

LaTeX intermédiaires (L-SHS)

Grec, Arabe, Hébreu avec pdfLaTeX

Jean Hare

Sorbonne Université
Laboratoire Kastler Brossel
jean.hare@lkb.ens.fr

Avril 2023

Table des matières

- 1 Les polices en TeX, bref historique
- 2 Textes en grec
- 3 Textes en arabe
- 4 Textes en hébreu

Table des matières

1 Les polices en TeX, bref historique

2 Textes en grec

3 Textes en arabe

4 Textes en hébreu

Le codage d'entrée (inputenc)

- Historiquement T_EX n'utilise que 7 bits ($2^7 - 1 = 127$ car. ASCII). Suffit pour l'anglais, pas de diacritiques $\acute{e}=\backslash'e$, $\tilde{n}=\backslash~n$, $\zeta=\backslash c c$.
- Plus tard on est passé à 8 bits (jeu de $2^8 - 1 = 255$ caractères) mais dépendant de la plate-forme, (dont [macroman](#), [Win1252](#), [ANSI](#), etc).
- Puis apparition des normes [ISO 8859](#). Notamment :
 - ISO 8859-1 = [latin1](#) ou -15 = [latin9](#) (europe occidentale)
 - ISO-8859-5 (cyrillique), etc.
- [Unicode](#) (à partir de 1992) :
 - jeu de $2^{16} - 1 = 65\,535$ caractères visant à contenir tous les caractères de toutes les langues, y compris les langues mortes dans leurs variantes, et tous les idéogrammes des langues de l'Extrême-Orient.
- Windows « occidental » utilise un codage sur 2 octets appelé UTF-16

Le codage d'entrée UTF-8

- L'encodage UTF-8¹ introduit en 1996, est le standard depuis ~ 2000.
- Les « points de code » sont donnés par 1 à 4 octet selon leur classe .
- La première page contient les caractères ASCII seulement :
 - les U+0128 à U+0255 (en hexa U+x0080 à U+x00FF) sont vides,
 - le caractère étendu « À » est codé U+x00C0 (code décimal 192).
- UTF-8 fournit un jeu de caractères plus étendu, si on sait les saisir
- Depuis avril 2018, le codage par défaut des fichier de latex est UTF-8.
En conséquence, la ligne suivante est devenu facultative :

```
\usepackage[utf8]{inputenc}
```

- Il est toutefois prudent de la conserver, pour la portabilité, et aussi pour la commande `\DeclareUnicodeCharacter` qu'il définit :

```
\DeclareUnicodeCharacter{00AB}{\og{}\ignorespaces\,}
\DeclareUnicodeCharacter{00BB}{\unskip\,\fg{}}
```

- Le paquet `newunicodechar` est toutefois plus commode :

```
\newunicodechar{\alpha}
```

Le codage de sortie fontenc

- De même, les polices T_EX de sortie (dans le fichier `.dvi` , `.ps` (et plus tard `.pdf`) les polices ont d'abord été encodées sur 7 bits, mais bien plus tard sur 8 bits (256 glyphes).
- Même pour un texte latin sans diacritiques, il faut bien plus de glyphes par ex. pour les variantes `\textbf`, `\textsf`, `\textit`, `\textsc`, etc.
- Nécessité de charger de très nombreuses polices et/ou définir des polices virtuelles `.vf`.
- Les polices étaient jadis en bitmap `.pk`, produites par METAFONT mais progressivement le format PostScript (polices vectorielles) s'est imposé : polices Adobe-`Type1`, toujours sur 8 bits.
- L'encodage ainsi défini appelé OT1.
- Mais les glyphes avec diacritiques sont toujours formées par composition diacritique-lettre, et cette structure est incompatible la césure (`\hyphenation`) et ne permet pas de chercher ou de copier coller dans le PDF.

Le codage de sortie fontenc

- En 1990 à **Cork (IR)**, le TUG définit un nouvel encodage (EC ou T1) contenant 256 caractères, la plupart de ceux des langues européennes utilisant un alphabet dérivé de l'alphabet latin.
- Ces polices contiennent des glyphes pré-composés et moins disgracieux, permettent la césure (si les pattern de la langue ont été intégrés au format).
- Les polices EC toujours produites par METAFONT au format **.pk** contiennent tous les glyphes historiquement accessibles en `COMPUTER MODERN` mais donnent des horreurs dans le PDF.
- Les polices `LATIN MODERN` sont une version vectorielle des polices EC, de meilleure qualité, mais ne contiennent pas les polices grecques ni cyrilliques.
- La solution finale est dans les polices `OpenType`, qui peuvent (virtuellement) implémenter tous les caractères Unicode. Mais leur utilisation requiert de passer aux moteurs $X\TeX$ ou $\text{Lua}\TeX$, avec le paquet **fontspec** à la place de **fontenc**.

Table des matières

- 1 Les polices en TeX, bref historique
- 2 Textes en grec
- 3 Textes en arabe
- 4 Textes en hébreu

Écrire du grec

- La solution générale serait de passer à Xe_{La}T_EX ou Lua_{La}T_EX, mais le paquet `cbgreek` fournit des polices grecques en Type1 pour pdf_{La}T_EX.
- On doit cependant installer et charger d'autres paquets :
 - L'encodage LGR doit être obtenu avec le paquet `greek-fontenc`.
 - Ni `cbgreek` ni `greek-fontenc` n'utilisent de `\usepackage`, ils sont activés par la commande : `\usepackage[LGR,T1]{fontenc}`.
 - L'extension `babel-greek`, qui permet d'écrire, par exemple : `\usepackage[greek.polutoniko,french]{babel}`
- Ces deux lignes de code suffisent pour pouvoir écrire :

Πάντες ἄνθρωποι τοῦ εἰδέναι ὀρέγονται φύσει

 le texte en grec étant :
 - dans les macros `\textgreek{...}`, ou `\foreignlanguage{greek}{...}`
 - après `\selectlanguage{greek}` ou l'env. `\begin{otherlanguage}{greek}`
- Dans le même contexte, on obtient la même chose par translittération : `\textgreek{P'antes >'anjrwpoi to\~u e>idénai >orégontai f'usei}`

Copie d'écran pour MiKTeX

File Tasks Help

Overview






Updates

Documentation

Packages

Retrieve from: Change...

Install in:

Name	Category	Size	Packaged on	Installed on	Installed by
babel-greek	\Formats\LaTeX...	1,25 Mio	24/11/2020	03/03/2023	Admin
cbgreek	\Language Sup...	73,58 Mio	27/02/2013	03/03/2023	Admin
greek-fontenc	\Uncategorized	192,90 Kio	26/06/2022	03/03/2023	Admin
greek-inputenc	\Language Sup...	43,51 Kio	23/02/2023	03/03/2023	Admin
greek-fontenc_doc	\Uncategorized	2,86 Mio	19/03/2023	11/04/2023	Admin
greek-inputenc_doc	\Uncategorized	730,06 Kio	27/03/2023	11/04/2023	Admin
beginigreek	\Formats\LaTeX...	4,30 Kio	15/05/2022		
beginigreek_doc	\Formats\LaTeX...	783,90 Kio	15/05/2022		
beginigreek_source	\Formats\LaTeX...	22,23 Kio	15/05/2022		
betababel	\Formats\LaTeX...	185,07 Kio	12/07/2005		
bgreek	\Uncategorized	1,23 Mio	21/02/2009		
boldtensors	\Formats\LaTeX...	15,69 Kio	04/07/2007		
chemgreek	\Formats\LaTeX...	752,02 Kio	17/01/2020		
datetime2-greek	\Formats\LaTeX...	95,97 Kio	19/04/2018		
greek-fontenc_source	\Uncategorized	122,57 Kio	26/06/2022		

Installs the selected packages.

Update Font map

- Il faut mettre à jour la liste des polices connues de T_EX.
- Pour MiKTeX :
 - On a dans le menu Tasks, la ligne Refresh font map files.
Ou en ligne de commande : `initexmf -update-fndb` .
 - Le fichier maître est `C:\Program Files\MiKTeX\miktex\config`, et
devrait contenir les lignes :


```
# cbgreek
MixedMap cbgreek-full.map
```
 - Sinon `initexmf -edit-config-file updmap` qui ouvre un fichier `updmap.cfg` où on peut ajouter ces lignes (fichier ici dans `D:\Jean\AppData\Roaming\MiKTeX\miktex\config`)
- Pour TeXLive :
 - Exécuter la commande `sudo updmap-sys -force -enable Map=cbgreek-full.map`,
ou mieux `updmap -enable Map=cbgreek-full.map` .
- Pour Tous : on peut aussi mettre dans la code source la commande `\pdfmapfile{=cbgreek-full.map}`, mais c'est déconseillé.

Exemples de textes grecs

- ```

\documentclass[a5paper]{article}
%\usepackage[utf8]{inputenc}
\usepackage[LGR,T1]{fontenc}
\usepackage{lmodern}
\usepackage[greek.polutoniko,french]{babel}
\languageattribute{greek}{ancient}
\begin{document}
un · texte · en · alphabet · latin\par
\textgreek{· Σώστρατε, · ἐρῶν · ἀπῆλθες · εὐθύς}
un · autre
\end{document}

```

- un texte en alphabet latin  
Σώστρατε, ἐρῶν ἀπῆλθες εὐθύς un autre

- Ici, on voit que :
  - `\textgreek` n'interrompt pas le flot du texte, (comme un `\textbf`)
  - le texte littéral en grec est correctement interprété dans la macro
  - cela ne marcherait pas sans elle (on s'en doute).

# Exemples de textes grecs (2)

- ```

\documentclass[a5paper]{article}
\usepackage[LGR,T1]{fontenc}
\usepackage{lmodern}
\usepackage[greek.polutoniko,french]{babel}
\languageattribute[greek]{ancient}
\usepackage{textalpha}
%\usepackage{alphabeta}
\begin{document}
\noindent un·texte
\begin{otherlanguage}{greek}
Σώστρατε, UN·INTRUS·έρῳν·ἀπῆλθες·εὐθύς
'Ιδὼν·ἐνθέδε·παῖδ'·ἐλευθέραν·τὰς·πλησίον·Νύμφας·
στεφανοῦσαν, Σώστρατε, ἐρῳν·ἀπῆλθες·εὐθύς
\end{otherlanguage}
un·autre.\

```

Σώστρατε, UN·AUTRE·INTRUS·έρῳν·ἀπῆλθες·εὐθύς
`\end{document}`

un texte Σώστρατε, ὩΝ ΙΝΤΡΥΣ ἐρῳν ἀπῆλθες εὐθύς Ἰδὼν ἐν-
 θεδε παῖδ' ἐλευθέραν τὰς πλησίον Νύμφας στεφανοῦσαν, Σώστρα-
 τε, ἐρῳν ἀπῆλθες εὐθύς un autre.

Σώστρατε, UN AUTRE INTRUS ἐρῳν ἀπῆλθες εὐθύς

- Ici, on a remplacé `\textgreek` par l'environnement `\begin{otherlanguage}{greek}`, et on voit qu'il n'interrompt pas non plus le flot du texte, mais le texte grec est justifié.
- Le texte latin dans l'environnement est translittéré.
- Avec le paquet `textalpha`, on peut écrire du grec littéral sans aucune macro, et le texte latin qu'il contient est conservé. En plus ou à la place, on peut charger `alphabeta`. La différence est peu visible dans un texte.

Table des matières

- 1 Les polices en TeX, bref historique
- 2 Textes en grec
- 3 Textes en arabe
- 4 Textes en hébreu

Écrire de l'arabe

- Pour écrire de l'arabe sans passer à Xe_{La}TeX ou Lua_{La}TeX, on dispose de deux paquets distincts, qui tous deux permettent la translittération à partir d'une saisie en alphabet latin :
 - **arabtex**, qui utilise l'UTF-8, est indépendant de **babel**, produit aussi de l'hébreu, et vient avec sa police Type 1 **xnsh14**
 - **arabi**, qui utilise aussi l'UTF-8, utilise **babel** pour sa gestion des écritures « R to L ». Ses polices Type 1, notamment la famille **ae_AlMohanad**, ne sont *actuellement* pas incluses dans la version de MiKTeX. Il faut donc télécharger le ZIP depuis CTAN, et l'installer dans un TEXMF personnel.
 - Pour cette installation :
 - Le localiser avec `kpsewhich -all -var-value=TEXMFHOME`.
 - Si la réponse est vide, créer un répertoire, et dire à T_EX où il est.
 - Ensuite, il faut y de-zipper les sous-dossier de texmf, puis rafraîchir la liste de fichiers et mettre à jour updmap comme précédemment.
- Je préfère **arabi** car il réduit fortement l'*overhead* et risque moins de créer des incompatibilités, et aussi parce que **arabtex** remonte à près de 20 ans, mais d'un autre côté, la police de **arabtex** est plus élégante...

Copie d'écran pour MiKTeX

File Tasks Help

Overview

Updates


Documentation

Packages

Diagnose

Cleanup


Settings



Packages

Retrieve from: Change...

Install in:

✕ 

Name	Category	Size	Packaged on	Installed on	Installed by
arabi	\Uncategorized	1,01 Mio	15/12/2021	13/04/2023	Admin
arabtex	\Uncategorized	900,21 Kio	26/06/2022	13/04/2023	Admin
arabtex_doc	\Uncategorized	643,92 Kio	12/06/2022	13/04/2023	Admin
arabi_doc	\Uncategorized	2,68 Mio	06/06/2021	14/04/2023	Admin
arabi-add	\Uncategorized	198,21 Kio	01/07/2015		
arabic-book	\Uncategorized	7,76 Kio	15/06/2021		
arabi_source	\Uncategorized	2,79 Mio	15/12/2021		
arabic-book_doc	\Uncategorized	530,07 Kio	15/06/2021		
romannum	\Formats\LaTeX\LaTeX ...	156,19 Kio	11/09/2009		
arabic-book_source	\Uncategorized	3,03 Kio	15/06/2021		
arabicfront	\Formats\LaTeX\LaTeX ...	3,89 Kio	05/07/2019		
arabluatex	\Uncategorized	1,10 Mio	25/03/2020		
arabtex_source	\Uncategorized	1,64 Mio	26/06/2022		
arabxetex	\Uncategorized	1 006,59 Kio	08/09/2015		
sarabian	\Fonts\Outline fonts	122,25 Kio	12/02/2006		
hvarabic	\Formats\LaTeX\LaTeX ...	173,58 Kio	03/06/2021		
sexam	\Uncategorized	1,36 Mio	14/02/2018		

Exemple avec arabtex

```
\documentclass[a5paper]{book}
%\usepackage[utf8]{inputenc}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
\usepackage{arabtex,utf8}
\setlength{\parindent}{0pt}
\setcode{utf8}
\begin{document}
Citation coranique : Sourate 2 (versets 2 à 4)
\begin{arabtext}
ذَٰلِكَ الْكِتَٰبُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ
وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ وَالَّذِينَ يُؤْمِنُونَ
بِمَا أُنزِلَ إِلَيْكَ وَمَا أُنزِلَ مِن قَبْلِكَ
وَبِآخِرَةِ هُمْ يُوقِنُونَ
\end{arabtext}
```

Citation coranique : Sourate 2 (versets 2 à 4)

ذَٰلِكَ الْكِتَٰبُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ
وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ وَالَّذِينَ يُؤْمِنُونَ
بِمَا أُنزِلَ إِلَيْكَ وَمَا أُنزِلَ مِن قَبْلِكَ
وَبِآخِرَةِ هُمْ يُوقِنُونَ

Le mot الْكِتَٰبُ (prononcer «alkitabou») désigne Le Kitab (l'Écrit, la prescription)

```
Le mot \RL{الْكِتَٰبُ} (prononcer «alkitabou») désigne Le Kitab  
(l'Écrit, la prescription)
\end{document}
```

On note que :

- l'environnement `\begin{arabtext}` (qu'est un alias de `\begin{RLtext}`) permet l'alignement à droite et la justification.
- Contrairement au cas du grec, il forme un paragraphe complet
- Pour mettre quelques mots dans un paragraphe en alphabet latin, on utilise la macro `\RL{...}`.

Exemple avec arabi

```
\documentclass[a5paper]{article}
\usepackage[utf8]{inputenc}
\usepackage[LAE,T1]{fontenc}
\usepackage[arabic,french]{babel}
\usepackage{lmodern}
\setlength{\parindent}{0pt}
\begin{document}
Citation coranique : Sourate 2 (versets 2 à 4)
\begin{otherlanguage}{arabic}
لَا إِلَهَ إِلَّا اللَّهُ الْمَنَّانُ الَّذِي يُمْنُونَ بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنْفِقُونَ
وَالَّذِينَ يُؤْمِنُونَ بِمَا أُنزِلَ إِلَيْكَ وَمَا أُنزِلَ مِنْ قَبْلِكَ وَبِآخِرَةِ هُمْ يُوقِنُونَ
\end{otherlanguage}
\end{document}
```

Le mot `\AR{الْكِتَابُ}` (prononcer «alkitabou») désigne Le Kitab (l'Écrit, la prescription)

On note que :

- Pas `\usepackage` d'un paquet spécial à charger (seulement encodage LAE)
- L'environnement `\begin{otherlanguage}{arabic}` (bascule standard de `babel`) fait l'alignement à droite et la justification, sans interrompre le flot du texte.
- Pour mettre quelques mots arabes dans un paragraphe en alphabet latin, on utilise la macro `\AR{...}`, ou la bascule de `babel` `\foreignlanguage{arabic}{...}` (`\textarabic` ne marche pas). Et `\textlatin{...}` pour l'inverse.

Citation coranique : Sourate 2 (versets 2 à 4) هَيْفَ بَرِّ الْبَتَكُلَا كَلْذْ مَهَنْقَزَرِ أَمَمُو قَوْلَصِدَا نَوْمِيقِيُو بِيْعَلَا نُونْمُوِي نِيدَلَا نِيْقَتْمَلْ، يْدَهْ قَرْخَاءَلَايُو كَلْبِقْنِم لَزْدَا أَمُو كِيْلَا لَزْدَا أَمِبْ نُونْمُوِي نِيدَلَاو نُونْقَضِي نُونْقَوِي مَهْ

Le mot `الْكِتَابُ` (prononcer «alkitabou») désigne Le Kitab (l'Écrit, la prescription)

Ce code donne des erreurs de caractères unicodes non définis. J'ai fait

```
\DeclareUnicodeCharacter{0653}{}% Madda en chef
\DeclareUnicodeCharacter{0671}{i}% Alif wasla
\DeclareUnicodeCharacter{0670}{}% Diacr voy alif en chef
```

Table des matières

- 1 Les polices en TeX, bref historique
- 2 Textes en grec
- 3 Textes en arabe
- 4 Textes en hébreu