Arab\TeX
Typesetting Arabic and Hebrew\textsuperscript{1}
User Manual Version 4.00\textsuperscript{2}

Klaus Lagally

March 11, 2004

\textsuperscript{1}Report Nr. 2004/03, Universität Stuttgart, Fakultät Informatik
\textsuperscript{2}This Report supersedes Reports Nr. 1992/06, 1993/11 and 1998/09
Abstract

ArabTEX is a package extending the capabilities of \TeX/\LaTeX\ to generate the Perso-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. ArabTEX will run with Plain \TeX\ and also with \LaTeX\2e. It is compatible with Babel, CJK, LyX, the EDMAC package, and Pic\TeX\ (with some restrictions); other additions to \TeX\ may work, but have not been tried.

ArabTEX 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 extensions of the Perso-Arabic script, some limited support is available.

ArabTEX defines its own input notation which is both machine, and human, readable, and suited for electronic transmission and E-Mail communication. However, texts in many of the Arabic standard encodings can also be processed. ArabTEX also provides support for fully vowelized Hebrew, both in its private ASCII input notation and in several other popular encodings.

ArabTEX is copyrighted by the author, but free. If you use the system for scientific work please give appropriate credit to the software and the author (e.g. in the colophon of a monograph.) We also appreciate a complimentary copy of any scientific work produced with ArabTEX.

ArabTEX may be redistributed and/or modified under the terms of the LPPL (\LaTeX\ Project Public License) distributed from CTAN archives in the directory \texttt{ftp://ftp.dante.de/tex-archive/macros/latex/base/lppl.txt}, either version 1 of the License, or (at your option) any later version. 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
Universität Stuttgart
Institut für Formale Methoden der Informatik
Universitätstraße 38
70569 Stuttgart
GERMANY
lagally@informatik.uni-stuttgart.de

Copyright © 1992–2004, Klaus Lagally

Note: This manual describes version 4.00 of ArabTEX. The current version 3.11 is supposed not to change except for error corrections.
Contents

1 Introduction to Arab\TeX 7

2 Input to Arab\TeX 11
  2.1 Arabic text elements 12
  2.2 Commands within an Arabic context 13

3 Running Arab\TeX 15
  3.1 Activating Arab\TeX 15
  3.2 Language selection 16
  3.3 Font selection 16

4 Input encoding conventions 18
  4.1 ASCII Transliteration encoding 18
    4.1.1 Standard Arabic and Persian characters 18
    4.1.2 Vowelization 21
    4.1.3 Quoting 22
    4.1.4 Ligatures 23
    4.1.5 Coding examples for Arabic 23
  4.2 Verbatim input 28
  4.3 Alternate input encodings 29
    4.3.1 ASMO 449 = ISO 9036 29
    4.3.2 ASMO 449E = ISO 8859-6, MacOS Arabic 31
    4.3.3 CP 1256 = Arabic Windows Encoding 34
    4.3.4 UNICODE Arabic 34
CONTENTS

4.3.5 ISIRI 3342 .............................................. 34
4.3.6 Buckwalter Transliteration .......................... 38

5 Transliteration .............................................. 39
  5.1 ZDMG transliteration style ............................. 39
  5.2 Other transliteration styles ............................ 40
  5.3 Capitalization ........................................... 40

6 Support for other languages ................................ 41
  6.1 Persian (Farsi, Dari) .................................... 41
    6.1.1 Coding examples for Persian ........................ 42
  6.2 Maghribi .................................................. 44
  6.3 Urdu ...................................................... 44
    6.3.1 Coding examples for Urdu ............................ 45
  6.4 Pashto (Afghanic) ........................................ 48
  6.5 Sindhi .................................................... 49
  6.6 Kashmiri .................................................. 50
  6.7 Uighuric .................................................. 50
  6.8 Old Malay ................................................ 52
  6.9 Other extensions of the Perso-Arabic script ......... 52

7 Hebrew mode ............................................... 53
  7.1 Language switching ..................................... 53
  7.2 Standard Hebrew encoding ............................. 54
  7.3 ISO 8859-8 and Hebrew MS-Windows ................... 55
  7.4 HebrewTEX “oldcode” and “newcode” .................... 55
  7.5 BHS encodings .......................................... 59
  7.6 UNICODE Hebrew .......................................... 60
  7.7 Hebrew transcription systems .......................... 60
  7.8 Hebrew fonts .............................................. 60
  7.9 Judeo-Arabic ............................................. 62
  7.10 Yiddish ................................................ 63
## CONTENTS

8 Miscellaneous features 64

8.1 Additional codings 64
8.2 Dots on ya’ 65
8.3 Vowel positioning 65
8.4 Abjad numerals 65
8.5 Automatic stretching 65
8.6 Uniform baselines 66
8.7 Verbatim copy of the input 66
8.8 Progress report 66
8.9 Module Reporting 66

9 Compatibility issues 67

9.1 Arabic document classes 68
9.2 Using ArabTEX with EDMAC 68
9.3 Using ArabTEX with Babel 69
9.4 Using ArabTEX with PicTEX 69
9.5 Using ArabTEX with CJK 69

10 Acknowledgments 70

A Obtaining and installing ArabTEX 74

A.1 Obtaining ArabTEX 74
A.2 Installing ArabTEX 75

B Release history 76

C Miscellaneous utilities 79

C.1 twoblks.sty 79
C.2 verses.sty 80
C.3 raw.sty 80
List of Figures

1.1 Sample Arab\TeX\ input ............................................. 9
1.2 Sample Arab\TeX\ output ........................................... 10
7.1 BHS Hebrew example .................................................. 59
C.1 Example of Arabic poetry (source unknown) ...................... 80
List of Tables

4.1 Standard encodings for Arabic and Persian consonants. . . . . . . 19
4.2 Additional encodings generally available. . . . . . . . . . . . . . 20
4.3 Verbatim encodings for the carrier of hamza . . . . . . . . . . . 29
4.4 ASMO 449 code table . . . . . . . . . . . . . . . . . . . . . . . 30
4.5 ISO 8859-6 and MacOS Arabic code table . . . . . . . . . . . . . 32
4.6 Windows CP1256 code table . . . . . . . . . . . . . . . . . . . . 33
4.7 UNICODE Arabic, Part 1 . . . . . . . . . . . . . . . . . . . . . . 35
4.8 UNICODE Arabic, Part 2 . . . . . . . . . . . . . . . . . . . . . . 36
4.9 ISIRI 3342 code table . . . . . . . . . . . . . . . . . . . . . . . 37
4.10 Buckwalter transliteration . . . . . . . . . . . . . . . . . . . . . 38

6.1 The Urdu Alphabet . . . . . . . . . . . . . . . . . . . . . . . . . . 46
6.2 Additional codings for Pashto. . . . . . . . . . . . . . . . . . . . 48
6.3 The Sindhi Alphabet . . . . . . . . . . . . . . . . . . . . . . . . 49
6.4 The Kashmiri Alphabet . . . . . . . . . . . . . . . . . . . . . . . 50
6.5 Arab\TeX encoding of Uighuric . . . . . . . . . . . . . . . . . . . 51

7.1 ISO 8859-8 and Windows CP 1255 code table . . . . . . . . . . . . 56
7.2 MAC Hebrew code table . . . . . . . . . . . . . . . . . . . . . . 57
7.3 Hebrew\TeX, CP 1255 and ISO 8859-8 code table . . . . . . . . . . 58
7.4 UNICODE Hebrew . . . . . . . . . . . . . . . . . . . . . . . . . 61
7.5 Judeo-Arabic encoding . . . . . . . . . . . . . . . . . . . . . . . 63
7.6 Yiddish YIVO transcription . . . . . . . . . . . . . . . . . . . . . 63
8.1 Additional codings for special purposes. . . . . . . . . . . 64
Chapter 1

Introduction to ArabTEX

*Note: This manual describes Version 4 of ArabTEX, which in general is upwards compatible to earlier versions. Incompatible features and extensions are flagged by an asterisk in the margin.*

ArabTEX is a package extending the capabilities of \TeX/\LaTeX to generate an extended Perso-Arabic and/or Hebrew script in addition to the customary left-to-right scripts (called Roman in the sequel). Besides Arabic and Hebrew there are provisions for many other languages that use some extensions of the Arabic script; we shall use the term Arabic generically to denote any supported right-to-left script, without any cultural or political connotation.

An ArabTEX document is usually multi-lingual and may contain right-to-left insertions within left-to-right paragraphs, and vice versa. There are various possibilities to encode right-to-left insertions: in addition to several standard encodings there are also 7-bit ASCII encodings modelled after various transliteration standards (there is more than one convention, so the intended language must be specified).

ArabTEX, like \TeX and \LaTeX, is not a substitute for a word processor, and does not follow the WYSIWYG paradigm (What You See Is What You Get) where the author has complete control on, and therefore is also completely responsible for, every detail of the visual representation of her text. Instead it is an off-line system mirroring the classical publication process where an author, concentrating on the content and the structure of her paper only, will deliver a manuscript to the publisher, who will take care of a reasonable visual rendering of the text.

The publisher involved in our case is \TeX, the famous typesetting program by D. E. Knuth; and the amount of technical typesetting knowledge embodied in the algorithms of \TeX is such that, in general, the author will tacitly accept the resulting output if she had stated her preferences in a sufficiently precise way;
otherwise, the formatting task has to be repeated under changed specifications, as also in the classical case.

\TeX{} is easily extensible by providing its own macro programming language, enabling the user to implement packages that contain algorithms embodying additional typesetting knowledge. A prominent example is \LaTeX{} by L. Lamport, a package providing standard formatting styles for several common document classes. One of the basic concepts of \LaTeX{}, but already contained in \TeX{} itself, is “structured markup”: the user generally only indicates the logical structure of the document, whereas the formatting rules are laid down within descriptions of standard document classes.

\ArabTeX{}, at the user interface, needs to add only very few features: we provide a way to indicate which parts of the input text are considered to be in Arabic (or some other supported language written from right to left) and thus have to be rendered accordingly, and a way of setting parameters influencing the rendering of Arabic text. Additionally we have to support a representation of the Arabic text that can be handled using the available standard equipment.

Internally, of course, there is much more to it. \TeX{} may be very good at rendering left-to-right texts, but Arabic runs from right to left, so we have to instruct \TeX{} how to do it the other way; and if we want to mix Arabic and traditional left-to-right text within the same paragraph, things can get extremely complicated. This shows indeed: \ArabTeX{} is very large and comparatively slow. Fortunately, computer technology is advancing at a high pace, so that the penalty of using very involved algorithms will decrease just by simple waiting; and we thus never were too much concerned about efficiency (even if we tried to avoid the worst blunders).

In order to see how \ArabTeX{} works, let us consider a very simple example: Figure 1.1 contains the complete \TeX{} input text, Figure 1.2 the corresponding output. In this example (using \LaTeX{}), we use a standard “article” format, we activate the \ArabTeX{} package, and we set a few options: our text is supposed to be in Arabic, we want to see all vowel marks (this is not always done in Arabic printing), we additionally want to get the scientific transcription, and, of course, the Arabic writing. The document proper consists of a centered headline and a sequence of short Arabic paragraphs, separated in the input by blank lines.\footnote{If you happen to be curious and cannot read Arabic: the text contains a traditional story about a somewhat silly person named \Gu ha, trying not to lend his donkey to a friend, and failing.}

This example is not typical insofar as we produce both the scientific transcription of the text, and the Arabic writing, from the same input at the same time, interleaving them. This is rarely required; but in our example it allows us to demonstrate that the input notation is very closely related to the transcription. Details about this correspondence are covered in Chapter 4. What indeed is typical is the fact that, apart from the centered headline, we supplied no formatting information whatever; \TeX{} will take care of that.
\documentclass[12pt]{article}
\usepackage{arabtex}

\begin{document}

\setarab % choose the language specific conventions
\vocalize % switch diacritics for short vowels on
\transtrue % additionally switch on the transliteration
\arabtrue % print arabic text ... is on by default anyway

\centerline {\RL{^gu.hA wa-.himAruhu}}

\begin{RLtext}
\'at_A .sadiquN 'il_A "gu.hA ya.tlubu minhu .himArahu li-yarkabahu
ff safraTiN qa.sIraTiN wa-qAla lahu:
sawfa '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 yanhaqu fI i.s.tablihi.

fa-qAla lahu .sadiquhu:

\'innI 'asma'u .himAraka yA ^gu.hA yanhaqu.

fa-qAla lahu ^gu.hA:

.garIbuN 'amruka yA .sadiqI!
\'a-tu.saddiqu al-.himAra wa-tuka_d_dibunI?
\end{RLtext}

\end{document}

Figure 1.1: Sample ArabTEX input


guhat wa-himāruhu

ratā ṣadīqun ilā guhā yatlabu minhu ḥimāruhu li-yarkabahu fisaffratin qaṣīratin wa-qāla lahu:

أَتَى صَدِيقٌ إِلَى جِحَادا يَتَلَبَّ مِنْهُ حِمَارٍ لِتَرَكَّبِهِ فِي سَفْرَةٍ فِي صَفْرَةٍ فَقَالَ لَهُ:

sawfa ṣiddu ṣiddu fi 'l-masāfī, wa-yafaru laka wugrata.

سُوْفَ أُعِيدُهُ إِلَيْهِ فِي الْحُمَارِ، وَأَفْقَعُ لَهُ أَجْرَهُ.

faxāla guhā:

فَقَالَ جِحَادٌ:

rānā ṣasifun giddan rannī lā rastāfīrū rā ṣāhagīqā laka raqbatanka, faxhimāru layasa hunā 'l-yawma.

أَنَا أَلْبَسْ جُهَادًا أَنَّى لَا أَسْتَطِيعُ أَنْ أَحْقِقَ لَهُ رَغْبَتِهِ، فَحَمَارُ لَيْسَ هَذَا الْيَوْمُ.

wa-gablā rān yutimmu guhā kalāmahu badar 'l-himāru yanhaqu fi 'stablīhi.

وَقَالَ أَنْ يَتِمُّ جِحَادٌ كَلاَمَهُ بَدَا الْحِمَارُ بَنِيَّقُ فِي اسْتَبْلِيْهِ.

faxāla lahu ṣadīquhu:

فَقَالَ لَهُ صَدِيقُهُ:

inna 'asmaru himāra yā guhā yanhaqu.

إِنِّي أَنْحِجُ حِمَارَكَ بِهَا جَمَا بَنِيَّقُ.

faxāla lahu guhā:

فَقَالَ لَهُ جِحَادٌ:

jarībun ṣumraka yā ṣadīqī 'a-tuṣaddiqu 'l-himāra wa-tukaddībūni?

حَرِيبٌ أُمِّرْكَ بِهَا صَدِيقَيْنِ! أَتْصَدَقَ الْحِمَارُ وَتَكَّذَبْ بِئْنِي؟

Figure 1.2: Sample ArabTeX output
Chapter 2

Input to Arab\TeX

Arab\TeX is activated by the command $\texttt{\input arabtex}$ (with Plain \TeX) or $\texttt{\usepackage{arabtex}}$ (with \LaTeX). After activating Arab\TeX, select one of the language-specific Arabic writing styles, e.g., $\texttt{\setarab}$ (see Section 3.2). Your modified \TeX/\LaTeX system will recognize the following items:

- standard \TeX/\LaTeX text and commands,
- \textit{Arabic quotations}\footnote{The former restriction that quotations must fit on the current line no more applies.} as arguments to the command $\texttt{\RL{}}$ (read: “right-to-left”) within a \textit{Roman} paragraph. A quotation may also be bracketed by $<$ and $>$ (deprecated!) or by $<$ and $>$ except inside a \LaTeX \texttt{tabbing} environment\footnote{Quotations closed by $>$ may not contain nested insertions; also observe that $\langle$ must be matched by $>$, \textit{not} by $\rangle$!}. An Arabic quotation may not contain more than a single paragraph of right-to-left text and forms a new \TeX group, so any assignments will be local by default.
- longer Arabic text segments called \textit{Arabic environments}, which are bracketed by the commands $\texttt{\begin{RLtext}}$ and $\texttt{\end{RLtext}}$ (or also $\texttt{\begin{arabtext}}$ and $\texttt{\end{arabtext}}$ with the same meaning) (even when using Plain \TeX!). An \textit{Arabic environment} consists of one or more paragraphs separated by blank lines or $\texttt{\par}$ commands. It forms a group, so assignments will be local by default.

\textit{Arabic quotations} and \textit{Arabic environments} will be called \textit{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 white space or single newlines:

- isolated punctuation marks, interpreted as the corresponding Arabic punctuation mark;
- “numbers”, i.e. character sequences starting with a digit, and possibly continued by digits, commas, hyphens, or slashes. A “number” will be processed using the normal writing sequence from left to right; however, a final punctuation mark will be split off and processed separately.
- “Arabic quotes” coded as two left quotes or two right quotes each, or as `\lq` and `\rq`. They should 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 Arabic text elements (words, numbers, and special characters) enclosed in curly braces `{}`. 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 Arabic Environment, or in an Arabic quotation, you may also have:

- \Arab\TeX\ commands with or without parameters. These will be executed immediately.
- Some, but not all, \TeX/L\Au\TeX\ commands (see below). These will be executed immediately.
- Short mathematical insertions, bracketed by `\(` and `\)` or by two `\$` signs. They must fit on one output line and are processed as usual.
- Displayed equations bracketed by `\[` and `\]` or by `$$`.
- short left-to-right (“Roman”) quotations, containing text and possibly also \TeX/L\Au\TeX\ commands, as argument to `\LR{}` (read: “left-to-right”) or bracketed by `<` and `>` (deprecated!)

These introduce a new level of grouping, so if they contain any \TeX/L\Au\TeX\ assignments the effects of these will be local by default. When using `<` and `>` this feature is not available within an Arabic quotation. The alternate notation `\<` is not provided.

\footnote{Quotations closed by `>` may not contain nested insertions.}
2.2 Commands within an Arabic context

ArabTEX will process all input text contained within an Arabic context, even all \TeX/\LaTeX commands. These must thus be known to ArabTEX, since their meaning, or their implementation, might be different there from the usual interpretation. Thus ArabTEX has to provide a private implementation for any admissible command, and will normally try to reuse parts of the existing mechanisms. As a consequence, any incompatible changes to Plain \TeX and/or \LaTeX might require corresponding changes to ArabTEX too.

A control sequence inside an Arabic context may be of the following kinds. Note that commands which are intrinsically vertical, like e.g., \vskip, make no sense within Arabic quotations, and may lead to unpredictable effects there.

- ArabTEX option changing commands. These may also be used outside an Arabic Context, and usually follow \TeX’s grouping rules. Changes of language, Arabic and Hebrew font, input encoding, and transliteration conventions are global.

- \ for a line break; the current line will be padded out on the left with spaces.

- \ or \break for a line break; the current line will be spread out. If it comes out very badly spaced, automatic stretching might help (see Section 8). The current paragraph will not be reformatted.

- \indent or \par (or a blank line) for a new paragraph, \noindent for a new paragraph without indentation.

- \emph{\{group of Arabic items\}} or \textbf{\{items\}} will put a bar over the indicated group of Arabic items.

- \setnash, \setnashbf, \setnastaliq and other font selection commands, see Section 3.3.

- size changing \LaTeX commands like \large etc., only if \LaTeX is used!

- the following commands: \footnote (observe that the syntax for Plain \TeX and \LaTeX is different!), \marginpar (even with Plain \TeX, analogous to the \LaTeX usage).

- the \TeX/\LaTeX commands \smallskip, \medskip, \bigskip, \input, \hfill, \hfil, \vfill, \vfil, \null (for a space), \space, \, (small space), \newpage, \clearpage, \pagebreak with their usual meaning.

- \mbox{\text} puts the text argument into a box that will not be split.

- \nospace, \unskip will place the adjacent items in the output in direct contact, without any intervening space, except in case of a line break.
\hspace{width} will introduce the indicated amount of spacing in the output. The same is true for \vspace, \hskip, and \vskip.

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

- \spreadline{text} will start a new line whose contents are spread out over the whole width of the page. It is approximately equivalent to \spreadbox{\hsize}{text}.

User defined commands, even if they require parameters, may be called directly within an Arabic context if they have been previously announced to the Arab\TeX processor by \allowarab{command name}. After substituting the parameters they are expanded exactly once, and the expansion text will be processed by Arab\TeX again, so it must be legal Arab\TeX input text.

Even if not announced previously, user defined commands may also be called by \docommand{command name and parameters}. The command is expanded exactly once, and the expansion text, after suitable substitution of parameters, will be processed by Arab\TeX again.

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 normally no further \LaTeX or Arab\TeX environment may be nested; this restriction does not apply to the experimental \LaTeX document classes arabrep.cls, arabart.cls, arabbook.cls (see Section 9.1) which are provided for right-to-left documents.

For a list of all available commands, consult the Index to this report. As a reminder, the command \arabstat will cause a list of all commands that are presently valid inside Arabic text to appear in the \TeX log file.

\footnote{This is no strong restriction as the expansion may contain \docommand calls again.}
Chapter 3

Running ArabTEX

3.1 Activating ArabTEX

ArabTEX can be used both with Plain TEX and with LATEX, but is activated differently in both cases.

With Plain TEX, a small loader program is activated by the command \input arabtex at the beginning of an input text. It will define a default font, prepare a minimal environment simulating the (very few) LATEX-like features needed, and load the ArabTEX macro files. With LATEX2ε the command \usepackage{arabtex} will do all the loading.

If Hebrew mode (see Section 7) is required, use \input hebtex.tex resp. \usepackage{hebtex} instead. The required modules of ArabTEX will be loaded automatically.

With Plain TEX only, add \input apatch.sty at the beginning of your document proper, after reading any other auxiliary files, to get the most recent corrections.

ArabTEX will load many internal files automatically, and defines a large numbers of internal commands. These all contain an “at”-sign (@) within their names and thus should not interfere with user defined commands. Collisions with other macro packages are possible, however, and may lead to surprises and interesting effects.

ArabTEX tries to diagnose the presence of some other packages with which it could run into conflicts, and sometimes locally modifies itself accordingly. For this to be possible, in case of doubt the ArabTEX package should be loaded last.

Users still running LATEX 2.09 (horror!) should either add the option arabtex to the \documentstyle command, or rather upgrade to LATEX2ε.
3.2 Language selection

The processing of input text in ASCII transliteration encoding 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 language selection commands \setarab, \setfarsi, \seturdu, \setpashto, \setmaghribi, etc.,\footnote{We would prefer to use a single switching command like \texttt{language}, \texttt{setlanguage}, or \texttt{selectlanguage}, but these names have already been preempted by \TeX\!3 and the Babel package.} or \setverb (no special processing; see however Section 4.2). The processing mode may be changed at any time, even inside an Arabic environment or an Arabic quotation.

Arabic insertions are generally included in \texttt{\RL{\{}} or bracketed by \texttt{\textless} and \texttt{\textgreater}. By selecting a language, the symbols \texttt{\textless} and \texttt{\textgreater} are also activated as shorthands to bracket short insertions in the chosen language. Whereas this is usually convenient, it also has some drawbacks: the angle brackets can thus no more be used for other purposes, except in mathematical mode where they retain their normal meaning as relational operators. To return them to their normal mode of operation, you can deselect them by \texttt{\setnone}.

For further details on supported languages, see Section 6.

3.3 Font selection

For producing the extended Arabic script Arab\TeX\ uses a special strategy to build up character shapes from a collection of fragments, which normally do not correspond to individual character glyphs. Therefore none of the available free or commercial Arabic fonts can be used; we provide our own “pseudo-fonts”.

Presently the following pseudo-fonts are available:

- "xnsh14" is the default,
- "xnsh14bf" is a bold-face version of "xnash14",
- "nash14" is an older version of "xnsh14" containing less glyphs and only provided for compatibility with existing documents.
- "nash14bf" is a bold-face version of "nash14".

The font "xnsh14" the default; use \texttt{\setnashbf} or \texttt{\bf} to switch to bold-face. \texttt{\setnash} or \texttt{\rm (!)} will switch back.

\footnote{Note for advanced \TeX\! users: All language selecting commands except \texttt{\setnone} set the character \texttt{\textless} to be active. If Arabic insertions are not needed, or are always started with \texttt{\textless} or \texttt{\RL{\textless}}, the user may reuse the command \texttt{\textless} for other purposes, or deactivate it by \texttt{\setnone} or \texttt{\catcode `\textless=12} to return it to its normal meaning.}
3.3. FONT SELECTION

With Plain \TeX{} the fonts are available by default at 14 point size only, which cooperates well with the "cm" fonts at 10 points. Additional sizes are defined within the file "arabtex.tex"; they can be activated whenever needed by the command \texttt{\setarabfont(font)}.

With \LaTeX{} the font size changing commands will also operate on the Arabic fonts.

We strongly recommend migrating to "xnsh14" and "xnsh14bf"; the command \texttt{\oldarabfont} will switch to the old fonts, if necessary, and \texttt{\newarabfont} will restore the default. The fonts "nash14" and "nash14bf" are of inferior quality \textbullet{} and will be phased out gradually.

All fonts indicated presently are in the Naskhi style; we had started to write a Nastaliq font for Persian and Urdu, but ran into grave implementation problems, yet unsolved.

Due to a donation by Taco Hoekwater, the fonts "xnsh14" and "xnsh14bf" are also available in Postscript T1 format. Their use is highly recommended when using a Postscript interpreter, or converting the output to PDF format; readability is dramatically improved.

In Hebrew mode (see Section 7) you can use the standard fonts available on CTAN (after installing them locally). As defaults the fonts "hclassic" and "hcpaption" are provided with Arab\TeX{}; these fonts contain vowel points. Also \textbullet{} for the other fonts vowel points are provided.
Chapter 4

Input encoding conventions

4.1 ASCII Transliteration encoding

The ASCII input notation for Arabic text has been modelled closely after the transliteration standards ISO/R 233 and DIN 31635. These standards do not guarantee unique re-transliteration and are also not 7-bit ASCII compatible, therefore 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, and the encoding could become ambiguous).
- use capital letters for writing variants.

4.1.1 Standard Arabic and Persian characters

The standard encodings for Arabic and Persian consonants are given in Table 4.1 and Table 4.2.

- For long vowels, we use the capital letters <A>, <I>, <U> or also <aa>, <iy>, <uw>, with the same meaning.
- To get the defective writing of long vowels, use <_a>, <_i>, <_u>.
- 'alif maqṣūra is <_A> or <Y>.
4.1. ASCII TRANSLITERATION ENCODING

<table>
<thead>
<tr>
<th>a</th>
<th>ً</th>
<th>a</th>
<th>'alif</th>
<th>b</th>
<th>َ</th>
<th>b</th>
<th>ba’</th>
<th>p</th>
<th>َ</th>
<th>p</th>
<th>pā’</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>ت</td>
<td>t</td>
<td>tā’</td>
<td>t</td>
<td>ت</td>
<td>tā’</td>
<td>נּ</td>
<td>נּ</td>
<td>ِ</td>
<td>ِ</td>
<td>ِ</td>
</tr>
<tr>
<td>h</td>
<td>רּ</td>
<td>h</td>
<td>ḥā’</td>
<td>h</td>
<td>רּ</td>
<td>ḥā’</td>
<td>d</td>
<td>ِ</td>
<td>d</td>
<td>d</td>
<td>dāl</td>
</tr>
<tr>
<td>d</td>
<td>ذ</td>
<td>d</td>
<td>dāl</td>
<td>r</td>
<td>r</td>
<td>rā’</td>
<td>z</td>
<td>َ</td>
<td>z</td>
<td>z</td>
<td>zāy</td>
</tr>
<tr>
<td>s</td>
<td>س</td>
<td>s</td>
<td>sin</td>
<td>s</td>
<td>ش</td>
<td>sh</td>
<td>s</td>
<td>ص</td>
<td>s</td>
<td>s</td>
<td>šād</td>
</tr>
<tr>
<td>t</td>
<td>ض</td>
<td>d</td>
<td>dād</td>
<td>t</td>
<td>َ</td>
<td>tā’</td>
<td>.z</td>
<td>َ</td>
<td>z</td>
<td>z</td>
<td>zā’</td>
</tr>
<tr>
<td>q</td>
<td>ق</td>
<td>q</td>
<td>qāf</td>
<td>v</td>
<td>ش</td>
<td>sh</td>
<td>v</td>
<td>vā’</td>
<td>k</td>
<td>َ</td>
<td>k</td>
</tr>
<tr>
<td>g</td>
<td>گ</td>
<td>g</td>
<td>gāf</td>
<td>l</td>
<td>ل</td>
<td>lām</td>
<td>m</td>
<td>م</td>
<td>m</td>
<td>mīm</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>ن</td>
<td>n</td>
<td>nun</td>
<td>h</td>
<td>ه</td>
<td>ḥā’</td>
<td>w</td>
<td>و</td>
<td>w</td>
<td>wāw</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>ی</td>
<td>y</td>
<td>yā’</td>
<td>َ</td>
<td>َ</td>
<td>َ</td>
<td>T</td>
<td>َ</td>
<td>h</td>
<td>tā’</td>
<td>marbuṭa</td>
</tr>
</tbody>
</table>

Table 4.1: Standard encodings for Arabic and Persian consonants.

- The short vowels fatha, kasra, damma are coded <a>, <i>, <u> and need not normally be written except in the following cases:
  - 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 \vocalize and \fullvocalize modes.

- tanwin is coded <aN>, <IN>, or <uN>. A silent ‘alif, if required, is supplied automatically; it may also be explicitly written: <aNA>. Likewise, a silent wāw may be written <NU> as in <‘amruNU>.

- hamza is denoted by a single right quote <’>. After selecting the language by \setarab the carrier of hamza will be determined from the context according to the rules for writing Arabic words; if that is not wanted, “quote” the hamza (see Section 4.1.3 below). In the \setverb mode, the carrier of hamza is determined by the following input character; see Section 4.2.
<table>
<thead>
<tr>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>‘hā’ with hamza</td>
</tr>
<tr>
<td>١c</td>
<td>ǧīm with three dots (below)</td>
</tr>
<tr>
<td>١c</td>
<td>ḥā’ with three dots (above)</td>
</tr>
<tr>
<td>١z</td>
<td>zāy with three dots (above)</td>
</tr>
<tr>
<td>١n</td>
<td>kāf with three dots (Ottoman)</td>
</tr>
<tr>
<td>١l</td>
<td>lām with a bow accent (Kurdish)</td>
</tr>
<tr>
<td>١r</td>
<td>rā’ with a bow below (Kurdish)</td>
</tr>
</tbody>
</table>

Table 4.2: Additional encodings generally available.

- **madda** on ‘alif is generated by a right quote (hamza) before `<A>`: `<’A>`.
  - It may also be written `<^A>`; likewise, `<^I>` and `<^U>` will produce madda on ya’ and on wāw, as required in some older writing conventions.\footnote{The former use of `<^A>`, `<^I>`, and `<^U>` has been discontinued in Arab\TeX\, version 4.}

- 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 4.1.3 below). The adjacent characters will not be connected.

- šadda is indicated by doubling the appropriate letter coding. Therefore two equal consonants in sequence have to be separated by a short vowel indicator or `<|>` even in \n\texttt{novocalize} mode.

- The definite article is separated from the following word by a hyphen. It may be written in the assimilated form (if it exists): `<as-salāmu>`, or always as `<al->`; in that case a subsequent “sun letter” must be doubled: `<al-ssalāmu>`, to receive a šadda, and to prevent a sukūn on the lām. The transliteration in both cases is identical.

- Hyphens `<->` are used for tying words together and for separating prefixes and the article; in these cases they start a new word. Hyphens can also be used to indicate inflectional endings, a connecting vowel in Arabic, or an
izāfet connection in Persian. Hyphens will show up in the transliteration. Additionally, at the beginning and/or 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 h>.

- A double hyphen <--> between two otherwise joining letters will break any ligature and will insert a horizontal stroke (tatwil, kašida) without appearing in the transliteration. It may be used repeatedly. See also Section 8.5: automatic stretching.

For special applications, it can also be coded <B>; and <|B> will behave like any ordinary consonant and may carry vowel indicators, tanwin, sukūn, and, in the combination <|BB>: šadda.

4.1.2 Vowelization

There are three modes of rendering short vowels:

- \fullvocalize:
  - Every short vowel written will generate the corresponding diacritical mark fatha, kasra, damma, except if quoted.
  - If <N> follows a short vowel, the corresponding form of tanwin is generated instead.
  - Defective writing: The coding <_a> will produce a Qur’an ‘alif accent (also called dagger ‘alif) instead of an explicit ‘alif character which would be coded <A> or <aa>. Likewise, <_i> will produce a small ‘alif below the preceding consonant in place of <I> (<iy>), <_u> will produce an inverted damma in place of <U> (<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, sukūn will be generated except if a double quote precedes the next consonant. The lām of the definite article receives no sukūn if a doubled “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.

- \vocalize: As above, but sukūn and wasla will not be generated except if explicitly indicated by “quoting” (see section 4.1.3).

- \novocalize: No diacritics will be generated except if explicitly asked for by “quoting”(see section 4.1.3).
In all modes, a doubled consonant will generate šadda, and <'A> always generates madda on 'alif.

After <aN> the silent 'alif character is generated automatically if required. The silent 'alif may also be explicitly indicated by <aNa>, or coded literally as <A> in \novocalize mode. If a silent 'alif maqsūra is wanted instead, write <aNa_A>, <aNY>, <_A> or <Y>.

The tanwin fatḥa is normally positioned on the last consonant of the word, even if a silent 'alif follows. If it is instead supposed to go onto the 'alif as required by some modern Arabic writing conventions, or in Persian, this behaviour can be achieved by the option \newtanwin. The option \oldtanwin will restore the classical behaviour.

A silent 'alif after wāw is indicated by <UA> or <WA> (with a capital <W>!).

4.1.3 Quoting

In \novocalize mode (see Section 4.1.2), a double quote <"> will modify the meaning of the following character as follows:

- if a short vowel follows, the appropriate diacritical mark fatḥa, kasra, ḍamma will be put on the preceding character.
  - If <N> follows the short vowel, the appropriate form of tanwin will be generated instead.
  - At the beginning of a word, 'alif is assumed as the first character.

- if the following character is a single right quote, a hamza mark will be put on the preceding character even if in conflict with the hamza rules.
  At the beginning of a word, <"'> will generate an isolated hamza.

- if the following character is the “invisible consonant” <|>, 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, 'alif with waṣla will be generated.

- otherwise: a sukūn 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 \vocalize mode, (see Section 4.1.2), quoting will turn a short vowel off; likewise, in \fullvocalize mode, quoting will also turn a sukūn off. Put in other words: quoting will toggle the generation of short vowel indicators and sukūn on and off.
4.1. ASCII TRANSLITERATION ENCODING

4.1.4 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 consonant «» between the two letters otherwise combined into a ligature. After `\ligsf`, 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 `\ligstr`.

4.1.5 Coding examples for Arabic

Short vowels *fatha, kasra, damma* are denoted, as in the transliteration, by the small letters a, i, u:

\[
\begin{align*}
\text{mana’a} & \quad \text{manāa}, \\
\text{dhaba} & \quad \text{dhabā}, \\
\text{qabil} & \quad \text{qabil}, \\
\text{ni’ma} & \quad \text{nīma},
\end{align*}
\]

Long vowels ō, ĩ, ū are denoted by capitals A, I, U or by aa, iy, uu:

\[
\begin{align*}
\text{qAtala} & \quad \text{qātala}, \\
\text{sIrI} & \quad \text{sīrī}, \\
\text{Alif maqsūra} & \text{ coded as } _A \text{ (or Y.)}
\end{align*}
\]

Silent ‘alif: The suffixes -ū, -aw of the verb are denoted UA, aw or aWA:

\[
\begin{align*}
\text{katabUA} & \text{ katabū}, \\
\text{rAmaWA} & \text{ rāmu}, \\
\text{YaqūNA} & \text{ yactubu},
\end{align*}
\]

The defective notation of ō, ĩ, ū can be indicated by _a, _i, _u and leads to the appropriate spelling:

\[
\begin{align*}
\text{dAu-h_u} & \quad \text{dāra-hū}, \\
\text{rI-gli-h_i} & \quad \text{riḡli-hī}, \\
\text{rabb_i} & \quad \text{rabbī}, \\
\text{hAt_i} & \quad \text{hāti}, \\
\text{haDi_i} & \quad \text{hādi}, \\
\text{tih_i} & \quad \text{tihī}, \\
\text{hāz_u} & \quad \text{hāzū}, \\
\text{hāzA_u} & \quad \text{hāzā}, \\
\text{hāzA_i} & \quad \text{hāzāī}, \\
\text{rabb_i} & \quad \text{rabbī}, \\
\text{hAl_i} & \quad \text{ṣālī}, \\
\text{hum_u} & \quad \text{humū}.
\end{align*}
\]

\footnote{Most of the examples are taken from: Wolfdietrich Fischer, Grammatik des Klassischen Arabisch, 2. Auflage, Verlag Otto Harrassowitz, Wiesbaden 1987.}
CHAPTER 4. INPUT ENCODING CONVENTIONS

qiy_amaTuN ❰质量管理❯, 'il_ahuN ❰伊斯兰❯ ilahun,
sam_awAtuN ❰单词❯ samawatun, _tal_a_tuN ❰时态❯ talatun,
_akin ❰亲近❯ lakin, h_a_dA ❰真主❯ hadda, 'al-11_ahu ❰真主❯ al-lahu,
'al-rra.hm_anu ❰真宰❯ ar-rahamanu, _d_alika ❰亲切❯ dalka.

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

.saLaTuN ❰察尔❯ salatun, .hay_aUTuN ❰受❯ hayatun,
zak_aUTuN ❰作❯ zakatun, mi_lK_aUTuN ❰施舍❯ miskatun,
ar-rib_aU ❰里巴❯ arriba, tawr_aITuN ❰塔瓦特❯ tawratun,
ram_aYhu ❰拉曼❯ ramahu, sIm_aYhum ❰史密斯❯ smahum.

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

_/U_aL_ka ❰乌尔喀❯ _ulaka, _/U_aL'i ❰乌尔米❯ _ulai,
_/U_aL_a ❰乌尔阿❯ _ulala, /_U_aL'a ❰乌尔阿卡❯ _ulaka.

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

ra_gulun ❰拉谷伦❯ raqulun, ra_guliN ❰拉谷尼❯ raqulun, ra_gulaN ❰拉谷拉❯ raqulun,
madInaTaN ❰麦地那❯ madinatan, _gamilaTaN ❰加米拉塔❯ gamilatan,
'i_daN ❰伊丁❯ _idan, sama'a'AN ❰茶马安❯ sama'an.

There is a special case:
ribaNU ❰里巴努❯ riban; 'amruNU ❰阿姆鲁努❯ _amrun, 'amriNU ❰阿姆里努❯ _amrinn,
however: 'amraN ❰阿姆劳❯ _amran.

Tanwin 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.
ra_gulaN ❰拉谷拉❯ raqulun, _i_daN ❰伊达❯ _idan.

A silent 'alif maqṣūra after tanwin is written aNY or aN_A:

hudaNY ❰涂达❯ hudan, fataN_A ❰法塔❯ fatan;
compare: al-hud\Y ❰阿卢达❯ al-huday, 'al-fat_8 ❰阿法塔❯ al-fatay.
4.1. ASCII TRANSLITERATION ENCODING

Tā’ marbuta is denoted by T:
kalimaTuN كَلِّمَةٍ kalimatun, kalimaTiN كَلِّمَة١٠ kalimatin,
kalimaTaN كَلِّمَة٢٠ kalimatun; fatATuN فَذَّة١٠ fatâtan,
fatATiN فَذَّة٢٠ fatâtin, fatATaN فَذَّة٢٠ fatâtan.

Hamza is indicated by ’; its carrier is determined by the context:
‘amruN أَمْرٍ amrun, ‘ibiluN أَبْيَلَ ibilun, ‘u.htuN أَهْتَتَ uhtun;
ra’suN رَأْسَهُ rasun, ‘ar’asu أَرْأَسَ arasu, sa’ala saala;
qara’a قَرَاءَ qarâ; bu’suN بَعْسَهُ busun, ’ab’usuN عَبْسَعَ abusun,
ra’ufa رَأْفَتَ ra’ufa, ru’asA’u رَعْشَا’و rasâ’u; bi’ruN بَيْرَتَ birun,
as’ilaTuN كَيْبَة١٠ کَیْبَتَ as ilaTaN كَیْبَتَ ۱٠ as ilatun,
su’ila كَیْبَة٢٠ کَیْبَتَ su ila, sa’ala saala;
su’ila كَیْبَة٢٠ کَیْبَتَ su ila, mas’alaTuN كَیْبَة٢٠ کَیْبَتَ mas alatun,
saw’aTuN كَیْبَة٢٠ کَیْبَتَ sawatun, ha.tI’aTuN حَاطِیَتَتَنَّ ha’tatun.

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 \oldhamza (and off
again by \newhamza):
as’ilaTuN كَیْبَة١٠ کَیْبَتَ as ilaTaN کَیْبَتَ ۱٠ as ilatun,
su’ila كَیْبَة٢٠ کَیْبَتَ su ila, sa’ala saala;
su’ila كَیْبَة٢٠ کَیْبَتَ su ila, mas’alaTuN كَیْبَة٢٠ کَیْبَتَ mas alatun,
saw’aTuN كَیْبَة٢٠ کَیْبَتَ sawatun, ha.tI’aTuN حَاطِیَتَتَنَّ ha’tatun.

Madda in the context ’a is generated automatically:
‘AkiluN أَکِیَلَ un akilun, qur’AnuN قُرْآنَ qur’ânun, ra’Ahu رَأْءَحَu ra’ahu.
To reproduce the historic writing correctly, it can also be explicitly indi-
cated by ‘A, ‘I, ‘U in other contexts:
a.sdqA’uh_u أَصْدَقَاوُاْ asdqâ’uè;
ya’g’I’u بَیْنَ الْبَیْنِ ya’g’u, s’U’ila سَوْیَلَ sâ’ila.
CHAPTER 4. INPUT ENCODING CONVENTIONS

\[ \text{Şadda:} \] A double consonant must be written twice, even if it is coded by more than one character:

\[
\text{nazzala} \quad \text{nazzala}, \quad \text{ba}-\text{s}-\text{sAruN} \quad \text{başşarun}, \quad \text{nawwara} \quad \text{nawwara},
\]

\[
\text{sayyiduN} \quad \text{sayyidun}, \quad \text{sa}-\text{i\text{AluN}} \quad \text{sarâlun},
\]

\[
\text{.sabiyuN} \quad \text{.sabyyun}, \quad \text{‘aduwwuN} \quad \text{‘aduwwun}.
\]

Instead of \text{iyy}, \text{uww one can also write} \text{Iy, Uw}:

\[
\text{.sabIyuN} \quad \text{.s.abyun}, \quad \text{‘adUwuN} \quad \text{‘aduwun}.
\]

\[ \text{Assimilation:} \] 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 \text{sukun} are adjusted accordingly:

\[
\text{‘al-ddAru} \quad \text{‘al-ddaru,} \quad \text{‘al-rra}^\text{gulu} \quad \text{‘ar-ra}^\text{gulu},
\]

\[
\text{‘al-ssanaTu} \quad \text{‘al-ssanatu,} \quad \text{‘al-nnAru} \quad \text{‘an-nâru;}
\]

\[
\text{‘al-}^\text{gAru} \quad \text{‘al-gâru,} \quad \text{‘al-bAbu} \quad \text{‘al-bâbu;}
\]

\[
\text{‘al-llaylaTu} \quad \text{‘al-laylatu,} \quad \text{‘al-llisAnu} \quad \text{‘al-lisanu},
\]

\[
\text{‘al-ll_ahu} \quad \text{‘al-lahu}.
\]

The article may also be written in the assimilated form, with identical result:

\[
\text{‘ad-dAru} \quad \text{‘ad-daru,} \quad \text{‘ar-ra}^\text{gulu} \quad \text{‘ar-ra}^\text{gulu},
\]

\[
\text{‘as-sanaTu} \quad \text{‘as-sanatu,} \quad \text{‘an-nnAru} \quad \text{‘an-nâru.}
\]

In some special cases the literal spelling must be used:

\[
\text{‘alla_dI} \quad \text{‘alla_dIna} \quad \text{‘alla_dIna,} \quad \text{‘allatI} \quad \text{‘allatI;}
\]

however:

\[
\text{‘al-lla_dAni} \quad \text{‘al-lad ani,} \quad \text{‘al-llatAni} \quad \text{‘al-latani,}
\]

\[
\text{‘al-llawAtI} \quad \text{‘al-lawat}.
\]

\[ \text{Wasl}a: \] an auxiliary vowel at the beginning of a word is always written, but in the middle of a sentence generally without \text{hamza}. If a vowel precedes the word, the auxiliary vowel will be omitted in the transcription, and the \text{wasl}a sign will be used in the spelling:

\[
\text{wa-ismuhu} \quad \text{wa-‘smuhu}, \quad \text{f--a-in.sarafa} \quad \text{fa-‘nsarafa}.
\]

\[ \text{In vowelized writing, it may sometimes be advisable to introduce a ka\text{ā}da to prevent the vowel marks from bumping into each other.} \]
This also works across word boundaries:

*yA ibnI*  
汉语 'bni, h_a_dA ibnI_u  
*qAla u_hru*g*  
قَالَ أَخْرَجَ qāla َ'hrūğ.

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

*qad-i in.sarafa*  
قد أَنْصَرَ qad-i 'nsarafa,

*ra'aw-u al-bAba*  
رَأِىَّ أَبْنَ يَافَرَ u 'l-bāba,

*min-i ibnih_i*  
منِّ أَبْنَهُ min-i 'bnihī.

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

*’al-i-ismu*  
‘اَلْيِسْمُ ‘al-i-’smu, *’al-i-i'stirA'u*  
‘اَلْیْسْتِرَأَّوْ ‘al-i-'stirāw,

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

*ra^guluN-i ibnIatuh_u *  
راُعُلُوْنْ-ْيِ أَبْنَتُهُ ُثْرَيْنُ raqulun-i 'bnatuhu  
*mu.hammaduN-i al-qura^sIyu*  
ْمُحَمَّدُ الْقُرْآنُ muhammadun-i 'l-qurasyu.

The particles *li-* and *la-* must be combined with the article except before *lām*:

*lil-rra^guli*  
لْيُرْجاَلُ لْیِرُجاَلَ lir-raquli, *lal-ma^gdu*  
لْيُلْمَغْدُ lal-maģdu;

however:

*li-llaylaTi*  
ْلِیَلَّاْتِ لی-للَّاَتِ, *li-ll_ahi*  
ْلِیَلَّاْتِ لی-للَاَتِ, *li-llahi*  
ْلِیَلَّاْتِ لی-للَاَتِ.

The Name of God is written with a special ligature if it is recognized from the input sequence *ll_ah*:

*'al-ll_ahu*  
ْلَاْلَّاْحُ ‘al-lahu, *ta-al-ll_ahi*  
ْلَاْلَّاْحُ 'ta-'l-lahu.

Increased spacing (*Tatwil*) between adjoining characters may be produced by a double hyphen --;

*qabila*  
قَبِيلَ qabil, qa--bi--la  
قُبِيلَ qabil, q--ab--ila qabil.

This feature should be used with discretion; automatic spreading usually leads to a better result.

Ties between words are indicated by a single hyphen:

*bi-baladiN*  
بَيْنَ الْبَلَدِ bi-baladin, *ta-al-ll_ahi*  
ْبَيْنَ الْبَلَدِ 'ta-'l-lahu,

*sa-ya'tI*  
سَابِقٌ sa-yātī, *li-yafra.ha*  
لِیَفْرَحُ li-yafrahā.
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>h</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>-s</td>
<td>-h</td>
<td>-d</td>
</tr>
<tr>
<td>-S</td>
<td>-H</td>
<td>-D</td>
</tr>
</tbody>
</table>

Digit sequences are written in the natural order:

1234567890 1234567890

Hyphen and comma as a decimal separator do not terminate the number:

123-456,789 123-456,789

Ligatures are generated automatically; they can be suppressed by \lvert:

'ال-'islAmu

Abbreviations and emphasis are indicated by \emph or \emphasize:

\emph {\`alayhi as-salAmu} \emphasize {\s.lm}
4.3. ALTERNATE INPUT ENCODINGS

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 re-coding existing machine-readable text files that have been encoded according to other standards. Thus, some alternate reading modules have been written and a general code switching procedure has been provided.

An alternate reading module, e.g. \texttt{asmo449.sty} for the ASMO 449 code, is installed by \texttt{\usepackage{asmo449}} or by \texttt{\input asmo449.sty}. Afterwards, \texttt{\setcode{asmo449}} or \texttt{\setcode{iso9036}} activates the module. All following Arabic text will be considered to be coded according to the ASMO 449 standard.

### 4.3.1 ASMO 449 = ISO 9036

ASMO 449 (see Table 4.4) is a 7-bit code, differing from ASCII (ISO 646) mainly by replacing the Roman 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. ASMO 449 is supported by Arabic MS-DOS.

The file \texttt{asmo449.sty} contains a reading module for the ASMO 449 code (identical to ISO 9036). It is installed by the \LaTeX{} command \texttt{\usepackage{asmo449}} or by \texttt{\input asmo449.sty}. The module is activated by \texttt{\setcode{asmo449}} or \texttt{\setcode{iso9036}}; all following Arabic text will be considered to be coded according to the ASMO 449 standard.
### Table 4.4: ASMO 449 code table

<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>0</td>
<td>NUL</td>
<td>DLE</td>
<td>SP</td>
<td>0</td>
<td>ฉ</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>1</td>
<td>SOH</td>
<td>DC1</td>
<td>!</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>2</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>3</td>
<td>ETX</td>
<td>DC3</td>
<td>#</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>4</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>5</td>
<td>ENQ</td>
<td>NAK</td>
<td>%</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>6</td>
<td>ACK</td>
<td>SYN</td>
<td>&amp;</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>7</td>
<td>BEL</td>
<td>ETB</td>
<td>'</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>8</td>
<td>BS</td>
<td>CAN</td>
<td>)</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>9</td>
<td>HT</td>
<td>EM</td>
<td>(</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>10</td>
<td>LF</td>
<td>SUB</td>
<td>*</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>11</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>12</td>
<td>FF</td>
<td>IS4</td>
<td>,</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>13</td>
<td>CR</td>
<td>IS3</td>
<td>=</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>14</td>
<td>SO</td>
<td>IS2</td>
<td>&lt;</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
<tr>
<td>15</td>
<td>SI</td>
<td>IS1</td>
<td>/</td>
<td>ค</td>
<td>ฟ</td>
<td>ะ</td>
<td>ะ</td>
<td>ะ</td>
</tr>
</tbody>
</table>
4.3. ALTERNATE INPUT ENCODINGS

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 'a' and 'alif maṣṣūra.

A minimal driver file for processing existing ASMO 449 text, e.g. in a file `asmotext.dat`, could look as follows:

\documentclass{article}
\usepackage{arabtex}
\usepackage{asmo449}
\begin{document}
\setcode{asmo449}
\begin{RLtext}
\input asmotext.dat
\end{RLtext}
\end{document}

4.3.2 ASMO 449E = ISO 8859-6, MacOS Arabic

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

ISO 8859-6 (see Table 4.5) 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.

This code, plus additional Persian, Urdu, graphic, and control characters, is supported by MacOS in Arabic mode.

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 could 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.
### Table 4.5: ISO 8859-6 and MacOS Arabic code table

<table>
<thead>
<tr>
<th></th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>NUL</td>
<td>DLE</td>
<td>SP</td>
<td>0</td>
<td>@</td>
<td>P</td>
<td>`</td>
<td>p</td>
<td>ASP</td>
<td>0</td>
<td>*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>SOH</td>
<td>DC1</td>
<td>!</td>
<td>1</td>
<td>A</td>
<td>Q</td>
<td>a</td>
<td>q</td>
<td>NSP</td>
<td>!</td>
<td>ً</td>
<td>ً</td>
<td>ً</td>
<td>ً</td>
<td>ً</td>
<td>ً</td>
</tr>
<tr>
<td>02</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>2</td>
<td>B</td>
<td>R</td>
<td>b</td>
<td>r</td>
<td>٢</td>
<td>&quot;</td>
<td>٢</td>
<td>٢</td>
<td>٢</td>
<td>٢</td>
<td>٢</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>ETX</td>
<td>DC3</td>
<td>#</td>
<td>3</td>
<td>C</td>
<td>S</td>
<td>s</td>
<td>٣</td>
<td>٥</td>
<td>#</td>
<td>٥</td>
<td>٥</td>
<td>٥</td>
<td>٥</td>
<td>٥</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>4</td>
<td>D</td>
<td>T</td>
<td>d</td>
<td>t</td>
<td>٤</td>
<td>$</td>
<td>٤</td>
<td>٤</td>
<td>٤</td>
<td>٤</td>
<td>٤</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>ENQ</td>
<td>SYN</td>
<td>%</td>
<td>5</td>
<td>E</td>
<td>U</td>
<td>e</td>
<td>u</td>
<td>٥</td>
<td>%</td>
<td>٥</td>
<td>٥</td>
<td>٥</td>
<td>٥</td>
<td>٥</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>ACK</td>
<td>&amp;</td>
<td>6</td>
<td>F</td>
<td>V</td>
<td>F</td>
<td>v</td>
<td>&amp;</td>
<td>٦</td>
<td>٦</td>
<td>٦</td>
<td>٦</td>
<td>٦</td>
<td>٦</td>
<td>٦</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>BEL</td>
<td>ETB</td>
<td>'</td>
<td>7</td>
<td>G</td>
<td>W</td>
<td>g</td>
<td>w</td>
<td>٧</td>
<td>'</td>
<td>٧</td>
<td>٧</td>
<td>٧</td>
<td>٧</td>
<td>٧</td>
<td>٧</td>
</tr>
<tr>
<td>08</td>
<td>BS</td>
<td>CAN</td>
<td>(</td>
<td>8</td>
<td>H</td>
<td>X</td>
<td>h</td>
<td>x</td>
<td>٨</td>
<td>(</td>
<td>٨</td>
<td>٨</td>
<td>٨</td>
<td>٨</td>
<td>٨</td>
<td>٨</td>
</tr>
<tr>
<td>09</td>
<td>HT</td>
<td>EM</td>
<td>)</td>
<td>9</td>
<td>I</td>
<td>Y</td>
<td>i</td>
<td>y</td>
<td>٩</td>
<td>)</td>
<td>٩</td>
<td>٩</td>
<td>٩</td>
<td>٩</td>
<td>٩</td>
<td>٩</td>
</tr>
<tr>
<td>10</td>
<td>LF</td>
<td>SUB</td>
<td>:</td>
<td>J</td>
<td>Z</td>
<td>j</td>
<td>z</td>
<td>:</td>
<td>ز</td>
<td>:</td>
<td>ز</td>
<td>:</td>
<td>ز</td>
<td>:</td>
<td>ز</td>
<td>:</td>
</tr>
<tr>
<td>11</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>K</td>
<td>{</td>
<td>k</td>
<td>{</td>
<td>ن</td>
<td>+</td>
<td>٣</td>
<td>+</td>
<td>٣</td>
<td>+</td>
<td>٣</td>
<td>+</td>
<td>٣</td>
</tr>
<tr>
<td>12</td>
<td>FF</td>
<td>IS4</td>
<td>,</td>
<td>L</td>
<td>\</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CR</td>
<td>IS3</td>
<td>-</td>
<td>M</td>
<td>}</td>
<td>m</td>
<td>}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>SO</td>
<td>IS2</td>
<td>&gt;</td>
<td>N</td>
<td>&quot;</td>
<td>n</td>
<td>&quot;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>15</td>
<td>SI</td>
<td>IS1</td>
<td>/</td>
<td>?</td>
<td>O</td>
<td>o</td>
<td>DEL</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>
### 4.3. ALTERNATE INPUT ENCODINGS

<table>
<thead>
<tr>
<th></th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>NUL</td>
<td>DLE</td>
<td>SP</td>
<td>0</td>
<td>@</td>
<td>P</td>
<td>&quot;</td>
<td>p</td>
<td>گ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>SOH</td>
<td>DC1</td>
<td>!</td>
<td>1</td>
<td>A</td>
<td>Q</td>
<td>a</td>
<td>q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>2</td>
<td>B</td>
<td>R</td>
<td>b</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>ETX</td>
<td>DC3</td>
<td>#</td>
<td>3</td>
<td>C</td>
<td>S</td>
<td>c</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>4</td>
<td>D</td>
<td>T</td>
<td>d</td>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>ENQ</td>
<td>NAK</td>
<td>%</td>
<td>5</td>
<td>E</td>
<td>U</td>
<td>e</td>
<td>u</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>ACK</td>
<td>SYN</td>
<td>&amp;</td>
<td>6</td>
<td>F</td>
<td>V</td>
<td>f</td>
<td>v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>BEL</td>
<td>ETB</td>
<td>’</td>
<td>7</td>
<td>G</td>
<td>W</td>
<td>g</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>BS</td>
<td>CAN</td>
<td>(</td>
<td>8</td>
<td>H</td>
<td>X</td>
<td>h</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09</td>
<td>HT</td>
<td>EM</td>
<td>)</td>
<td>9</td>
<td>I</td>
<td>Y</td>
<td>i</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LF</td>
<td>SUB</td>
<td>*</td>
<td>:</td>
<td>J</td>
<td>Z</td>
<td>j</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>;</td>
<td>K</td>
<td>[</td>
<td>k</td>
<td>{</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>FF</td>
<td>IS4</td>
<td>,</td>
<td>&lt;</td>
<td>L</td>
<td>\</td>
<td>l</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CR</td>
<td>IS3</td>
<td>-</td>
<td>=</td>
<td>M</td>
<td>]</td>
<td>m</td>
<td>}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>SO</td>
<td>IS2</td>
<td>&gt;</td>
<td>&gt;</td>
<td>N</td>
<td>”</td>
<td>n</td>
<td>”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>SI</td>
<td>IS1</td>
<td>/</td>
<td>?</td>
<td>O</td>
<td>_</td>
<td>o</td>
<td>DEL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: Windows CP1256 code table
CHAPTER 4. INPUT ENCODING CONVENTIONS

4.3.3 CP 1256 = Arabic Windows Encoding

The file `cp1256.sty` contains a reading module for the Arabic part of the code page CP1256 used within *MS Arabic Windows*. It is installed by the \LaTeX command \texttt{\usepackage{cp1256}} or by \texttt{\input cp1256.sty}. The module is activated by \texttt{\setcode{arabwin}} or \texttt{\setcode{cp1256}}; all following Arabic text will be considered to be coded according to the *MS Arabic Windows* standard. The Arab\TeX notation may be reactivated by \texttt{\setcode{arabtex}}.

The code page 1256 used in *MS Arabic Windows* (see Table 4.6) is an 8-bit code closely related to 7-bit ASCII; whereas the lower 128 positions are identical to ASCII (ISO 646), some of the upper 128 positions contain the Arabic characters plus additional Persian, Urdu, graphic, and control characters. The notes on vowelization and transliteration of ASMO 449 apply also.

4.3.4 UNICODE Arabic

The file `utf8.sty` contains a reading module for the Arabic and the Hebrew segment of UNICODE in UTF-8 encoding. It is installed by the \LaTeX command \texttt{\usepackage{utf8}} or by \texttt{\input utf8.sty}.

UTF-8 (UNICODE Transmission Format, see tables 4.7, 4.8, and 7.4) is a multi-byte encoding which, for Arabic and Hebrew, uses two bytes per character whereas ASCII characters use a single byte. Far-eastern languages are encoded in three bytes per character. This is in contrast to UNICODE itself which always uses two bytes per character.

The module is activated by \texttt{\setcode{utf8}}; all following Arabic and Hebrew text will be considered to be coded according to the UTF-8 encoding standard. To use the correct font, select the appropriate language. The Arab\TeX notation may be reactivated by \texttt{\setcode{arabtex}}.

4.3.5 ISIRI 3342

The file `isiri.sty` contains a reading module for the ISIRI 3342 Persian Standard Code. It is installed by the \LaTeX command \texttt{\usepackage{isiri}} or by \texttt{\input isiri.sty}. The module is activated by \texttt{\setcode{isiri}}; all following Arabic text will be considered to be coded according to the ISIRI 3342 standard. The Arab\TeX notation may be reactivated by \texttt{\setcode{arabtex}}.

The ISIRI 3342 code (see Table 4.9) is an 8-bit code closely related to 7-bit ASCII; whereas the lower 128 positions are identical to ASCII (ISO 646), some of the upper 128 positions contain the Arabic/Persian characters plus additional graphic and control characters. The notes on vowelization and transliteration of ASMO 449 apply also.
### Table 4.7: UNICODE Arabic, Part 1

<table>
<thead>
<tr>
<th>Code</th>
<th>060</th>
<th>061</th>
<th>062</th>
<th>063</th>
<th>064</th>
<th>065</th>
<th>066</th>
<th>067</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D880</td>
<td>D890</td>
<td>D8A0</td>
<td>D8B0</td>
<td>D980</td>
<td>D990</td>
<td>D9A0</td>
<td>D9B0</td>
</tr>
<tr>
<td>0</td>
<td>ꏊ</td>
<td>ꏌ</td>
<td>ꏍ</td>
<td>ꏎ</td>
<td>ꏐ</td>
<td>ꏑ</td>
<td>ꏒ</td>
<td>ꏓ</td>
</tr>
<tr>
<td>1</td>
<td>ꏔ</td>
<td>ꏕ</td>
<td>ꏖ</td>
<td>ꏗ</td>
<td>ꏘ</td>
<td>ꏙ</td>
<td>ꏚ</td>
<td>ꏛ</td>
</tr>
<tr>
<td>2</td>
<td>ꏜ</td>
<td>ꏝ</td>
<td>ꏞ</td>
<td>ꏟ</td>
<td>ꏠ</td>
<td>ꏡ</td>
<td>ꏢ</td>
<td>ꏣ</td>
</tr>
<tr>
<td>3</td>
<td>ꏤ</td>
<td>ꏥ</td>
<td>ꏦ</td>
<td>ꏧ</td>
<td>ꏨ</td>
<td>ꏩ</td>
<td>ꏪ</td>
<td>ꏫ</td>
</tr>
<tr>
<td>4</td>
<td>ꏬ</td>
<td>ꏭ</td>
<td>ꏮ</td>
<td>ꏯ</td>
<td>ꏰ</td>
<td>ꏱ</td>
<td>ꏲ</td>
<td>ꏳ</td>
</tr>
<tr>
<td>5</td>
<td>ꏴ</td>
<td>ꏵ</td>
<td>ꏶ</td>
<td>ꏷ</td>
<td>ꏸ</td>
<td>ꏹ</td>
<td>ꏺ</td>
<td>ꏻ</td>
</tr>
<tr>
<td>6</td>
<td>ꏼ</td>
<td>ꏽ</td>
<td>ꏾ</td>
<td>ꏿ</td>
<td>ꇀ</td>
<td>ꇁ</td>
<td>ꇂ</td>
<td>ꇃ</td>
</tr>
<tr>
<td>7</td>
<td>ꇄ</td>
<td>ꇅ</td>
<td>ꇆ</td>
<td>ꇇ</td>
<td>ꇈ</td>
<td>ꇉ</td>
<td>ꇊ</td>
<td>ꇋ</td>
</tr>
<tr>
<td>8</td>
<td>ꇌ</td>
<td>ꇍ</td>
<td>ꇎ</td>
<td>ꇏ</td>
<td>ꇐ</td>
<td>ꇑ</td>
<td>ꇒ</td>
<td>ꇓ</td>
</tr>
<tr>
<td>9</td>
<td>ꇔ</td>
<td>ꇕ</td>
<td>ꇖ</td>
<td>ꇗ</td>
<td>ꇘ</td>
<td>ꇙ</td>
<td>ꇚ</td>
<td>ꇛ</td>
</tr>
<tr>
<td>A</td>
<td>ꇜ</td>
<td>ꇝ</td>
<td>ꇞ</td>
<td>ꇟ</td>
<td>ꇠ</td>
<td>ꇡ</td>
<td>ꇢ</td>
<td>ꇣ</td>
</tr>
<tr>
<td>B</td>
<td>ꇤ</td>
<td>ꇥ</td>
<td>ꇦ</td>
<td>ꇧ</td>
<td>ꇨ</td>
<td>ꇩ</td>
<td>ꇪ</td>
<td>ꇫ</td>
</tr>
<tr>
<td>C</td>
<td>ꇬ</td>
<td>ꇭ</td>
<td>ꇮ</td>
<td>ꇯ</td>
<td>ꇰ</td>
<td>ꇱ</td>
<td>ꇲ</td>
<td>ꇳ</td>
</tr>
<tr>
<td>D</td>
<td>ꇴ</td>
<td>ꇵ</td>
<td>ꇶ</td>
<td>ꇷ</td>
<td>ꇸ</td>
<td>ꇹ</td>
<td>ꇺ</td>
<td>ꇻ</td>
</tr>
<tr>
<td>E</td>
<td>ꇼ</td>
<td>ꇽ</td>
<td>ꇾ</td>
<td>ꇿ</td>
<td>ꀀ</td>
<td>ꀁ</td>
<td>ꀂ</td>
<td>ꀃ</td>
</tr>
<tr>
<td>F</td>
<td>ꀄ</td>
<td>ꀅ</td>
<td>ꀆ</td>
<td>ꀇ</td>
<td>ꀈ</td>
<td>ꀉ</td>
<td>ꀊ</td>
<td>ꀋ</td>
</tr>
</tbody>
</table>
### Table 4.8: UNICODE Arabic, Part 2

<table>
<thead>
<tr>
<th>068</th>
<th>069</th>
<th>06A</th>
<th>06B</th>
<th>06C</th>
<th>06D</th>
<th>06E</th>
<th>06F</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA80</td>
<td>DA90</td>
<td>DAA0</td>
<td>DAB0</td>
<td>DB80</td>
<td>DB90</td>
<td>DBA0</td>
<td>DBB0</td>
</tr>
<tr>
<td>0</td>
<td>﷼</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>ﷷ</td>
<td>ﷶ</td>
<td>ﷵ</td>
</tr>
<tr>
<td>1</td>
<td>ﷴ</td>
<td>ﷳ</td>
<td>ﷲ</td>
<td>ﷱ</td>
<td>ﷰ</td>
<td>﷯</td>
<td>ﷸ</td>
</tr>
<tr>
<td>2</td>
<td>ﷷ</td>
<td>ﷶ</td>
<td>ﷵ</td>
<td>ﷴ</td>
<td>ﷳ</td>
<td>ﷲ</td>
<td>ﷱ</td>
</tr>
<tr>
<td>3</td>
<td>ﷸ</td>
<td>ﷷ</td>
<td>ﷶ</td>
<td>ﷵ</td>
<td>ﷴ</td>
<td>ﷳ</td>
<td>ﷲ</td>
</tr>
<tr>
<td>4</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>ﷷ</td>
<td>ﷶ</td>
<td>ﷵ</td>
<td>ﷴ</td>
<td>ﷳ</td>
</tr>
<tr>
<td>5</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>ﷷ</td>
<td>ﷶ</td>
<td>ﷵ</td>
<td>ﷴ</td>
</tr>
<tr>
<td>6</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>ﷷ</td>
<td>ﷶ</td>
<td>ﷵ</td>
</tr>
<tr>
<td>7</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>ﷷ</td>
<td>ﷶ</td>
</tr>
<tr>
<td>8</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>ﷷ</td>
</tr>
<tr>
<td>9</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
</tr>
<tr>
<td>A</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
</tr>
<tr>
<td>B</td>
<td>ﷹ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
</tr>
<tr>
<td>C</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
</tr>
<tr>
<td>D</td>
<td>ﷹ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
</tr>
<tr>
<td>E</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
<td>ﷹ</td>
</tr>
<tr>
<td>F</td>
<td>ﷹ</td>
<td>ﷺ</td>
<td>ﷹ</td>
<td>ﷸ</td>
<td>﷼</td>
<td>ﷻ</td>
<td>ﷺ</td>
</tr>
</tbody>
</table>

*Note: The table shows the Unicode encoding for Arabic characters.*
### 4.3. Alternate Input Encodings

<table>
<thead>
<tr>
<th>Code</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>NUL, DEL, SP</td>
</tr>
<tr>
<td>01</td>
<td>SOH, DC1, !</td>
</tr>
<tr>
<td>02</td>
<td>STX, DC2, &quot;</td>
</tr>
<tr>
<td>03</td>
<td>ETX, DC3, #</td>
</tr>
<tr>
<td>04</td>
<td>EOT, DC4, $</td>
</tr>
<tr>
<td>05</td>
<td>ENQ, NAK, %</td>
</tr>
<tr>
<td>06</td>
<td>ACK, SYN, &amp;</td>
</tr>
<tr>
<td>07</td>
<td>BEL, ETC, '</td>
</tr>
<tr>
<td>08</td>
<td>BS, CAN, (</td>
</tr>
<tr>
<td>09</td>
<td>HT, EM, )</td>
</tr>
<tr>
<td>10</td>
<td>LF, SUB, :</td>
</tr>
<tr>
<td>11</td>
<td>VT, ESC, +</td>
</tr>
<tr>
<td>12</td>
<td>FF, IS4, ,&lt;</td>
</tr>
<tr>
<td>13</td>
<td>CR, IS3, -</td>
</tr>
<tr>
<td>14</td>
<td>SO, IS2, .&gt;</td>
</tr>
<tr>
<td>15</td>
<td>SI, IS1, /?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>A, Q, a, q</td>
</tr>
<tr>
<td>02</td>
<td>B, R, b, r</td>
</tr>
<tr>
<td>03</td>
<td>C, S, c, s</td>
</tr>
<tr>
<td>04</td>
<td>D, T, d, t</td>
</tr>
<tr>
<td>05</td>
<td>E, U, e, u</td>
</tr>
<tr>
<td>06</td>
<td>F, V, f, v</td>
</tr>
<tr>
<td>07</td>
<td>G, W, g, w</td>
</tr>
<tr>
<td>08</td>
<td>H, X, h, x</td>
</tr>
<tr>
<td>09</td>
<td>I, Y, i, y</td>
</tr>
<tr>
<td>10</td>
<td>J, Z, j, z</td>
</tr>
<tr>
<td>11</td>
<td>K, [, k, {</td>
</tr>
<tr>
<td>12</td>
<td>L, , l, l</td>
</tr>
<tr>
<td>13</td>
<td>M, ]/}, m</td>
</tr>
<tr>
<td>14</td>
<td>N, ^, n, &quot;</td>
</tr>
<tr>
<td>15</td>
<td>O, o, DEL, /</td>
</tr>
</tbody>
</table>

Table 4.9: ISIRI 3342 code table
4.3.6 Buckwalter Transliteration

The file buck.sty contains a reading module for the transliteration devised by the lexicographer Tim Buckwalter. It is installed by the \LaTeX command \texttt{\usepackage{buck}} or by \texttt{\input buck.sty}.

The Buckwalter System (see Table 4.10) is a 7-bit encoding, reusing the Roman letters and the special characters for encoding Arabic letters and diacritical marks; the Arabic digits share their positions with the ASCII digits.

Text coded in ASMO449 or ISO-8859-6 can be converted trivially and unambiguously into the Buckwalter transliteration; and except for the rarely used dagger ‘alif and wašla-on-‘alif, which are missing from ASMO449 and ISO-8859-6, the reverse mapping is equally straightforward. Buckwalter transliteration can be trivially converted to and from UNICODE, which contains the extra characters.

The Buckwalter System is a true or strict orthographical transliteration, as opposed to the majority of “transcription” romanizations of Arabic that are intended to convey phonological or morphophonological information; it has the advantages of being complete, typable and displayable on common ASCII terminals and printers. Because the Roman character equivalents are chosen to be reasonably mnemonic, Arabists can learn to read and write Buckwalter texts with little difficulty.

The Buckwalter convention is activated by \texttt{\setcode{buck}}. Since it preempts the curly braces \{ and \}, as well as the angle braces \< and \>, for encoding purposes, short Arabic insertions must be denoted specially, as arguments of the command TB( ) (with round parentheses!) On the other hand, Arabic environments work as expected.

![Table 4.10: Buckwalter transliteration](image)
Chapter 5

Transliteration

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{\transtrue} and may be switched off again by \texttt{\transfalse}.

If only the transliteration is wanted, you can deactivate the Arabic writing by \texttt{\arabfalse}; it can be reactivated by \texttt{\arabtrue}. If both modes are active their output will be interleaved line by line. The font used for the transliteration is normally italic, it can be changed globally by \texttt{\settransfont{font}}.

5.1 ZDMG transliteration style

The “ZDMG transliteration” is in fact a family of closely related, but slightly different, transliteration conventions for several languages using the Perso-Arabic script. Therefore for producing it correctly, the appropriate language mode must have been selected.

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 wasla appears in the Arabic writing). If that is not wanted, start with hamza.
- a silent 'alif or 'alif maqsura after <q> (tanwin) and <u> is omitted in the transliteration. The same happens after wāw if it is written as a capital <W>.
- To correctly reproduce some historical writings, a silent long vowel after <_a> is omitted in the transliteration. For examples, see section 4.1.5.

\footnote{The former option "atrans" is no more necessary.}
CHAPTER 5. TRANSLITERATION

For Persian texts, the Izafet connection is handled specially, and a final silent \( h \) will be omitted in the transliteration.

5.2 Other transliteration styles

Since there is no general agreement on transcriptions, a number of variants have been provided:

- \texttt{\settrans{standard}} or \texttt{\settrans{zdmg}} is the default ZDMG mode.
- \texttt{\settrans{english}} will switch to the style of the Encyclopedia of Islam which is close to the conventions of the Library of Congress.
- \texttt{\settrans{iranica}} will produce the style used in the Encyclopedia Iranica.
- \texttt{\settrans{farsi}} produces a variant of the style used in the Encyclopedia Iranica.
- \texttt{\settrans{lazard}} switches to the conventions of Gilbert Lazard: “La langue des plus anciens documents de la prose persane”.
- \texttt{\settrans{urdu}} switches to the conventions for Urdu used in the ALA-LC tables.
- \texttt{\settrans{kashmiri}} switches to the conventions for Kashmiri used in the ALA-LC tables; this mode is also chosen automatically by \texttt{\setkashmiri}.
- \texttt{\settrans{turk}} will produce a style similar to modern Turkish; it only makes sense for Ottoman texts.
- Transcription conventions for Hebrew are given in section 7.7.

The transliteration mode may be switched globally at any time. If the input text is not fully vowelized, the transcription cannot be expected to be correct.

5.3 Capitalization

If transcription output is used as part of a Roman text, it may be desirable to have some words start with a capital letter. This can be achieved by prefixing the command \texttt{\cap} to the word in question. If the first letter is \textit{hamza} or ‘\textit{ayn}, the next letter will be capitalized. This feature may also be used after the article or a prefix, and even in other arbitrary positions; \texttt{\cap} will only influence the following letter. The Arabic writing is not affected.
Chapter 6

Support for other languages using Perso-Arabic script

ArabTEX 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 using the Arabic alphabet or some extension of it.

In order to switch to the conventions for one of these languages, say \texttt{\textbackslash setfarsi}, \texttt{\textbackslash seturdu}, \texttt{\textbackslash setpashto}, \texttt{\textbackslash setmaghribi}, etc.; \texttt{\textbackslash setverb} will switch off any language specific processing. \texttt{\textbackslash setarab} can be used to switch back to the Arabic conventions.

After selecting the language, < and > are active as delimiters for quotations; \texttt{\textbackslash setnone} will return < and > to their normal TeX meaning. Quotations still can be bracketed by \texttt{\textbackslash < and \textbackslash >} or by using \texttt{\textbackslash RL{ }}.

This part of ArabTEX relies heavily on contributions from the user community; we want to especially mention Ivan Derzhanski who completely reimplemented the routines for processing Persian, and Anshuman Pandey for the Urdu part. As we extensively modified these contributions again while integrating the system, we are solely responsible for any remaining, or newly introduced, errors.

6.1 Persian (Farsi, Dari)

The Persian mode is activated by \texttt{\textbackslash setfarsi}.

- All characters needed for writing Farsi are available by default. The short vowels \texttt{\textbackslash e} and \texttt{\textbackslash o} are mapped to \texttt{\textbackslash i} and \texttt{\textbackslash u}, the long vowels \texttt{\textbackslash E} and \texttt{\textbackslash O} to \texttt{\textbackslash I} and \texttt{\textbackslash U} without a vowel indicator. \texttt{\textbackslash H} denotes final silent \texttt{\textasciitilde h}{	extbackslash a}. This \texttt{\textasciitilde h}{	extbackslash a} receives no \texttt{\textasciitilde sukun} even in fully vowelized mode.
• For *fatḥa* or *kasra* followed by a final silent ḥāʾ you can also write `<,a>` or `<,e>` in place of `<aH>` and `<eH>` (deprecated).

• The *izāfet* connection may always be written `<-i>` or `<-e` (with hyphen); then ArabTEX tries to determine the correct spelling from the context. Likewise the *yaʿ-i-wahdat* can always be written `<-i>` or `<-e`.

• The present tense forms of the copula are coded `<-am>`, `<-I>`, `<-ast>`, `<-Im>`, `<-Id>`, `<and>`. In the output they are written as separate words after a little space.

• The final *yaʿ* carries no dots. Farsi uses the Nasta`liq font when available, otherwise Naskh.

6.1.1 Coding examples for Persian

The short vowels *a* (*ā*), *e* (*ī*), *o* (*ū*) are denoted by the lowercase letters *a*, *e* or *i*, *o* or *u*:

*bar* ْبَر, *beh* ِبَه, *bon* بُون.

The long vowels *a* (*ā*), *i* (*ī*), *u* (*ū*), *ō* are denoted by the capital letters *A*, *I* or *E*, *U* or *O*.


Note that *I* yields a *ya-ye maʿruf* (with *zīr*), whilst *E* yields a *ya-ye maṣjīl* (without *zīr*). Similarly, *U* yields a *waw-e maʿruf* (with *pis*), whilst *O* yields a *waw-e maṣjūl* (without *pis*):


The diphthongs *ei* and *ōu* are written *ay* and *aw*:

*pay* ْپاَي, *nāw* نَوُن.

Intervocalic *ḥaemzeh* is written `ˈ`:

*ma`ūn* مَيْعُون, *miyā`a* مِيْيَأ, *miγu`a* مِيْγِعَأ, *ta`awān* تَأْوَان, *za`an`a* زَأْنَا, *za`an`a* زَأْنَا.

Silent word-final *wāw* is generated by `U` or `O`:

Waw-e ma’dul is written ۸; it is omitted in the transliteration and the preceding xe receives no jazm:

\( \_\\text{hwAb} \) ْب, \( \_\\text{hwI} \) ْی, \( \_\\text{hwod} \) ْد. 

Ha-ye haawwaz-e mażfi is generated by H, or optionally by ,e, ,a or ,A (the latter forms are deprecated). It does not receive a jazm even in fully vocalised mode and is not joined to a following letter:

\( \_\\text{hAneH} \) ْن، \( \_\\text{hwAb} \) ْب, \( \_\\text{hwI}^c \) ْی, \( \_\\text{hwod} \) ْد.

\( \_\\text{hAneHhA} \) ْنْه ْه, ْه. 

Short edafe is written -e or -i:

ketAb-e ْت كتاب أو, rAh-e ْت راه تو tu,
nAmeh-i ْن عمه man,
blIn-e ْن بن اهن mard,
pA-i ْت پا in zan,
bAzU-i ْت بژو in zan.

Long edafe is written -i:

dAr-i ْت دار man, _hU-i ْت hū-tu.

Hāmze as ya-ye waḥdat/nesbat/xetab is likewise written -i:

nAmeh- ْن عمه, sormeh- ْن سره,
gofteH- ْت گفته.

Ye-ye waḥdat is written -I or -E:

ketAb-I ْت کتاب-I, rAh-I ْت راه-I rāh-I, nAmeh-I ْن عمه-I,
dAnA-I ْت دان-I, pArU-I ْت پار-I,
dAnA-I-keH ْت دان-I-keh.

The present tense forms of the verb būdan and the pronominal clitics are written as they are spoken:

rafteH-am ْت رفته am, rafteH-Im ْت رفته-im,
rafteH-I ْت رفته-I, rafteH-Id ْت رفته-id,
rafteH-ast ْت رفته-ast, rafteH-and ْت رفته-and.
CHAPTER 6. SUPPORT FOR OTHER LANGUAGES

mard-Id  مَرْدِیذ mard-id, asb-etAn  اَشْبَان asb-etân;
An-gA-st  اَنْگَا-سْت anqi-st, U-st  اُ-سْت, t-U-st  تُوْسْت tu-st;

The preposition be- can be written with or without a hyphen:
be-man  بينْ, be-man, be-t_U  بينْ be-tu;
be-An  بِینْ be-ân, be-In  بِينْ be-in, beâU  بِيْ beü.

The components of compounds can be separated by \, or ":
. sa. heb\, hâneH  ضَاحِب حَانُه sâhebhâneh,
ta_ht-e="|_hwAb  تَحْت–هَاب taht-e-ḥâb;
pas\, andAz  پَس أَنْذَآر pasandâz, naw"|AmUz  نَوْ آمُوْز nawâmûz,
bI\, hwod  بَیْ حَوْد bihod.

Digit sequences are written in their natural order:
1234567890 ۱۲۳۴۵۶۷۸۹۰ 1234567890

6.2 Maghribi

This works nearly like Arabic, but using a different writing convention. ٍّ’ is written with one dot below the letter, ٍّ with one dot above the normal letter form of ٍ. The three dots of ۝’ are put below the letter.

Switch to this mode by \setmaghribi.

6.3 Urdu

The Urdu mode is activated by \seturdu.

- For Urdu, additional codings are available, see Table 6.1. Some of the given codings also occur in Pashto but with a different meaning, see Section 6.4.
- Even in fully vowelized mode, an aspirated consonant before <h> receives no sukân since the two are technically a single letter.
- Urdu uses the Nasta’liq font when available, otherwise Naskh.
6.3. URDU

6.3.1 Coding examples for Urdu

The short vowels \(a\), \(i\), and \(u\) are encoded by the lowercase letters \(a\), \(i\), and \(u\), and are marked respectively by zabar, zer, and pes:

\[
\begin{align*}
\text{par} & \quad \text{dam} & \quad \text{fır} & \quad \text{din} \\
\text{par} & \quad \text{dam} & \quad \text{fır} & \quad \text{din}
\end{align*}
\]

The long vowels \(a\), \(i\), \(u\), \(e\), and \(o\) are encoded by the capital letters \(A\), \(I\), \(U\), \(E\), and \(O\):

\[
\begin{align*}
\text{Note: 'alif madda is automatically generated for word-initial } a: \ & \quad \text{Ap} & \quad \text{Am} \\
\text{tIn} & \quad \text{la}, \text{rkI} & \quad \text{lärkI} & \quad \text{dúr} & \quad \text{dEr} & \quad \text{dér}, \\
\text{ba}, \text{rE} & \quad \text{baré} & \quad \text{mOr} & \quad \text{môr} \\
\text{tIn} & \quad \text{rItI} & \quad \text{rittI} & \quad \text{mérI} & \quad \text{la}, \text{rkE} & \quad \text{lärkI}
\end{align*}
\]

The diphthongs \(ae\) and \(ao\) are encoded ae, and ao:

\[
\begin{align*}
\text{kaesA} & \quad \text{paodA} & \quad \text{paodá} \\
\text{Note: } U \text{ yields a wāw-e-ma’ruf (with pes), while } O \text{ yields a wāw-e-maghul (without zér).} & \quad \text{pUr} & \quad \text{pOtA} & \quad \text{pótá} & \quad \text{paodA} & \quad \text{paodá}
\end{align*}
\]

Intervocalic \(h\). amza is written ‘:

\[
\begin{align*}
\text{cA’E} & \quad \text{ma’I} & \quad \text{mařI} & \quad \text{kO’I} & \quad \text{kóřI}
\end{align*}
\]

Aspiration is produced by coding a \(h\) after the consonant to be aspirated. Aspiration in Urdu is produced by adding \(dō čašmi he\) after the consonant:

\[
\begin{align*}
\text{khEt} & \quad \text{khet} & \quad \text{ghar} & \quad \text{ghar} & \quad \text{mujhE} & \quad \text{muğhé} & \quad \text{dharm}
\end{align*}
\]

Nasalization is indicated by \(nūn-e-gunnah\) coded as \(n\). Note that the nuqta for \(nūn\) is not written when it is used to represent nasalization:

\[
\begin{align*}
\text{mae.n} & \quad \text{maen} & \quad \text{a,hi.nS} & \quad \text{ahişsá}
\end{align*}
\]

---

2Please contact Anshuman Pandey [apandey@u.washington.edu] with questions or comments regarding this section.
### Table 6.1: The Urdu Alphabet

<table>
<thead>
<tr>
<th>URDU</th>
<th>NAME</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>'ašf</td>
<td>a</td>
</tr>
<tr>
<td>b</td>
<td>be</td>
<td>b</td>
</tr>
<tr>
<td>bh</td>
<td>bhe</td>
<td>bh</td>
</tr>
<tr>
<td>p</td>
<td>pe</td>
<td>p</td>
</tr>
<tr>
<td>ph</td>
<td>phe</td>
<td>ph</td>
</tr>
<tr>
<td>t</td>
<td>te</td>
<td>t</td>
</tr>
<tr>
<td>th</td>
<td>the</td>
<td>th</td>
</tr>
<tr>
<td>l</td>
<td>le</td>
<td>t</td>
</tr>
<tr>
<td>dh</td>
<td>dāl</td>
<td>d</td>
</tr>
<tr>
<td>d</td>
<td>dāl</td>
<td>d</td>
</tr>
<tr>
<td>zh</td>
<td>dhal</td>
<td>dh</td>
</tr>
<tr>
<td>d</td>
<td>dāl</td>
<td>d</td>
</tr>
<tr>
<td>jh</td>
<td>dhal</td>
<td>dh</td>
</tr>
<tr>
<td>z</td>
<td>dāl</td>
<td>d</td>
</tr>
<tr>
<td>r</td>
<td>re</td>
<td>r</td>
</tr>
<tr>
<td>rh</td>
<td>rhe</td>
<td>rh</td>
</tr>
<tr>
<td>š</td>
<td>še</td>
<td>z</td>
</tr>
<tr>
<td>ŝ</td>
<td>še</td>
<td>z</td>
</tr>
<tr>
<td>s</td>
<td>sin</td>
<td>s</td>
</tr>
<tr>
<td>š</td>
<td>šin</td>
<td>ša</td>
</tr>
<tr>
<td>š</td>
<td>šad</td>
<td>ša</td>
</tr>
<tr>
<td>d</td>
<td>dād</td>
<td>d</td>
</tr>
<tr>
<td>ť</td>
<td>ťoi</td>
<td>t</td>
</tr>
<tr>
<td>t</td>
<td>toi</td>
<td>t</td>
</tr>
<tr>
<td>z</td>
<td>zo</td>
<td>z</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URDU</th>
<th>NAME</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ّ</td>
<td>'ain</td>
<td>r</td>
</tr>
<tr>
<td>﷙</td>
<td>ġāin</td>
<td>ġ</td>
</tr>
<tr>
<td>ّ</td>
<td>fe</td>
<td>f</td>
</tr>
<tr>
<td>ّ</td>
<td>qāf</td>
<td>q</td>
</tr>
<tr>
<td>ّ</td>
<td>kāf</td>
<td>k</td>
</tr>
<tr>
<td>ّ</td>
<td>khāf</td>
<td>kh</td>
</tr>
<tr>
<td>ّ</td>
<td>gāf</td>
<td>g</td>
</tr>
<tr>
<td>ّ</td>
<td>ghāf</td>
<td>gh</td>
</tr>
<tr>
<td>ّ</td>
<td>lām</td>
<td>l</td>
</tr>
<tr>
<td>ّ</td>
<td>lham</td>
<td>lh</td>
</tr>
<tr>
<td>ّ</td>
<td>mim</td>
<td>m</td>
</tr>
<tr>
<td>ّ</td>
<td>mhīm</td>
<td>mh</td>
</tr>
<tr>
<td>ّ</td>
<td>nūn</td>
<td>n</td>
</tr>
<tr>
<td>ّ</td>
<td>nhūn</td>
<td>nh</td>
</tr>
<tr>
<td>ّ</td>
<td>wāw</td>
<td>w / U, 0, ao</td>
</tr>
<tr>
<td>ّ</td>
<td>hīfah</td>
<td>h</td>
</tr>
<tr>
<td>ّ</td>
<td>do ʃaʃi</td>
<td>h</td>
</tr>
<tr>
<td>ّ</td>
<td>ʃhīfah</td>
<td>y / I, E</td>
</tr>
<tr>
<td>ّ</td>
<td>ʃhīfah</td>
<td>E / ae</td>
</tr>
<tr>
<td>ّ</td>
<td>ʃhīfah</td>
<td>.n</td>
</tr>
<tr>
<td>ّ</td>
<td>hamza</td>
<td>,</td>
</tr>
<tr>
<td>ّ</td>
<td>te marbuṭa</td>
<td>T</td>
</tr>
<tr>
<td>ّ</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>ّ</td>
<td>ā</td>
<td>A</td>
</tr>
<tr>
<td>ّ</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>ّ</td>
<td>i</td>
<td>I</td>
</tr>
<tr>
<td>ّ</td>
<td>u</td>
<td>u</td>
</tr>
<tr>
<td>ّ</td>
<td>ū</td>
<td>U</td>
</tr>
<tr>
<td>ّ</td>
<td>ē</td>
<td>E</td>
</tr>
<tr>
<td>ّ</td>
<td>ae</td>
<td>ae</td>
</tr>
<tr>
<td>ّ</td>
<td>ō</td>
<td>0</td>
</tr>
<tr>
<td>ّ</td>
<td>ao</td>
<td>ao</td>
</tr>
</tbody>
</table>
The “hanging he” or Ha-ye hawwaz-e-mahfi is generated by H. It does not receive a jazm even in fully vocalised mode and is not joined to a following letter: \rajaH / rāghah / kīh / nāh / waH / wah /

nāmāh / yāH

Tanwin is coded by aN:

\taqriBaN / taqrīban / faoraN / fuorra

Taśdid is produced by coding the consonant twice:

\mi,t,tI / mīlīs / unnIs / unnīs

Note: Double consonants in Urdu verbs are written without taśdid. Taśdid is disabled by inserting | between the double consonants:

bann|nA / banna / gin|nA / ginnā / jAn|nA / ġānnā

Idāfat is written -e or -i:

šēr-e pañJāb / šēr-e pañjab

Ab-e .hayat / āb-e ḥayāt

Wāw-e ma‘dūla, or the “wāw which is passed over”, is written w; it is omitted in the transliteration and the preceding he receives no jazm:

hāb / hwAja / hāga / hwud / ḥud

Wāw-e ‘afī, or the “wāw of conjunction” is coded as -0

gul-0 bulbul / gul-ō bulbul

sarw-0 sanōbar / sarw-ō sanōbar

tar-0 tāzaH / tar-ō tāzah

’Aļif maqṣūra is encoded as _A or Y:

fatw_A / da’wā / dawā

Digit sequences are written in the natural order:

0123456789 0123456789
6.4 Pashto (Aghanic)

<table>
<thead>
<tr>
<th>symbol</th>
<th>transcription</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>١</td>
<td>ا</td>
<td>tā' with a small loop</td>
</tr>
<tr>
<td>٢</td>
<td>د</td>
<td>dāl with a small loop</td>
</tr>
<tr>
<td>٣</td>
<td>ر</td>
<td>rā' with a small loop</td>
</tr>
<tr>
<td>٤</td>
<td>ن</td>
<td>nūn with a small loop</td>
</tr>
<tr>
<td>٥</td>
<td>ج</td>
<td>gāf with a small loop instead of a bar</td>
</tr>
<tr>
<td>٦</td>
<td>ز</td>
<td>rā' with one dot above and one below</td>
</tr>
<tr>
<td>٧</td>
<td>س</td>
<td>sīn with one dot above and one below</td>
</tr>
<tr>
<td>٨</td>
<td>أ</td>
<td>the diphong ae</td>
</tr>
<tr>
<td>٩</td>
<td>ۓ</td>
<td>the diphong ey</td>
</tr>
<tr>
<td>٩</td>
<td>ى</td>
<td>the diphong ey</td>
</tr>
<tr>
<td>١</td>
<td>أ</td>
<td>the long vowel ŏ</td>
</tr>
<tr>
<td>١</td>
<td>أ</td>
<td>the long vowel ŏ</td>
</tr>
</tbody>
</table>

Table 6.2: Additional codings for Pashto.

Switch to this mode by \setpashto. For writing some Pashto words in the Urdu style, write the command \seturdu and afterwards switch back.

- For Pashto, additional codings are available, see Table 6.2. Some of the given codings also occur in Urdu but with a different meaning.
- The codings \ll, \la, \le and \le are used as in Persian. The rules for izāfet and ya‘-i-wahdat apply.
- The short vowel \ae is indicated by a zwarakay, \o by an inverted damma.
6.5 Sindhi

To activate the Sindhi mode, select the language by \texttt{setsindhi}. Sindhi input texts are encoded in a modification of the standard Arab\texttt{TEX} encoding. The alphabet is given in Table 6.3.

- Use hyphens to resolve ambiguities with aspired consonants.
- There are two special codings: $\langle .\text{mIN} \rangle$ or $\texttt{\textbackslash MIN}$, $\langle '|\text{IN} \rangle$ or $\texttt{\textbackslash IN}$. 
- The user might want to break some ligatures by inserting a vertical bar to get the correct writing, or just for a better appearance of the script.

<table>
<thead>
<tr>
<th>A</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>C</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</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>
<td>d</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>E</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
</tr>
<tr>
<td>F</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
</tr>
<tr>
<td>G</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
</tr>
<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>
<td>h</td>
<td>h</td>
<td>h</td>
</tr>
<tr>
<td>I</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>J</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
</tr>
<tr>
<td>K</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
</tr>
<tr>
<td>L</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>M</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>N</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>O</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
<td>q</td>
</tr>
<tr>
<td>R</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>S</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td>T</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>U</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
<td>u</td>
</tr>
<tr>
<td>V</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>W</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>X</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>Z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
</tr>
</tbody>
</table>

Table 6.3: The Sindhi Alphabet
6.6 Kashmiri

Select Kashmiri by $\texttt{setkashmiri}$. The input codes are given in Table 6.4. The transcription follows the ALA-LC romanization conventions.

6.7 Uighuric

Switch to this mode by $\texttt{setuighur}$.

Uighuric input texts are encoded in a modification of the standard ArabTEX encoding, see column 5 of Table 6.5. Please observe that in Uighuric all characters are coded verbatim.
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (6) 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>02</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>03</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>04</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>05</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>06</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>07</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>08</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>09</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>10</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>11</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>12</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>13</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>14</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>15</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>16</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>17</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (6) 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>19</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>20</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>21</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>22</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>23</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>24</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>25</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>26</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>27</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>28</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>29</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>30</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>31</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>32</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>33</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
<tr>
<td>34</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
<td>ئ</td>
</tr>
</tbody>
</table>

1. initial shape 
2. medial shape 
3. final shape 
4. isolated shape 
5. external encoding 
6. sorting position 
7. name 

Table 6.5: ArabTEX encoding of Uighuric
6.8 Old Malay

The preliminary ArabTeX language mode \texttt{\setmalay} is provided for processing Old Malay texts in the extended Arabic writing.

Old Malay (Jawi) input texts are encoded in a modification of the standard ArabTeX encoding, see below.

Additional encodings:

\begin{verbatim}
\begin{tabular}{|c|c|}
\hline
p & ف \\
\hline
ng & ع \\
\hline
\end{tabular}
\end{verbatim}

\begin{verbatim}
\begin{tabular}{|c|c|}
\hline
v & ن \\
\hline
ny & در \\
\hline
\end{tabular}
\end{verbatim}

\begin{verbatim}
\begin{tabular}{|c|c|}
\hline
\ & گ \\
\hline
\ & چ \\
\hline
\end{tabular}
\end{verbatim}

This language mode is strictly experimental and expected to contain many errors; it will be adapted to the users’ requirements. Please report your experience and suggestions for changes and improvements to the author.

6.9 Other extensions of the Perso-Arabic script

This is up to experimentation by the user. If \texttt{\setarab} or \texttt{\setfarsi} will not produce the desired result, try \texttt{\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 directly.
Chapter 7

Hebrew mode

On the request of some users, starting with Version 3.02 ArabTEX has been extended by some modules adding support for Hebrew. Whereas the initial applications only called for short Hebrew quotations within Roman texts, possibly containing Arabic insertions too, adding “Hebrew environments” proved comparatively easy. We also added most commands provided by the HebrewTEX package (an alternative TeX extension developed in Israel, that requires TeX--XeT). The Hebrew date quite probably will not work correctly.

To process Hebrew input with ArabTEX, proceed as follows:

- for use with Plain TeX, say `\input hebtex`; a small loader module will load both ArabTEX and the Hebrew extension.
- with \LaTeX\ 2\epsilon{} say: `\usepackage{hebtex}`.

The extension provides a language mode `\sethebrew`, and several common encodings of texts in Hebrew, that may be switched by the `\setcode` command. One (nameless) encoding is compatible with Dov Grobgeld’s editor HED, so files prepared for HebrewTEX are supposed to be compatible. In addition, the standard ArabTEX encoding has been adapted to cater for Hebrew too.

7.1 Language switching

`\sethebrew` switches to Hebrew mode, `\setarab` back to Arabic.

Remember to switch the encoding and the vowelization mode too!
CHAPTER 7. HEBREW MODE

7.2 Standard Hebrew encoding

\setcode{standard} or \setcode{arabtex} will switch to the standard Arab\TeX{} Hebrew encoding, defining the consonants as follows:

<table>
<thead>
<tr>
<th>Consonant</th>
<th></th>
<th>Consonant</th>
<th></th>
<th>Consonant</th>
<th></th>
<th>Consonant</th>
</tr>
</thead>
<tbody>
<tr>
<td>aleph</td>
<td>&quot;</td>
<td>beth</td>
<td>&quot;</td>
<td>gimel</td>
<td>&quot;</td>
<td>daleth</td>
</tr>
<tr>
<td>heh</td>
<td>&quot;</td>
<td>yod</td>
<td>&quot;</td>
<td>zayin</td>
<td>&quot;</td>
<td>chet</td>
</tr>
<tr>
<td>mem</td>
<td>&quot;</td>
<td>nun</td>
<td>&quot;</td>
<td>kaph</td>
<td>&quot;</td>
<td>lamed</td>
</tr>
<tr>
<td>peh</td>
<td>&quot;</td>
<td>sade</td>
<td>&quot;</td>
<td>samekh</td>
<td>&quot;</td>
<td>ayin</td>
</tr>
<tr>
<td>sin</td>
<td>&quot;</td>
<td>shin</td>
<td>&quot;</td>
<td>qof</td>
<td>&quot;</td>
<td>resh</td>
</tr>
</tbody>
</table>

Note: without punctuation, the characters sin, shin and s(h)in look identical; otherwise sin \[ \text{ש} \] has a dot to the left, shin \[ \text{ש} \] has a dot to the right, s(h)in \[ \text{ס(ה)ינ} \] is the form without a dot.

Soft consonants are marked by raphe: \(<v>\) for \[ \text{כ} \], \(<f>\) for \[ \text{כ} \].

Vowels are encoded as follows:

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Short Vowel</th>
<th>Long Vowel</th>
<th>Defective</th>
<th>Half Vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>[ \text{א} ] patach</td>
<td>[ \text{א} ] qames</td>
<td>[ \text{א} ] chateph</td>
<td>[ \text{א} ] patach</td>
</tr>
<tr>
<td>e</td>
<td>[ \text{א} ] segol</td>
<td>[ \text{א} ] sere</td>
<td>[ \text{א} ] sere</td>
<td>[ \text{א} ] chateph</td>
</tr>
<tr>
<td>i</td>
<td>[ \text{א} ] chireq</td>
<td>[ \text{א} ] chireq</td>
<td>[ \text{א} ] chireq</td>
<td>[ \text{א} ] shewa</td>
</tr>
<tr>
<td>o</td>
<td>[ \text{א} ] qames</td>
<td>[ \text{א} ] cholem</td>
<td>[ \text{א} ] cholem</td>
<td>[ \text{א} ] chateph</td>
</tr>
<tr>
<td>u</td>
<td>[ \text{א} ] qibbus</td>
<td>[ \text{א} ] shureq</td>
<td>[ \text{א} ] shureq</td>
<td>[ \text{א} ] no vowel</td>
</tr>
</tbody>
</table>

The matres lectionis can also be written explicitly, e.g., \(<_ey>\) or \(<E>\) for \[ \text{י} \], \(<iy>\) or \(<I>\) for \[ \text{י} \], \(<_ow>\) or \(<O>\) for \[ \text{ו} \].

- \texttt{\vocalize} (default) activates vowel points and special punctuation; \texttt{\novocalize} switches them off again.
- \texttt{patach furtivum} is written \(<.a>\) \textit{before} its carrier: \(<rU.a_.h>\) gives \[ \text{ו} \].
- \texttt{dagesh lene} with \(<b g d k p t>\) is expressed by prefixing a dot: \(<.b>\) \[ \text{ב} \], \(<.g>\) \[ \text{ג} \], \(<.d>\) \[ \text{ד} \], \(<.k>\) \[ \text{כ} \], \(<.p>\) \[ \text{פ} \], \(<.t>\) \[ \text{ת} \]; the same device is used for \texttt{mappiq} with \(<b>\): \(<.h>\) \[ \text{ה} \].
- \texttt{dagesh forte} and likewise \texttt{dagesh orthophonicum} are expressed by doubling the consonant; thus two equal consonants in sequence (even in
\novocalize mode) must be separated by some short vowel indicator (or <.u>), if the standard encoding is used.

- \meteg is indicated by <1> after the vowel.
- \maggef is <--> (en-dash; a single hyphen will be ignored)
- Prefixes may be separated by a single hyphen, which appears in the transcription without changing the Hebrew writing.
- For those rare cases where a consonant is missing, input <"> (bar quote) or <N> to produce a small space; this may also carry vowels.
- Accents and cantillation marks are not supported.
- For special applications, as in the table above, the input encoding <M> will produce a box □ which may carry vowels; the \LaTeXX package \texttt{latexsym} is required to show the box.
- Abbreviations are marked by \texttt{abbr}.

7.3 ISO 8859-8 and Hebrew MS-Windows

ISO 8859-8 is an 8-bit encoding extending 7-bit ASCII with Hebrew letters. Within Hebrew MS-Windows, the code page CP 1255 provides a superset of ISO 8859-8, containing a full complement of Hebrew vowels, and a host of miscellaneous special characters.

\texttt{ArabTeX} provides within the package \texttt{cp1255.sty} a reading module for a subset of CP1255, containing all Hebrew characters from ISO 8859-8 and all Hebrew vowels of CP1255, but omitting the extra special characters (see Table 7.1). This encoding is activated by the switching command \texttt{\setcode{cp1255}} or \texttt{\setcode{hwin}}.

7.4 Hebrew\TeX \texttt{ “oldcode” and “newcode”}

There is a default reading module for the Hebrew characters in code positions 96\ldots 122 (Hebrew\TeX \texttt{ “pccode”}), in code positions 128\ldots 154 as generated by the editor HED, and also in code positions 224\ldots 250 (Hebrew\TeX \texttt{ “newcode”}, ISO 8859-8). In fact these are three different encodings catered for by a single reading module; see the code assignments in Table 7.3. Observe that the Hebrew characters in ISO 8859-8 are supported both by this reading module and by the Hebrew MS-Windows encoding mentioned in Section 7.3. The code switching commands \texttt{\setcode{hed}}, \texttt{\setcode{newcode}}, \texttt{\setcode{pccode}}, and \texttt{\setcode{iso8859-8}} all activate the default reading module.
<table>
<thead>
<tr>
<th></th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>NUL</td>
<td>DLE</td>
<td>SP</td>
<td>0</td>
<td>@</td>
<td>&quot;</td>
<td>p</td>
<td>NSP</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>SOH</td>
<td>DC1</td>
<td>!</td>
<td>1</td>
<td>A</td>
<td>Q</td>
<td>a</td>
<td>q</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>2</td>
<td>B</td>
<td>R</td>
<td>b</td>
<td>r</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>ETX</td>
<td>DC3</td>
<td>#</td>
<td>3</td>
<td>C</td>
<td>S</td>
<td>c</td>
<td>s</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>4</td>
<td>D</td>
<td>T</td>
<td>d</td>
<td>t</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>ENQ</td>
<td>NAK</td>
<td>%</td>
<td>5</td>
<td>E</td>
<td>U</td>
<td>e</td>
<td>u</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>ACK</td>
<td>SYN</td>
<td>&amp;</td>
<td>6</td>
<td>F</td>
<td>V</td>
<td>f</td>
<td>v</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>BEL</td>
<td>ETB</td>
<td>'</td>
<td>7</td>
<td>G</td>
<td>W</td>
<td>g</td>
<td>w</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>BS</td>
<td>CAN</td>
<td>(</td>
<td>8</td>
<td>H</td>
<td>X</td>
<td>h</td>
<td>x</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09</td>
<td>HT</td>
<td>EM</td>
<td>)</td>
<td>9</td>
<td>I</td>
<td>Y</td>
<td>i</td>
<td>y</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LF</td>
<td>SUB</td>
<td>*</td>
<td>J</td>
<td>Z</td>
<td>j</td>
<td>z</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>K</td>
<td>l</td>
<td>k</td>
<td>l</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>FF</td>
<td>IS4</td>
<td>,</td>
<td>L</td>
<td>\</td>
<td>l</td>
<td>l</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CR</td>
<td>IS3</td>
<td>-</td>
<td>M</td>
<td>]</td>
<td>m</td>
<td>}</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>SO</td>
<td>IS2</td>
<td>.</td>
<td>&gt;</td>
<td>N</td>
<td>n</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>SI</td>
<td>IS1</td>
<td>/</td>
<td>?</td>
<td>O</td>
<td>o</td>
<td>DEL</td>
<td>`</td>
<td>`</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1: ISO 8859-8 and Windows CP 1255 code table
### Table 7.2: MAC Hebrew code table

<table>
<thead>
<tr>
<th></th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>NUL</td>
<td>DLE</td>
<td>SP</td>
<td>0</td>
<td>@</td>
<td>P</td>
<td>‘</td>
<td>p</td>
<td>LSP</td>
<td>0</td>
<td>נ</td>
<td>א</td>
<td>–</td>
<td>ל</td>
<td>ד</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>SOH</td>
<td>DC1</td>
<td>!</td>
<td>1</td>
<td>A</td>
<td>Q</td>
<td>a</td>
<td>q</td>
<td>”</td>
<td>1</td>
<td>ס</td>
<td>ב</td>
<td>–</td>
<td>ט</td>
<td>ל</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>2</td>
<td>B</td>
<td>R</td>
<td>b</td>
<td>r</td>
<td>“</td>
<td>2</td>
<td>פ</td>
<td>ז</td>
<td>•</td>
<td>פ</td>
<td>ז</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>ETX</td>
<td>DC3</td>
<td>#</td>
<td>3</td>
<td>C</td>
<td>S</td>
<td>c</td>
<td>s</td>
<td>#</td>
<td>3</td>
<td>ח</td>
<td>י</td>
<td>•</td>
<td>ח</td>
<td>י</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>4</td>
<td>D</td>
<td>T</td>
<td>d</td>
<td>t</td>
<td>$</td>
<td>4</td>
<td>פ</td>
<td>ז</td>
<td>•</td>
<td>פ</td>
<td>ז</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>ENQ</td>
<td>NAK</td>
<td>%</td>
<td>5</td>
<td>E</td>
<td>U</td>
<td>e</td>
<td>u</td>
<td>%</td>
<td>5</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>ACK</td>
<td>SYN</td>
<td>&amp;</td>
<td>6</td>
<td>F</td>
<td>V</td>
<td>f</td>
<td>v</td>
<td>&amp;</td>
<td>6</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>BEL</td>
<td>ETB</td>
<td>’</td>
<td>7</td>
<td>G</td>
<td>W</td>
<td>g</td>
<td>w</td>
<td>’</td>
<td>7</td>
<td>כ</td>
<td>ז</td>
<td>•</td>
<td>כ</td>
<td>ז</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>BS</td>
<td>CAN</td>
<td>(</td>
<td>8</td>
<td>H</td>
<td>X</td>
<td>h</td>
<td>x</td>
<td>(</td>
<td>8</td>
<td>ר</td>
<td>ט</td>
<td>•</td>
<td>ר</td>
<td>ט</td>
<td></td>
</tr>
<tr>
<td>09</td>
<td>HT</td>
<td>EM</td>
<td>)</td>
<td>9</td>
<td>I</td>
<td>Y</td>
<td>i</td>
<td>y</td>
<td>)</td>
<td>9</td>
<td>ש</td>
<td>ו</td>
<td>•</td>
<td>ש</td>
<td>ו</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LF</td>
<td>SUB</td>
<td>*</td>
<td>9</td>
<td>J</td>
<td>Z</td>
<td>j</td>
<td>z</td>
<td>*</td>
<td>NSP</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>K</td>
<td>[</td>
<td>k</td>
<td>{</td>
<td>+</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>FF</td>
<td>IS4</td>
<td>,</td>
<td>L</td>
<td>\</td>
<td>l</td>
<td></td>
<td>,</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CR</td>
<td>IS3</td>
<td>–</td>
<td>M</td>
<td>]</td>
<td>m</td>
<td>}</td>
<td>–</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>SO</td>
<td>IS2</td>
<td>.</td>
<td>N</td>
<td>`</td>
<td>n</td>
<td>~</td>
<td>.</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>SI</td>
<td>IS1</td>
<td>/</td>
<td>?</td>
<td>O</td>
<td>o</td>
<td>DEL</td>
<td>/</td>
<td>נ</td>
<td>י</td>
<td>•</td>
<td>נ</td>
<td>י</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.3: HebrewTeX, CP 1255 and ISO 8859-8 code table
7.5. BHS ENCODINGS

In this encoding vowel points, dagesh and meteg cannot be used, as they cannot be represented in the input. Abbreviations may be expressed by a single or double apostrophe (right quote). The final and the medial forms of characters are equivalent; ArabTEX will choose the appropriate shape automatically.

7.5 BHS encodings

The package bhs.sty provides support for the encoding that is used in the machine-readable version of BHS (Biblia Hebraica Stuttgartensia). After loading the package, you can switch to this encoding by the command \setcode{bhs}.

The patach furtivum is not coded explicitly in BHS; ArabTEX tries to guess it. For an example, see Figure 7.1.

BHS line numbers and comments are only partially supported. The line-breaks of the source are ignored. To respect them set \bhsmode=2 or \bhsmode=1; adjusting \textwidth might be necessary.

Figure 7.1: BHS Hebrew example

The package bhs.sty provides support for the encoding that is used in the machine-readable version of BHS (Biblia Hebraica Stuttgartensia). After loading the package, you can switch to this encoding by the command \setcode{bhs}.

The patach furtivum is not coded explicitly in BHS; ArabTEX tries to guess it. For an example, see Figure 7.1.

BHS line numbers and comments are only partially supported. The line-breaks of the source are ignored. To respect them set \bhsmode=2 or \bhsmode=1; adjusting \textwidth might be necessary.
\setcode{witbhs} switches to a variant of the BHS encoding that was developed by the Werkgroep Informatica of the Vrije Universiteit Amsterdam. Activate it by \usepackage{witbhs}. We have slightly extended the latter encoding for special purposes: \texttt{<=} after a consonant forces the use of its medial shape unconditionally, even at the end of a word; \texttt{<=} produces a small space that may carry vowel points, and \texttt{<>} shows as a box that may carry vowel points. To use this feature, the \LaTeX{} package latexsym is required.

### 7.6 UNICODE Hebrew

The file utf8.sty contains a reading module for the Arabic and the Hebrew segment of UNICODE in UTF-8 encoding. It is installed by the \LaTeX{} command \usepackage{utf8} or by \input{utf8.sty}.

UTF-8 (UNICODE Transmission Format, see table 7.4) is a multi-byte encoding which, for Arabic and Hebrew, uses two bytes per character whereas ASCII characters use a single byte. Far-eastern languages are encoded in three bytes per character. This is in contrast to UNICODE itself which always uses two bytes per character.

The module is activated by \setcode{utf8}; all following Arabic and Hebrew text will be considered to be coded according to the UTF-8 encoding standard. To use the correct font, select the appropriate language. The Arab\TeX{} notation may be reactivated by \setcode{arabtex}.

### 7.7 Hebrew transcription systems

\transtrue activates the standard ZDMG transcription, and there are provisions for additional transcription systems:

- \settrans{zaw} switches to the conventions of "Zeitschrift für die Altes-
tamentliche Wissenschaft" (recommended);

- \settrans{gesenius} activates the system used in W. Gesenius' Hebrew Grammar, 26th edition (deprecated).

- \settrans{standard} restores the standard ZDMG transcription.

### 7.8 Hebrew fonts

- As a default, the fonts "hclassic" (default) and "hcaption" are distributed with Arab\TeX{}. Switch to "hcaption" by \hp, and back by \hc.

### Table 7.4: UNICODE Hebrew

<table>
<thead>
<tr>
<th>Unicode</th>
<th>058</th>
<th>059</th>
<th>05A</th>
<th>05B</th>
<th>05C</th>
<th>05D</th>
<th>05E</th>
<th>05F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D680</td>
<td>D690</td>
<td>D6A0</td>
<td>D6B0</td>
<td>D780</td>
<td>D790</td>
<td>D7A0</td>
<td>D7B0</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These fonts have been designed and donated by Joel Hoffman, who also wrote several macro packages from which we took a few ideas for positioning punctuation. There is a variety of other usable fonts on the CTAN archives.

- If no vowel points are required, all the standard fonts "DeadSea", "OldJaffa", "TelAviv", and "Jerusalem" can be used if locally available. They are activated by the commands \ds, \oj, \ta, \jm; the command \hc switches back to the default "hclassic" font.

- The "Shalom" family of fonts, if available, can be activated by \shlmold, \shlmscr, and \shlmstk. Their vowel points presently do not work since they must be handled differently from the ArabITeX strategy.

- In case a font appears to be locally available but is not found, check and, if required, correct the exact spelling of the font name within the file "uhev.fd". We have seen various incompatible variants on CTAN and on the InterNet.

- Activating other Hebrew fonts by the command \sethebfont{font} might work.

Note: We recommend to set \lineskiplimit -20pt whenever Hebrew and Roman script are used within the same paragraph; this will lead to uniform line spacing. The value of \baselineskip may have to be adjusted.

### 7.9 Judeo-Arabic

Some medieaval manuscripts contain Arabic texts written using an extended Hebrew notation. There appear to be several conventions; one of them is given in Table 7.5.

Judeo-Arabic mode is supported by default. Hebrew output can be globally activated by the language selection command \setjudarab; switch back to Arabic by \setarab. Input text is assumed to be in one of the supported input notations for Arabic.

This mode is still experimental; neither Hebrew vowel points nor Arabic diacritics are presently provided. Prospective users of this mode are invited to contact the author for discussing some open issues.

\footnote{Unfortunately we cannot read Hebrew, so the relevant literature is not accessible to us.}
7.10. **Yiddish**

Yiddish, originally a dialect of Mediaeval German containing many Hebrew loan words, is commonly written phonetically using an extension of the Hebrew character set whenever available, otherwise using Roman (ASCII) letters in the YIVO standard (which we were yet unable to locate).

For writing Yiddish using Hebrew letters, load the package `yiddish.sty`, and globally activate the Yiddish mode by `\setcode{yiddish}`. The character assignment is given in Table 7.6.

Since all required vowels are provided, there will be no Hebrew vowel points, nor *dagesh*, *meteg*, and *raphe* except for denoting soft consonants.

Some words of Hebrew origin may not be rendered correctly this way; in these cases switch temporarily to Hebrew mode by `\setcode{standard}` (and switch back later by `\setcode{yiddish}`) and use the standard ArabTeX encoding.

![Table 7.5: Judeo-Arabic encoding](image)

![Table 7.6: Yiddish YIVO transcription](image)
Chapter 8

Miscellaneous features

<table>
<thead>
<tr>
<th>.k</th>
<th>k</th>
<th>kāf in the final position without a mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>ُd</td>
<td>d</td>
<td>dāl with a dot below</td>
</tr>
<tr>
<td>.f</td>
<td>f</td>
<td>fā’ without a dot</td>
</tr>
<tr>
<td>.b</td>
<td>b</td>
<td>bā’ without a dot</td>
</tr>
<tr>
<td>.n</td>
<td>n</td>
<td>nūn without a dot</td>
</tr>
<tr>
<td>ُa</td>
<td>ā</td>
<td>’alif maqṣūra; gā’ without dots in all positions</td>
</tr>
</tbody>
</table>

Table 8.1: Additional codings for special purposes.

8.1 Additional codings

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

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 outright impossible. ArabTEX is rather flexible, but there are also some technical limitations.

8.2 Dots on \textit{ya’}

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

8.3 Vowel positioning

In vowelized Arabic text, the short vowel marks are by default positioned close to the basic character glyphs. If this is not wanted, they may be raised to approximately uniform height by the command \texttt{\accentshigh}. You may revert to the default strategy by \texttt{\accentslow}.

8.4 Abjad numerals

The command \texttt{\abjad \{#1\}} will convert its argument, which has to be a legal representation of a number between 1 and 1999, to the Arabic ‘\textit{abjad}’ notation used in some mediaeval manuscripts. The result of the conversion will not look perfect, and the legal ‘\textit{abjad}’ number 0 can presently not be generated. The command \texttt{\abjad \{#1\}} can be used inside and outside of an\textit{ Arabic context}.

This routine profited greatly from suggestions by Dr. Benno van Dalen (Utrecht University).

8.5 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 \textit{kaşida} is inserted whose length will be adjusted automatically to uniformly fill the output line.

This feature increases the already high storage demands of Arab\TeX, and should therefore be used sparingly. It can be switched on with \texttt{\spreadtrue} and switched off again with \texttt{\spreadsfalse}. Inside an\textit{ Arabic Environment}, it will also be switched off automatically at the end of every paragraph.
8.6 Uniform baselines

The Arabic and Hebrew fonts are optically compatible with the standard Roman fonts, but have larger ascenders and descenders; this will lead to unequal distances between the baselines of consecutive lines, especially if Roman and non-Roman text are mixed within the same Roman paragraph.

Typesetting on a grid will improve line spacing. We recommend to set \lineskip\limit -20pt whenever Roman script and Arabic and/or Hebrew are used within the same Roman paragraph. The value of \baselineskip may have to be adjusted; with \TeX use \baselinestretch.

Also within an Arabic environment typesetting on a grid may lead to a better result.

8.7 Verbatim copy of the input

For testing 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 examples above.

8.8 Progress report

Since Arab\TeX is still rather slow (due to evolving technology it is getting faster every year), 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 computer system, or while running a batch job, say \quiet or \tracingarab = 0 (outside an Arabic Environment; otherwise say \doassign {\tracingarab }{0}). The setting \tracingarab = 1 will only report Arabic paragraphs, a value of 2: Arabic lines and insertions, a value of 3 or more: individual Arabic items.

8.9 Module Reporting

A complete list of the modules loaded in a particular run will be put into the \TeX log file (before the run statistics), if \TEX is used. This is believed to be useful when tracing down errors. This list is also available to the user, even with Plain \TeX, as the contents of the control sequence \arabtexconfig.
Chapter 9

Compatibility issues

Arab\TeX relies only on part of the powerful features of the \TeX typesetting engine (neither mathematical mode nor the alignment mechanism are used), and few of the features provided by the Plain \TeX package and none of \LaTeX are required, but may be necessary in other parts of a multi-lingual document. Of course, \TeX’s macro processor is very heavily used.

It turned out that Arab\TeX could be made to cooperate with a number of other macro packages, sometimes after some adjustments to Arab\TeX when detecting the presence of another system, and sometimes by compatible adjustments to the other system. However there are some problem areas:

- The resource requirements of Arab\TeX and usually also of the other packages are very high, and might reach the limits of a small \TeX system. Fortunately, nowadays very large \TeX implementations are available.

- The running time is not negligible (however, computers are still becoming faster, and typesetting this very document takes only about 20 seconds on a Pentium 233 PC running em\TeX).

- Tracking down errors in a combination of several large macro packages might be difficult and time consuming.

- There might be conflicts between the names of internal commands of several packages. The resulting effects can be very obscure; there seems to be no easy solution.

- Arab\TeX assumes that the special and punctuation characters have their original category codes both when it is loaded, and when Arabic processing begins. If some macro package changes these codes, Arabic processing will usually be broken. This does not apply to Babel nor to "german.sty"; these packages are specially handled.
Conversely ArabTEX changes the category code of < which might break other packages. Loading ArabTEX as the last module usually helps, and enables ArabTEX to detect the presence of other packages.

9.1 Arabic document classes

The experimental LATEX2e classes "arabart", "arabbook", "arabrep" extend the standard classes "article", "book", and "report" in several respects: The overall document layout has been "arabized": page numbers are in Indic numerals, and columns run from right to left. The format of running heads depends on the context of the corresponding sectioning commands.

Within Arabic environments which are bracketed by \begin{RLtext} and \end{RLtext} most LATEX commands and environments are allowed, including all sectioning commands, \tabular, \tabbing, even \tableofcontents, and use an “Arabic looking” format. All arguments that denote text to be typeset are interpreted according to the currently activated Arabic encoding. Other arguments keep their LATEX standard meaning, including the preamble of \tabular, whose columns are processed from left to right (visual formatting). Generally only the basic functionality is available; optional arguments in brackets are not yet supported.

The commands \pagenumbering{abj} and \abj{ctr} generate ‘abjad” numerals for page numbers and/or arbitrary LATEX counters.

The document will start out in Roman mode, but may even be made into a single Arabic environment. Outside of Arabic environments the LATEX commands revert to their standard meaning. The picture environment and mathematical displays presently only work in Roman mode, but may contain Arabic insertions.

9.2 Using ArabTEX with EDMAC

ArabTEX 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 ArabTEX is loaded, the EDMAC commands will, after suitable local 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.

EDMAC has been extended to work with LATEX too, and ArabTEX still cooperates most of the time. However the three macro packages involved are very complicated and interact in very subtle ways, so the user may sometimes get a surprise. In this case, please contact the author.
9.3 Using ArabTEX with Babel

The Babel package by Johannes Braams provides support for multi-lingual texts in a large number of, mostly European, languages. ArabTEX does not use the language-switching facilities provided, but is otherwise compatible.

If ArabTEX is used in a Babel document, “Roman insertions” within an Arabic context are interpreted according to the presently active Babel language mode. Conversely, a “Roman paragraph” in a Babel document may contain Arabic insertions.

9.4 Using ArabTEX with PicTEX

With some caution, ArabTEX can be used together with PicTEX. However, PicTEX uses the angle brackets < and > for labeling diagrams, and this requires switching off their special meaning within ArabTEX by the command \setnone. Therefore short Arabic insertions must be included as arguments of \RL{} or bracketed with \< and >.

9.5 Using ArabTEX with CJK

The CJK package by Werner Lemberg, supporting typesetting of texts in Chinese, Japanese, and Korean, to our surprise proved to be compatible with ArabTEX (after a very small adjustment). Due to the high resource requirements of both packages, a Very Big \TeX{} may be required for processing texts of substantial size.
Chapter 10

Acknowledgments

The development of ArabiT\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 Schlebbe, helpful advice came, among others, from Chahriar Assad, Benno van Dalen, Ivan Derzhanski, Wolfdietrich Fischer, Ahmed El-Hadi, Yannis Haralambous, Abdelsalam Heddaya, Nicholas Heer, Taco Hoekwater, Yussuf Jabri, Iqbal Khan, Tom Koornwinder, Eberhard Krüger, Asif Lakehsar, Jan Lodder, Richard Lorch, Pierre MacKay, Eberhard Mattes, Fathy Neamat-Allah, Anshuman Pandey, Bernd Raichle, Ulrich Rebstock, Adrian Rezus, Paul Roochnik, Mohamed Saba, Waheed Samy, Annemarie Schimmel, Nariman Shehab, Arian Verheij, Dominik Wujastyk, and Michio Yano. We also have to thank all users who sent error reports, comments, and suggestions.

References


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


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

• هاشم محمد الحفصي، قواعد الخط العربي


**Publications**

The following list contains some examples of publications heavily relying on ArabTE\textup{X} for typesetting.

The list is neither complete, nor do we have any judgment on the content.


• *Kusyar ibn Labban’s Introduction to Astrology.*  
Edited and Translated by Michio Yano. Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo 1997.


• Wolfdietrich Fischer: *A Grammar of Classical Arabic.*  

• Tābit ibn Qurrah: *On the Sector-Figure and Related Texts.*  

Appendix A

Obtaining and installing ArabTEX

A.1 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

- ftp.dante.de/tex-archive/language/arabtex
- ftp.tex.ac.uk/tex-archive/language/arabtex
- ctan.tug.org/tex-archive/language/arabtex

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 file arabtex.htm or readme.txt first. Successfull operation on the Apple Macintosh in conjunction with OzTEX has also been reported.

At the time of this writing, version 4.00 is current. The Nasta’liq font is still under development; Naskh will be substituted automatically.

ArabTEX is being maintained and extended, if required, on a regular schedule. The current status can be found in the file arabtex.htm or by sending an EMail message (with arbitrary content) to

arabtex@informatik.uni-stuttgart.de

74
A.2 Installing ArabTEX

The installation procedure is strongly system dependent, and we recommend securing the assistance of a local TEXpert. You have to install the fonts provided ("nash14", "nash14bf", "xnsh14", "xnsh14bf", "hclassic", "hcaption") with their "*.pk" and "*.tfm" files on the font search path of your TEX system, and the "*.sty" files, "arabtex.tex", and "hebtex.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.

ArabTEX has been found to cooperate well with TEX versions 3.xxx, LATEX2e, LATEX versions 2.09 of 1991 or later, MiTEX, NFSS and NFSS2 (not required), and previewers that can handle fonts of more than 128 characters. TEX--XET or TEX--XeT are not required, and their additional features are presently not exploited.

TheTEX “hash size” should be at least 3000 to 3500, especially when using ArabTEX in conjunction with LATEX, and if the transliteration module is used. Use of a BIG TEX may be necessary when using the NFSS2 due to the latter’s high demand on string storage. Space and time requirements are not negligible, and have increased during development; however, ArabTEX currently still runs, albeit slowly, even on a PC XT standard configuration.
Appendix B

Release history

The development of the ArabTEX system began around 1991 as a private project of the author, for his personal use. However it turned out soon that the package, if at all feasible, could be of use for others also who see the need of printing Arabic text without involving a special publishing agency. As prospective users we mainly considered Orientalists which we believed very short on funding (this proved to be a drastic understatement). There was no Arabic word processing available at that time, and using \TeX as a platform looked like the only remaining possibility except perhaps implementing some complete system from scratch, which probably would necessitate building multiple versions for various computer platforms and operating systems, which we did neither dare nor could afford.

Basing the design on \TeX required the minimal user interface to become extremely lean, to facilitate the use by non-programmers. In fact, only three commands and the input notation conventions have to be learned, once the user is familiar with \TeX or \LaTeX, to be able to use ArabTEX for a standard Arabic document. Additional features can be looked up as required.

Using the \TeX typesetting engine as a machine independent platform suggested implementing the internal algorithms in \TeX’s powerful internal macro language. Unfortunately, it is not easy to use, and errors can be very hard to find and eliminate.

ArabTEX version 1.00

ArabTEX version 1.00 was a prototype, to check the basic feasibility of our approach, and to get some operating experience. Many of its features were only available in a very primitive form. It is no more supported.
ArabTeX version 2.00

ArabTeX Version 2 was the first stable version of ArabTeX. It 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 \LaTeX{}.

The main differences between versions 1 and 2 were:

- The font size was increased, so the document layout changed. The old font "nash10" was abolished and replaced by "nash14"; the character locations have been assigned differently.
- Some Arabic characters were now coded differently: ‘ayn is denoted by a left quote, and \texttt{<c>}, \texttt{<z>}, \texttt{<t>}, and \texttt{<.n>} have been assigned new meanings in order to better conform to the standard transliteration.
- Many more ligatures than before were supplied. This normally did not concern the user.
- \texttt{\vocalize} no more generated sukūn and waslā except if explicitly indicated by quoting. See \texttt{\fullvocalize}.
- Arabic Environments are now always bracketed by the new control sequences \texttt{\begin{arabtext}} and \texttt{\end{arabtext}} even if only the transliteration is wanted.

ArabTeX version 3.00

The changes introduced in 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.txt" that is part of the distribution package of ArabTeX. The earlier change history up to Version 3.00 is described in the text file "changes2.txt".

Version 3 is upwards compatible with version 2. However, many new features were introduced gradually, among them support for additional input encodings and a multitude of languages that use the Perso-Arabic script. We gratefully acknowledge the cooperation of several users who contributed information, documentation, and even helped with the coding.

On some users’ request, a Hebrew mode was added, as well as support for nearly all the Hebrew \TeX{} fonts that are available on the CTAN server network.
ArabTEX version 4.00

Version 4 is an upwards compatible extension of version 3, and many modules have been rewritten. All presently supported features are documented in this manual. It proved impractical to indicate every extension explicitly; the basic user interface is still the same.

In a few instances we had to abolish certain old features that, to our knowledge, were rarely used or not at all, because of ambiguities or conflicts with the extensions. These places are indicated in the manual, and are flagged by an asterisk in the margin.

The most important incompatible changes are:

* Tilde (\textasciitilde) is no more used as a prefix in the transliteration encoding, because of conflicts with \TeX’s use of tilde for a stable space. Caret (\textasciicircum) is now used instead in all cases.

* The double bar (\textbackslash{}\textbackslash{}) for indicating a small unbreakable space has been replaced by \textbackslash{}\textbackslash{}\textbackslash{}.

Users who still need the old features should contact the author; there might be a workaround.
Appendix C

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.

C.1 \texttt{twoblks.sty}

This \LaTeX{} option will define a command \texttt{\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{\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 contrasted. They may be in different languages, and one of them might be in Arabic (if enclosed in \texttt{\begin{arabtext} \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{\verbatim} text inside the parameters of \texttt{\twoblocks}, nor any \texttt{\catcode} changes; and all \TeX{} groups and \texttt{\if \ldots \fi} sequences must be properly nested.
C.2 verses.sty

This is a small utility for typesetting classical Arabic poetry in two parallel blocks, such that every line contains two half-verses. For its use, see the file itself.

Figure C.1: Example of Arabic poetry (source unknown)

C.3 raw.sty

This is a small utility to ease the processing of input files that have been produced by some OCR reading program. It will deactivate most of \TeX's special characters.

This package depends strongly on the special application; if you need it or a variant of it, enquire with the author.
Index

" (quoting), 21
", 19, 21
$\$, 11
--, 20, 54
\<, 12
\, 19, 21
\<, 10, 11, 15, 40
\LR, 11
\RL, 10, 15, 40
\, 12
\abjad, 64
\{ctr}, 67
\accenthigh, 64
\accentlow, 64
\allowarab, 13
\arabfalse, 38
\arabstat, 13
\arabtexconfig, 65
\arabtrue, 38
\baselineskip, 65
\baselinestretch, 65
\begin{RLtext}, 10
\begin{arabtext}, 10, 76
\bigskip, 12
\cap, 39
\centerline, 13
\clearpage, 12
\colsep, 78
\doassign, 13
\docommand, 13
\ds, 61
\emphasize, 12
\end{RLtext}, 10
\end{arabtext}, 10, 76
\footnote, 12
\fullvocalize, 18, 20, 21, 76
\hc, 61
\hfil, 12
\hfill, 12, 13
\hp, 61
\hskip, 13
\hspace, 13
\indent, 12
\input, 12
\ligsfalse, 22
\ligstrue, 22
\lineskiplim, 65
\lq, 11
\marginpar, 12
\mbox, 12
\medskip, 12
\medskipmark, 16
\newarabfont, 24
\newpage, 12
\newtanwin, 21, 23
\noindent, 12
\nospace, 12
\novocalize, 20, 21, 53
\oj, 61
\oldarabfont, 16
\oldhamza, 24
\oldtanwin, 21, 23
\pagebreak, 12
\pagemark, 67
\par, 10, 12
\quiet, 65
\rq, 11
\setarab, 10, 15, 18, 40, 52
\setarabfont, 16
\setcode, 28
\setcode{arabtex}, 28, 30, 33, 53
\setcode{asmo449}, 28
\setcode{bhs}, 58
INDEX

\setcode{cp1255}, 54
\setcode{cp1256}, 33
\setcode{hed}, 54
\setcode{hwin}, 54
\setcode{isiri}, 33
\setcode{iso8859-6}, 30
\setcode{iso8859-8}, 54
\setcode{iso9036}, 28
\setcode{newcode}, 54
\setcode{pccode}, 54
\setcode{standard}, 28, 53
\setcode{utf8}, 33, 59
\setcode{witbhs}, 59
\shlmold, 61
\shlmscr, 61
\shlmstk, 61
\shlm
\showfalse, 65
\showtrue, 65
\smallskip, 12
\space, 12
\spreadbox, 13
\spreadfalse, 64
\spreadline, 13
\spreadtrue, 64
\ta, 61
\tabular environment, 67
\tracingarab, 65
\transfalse, 38
\transtrue, 38
\twoblocks, 78
\usepackage{hebtex}, 52
\vfil, 12
\vfill, 12
\vocalize, 20, 21, 53, 76
\vskip, 13
\vspace, 13
\yahdots, 64
\yahnodots, 64
<, 10, 11, 15, 40, 67
>, 10, 11, 15, 40
\\, 12
“hanging he”, 46
\|, 19, 21, 22
\", 54
\|B, 20
\|BB, 20
11, 19, 21
‘ (‘ayn), 19
’ (hamza), 18
A, 17, 22, 27
‘A, 19, 21, 24
,A, 42
’a, 17, 21, 22
\_a, 41, 42, 47
\_a, 17, 20, 22
a (fatha), 18, 22
aa, 17, 22
abbreviation, 27
‘abjad numbers, 67
abjad.sty, 64
’abjad numbers, 64
accents, 54
ae, 44
Afghanic, 47
aH, 41
‘ayn, 19
INDEX

al-, 19, 38
'alif, 27
dagger, 17, 20, 22
initial, 27
maqṣūra, 17, 21–23, 46
silent, 21, 23
Qur’an, 20, 22
silent, 21–23, 38
small, 20, 22
below, 20, 22
‘Allah (spelling), 26
aN, 18, 21, 23
aN, A, 21, 23
aN, A, 18, 21, 23
aN, NY, 23
ao, 44
arabart.cls, 13, 67
arabbook.cls, 13, 67
Arabic, 6
generic term, 6
Arabic \LaTeX{} classes, 67
Arabic context, 10–12
Arabic environment, 10
Arabic fonts, 12, 15
Arabic group, 11
Arabic item, 11
Arabic MS Windows, 33
Arabic MS-DOS, 28
Arabic number, 11
Arabic quotation, 10
Arabic quotes, 11
Arabic script, 6
Arabic word, 11
arabrep.cls, 13, 67
\ArabLaTeX{} commands, 11, 12
archaic text, 63
ASCII, 28, 30, 33
ASMO 449, 28, 30
aspirated consonant, 43
aspiration, 44
Assad, Chahriar, 69
assignment, 13
global, 13
assimilation, 19, 20, 25, 38
automatic stretching, 64
aW, 22
aw, 41
aWA, 22
ay, 41
B, 20
Babel, 68
baselines
uniform, 65
be-, 43
bgdkpt, 53
boxing commands, 13
Braams, Johannes, 68
breaking connections, 21
cantillation, 54
capital letter, 39
category codes, 66, 67
CJK, 68
code
7-bit, 28
8-bit, 30, 33
arabtex, 28
ASCII, 28, 30, 33
ASMO 449, 28, 30
Buckwalter, 37
ISIRI 3342, 28, 33
ISO 646, 28, 30
ISO 8859-6, 28, 30
ISO 8859-8, 54
ISO 9036, 28
MS-Windows, 33
UNICODE, 33, 59
UTF-8, 33, 59
coding conventions, 17, 76
commands
\ArabLaTeX{}, 11, 12
boxing, 13
illegal, 13
\LaTeX{}, 11, 12
overview, 13
size changing, 12, 16
\TeX{}, 11, 12
user defined, 13
compatibility, 66
Babel, 68
CJK, 68
<table>
<thead>
<tr>
<th>Index Item</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDMAC, 67</td>
<td></td>
</tr>
<tr>
<td>PicTeX, 68</td>
<td></td>
</tr>
<tr>
<td>compounds, 43</td>
<td></td>
</tr>
<tr>
<td>connecting form, 20</td>
<td></td>
</tr>
<tr>
<td>copyright, 0</td>
<td></td>
</tr>
<tr>
<td>CP 1255, 54</td>
<td></td>
</tr>
<tr>
<td>CP 1256, 33</td>
<td></td>
</tr>
<tr>
<td>CTAN, 16</td>
<td></td>
</tr>
<tr>
<td>dagesh, 53, 58</td>
<td></td>
</tr>
<tr>
<td>forte, 54</td>
<td></td>
</tr>
<tr>
<td>lene, 53</td>
<td></td>
</tr>
<tr>
<td>orthophonicum, 54</td>
<td></td>
</tr>
<tr>
<td>dagger 'alif, 17, 20</td>
<td></td>
</tr>
<tr>
<td>damma, 18, 20, 21</td>
<td></td>
</tr>
<tr>
<td>inverted, 20, 22</td>
<td></td>
</tr>
<tr>
<td>Dari, 40</td>
<td></td>
</tr>
<tr>
<td>date, 20</td>
<td></td>
</tr>
<tr>
<td>default font, 15</td>
<td></td>
</tr>
<tr>
<td>defective writing, 17, 20, 22</td>
<td></td>
</tr>
<tr>
<td>definite article, 19, 25, 38</td>
<td></td>
</tr>
<tr>
<td>Derzhanski, Ivan, 41, 69</td>
<td></td>
</tr>
<tr>
<td>diacritics, 20</td>
<td></td>
</tr>
<tr>
<td>diphthongs, 41, 44</td>
<td></td>
</tr>
<tr>
<td>display mode, 11</td>
<td></td>
</tr>
<tr>
<td>đō čašmi ke, 44</td>
<td></td>
</tr>
<tr>
<td>document classes, 67</td>
<td></td>
</tr>
<tr>
<td>dots on yā’, 41, 64</td>
<td></td>
</tr>
<tr>
<td>E, 40</td>
<td></td>
</tr>
<tr>
<td>-E, 41</td>
<td></td>
</tr>
<tr>
<td>̣e, 41, 42, 47</td>
<td></td>
</tr>
<tr>
<td>̄e, 41</td>
<td></td>
</tr>
<tr>
<td>EDMAC, 67</td>
<td></td>
</tr>
<tr>
<td>eH, 41</td>
<td></td>
</tr>
<tr>
<td>El-Hadi, Ahmed, 69</td>
<td></td>
</tr>
<tr>
<td>emphasis, 27</td>
<td></td>
</tr>
<tr>
<td>Encyclopedia Iranica, 39</td>
<td></td>
</tr>
<tr>
<td>Encyclopedia of Islam, 39</td>
<td></td>
</tr>
<tr>
<td>ending, 20</td>
<td></td>
</tr>
<tr>
<td>environment</td>
<td></td>
</tr>
<tr>
<td>Arabic, 10</td>
<td></td>
</tr>
<tr>
<td>arabtext, 10</td>
<td></td>
</tr>
<tr>
<td>BiTeX, 67</td>
<td></td>
</tr>
<tr>
<td>picture, 67</td>
<td></td>
</tr>
<tr>
<td>RLText, 10</td>
<td></td>
</tr>
<tr>
<td>Roman, 10</td>
<td></td>
</tr>
<tr>
<td>tabbing, 10</td>
<td></td>
</tr>
<tr>
<td>tabular, 67</td>
<td></td>
</tr>
<tr>
<td>extra characters, 51, 64</td>
<td></td>
</tr>
<tr>
<td>Farsi, 40</td>
<td></td>
</tr>
<tr>
<td>fatha, 18, 20, 21</td>
<td></td>
</tr>
<tr>
<td>Fischer, Wolfdietrich, 22, 69</td>
<td></td>
</tr>
<tr>
<td>font</td>
<td></td>
</tr>
<tr>
<td>additional, 16</td>
<td></td>
</tr>
<tr>
<td>Arabic, 12, 15</td>
<td></td>
</tr>
<tr>
<td>nash10, 76</td>
<td></td>
</tr>
<tr>
<td>nash14, 15, 73, 74, 76</td>
<td></td>
</tr>
<tr>
<td>nash14bf, 15, 74</td>
<td></td>
</tr>
<tr>
<td>Naskh, 15, 73, 74</td>
<td></td>
</tr>
<tr>
<td>Nasta’liq, 15, 73</td>
<td></td>
</tr>
<tr>
<td>xnsh14, 15, 74</td>
<td></td>
</tr>
<tr>
<td>xnsh14bf, 15, 74</td>
<td></td>
</tr>
<tr>
<td>bold, 15</td>
<td></td>
</tr>
<tr>
<td>commercial, 15</td>
<td></td>
</tr>
<tr>
<td>default, 15</td>
<td></td>
</tr>
<tr>
<td>Hebrew, 16, 61</td>
<td></td>
</tr>
<tr>
<td>DeadSea, 61</td>
<td></td>
</tr>
<tr>
<td>hcaption, 61, 74</td>
<td></td>
</tr>
<tr>
<td>hhclassic, 61, 74</td>
<td></td>
</tr>
<tr>
<td>Jerusalem, 61</td>
<td></td>
</tr>
<tr>
<td>OldJaffa, 61</td>
<td></td>
</tr>
<tr>
<td>Shalom, 61</td>
<td></td>
</tr>
<tr>
<td>standard, 61</td>
<td></td>
</tr>
<tr>
<td>TelAviv, 61</td>
<td></td>
</tr>
<tr>
<td>installation, 74</td>
<td></td>
</tr>
<tr>
<td>nasta’liq, 41, 43</td>
<td></td>
</tr>
<tr>
<td>selection, 12</td>
<td></td>
</tr>
<tr>
<td>spelling, 61</td>
<td></td>
</tr>
<tr>
<td>standard, 16</td>
<td></td>
</tr>
<tr>
<td>transliteration, 38</td>
<td></td>
</tr>
<tr>
<td>grid, 65</td>
<td></td>
</tr>
<tr>
<td>Grobgeld, Dov, 52</td>
<td></td>
</tr>
<tr>
<td>grouping, 11, 27</td>
<td></td>
</tr>
<tr>
<td>H, 40, 42, 46, 47</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
</tr>
<tr>
<td>silent, 39</td>
<td></td>
</tr>
<tr>
<td>h-, 20</td>
<td></td>
</tr>
<tr>
<td>hamza, 18, 21, 24, 41, 42, 44</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Page(s)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Arabic</td>
<td>11</td>
</tr>
<tr>
<td>carrier</td>
<td>24, 27</td>
</tr>
<tr>
<td>old style</td>
<td>24</td>
</tr>
<tr>
<td>ḥarakāt</td>
<td>18, 20–22, 41</td>
</tr>
<tr>
<td>on tatwil</td>
<td>20</td>
</tr>
<tr>
<td>Haralambous, Yannis, 69</td>
<td>69</td>
</tr>
<tr>
<td>ha-yə hawwaz-e-mahfi, 46</td>
<td></td>
</tr>
<tr>
<td>hectaption</td>
<td>16</td>
</tr>
<tr>
<td>hcclassic</td>
<td>16</td>
</tr>
<tr>
<td>Hebrew consonants</td>
<td>53</td>
</tr>
<tr>
<td>Hebrew fonts, 61</td>
<td></td>
</tr>
<tr>
<td>Hebrew mode</td>
<td>52</td>
</tr>
<tr>
<td>Hebrew script, 6</td>
<td></td>
</tr>
<tr>
<td>Hebrew vowels</td>
<td>53</td>
</tr>
<tr>
<td>HebrewTEX</td>
<td>52</td>
</tr>
<tr>
<td>hebtex.tex</td>
<td>52</td>
</tr>
<tr>
<td>HED</td>
<td>52</td>
</tr>
<tr>
<td>Heddaya, Abdelsalam, 69</td>
<td>69</td>
</tr>
<tr>
<td>Heer, Nicholas, 69</td>
<td></td>
</tr>
<tr>
<td>Hoekwater, Taco, 69</td>
<td></td>
</tr>
<tr>
<td>hyphen, 20, 26, 27</td>
<td></td>
</tr>
<tr>
<td>I, 17, 22</td>
<td></td>
</tr>
<tr>
<td>-I, 41</td>
<td></td>
</tr>
<tr>
<td>^I, 19</td>
<td></td>
</tr>
<tr>
<td>-i, 41</td>
<td></td>
</tr>
<tr>
<td>_i, 17, 20, 22</td>
<td></td>
</tr>
<tr>
<td>i (kasra), 18, 22</td>
<td></td>
</tr>
<tr>
<td>implementation</td>
<td></td>
</tr>
<tr>
<td>Mac, 73</td>
<td></td>
</tr>
<tr>
<td>PC, 73</td>
<td></td>
</tr>
<tr>
<td>U*IX, 73</td>
<td></td>
</tr>
<tr>
<td>iN, 18, 23</td>
<td></td>
</tr>
<tr>
<td>input switching, 28</td>
<td></td>
</tr>
<tr>
<td>insertion</td>
<td></td>
</tr>
<tr>
<td>mathematical, 11</td>
<td></td>
</tr>
<tr>
<td>non-Arabic, 11</td>
<td></td>
</tr>
<tr>
<td>Roman, 11</td>
<td></td>
</tr>
<tr>
<td>installation, 74</td>
<td></td>
</tr>
<tr>
<td>inverted dāamma, 20, 47</td>
<td></td>
</tr>
<tr>
<td>invisible consonant, 19</td>
<td></td>
</tr>
<tr>
<td>ISIRI 3342, 33</td>
<td></td>
</tr>
<tr>
<td>ISO 646, 28, 30, 33</td>
<td></td>
</tr>
<tr>
<td>ISO 8859-6, 30</td>
<td></td>
</tr>
<tr>
<td>ISO 8859-8, 54</td>
<td></td>
</tr>
<tr>
<td>ISO 9036, 28</td>
<td></td>
</tr>
<tr>
<td>item</td>
<td></td>
</tr>
</tbody>
</table>
INDEX

MITeX, 74
Module list, 65
MS Arabic Windows, 33
MS-Windows, 33

N, 20, 21, 38
naming conflict, 66
nasalization, 44
Naskh, 15, 16, 73, 74
Nasta’liq, 16, 41, 43, 73
Neamat-Allah, Fathy, 69
nesting, 11, 13
NFSS, 74
nikudot, 16
no vowel, 54
non-Arabic insertion, 11
NU, 18, 23
numbers, 27, 43
‘abjad, 64
Arabic, 11
nun-e ‘gunnah, 44

O, 40, 41
¬0, 46
option
   abjad, 64
   asmo449, 28
   iso88596, 28
   twoblks, 78
Ottoman, 40

Pandey, Anshuman, 44, 69
Pashto, 43, 47
patach furtivum, 53
PC implementation, 73
Persian, 40
Persian copula, 41
PicTeX, 68
picture environment, 67
piš, 41
Progress report, 65
pseudo fonts, 15
punctuation, 11

quotation
   Arabic, 10
   non-Arabic, 11
   Roman, 11
quoting, 18, 20, 21
Qur’an ‘alif, 20

Raichle, Bernd, 69
raphe, 54
raw.sty, 79
reading module, 28
Rebstock, Ulrich, 69
Rezus, Adrian, 69
Roman, 6
generic term, 6
script, 6
Roman environment, 10
Roman insertion, 11
Roochnik, Paul, 69

Saba, Mohamed, 69
Samy, Waheed, 69
Schimmel, Annemarie, 69
Schlebbe, Heribert, 69

script
   Arabic, 6
   Hebrew, 6
   Roman, 6
šadda, 19, 21, 24, 25
   on tatwil, 20
Shehab, Nariman, 69
short vowels, 18
silent ‘alif, 21, 38
silent h, 39
size changing, 12, 16
space
   small, 19
   unbreakable, 19
special codings, 63
stretching, 12, 20, 64
   automatic, 64
sukün, 20, 21, 42, 43, 46, 76
   on lám, 19
   on tatwil, 20
sun letter, 19

T, 24
tabbing environment, 10
INDEX

87

tāʿ marbuta, 24
	tawâin, 18, 20, 21, 23, 38, 46
	fatḥa, 23
					ontawil, 20

taṣdid, 19, 21, 46
disabling, 46
Urdū verbs, 46
	tawil, 20, 26

TEX commands, 11, 12
TEX hash size, 74
text

archaic, 63
erroneous, 63
TEX--XeT, 52, 74
transliteration, 17, 38, 76

Encyclopedia Iranica, 39
Encyclopedia of Islam, 39
Farsi, 39
Kashmiri, 39
Lazard, 39
Library of Congress, 39
standard, 39
turkish, 39
Urdu, 39
ZDMG, 38, 39
twoblks.sty, 78

U, 17, 22, 38
_{U}, 19
^U, 23, 41
_u, 17, 20, 22
_u (damma), 18, 22
U^{*}IX implementation, 73
UA, 21, 22
uheb.fd, 61
uN, 18, 23
UNICODE, 33, 59

Arabic, 35, 36
Hebrew, 60
uniform baselines, 65
Urdu, 43, 47
user defined commands, 13
UTF-8, 33, 59
uw, 17, 22

van Dalen, Benno, 64, 69
verbatim, 27
Verheij, Arian, 69
verses.sty, 79
visual formatting, 67
vowel marks, 20
vowel points, 16
vowels

Hebrew, 53
invisible, 54
long, 17, 20, 22, 41, 44
positioning, 64
short, 18, 22, 41, 44
silent, 38

W, 38
WA, 21
wasla, 20, 21, 25, 38, 76
wāw-e ‘atf, 46
wāw-e ma’dīla, 46
wāw-e maqwul, 44
wāw-e ma’ruf, 44
Wujastyk, Dominik, 67, 69

Y, 17, 22

ya’
dots, 41, 64
ya’i-wahdat, 41, 42, 47
Yano, Michio, 69
ya-ye maqwul, 44
ya-ye ma’ruf, 44

zēr, 44
zīr, 41
zwarakay, 47