

NewCenturySchlbk-Italic: pncri
 NewCenturySchlbk-Roman: pncr
 Palatino-Bold: pplb
 Palatino-BoldItalic: pplbi
 Palatino-Italic: pplri
 Palatino-Roman: pplr
 Symbol: psyr
 Times-Bold: ptmb
 Times-BoldItalic: ptmbi
 Times-Italic: ptmri
 Times-Roman: ptmr
 ZapfChancery-MediumItalic: pzcmi
 ZapfDingbats: pzdr

Please contact me if you have any comments or additions.

◊ Karl Berry
 135 Center Hill Rd.
 Plymouth, MA 02360
 karl@cs.umb.edu

Arabic, Persian and Ottoman T_EX for Mac and PC

Yannis Haralambous

*¡El peor con T_EX,
 es la gente
 que lo gusta!*
 — Anonymous

The whole thing started when my friend Siavash Mirshams Shahshahani from the Sharif University of Technology, Tehran, asked me to find some arabic-alphabet T_EX for the PCs of his department. I looked around and found out that there is none, at least not in the public domain. Unfortunately, not even T_EX-X_ET, the right-to-left version of T_EX is available for Mac or PC. So I decided to fill the gap. For this I had to create the necessary font *yarb* — which is suitable for arabic, persian and ottoman as well — and a Pascal preprocessor *yarbtex*. This work has been done on a Mac SE/30 with OzT_EX, MacMETAFONT, and Think Pascal v3.0.

The arabic alphabet

The common arabic alphabet has the following 28 letters:

ذ د خ ح ج ث ت ب ا
 غ ع ظ ط ض ص ش س ز ر
 ي و ه ن م ل ك ق ف

To cover also persian and pre-1932 turkish, one has to consider the following 5

ئ گ ژ چ پ

Finally *ث* is used for the arabic transliteration of the sound “v”, as in “Vienna”: *فيينا*.

Let’s see now how letters are combined to form words; with the letters *ت*, *ك* and *ع* we can form the word *تَكَع* (right to left!). As you can see, letters are written in a different way according to their position inside the word. Consequently, arabic letters appear in four forms: *initial* (like *ت* for *ت*), *middle* (like *ك* for *ك*), *final* (like *ع* for *ع*) and *isolated*. The letters *ا*, *د*, *ذ*, *ر*, *ز*, *ث*, *و* have only final and isolated forms. So when they occur inside a word, the next letter is initial (as in *ولد*).

As in german, there are short and long vowels. The long ones are *ا*, *و* and *ي*. The short ones are not represented at all, except in classical or educative texts, where they are represented by accents: *ˆ* (fatha), *˘* (kasra), *˙* (damma), respectively for the sounds of a, i and u. In this case, the lack of vowel is represented by *ˆ* (tašdid). Also there are other accents (*˜*, *˝*, *˚*) for special purposes.

The preprocessor *yarbtex*

While typing the input, it would be too tedious to think in which form each letter should be written, and since this problem is much too complicated to be solved only by ligatures inherent to the font, it was necessary to make a preprocessor which converts an *xxx.arb* input file, into a *xxx.tex* T_EX-file. The usual procedure

$$\text{.tex} \xrightarrow{\text{T}_{\text{E}}\text{X}} \text{.dvi} \xrightarrow{\text{dviops}} \text{.ps}$$

is now one step longer

$$\text{.arb} \xrightarrow{\text{yarbtex}} \text{.tex} \xrightarrow{\text{T}_{\text{E}}\text{X}} \text{.dvi} \xrightarrow{\text{dviops}} \text{.ps}$$

Let’s take a look now at an *xxx.arb* file: to typeset in arabic, persian and/or ottoman (in short APO) you have to enter into “arabic mode” by typing a vertical bar *|*. The *|* will change to arabic mode regardless of whether you are in text or math mode, so you should use $\backslash\text{vert}$ to obtain *|* in non-arabic mode.

Once you are in arabic mode, you can choose between the following submodes:

- text and arabic numerals* (the numerals used in arabic typography),

- b. *transliterated text* (to get an output in latin alphabet, following your own transliteration),
- c. *comment mode* (to insert comments),
- d. *math mode* (to insert a *short* math sequence) and *display math mode*,
- e. *command mode* (to insert *some* commands which have to be executed inside arabic mode).

Text and arabic numerals. APO text has to be written in a special “machine-like” transliteration. By this I mean that you’ll have to type as few as possible ASCII characters and that this text won’t be very readable (at least until you get used to it). Since there are no uppercase letters in APO I used lower and uppercase ASCII characters to represent different APO letters: *s* stands for *س* and *S* for *ش*. I also used “hat + character” combinations: *ˆs* gives *ص* and *ˆS*, *ج*. Here is an example: to obtain

الأرنب و الفيل

you have to input

```
|^A^Lrnb U AlfI1|
```

which will look like

```
\arbword{\yarb{\char133}{\char237}%
{\char166}{\char12}%
{\char0}}
\arbword{\yarb{\char246}}
\arbword{\yarb{\char231}{\char250}%
{\char206}{\char229}{\char128}}
\arboff
```

in the corresponding *.tex* file. You can find the complete list of input codes in Appendix A (column 3). I suggest you make a copy of this list and keep it in front of you while you are typing; this because, unfortunately, not all characters transliterate into the right phonetic counterparts (like *c* which stands for *ع*).

APO words are never hyphenated. The way typographical spacing problems are resolved is by inserting straight line segments *_* between characters. Perhaps in some later version of *yarbtex* this will be done automatically. For the moment you’ll have to do it manually: just type an *_* to get a line segment with the length of half an *n*-dash. Here is some advice: if you are using accents, sometimes it looks better to place the accent over the middle of the line segments, *حَـ*, instead of one of the ends, *حـَ*. Note that all the “TeX-forbidden” ASCII characters we encountered (*ˆ*, *_*, etc.) and will encounter, disappear in the *xxx.tex* file.

When you type 0, 1, ..., 9 inside arabic mode you get the arabic numerals

٠, ١, ٢, ٣, ٤, ٥, ٦, ٧, ٨, ٩.

yarbtex automatically writes numbers left-to-right: 143/1962 becomes ١٩٦٢\١٤٣. The punctuation marks *.*, *:*, *!*, *?*, **, *-*, *—*, *() []* and special characters *ٲ*, *ٳ*, *ٴ*, *ٵ* are included in the font.

When you exit arabic mode (by typing again a vertical bar *|*), a new paragraph gets started. If you want just to insert a *short* arabic sequence inside latin text, you’ll have to type *\ins* before getting into arabic mode. We will return to this later in the discussion of the output file.

Transliterated text. As in the *AMS-TEX* cyrillic package, you can use the same input to obtain either a text in the arabic alphabet or a transliteration into the roman alphabet. The mechanism is simple: just type *@_* inside arabic mode, and everything up to the next *@_*, will be transliterated into latin alphabet.

The natural question to ask is: which transliteration should be chosen? Transliteration is most of the time based on phonetics and these can be very different from one country to the other. So it should be possible to use different transliterations (for example you could use characters from the Washington Computer Modern IPA font *wsuipa*), and switch easily from one to the other, without having to recompile *yarbtex*’s source. The solution I propose is the following: *yarbtex* is accompanied by a text-file, named *yarbtex.dat*. The first $57 \times 4 = 228$ lines of this file contain the transliterations of the 57 APO letters and accents, each in its 4 forms. The transliteration included in the package (and given also in Appendix A, column 2) comes from A. A. Ambros’s *Einführung in die moderne arabische Schriftsprache*. Here is how to change an item. Let’s say you want to transliterate *ث* (the final form of *ث*) by “th” instead of “t”. As you can see in column 4 of Appendix A, *ث* has the number 5. Go to line $23 = 5 \times 4 + 3$ (0 for isolated, 1 for initial, 2 for middle and 3 for final) of *yarbtex.dat* and replace *\d t* by *th*.

Comment mode. Just use *%* to insert a comment: everything on the right of *%* in the *.arb* file will stay on the right of it in the *.tex* file, in order to be considered as a comment by *TEX*.

Math mode. If you want to insert a *short* math sequence (such as *\$f(x)\$*), go ahead and type it. But you should try to avoid having math expressions split on an end-of-line, since it will be rather confusing to split them from left to right when the text goes from right to left. For longer math sequences use display math mode, as in *TEX*.

Command mode. Let us first describe the output file. Each APO character is written as `{\char xxx}`. The braces are there to allow the use of accents: `\fatha{\char160}` gives . Each arabic word gets into an `\arbword{...}` macro. Everytime `TEX` reads this macro, it first checks if its argument fits into the current line. If this is the case, it appends it on the left of the line and looks for the next `\arbword{...}`; else, it outputs the line and starts a new one. When you leave arabic mode, an `\arboff` macro appears in the output file. If the arabic text is to be inserted into a latin text (that is if you had typed `\ins` before entering into arabic mode), `\arboff` just outputs the remaining `\arbwords`; else, it also starts a new paragraph.

Since all these macros deal with `\hboxes`, one should be very careful which macros to insert into arabic mode and which not. For example, a change of arabic font must occur inside each `\arbword` and on the left of the first `{\char }`, but a macro like `\centerline` has to occur outside of arabic mode. The solution is the following: if you need a macro to affect the argument of `\arbword`, type an asterisk `*` and the macro in braces; for example if `\yarbbf` is the boldface arabic font, type `*{\yarbbf}`. This feature is similar to `TEX`'s `\everyhbox`; all forthcoming arabic words will be in boldface, until you type `*`, which is equivalent to `*{}`. If you want to add a second macro `\yyy` you have to rewrite the first: `*{\yarbbf\yyy}`, since each `*{...}` replaces the previous one.

On the other hand, for macros affecting the whole `\arbword` box, you'd better exit arabic mode. For example, the first line in the text of Appendix B (The Rabbit and the Elephant) has been typeset in the following way:

```
\noindent\centerline
{\ins|*{\yarbbig}^A^Lrnb U AlfIl{}}
```

The yarb font

The characters of this font belong to the nashī style, which is the most common in arabic typography. At this time there is not yet enough metaness in this font; in some later version, the same `.mf` sources will also produce typewriter and straight modern styles.

Calligraphic arabic has many ligatures and is much more difficult to typeset. The ligature problem could be solved, by using more than one table of 256 characters for each font and creating a special version of `yarbtex` which would, beside its regular tasks, detect the ligatures. For the moment,

the only ligature I include is the standard lām-ālif `ل`, which is input by one ASCII character, namely `L` (instead of `LA`).

The ornamentation of Appendix B comes from OSAMA EL NAHAS's beautiful book on *Islamic Decorative Elements* (Cairo 1985). The repetitive and symmetric character of islamic decoration makes it a real challenge for `TEX` and `METAFONT`.

The standard `yarb` font has 256 characters, but since there are still PCs which don't support the larger fonts, there is a second version of `yarbtex` which uses two different 128-character fonts `yarb` and `ysarb`.

Things remaining to do and availability

Besides further development on calligraphy and decoration, all other `TEX` features will be made available in APO. Compatibility with macro-packages like `AMS-TEX`, `LAMS-TEX`, `LATEX` will be tested, and fine points of APO typography studied. I would be really grateful for every suggestion and comment.

The status of the `YARB` package is postcardware (which means that each user should send me a postcard, for my collection). It is freely available at my address either by electronic mail, or by ordinary mail (please add some answering coupons).

◇ Yannis Haralambous
 Université de Lille 1
 59655 Villeneuve d'Ascq
 France
 Bitnet: yannis@frcit181

Output	Transl. Output	Input	N	Output	Transl. Output	Input	N
ا	ā	A	1	م	m	m	30
ب	b	b	2	ن	n	n	31
پ	p	p	3	ه	h	h	32
ت	t	t	4	ة	(tāz marbūṭa)	c	33
ث	ṭ	T	5	و	ū	U	34
ج	ġ	Z	6	ى	ī	I	35
چ	tsch	ˆS	7	لا	lā	L	36
ح	ḥ	H	8	ـ	(fatḥa)	a	37
خ	ḫ	j	9	ـ	(kasra)	i	38
د	d	d	10	ـ	(damma)	u	39
ذ	d	D	11	ـ	(tašdīd)	w	40
ر	r	r	12	ـ	(sukūn)	o	41
ز	z	z	13	ـ	(3-fatḥa)	ˆa	42
ژ	ž	ˆZ	14	ـ	(3-kasra)	ˆi	43
س	s	s	15	ـ	(3-damma)	ˆu	44
ش	š	S	16	ـ	(tašdīd-kasra)	ˆw	45
ص	ṣ	ˆs	17	ـ	3	ˆ_,ˆo	46
ض	ḍ	ˆd	18	ـ	(line segment)	-	47
ط	ṭ	ˆt	19	أ	3a	ˆA	48
ظ	ẓ	ˆz	20	إ	3i	ˆI	49
ع	ʿ	y	21	ؤ	3u	ˆU	50
غ	ġ	R	22	آ	(alif-madda)	E	51
ف	f	f	23	لا	(lām-alif-madda)	ˆl	52
ق	q	q	24	ـ	(upper wesla)	C	53
ق	v	v	25	لا	(lām-alif-3)	ˆL	54
ك	k	k	26	ي	(variant yāz)	Y	55
گ	g	g	27	ـ	(lower wesla)	ˆC	56
ن	n	ˆn	28	ـ	(tašdīd-fatḥa)	ˆe	57
ل	l	l	29				

Appendix A.

In the first column, the APO (Arabic-Persian-Ottoman) character is represented in its isolated form. Character ك comes from CARL FAULMANN'S *Das Buch der Schrift, enthaltend die Schriftzeichen und Alphabete aller Zeiten und aller Völker des Erdkreises*, Vienna 1880, p. 104, "Türkische Nesḫi", where it is transliterated as n. The transliteration in column two comes from ARNE A. AMBROS'S *Einführung in die moderne arabische Schriftsprache*. The input code of column three tries to be as short and as phonetic as possible; for example j stands

for خ because of its sound in Spanish, and y for ع because of its correspondence to the Cyrillic ы which has almost the same phonetic value. Unfortunately, this correspondence cannot always be achieved: ʿ has no phonetic relation to c, etc. In the case of hamza, ˆo is used for the "standing-alone hamza" ؤ and ˆ_ for the "hamza with carrier" ؤ (only initial and middle forms). Finally, in the fourth column you have a number which is purely inherent to `yarbtex` and will help you change the transliteration data in `yarbtex.dat`.

الأرنب و الفيل

زعموا أن أرضاً من أرض الفيلة ، تتابعت عليها السنون و أجذبت ، فقلّ الماء في تلك البلاد و غارت العيون ، و أصاب الفيلة عطش شديد . فشكت ذلك إلى ملكها . فأرسل الملك رسله ورواده في التماس الماء في كل ناحية . فرجع إليه بعض رسله فأخبروه بأنهم وجدوا في بعض الأمكنة عيناً تدعى القمريّة ، كثيرة الماء . فتوجه ملك الفيلة بفيلته إلى تلك العين ليشرّب منها . و كانت تلك الأرض أرض أرناب . فوطئت الفيلة الأرناب بأرجلها في جحرتها فأهلكت أكثرها . فاجتمع البقية منها إلى ملكها فقلن له : قد علمت ما أصابنا من الفيلة ، فاحتل لنا قبل رجوعهن علينا ؛ فإنهن راجعات لوردهن و مفنياتنا عن آخرنا . فقال ملكهن : ليحضرني كل ذي رأي برأيه . فتقدم خزر منها يقال له فيروز ، و قد كان الملك عرفه بالأذن و الرأي ، فقال : إن رأى الملك أن يبعثني إلى الفيلة و يبعث معي أميناً يرى و يسمع ما أقول و ما اصنع و يخبره به ، فليفعل . فقال له ملك الأرناب : أنت أميني ، و أنا أرضى رأيك ، و اصدق قولك ؛ فانطلق إلى الفيلة و بلغ عني ما أحببت ، و اعمل برأيك ، و اعلم أن الرسول ، به و برأيه و أدبه يعتبر عقل المرسل و كثير من شأنه ، و عليك باللين و المواتاة ؛ فإن الرسول هو يلين القلب إذا رفق ، و يخشن الصدر إذا خرق . فانطلق الأرنب في ليلة ، القمر فيها طالع ، حتى انتهى إلى موضع الفيلة . فكره أن يدنو منهم فيطأنه بأرجلهم ، و إن لم يردن ذلك ، فأشرف على تلّ فنادى ملك الفيلة باسمه و قال له : إن القمر أرسلني إليك ، و الرسول مبلغ غير ملوم ، و إن أغلظ في القول . فقال له ملك الفيلة : و ما الرسالة ؟ قال : يقول لك القمر إنه من عرف فضل قوته على الضعفاء فاغتر بذلك من الأقوياء ، كانت قوته حيناً و وبالاً عليه ؛ و إنك قد عرفت فضل قوتك على الدواب فغرك ذلك مني فعمدت إلى عيني التي تسمى بأسبي فشربت ماءها و كدرته أنت و أصحابك ؛ و إنني أتقدم إليك و أنذرك ألا تأتيها فأعشي بصرك و أتلف نفسك . و إن كنت في شك من رسالتي ، فهلم إلى العين من ساعتك ، فإني موافيك بها . فعجب ملك الفيلة من قول فيروز ، و انطلق معه إلى العين . فلما نظر إليها رأى ضوء القمر في الماء . فقال له فيروز : خذ بخروطومك من الماء و اغسل و جهك و اسجد القمر . ففعل . ولما أدخل خرطومه إلى الماء فحرّكه ، خيل إليه أن الماء يرتعد ، فقال ملك الفيلة : و ما شأن القمر يرتعد ؟ أترأه غضب من إدخال جحفتي في الماء ؟ قال : نعم ، فاسجد له . فسجد الفيل للقمر و تاب إليه مما صنع ، و شرط له ألا يعود هو و لا أحد من فيلته إلى العين .