all \TeX{} groups and \texttt{\textbackslash if \ldots \textbackslash fi} sequences must be properly nested.
I.2 abjad.sty

This file, loaded as a \LaTeX option, will define a command \abjad \#1 usable inside and outside of an Arabic context. It profited greatly from suggestions by Dr. Benno van Dalen (Utrecht University).

The command \abjad \#1 will convert its argument, which has to be a legal representation of a number between 1 and 999, to the Arabic 'abjad notation used in some mediaeval manuscripts. The result of the conversion will not look perfect, and the legal 'abjad number 0 can presently not be generated.

Improving this routine needs a font revision, which is hard and tedious; whenever this happens, the command might well become part of Arab\TeX proper.

I.3 MLS2ARAB

This is an UNIX SED script, written by Prof. Nicholas Heer (University of Washington), and released for free distribution. It will (almost) convert an ASCII file of Arabic text, produced by Multi-Lingual Scholar, to the Arab\TeX input notation. The conversion is not perfect so some manual corrections might be necessary.

For operating instructions, see the file itself.
Index

" (quoting), 15
"1, 14, 15
\$, 7
--, 15
\&, 8
\, 8
\abjad, 53
\arabfalse, 19
\arabtrue, 19
\begin{arabtext}, 6, 34
\begin{setcode}, 18
\bigskip, 8
\centerline, 9
\colsep, 52
\doassign, 9
\dcommand, 9
\emphsize, 8
\end{arabtext}, 6, 34
\end{setcode}, 18
\footnote, 8
\fullvocalize, 13, 15, 16, 34
\hfill, 8
\hspace, 8
\indent, 8
\input, 8, 18
\input arabtex.tex, 5
\input atrans.sty, 20
\input etrans.sty, 20
\ligsfalse, 16
\ligtrue, 16
\magnification, 5
\marginpar, 8
\mbox, 8
\medskip, 8
\newhamza, 41
\newtanwin, 17, 40
\noindent, 8
\nospace, 8
\novocalize, 15–17
\oldarabtex, 35
\oldhamza, 41
\oldtanwin, 17, 40
\par, 6, 8
\pnash, 5
\pnashbfbf, 5
\quiet, 26
\setarab, 6, 10, 13, 21
\setcode, 18
\setcode\arabtext, 18, 48, 50
\setcode\arith449, 48
\setcode\arabic8869–6, 50
\setcode\arabic9453, 48
\setfarsi, 10, 21
\setmaghribi, 10, 21
\setnash, 8, 11
\setnashbf, 8, 11
\setnastaliq, 8, 11
\setnone, 10, 21
\setpashto, 10, 21, 23
\seturdu, 10, 21, 23
\setverb, 10, 21, 24
\showfalse, 27
\showtrue, 27
\smallskip, 8
\space, 8
\spreadbox, 8
\spreadfalse, 25
\spreadfalse, 25
\spreadline, 9
\spreadtrue, 25
\tracingarab, 26
INDEX

\texttt{\transfalse}, 19
\texttt{\transtrue}, 19
\texttt{\twoblocks}, 52
\texttt{\vocalize}, 15, 16, 34
\texttt{\yahdots}, 25
\texttt{\yahnodots}, 25
>, 10, 21
\texttt{\|}, 8
1, 14–16
|B, 15
|BB, 15
|1, 14, 15
`
('ayn), 14
'(\emph{hamza}), 13
A, 12, 17, 38
\texttt{\textbackslash}A, 14, 16, 41
.A, 46
_\texttt{\textbackslash}A, 12, 17, 38
"\texttt{\textbackslash}A, 14
_\texttt{\textbackslash}A, 22, 23, 46
_\texttt{\textbackslash}a, 12, 16, 39
\texttt{a} (\emph{fatha}), 12, 38
\texttt{\textasciitilde}a, 12, 38
abbreviation, 44
abjad.sty, 53
'\texttt{\textbackslash}abjad\texttt{\textbackslash}numbers}, 53
Afghanic, 22
'\emph{ayn}, 14
al-, 14, 19
'alif, 17
dagger, 12, 16, 39
\emph{initial}, 17
\emph{maqṣūra}, 12, 17, 38, 40
silent, 17, 40
\emph{Qur'an}, 16, 39
\emph{silent}, 17, 19, 38–40
\emph{small}, 16, 39
\emph{\textbackslash}below, 16, 39
'\emph{Allah} (spelling), 43
\texttt{\textbackslash}A, 13, 17, 39
\texttt{\textbackslash}A\_A, 17, 40
\texttt{\textbackslash}A\_A, 13, 17, 39
\texttt{\textbackslash}aNY}, 40
Arabic context, 6, 7
Arabic environment, 6
Arabic group, 7
Arabic item, 6
Arabic number, 7
Arabic quotation, 6
Arabic quotes, 7
Arabic word, 7
\texttt{arabtex.tex}, 5
\texttt{\textbackslash}Arab\texttt{\textbackslash}TEX\texttt{\textbackslash}commands}, 7, 8
archaic text, 25
ASCI, 48, 50
ASMO 449, 18, 48, 50
aspirated consonant, 22
assignment, 9
assimilation, 14, 16, 19, 41
automatic stretching, 25
\texttt{aw}, 38
\texttt{\textbackslash}aw, 45
\texttt{\textbackslash}aWA, 38
ay, 45
B, 15
be-, 47
boxing commands, 8
breaking connections, 15
code
7-bit, 48
8-bit, 50
\texttt{arabtex}, 18
ASCII, 48, 50
ASMO 449, 18, 48, 50
ISO 646, 48, 50
ISO 8859-6, 18, 50
ISO 9036, 18, 48
coding conventions, 12, 34
commands
\texttt{\textbackslash}Arab\texttt{\textbackslash}TEX}, 7, 8
boxing, 8
illegal, 9
internal, 5
\texttt{\textbackslash}La\texttt{\textbackslash}TeX}, 7
INDEX

overview, 9
size changing, 5, 8, 11
\TeX, 7
user defined, 5, 9
compounds, 47
copyright, 0, 32
dagger 'alif, 12, 16
\damm\ae, 12, 15, 16
inverted, 16, 22, 39
Dari, 21
date, 15
default font, 5, 11
defective writing, 12, 16, 39
definite article, 14, 19, 41
Derzhanski, Ivan, 45
diacritics, 16
diphthongs, 45
display mode, 7
dots on \(\ddot{y}\), 22, 25
E, 21
-E, 22
.e, 22, 23, 46
-e, 22
EDMAC, 27
emphasis, 44
environment
- Arabic, 6, 18
arabtext, 6, 18
setcode, 18
tabbing, 6
Farsi, 21
fat\(\text{ha}\), 12, 15, 16
Fischer, Wolfdietrich, 38
font
bold, 11
default, 5, 11
installation, 33
nash10, 34
nash14, 32-34
nash14bf, 33
naskh, 11, 32, 33
nasta'\(\text{liq}\), 22, 32
selection, 11
unavailable, 11
grouping, 7, 44
H, 21-23, 46
-h-, 15
\ham\(\text{za}\), 13, 15, 22, 40, 45, 46
carrier, 17, 40
old style, 41
\(\hat{\text{h}}\ar\(\text{ar}\)k\(\hat{\text{a}}\)\), 12, 15, 16, 38, 45
on \(\text{tattwil}\), 15
Heer, Nicholas, 53
hyphen, 15, 43
I, 12, 38
-\(I\), 22
*\(I\), 14
-\(i\), 22
\(i\), 12, 16, 39
\(i\) (\(\text{kasra}\)), 12, 38
implementation
- Mac, 32
PC, 32
U*X, 32
iN, 13, 39
input switching, 18
insertion
- mathematical, 7
- non-Arabic, 7
- Roman, 7
installation, 33
internal commands, 5
inverted \(\damm\ae\), 16, 22
invisible consonant, 14
ISO 646, 48, 50
ISO 8859-6, 50
ISO 9036, 48
iy, 12, 38
\(\text{iz\(\text{f}\)e\(\text{t}\)}\), 15, 22, 23, 46
\(\text{kas\(\ddot{\text{d}}\)a}\), 15, 25, 43
\(\text{kasra}\), 12, 15, 16
Kurdish, 21

1a-, 43

language selection, 10

\texttt{\LaTeX} commands, 7

li-, 43

ligature, 16, 34, 44

  breaking, 14–16, 44

lists, 9

long vowels, 12, 16

Macintosh, 32

\texttt{mad{	extasciitilde}da}, 14, 16, 45

Maghribi, 23

mathematical insertion, 7

\texttt{METAFONT}, 33

\texttt{ML2ARAB}, 53

Multi-Lingual Scholar, 53

N, 15, 16, 19

naskh, 11, 32, 33

nasta'liq, 22, 32

nesting, 7, 9

NFSS, 33

NFSS2, 33

non-Arabic insertion, 7

NU, 13, 39

numbers, 43

\texttt{'abg{	extasciitilde}d}, 53

O, 21, 46

option

abjad, 53

arabtex, 5

asmo449, 18

atrans, 20

etrans, 20

iso88596, 18

nashbf, 11

nastaliq, 11

oldarabtex, 35

twoblks, 52

Ottoman, 21

Pashto, 22, 23

PC implementation, 32

Persian, 21

Persian copula, 22

\texttt{pis}, 45

punctuation, 6

quotation

  Arabic, 6

  non-Arabic, 7

  Roman, 7

quoting, 13, 15, 16

Qur'an \texttt{'alif}, 16

reading module, 18

Roman insertion, 7

\texttt{sad{	extasciitilde}da}, 14, 16, 41

  on \texttt{tatweel}, 15

short vowels, 12

silent \texttt{'alif}, 17, 19

size changing, 5, 8, 11

special codings, 25

stretching, 8, 15, 25

  automatic, 25

\texttt{suk{	extasciitilde}n}, 15, 16, 22, 34, 46

  on \texttt{\text{"a}m}, 14

  on \texttt{\text{"a}tm}, 15

sun letter, 14

T, 40

tabbing environment, 6

\texttt{\text{"a}m\text{"a}r\text{"a}ba\text{"a}t}, 40

\texttt{tatwe\text{"a}m}, 13, 15–17, 19, 39, 40

\texttt{fath{	extasciitilde}a}, 40

  on \texttt{\text{"a}tm}, 15

\texttt{tasd{	extasciitilde}d}, 14, 16

\texttt{\text{"a}tm}, 15, 43

\texttt{\LaTeX} commands, 7

\texttt{\LaTeX} hash size, 5, 33

text

  archaic, 25

  erroneous, 25

\texttt{\LaTeX-\text{"a}\text{"e}\text{"a}T}, 33

transliteration, 12, 19, 34
Encyclopaedia of Islam, 20
ZDMG, 19
twoblks.sty, 52

U, 12, 19, 38
_U, 39, 46
“U, 14
_u, 12, 16, 39
u (gamma), 12, 38
U\textsc{t} in X implementation, 32
UA, 17, 38
uN, 13, 39
unavailable font, 11
Urdu, 22, 23
user defined commands, 5, 9
uw, 12, 38

van Dalen, Benno, 53
verbatim, 17
vowel marks, 16
vowels
  long, 12, 16, 38, 45
  short, 12, 38, 45

W, 19
WA, 17
\textit{wa}sha, 15, 16, 19, 34, 42

Y, 12, 38
y\textdegree
  dots, 22, 25
y\textdegree-i\textasciitilde\textasciitilde\textasciitilde
  wa\textasciitilde\textasciitilde\textasciitilde, 22, 23, 46

z\textdegree\textasciitilde, 45
zwarakay, 22