

Unicode, OpenType et X_YL^AT_EX

X_YL^AT_EX : porte d'entrée des fontes OpenType dans le monde T_EX

Michel Goossens

UDS-PUB

Section édition scientifique

Département IT, CERN (Genève, Suisse)

michel.goossens@cern.ch

Présentation donnée à GUT2007 (Paris) le 8 octobre 2007



Plan de l'exposé

Fontes : quelques rappels

Fontes OpenType et leurs outils

X_YTeX : introduction et historique

X_YTeX : principes de base

X_YTeX : gestion des fontes

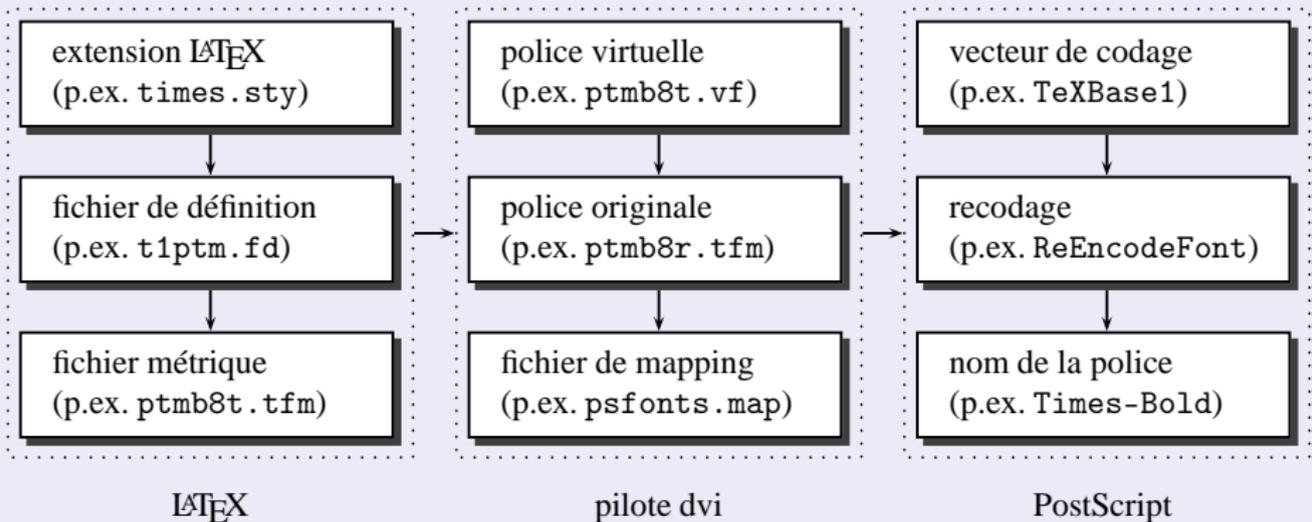
X_YTeX : modèle caractère/glyphe et algorithmes de composition

X_YLaTeX et ses extensions

X_YTeX et X_YLaTeX : exemples d'utilisation

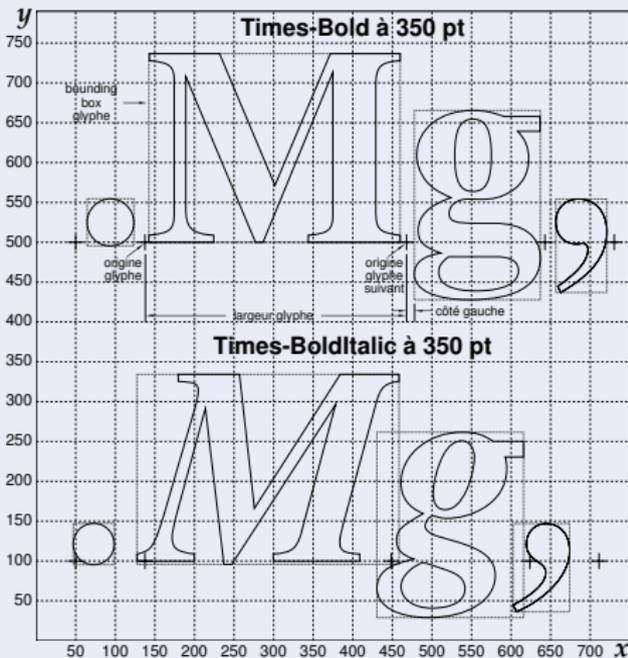
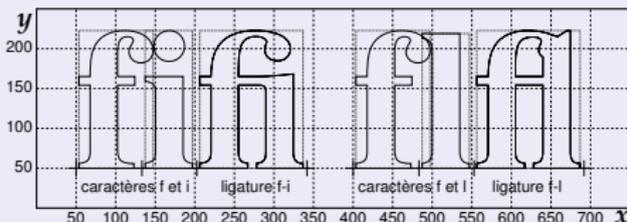
Conclusions

Un bref rappel : comment L^AT_EX gère les fontes



Les fichiers **TFM** et **AFM**

- ▶ informations générales
- ▶ paramètres globaux de la fonte
- ▶ les métriques des caractères
- ▶ les paires de ligatures et de crénage



OpenType : généralités

- ▶ conjointement développé par Adobe et Microsoft
 - store.adobe.com/type/opentype/main.html
 - www.microsoft.com/typography/default.mspx
- ▶ combine le meilleur des technologies TrueType et PostScript Type 1
- ▶ premières fontes OpenType en 1996
- ▶ depuis l'utilisation des fontes OpenType a fortement augmentée
- ▶ aujourd'hui pratiquement toutes les fontes sont en format OpenType

Une excellente source pour plus de détails :

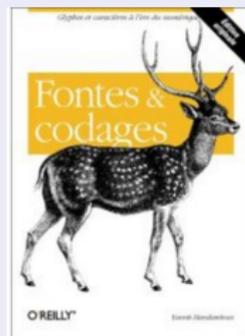
Yannis Haralambous

Fontes et codages

Editeur : O'Reilly (2004)

1012 pages

ISBN-13: 978-2841772735



OpenType : caractéristiques

- ▶ approche descriptive à la PostScript Type 1 avec ses spécifications des contours (*outlines*) et directives (*hints*)
- ▶ approche algorithmique à la TrueType avec ses tables et sous-programmes
- ▶ caractéristiques typographiques avancées, p.ex. tables pour le positionnement précis des glyphes et de leur substitution
- ▶ compatibilité avec toutes les plateformes informatiques
- ▶ excellente intégration avec Unicode
- ▶ bonne protection des données
- ▶ optimisation de la taille des fichiers (p.ex. fontes PostScript Type 1 codées en *Compact Font Format*)
- ▶ absence de limite pour le nombre de glyphes, c.-à-d. on peut avoir $n_{\text{glyphes}} > 65000$, ce qui est important pour ces fontes CJK

Tables OpenType

Tables obligatoires **cmap** (correspondances caractères/glyphes), **head** (information concernant la fonte), **hhea** (disposition horizontale), **hmtx** (métriques horizontales), **maxp** (besoins en mémoire), **name** (information supplémentaire), **OS/2** (informations métriques supplémentaires), **post** (information pour imprimantes PostScript).

Fontes de type TrueType **cvt**, **fpgm**, **glyf**, **loca**, **prep**.

Fontes de type PostScript Type 1 **CFF**, **VORG**.

Fontes bitmap **EBDT**, **EBLC**, **EBSC**.

Typographie avancée **BASE** (ligne de base), **GDEF** (définition des glyphes), **GPOS** (positionnement des glyphes), **GSUB** (données de substitution), **JSTF** (données de justification).

Autres tables **DSIG** (signature digitale), **gasp** (technique des rastérisation), **hdmx** (métriques prédéfinies pour imprimantes à résolution fixe), **kern** (crénage), **LTSH** (optimisation placement horizontal), **PCLT** (données pour imprimantes HP PCL 5), **VDMX** (optimisation placement vertical), **vhea** (métriques verticales globales), **vmtx** (métriques verticales des caractères).

Caractéristiques OpenType

- ▶ les « caractéristiques » (*features*) d'une fonte OpenType ou TrueType indiquent la façon dont une écriture ou une langue doit être rendue par le processeur
- ▶ p.ex. en arabe la forme d'une lettre dépend de sa position dans le mot, en écriture Kanji le positionnement des glyphes verticalement
- ▶ les caractéristiques de mise en page OpenType contiennent des données concernant le positionnement et la substitution des glyphes
- ▶ chaque caractéristique est identifiée par une *étiquette* (4 caractères ASCII), qui décrit sa fonction typographique et son effet

Pour une liste des (> 120) caractéristiques définies et leurs descriptions, voir :

www.microsoft.com/typography/otspec/featuretags.htm

partners.adobe.com/public/developer/opentype/index_tag3.html

Interroger les fontes OpenType

L'utilitaire `otfinfo` fait partie des outils `lcdf` d'Eddie Kohler (www.lcdf.org/type/)

```
> otfinfo --help
'Otfinfo' reports information about an OpenType font to standard output.
Options specify what information to print.

Usage: otfinfo [-sfzpg] [OTFFILES...]

Query options:
  -s, --scripts           Report font's supported scripts.
  -f, --features          Report font's GSUB/GPOS features.
  -z, --optical-size     Report font's optical size information.
  -p, --postscript-name  Report font's PostScript name.
  -a, --family           Report font's family name.
  -v, --font-version     Report font's version information.
  -i, --info             Report font's names and designer/vendor info.
  -g, --glyphs          Report font's glyph names.
  -t, --tables           Report font's OpenType tables.

Other options:
  --script=SCRIPT[.LANG] Set script used for --features [latn].
  -V, --verbose          Print progress information to standard error.
  -h, --help            Print this message and exit.
  -q, --quiet           Do not generate any error messages.
  --version             Print version number and exit.
```

```
> otfinfo --info texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf
Family:                Minion Pro
Subfamily:             Regular
Full name:             Minion Pro
PostScript name:      MinionPro-Regular
Version:              Version 2.012;PS 002.000;Core 1.0.38;makeotf.lib1.6.6565
Unique ID:            2.012;ADBE;MinionPro-Regular
Designer:             Robert Slimbach
Vendor URL:           http://www.adobe.com/type/
Trademark:            Minion is either a ...
Copyright:            © 2000, 2002, 2004 ...
License URL:          http://www.adobe.com/type/legal.html
```

Interroger les fontes OpenType

```
> otffinfo --script texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf
```

cyril	Cyrillic	latn.DEU	Latin/German (Standard)
grek	Greek	latn.MOL	Latin/Moldavian
latn	Latin	latn.ROM	Latin/Romanian
latn.AZE	Latin/Azeri	latn.SRB	Latin/Serbian
latn.CRT	Latin/Crimean Tatar	latn.TRK	Latin/Turkish

```
> otffinfo --tables texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf
```

64	BASE	54	head
132417	CFE	36	hhea
5228	DSIG	6652	hmtx
40074	GPOS	6	maxp
13872	GSUB	1533	name
96	OS/2	32	post
4048	cmap		

```
> otffinfo --features texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf
```

aalt	Access All Alternates	c2sc	Small Capitals From Capitals
case	Case-Sensitive Forms	csp	Capital Spacing
dlig	Discretionary Ligatures	dnom	Denominators
fin	Terminal Forms	frac	Fractions
hist	Historical Forms	kern	Kerning
liga	Standard Ligatures	lnum	Lining Figures
numr	Numerators	onum	Oldstyle Figures
ordn	Ordinals	ornm	Ornaments
pnun	Proportional Figures	salt	Stylistic Alternates
sinf	Scientific Inferiors	size	Optical Size
smcp	Small Capitals	ss01	Stylistic Set 1
ss02	Stylistic Set 2	sup	Superscript
tnun	Tabular Figures	zero	Slashed Zero

Visualiser le contenu d'une fonte OpenType

- ▶ `ttx` de Just van Rossum un utilitaire écrit en Python
- ▶ fait partie des outils `FontTools` (sourceforge.net/projects/fonttools)
- ▶ décompile le contenu d'une fonte OpenType et TrueType
- ▶ présente ses composants sous forme XML
- ▶ permet l'étude du contenu d'une fonte OpenType (p.ex. ses tables)
- ▶ représentation XML (`mafonte.ttx`) d'un fichier OpenType `mafonte.otf` :
> `ttx mafonte.otf`
- ▶ extraction uniquement des tables `GSUB` et `GPOS` :
> `ttx -t GSUB -t GPOS mafonte.otf`
- ▶ conversion d'un fichier XML `mafonte.ttx` en fonte OpenType ou TrueType :
> `ttx mafonte.ttx`
- ▶ introduction des changements explicités dans le fichier `mafonte2.ttx` dans la fonte `mafonte.otf` (en remplaçant certaines tables, si nécessaire)
> `ttx -m mafonte.otf mafonte2.ttx`

Visualiser le contenu d'une fonte OpenType

```
> ttx --help
usage: ttx [options] inputfile1 [... inputfileN]
```

TTX 2.0b1 -- From OpenType To XML And Back

If an input file is a TrueType or OpenType font file, it will be dumped to an TTX file (an XML-based text format).

If an input file is a TTX file, it will be compiled to a TrueType or OpenType font file.

Output files are created so they are unique: an existing file is never overwritten.

General options:

- h Help: print this message
- d <outputfolder> Specify a directory where the output files are to be created.
- v Verbose: more messages will be written to stdout about what is being done.

Dump options:

- l List table info: instead of dumping to a TTX file, list some minimal info about each table.
- t <table> Specify a table to dump. Multiple -t options are allowed. When no -t option is specified, all tables will be dumped.
- x <table> Specify a table to exclude from the dump. Multiple -x options are allowed. -t and -x are mutually exclusive.
- s Split tables: save the TTX data into separate TTX files per table and write one small TTX file that contains references to the individual table dumps. This file can be used as input to ttx, as long as the table files are in the same directory.

Compile options:

- m Merge with TrueType-input-file: specify a TrueType or OpenType font file to be merged with the TTX file. This option is only valid when at most one TTX file is specified.
- b Don't recalc glyph bounding boxes: use the values in the TTX file as-is.

Visualiser le contenu d'une fonte OpenType

```
> ttx -l texlive/2007/texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf
```

```
Listing table info for "/home/goossens/texlive/2007/texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf":
```

tag	checksum	length	offset	tag	checksum	length	offset
BASE	0x086729a7	64	199052	CFE	0x101232c2	132417	6032
DSIG	0x446dbd94	5228	199116	GPOS	0xx71552700	40074	158976
GSUB	0xx3bf7bcba	13872	145104	OS/2	0x40e57e9f	96	320
cmap	0x0cedc8f1	4048	1952	head	0xx2167aded	54	220
hhea	0x09140bb5	36	276	hmtx	0xx37425493	6652	138452
maxp	0x067f5000	6	312	name	0x3cf7b183	1533	416
post	0x0x47ffce	32	6000				

```
ttx -d. -t head texlive/2007/texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf
```

```
Dumping "texlive/2007/texmf-commercial/fonts/opentype/adobe/minionpro-regular.otf" to "/.minionpro-regular.ttx"...
```

```
Dumping 'head' table...
```

```
> less ./minionpro-regular.ttx
```

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<ttFont sfntVersion="OTTO" ttLibVersion="2.0b1">
```

```
<head>
```

```
<!-- Most of this table will be recalculated by the compiler -->
```

```
<tableVersion value="1.0"/>
```

```
<fontRevision value="2.0119934082"/>
```

```
<checksumAdjustment value="-0x107d913c"/>
```

```
<magicNumber value="0x5f0f3cf5"/>
```

```
<flags value="00000000 00000011"/>
```

```
<unitsPerEm value="1000"/>
```

```
<created value="Tue Jun 29 11:41:10 2004"/>
```

```
<modified value="Tue Jun 29 11:41:10 2004"/>
```

```
<xMin value="-290"/>
```

```
<yMin value="-360"/>
```

```
<xMax value="1684"/>
```

```
<yMax value="989"/>
```

```
<macStyle value="00000000 00000000"/>
```

```
<lowestRecPEM value="3"/>
```

```
<fontDirectionHint value="2"/>
```

```
<indexToLocFormat value="0"/>
```

```
<glyphDataFormat value="0"/>
```

```
</head>
```

```
</ttFont>
```

Les collections TrueType

- ▶ pour des raisons d'efficacité plusieurs fontes TrueType (OpenType) peuvent être regroupées dans une « collection » (`.ttc`)
- ▶ ces collections permettent aux différentes fontes la composant de partager des tables communes pour décrire les glyphes
- ▶ certains programmes ne parviennent pas à extraire les fontes individuelles d'une collection `ttc`
- ▶ un utilitaire `ttc2ttf` permet de créer des instances séparées `ttf`
- ▶ plusieurs programmes délèguent le traitement des différents formats de fontes à la bibliothèque `freetype` (<http://sourceforge.net/projects/freetype/>)
- ▶ le support des écritures complexes est délégué à `pango` (<http://www.pango.org/>)

Gérer les fontes OpenType

- ▶ sur le système Linux (*Cygwin* sur MS Windows) les fontes sont gérées par l'utilitaire `fontconfig` (<http://fontconfig.org/wiki/>)
- ▶ `fontconfig` est une bibliothèque qui permet la configuration, la personnalisation et la gestion des accès pour les applications des fontes installées sur le système
- ▶ il faut au moins `fontconfig` version 2.4 pour que X_YTeX fonctionne correctement
- ▶ l'information concernant les fontes est stockée sous forme d'un fichier XML (grammaire définie par la DTD `/etc/fonts/fonts.dtd`)
- ▶ ce fichier de configuration pour tout le système est `/etc/fonts/fonts.conf`
- ▶ ce fichier de configuration pour l'utilisateur est `~/.fonts.conf`.

Exemple d'un fichier `.fonts.conf` :

```
<?xml version="1.0"?>
<!DOCTYPE fontconfig SYSTEM "fonts.dtd">
<!-- /etc/fonts/fonts.conf file to configure system font access -->
<fontconfig>
<dir>/home/goossens/texlive/2007/texmf-update/fonts/opentype</dir>
<dir>/home/goossens/texlive/2007/texmf-commercial/fonts/opentype</dir>
<dir>/home/goossens/texlive/2007/texmf-dist/fonts/opentype</dir>
</fontconfig>
```

Gérer les fontes OpenType

La bibliothèque `fontconfig` vient accompagnée de trois programmes `fc-cache`, `fc-match`, `fc-list`.

```
> fc-list --help
```

```
usage: fc-list [-vV?] [--verbose] [--version] [--help] [pattern] element ...
```

```
List fonts matching [pattern]
```

```
-v, --verbose      display status information while busy
-V, --version      display font config version and exit
-?, --help         display this help and exit
```

```
> fc-cache --help
```

```
usage: fc-cache [-frsvV?] [--force|--really-force] [--system-only] [--verbose] [--version] [--help]
```

```
Build font information caches in [dirs]
```

```
(all directories in font configuration by default).
```

```
-f, --force        scan directories with apparently valid caches
-r, --really-force erase all existing caches, then rescan
-s, --system-only  scan system-wide directories only
-v, --verbose      display status information while busy
-V, --version      display font config version and exit
-?, --help         display this help and exit
```

```
> fc-match --help
```

```
usage: fc-match [-svV?] [--sort] [--verbose] [--version] [--help] [pattern]
```

```
List fonts matching [pattern]
```

```
-s, --sort          display sorted list of matches
-v, --verbose      display entire font pattern
-V, --version      display font config version and exit
-?, --help         display this help and exit
```

LaTeX et la fonte OpenType *Adobe Minion* (1)

- ▶ la famille complète *Minion Pro* vendue par Adobe se compose de cinq parties : *Minion Pro* (8 fontes), *Minion Pro Opticals* (32 fontes), *Minion Pro Condensed* (8 fontes), *Minion Pro Condensed Opticals* (32 fontes) et *Minion Std Black* (1 fonte)
- ▶ `acroread` d'Adobe vient avec quatre de ces fontes
`/usr/lib/acroread/Resource/Font/` (Linux)
`C:\Program Files\Adobe\Acrobat~7.0\Resource\Font` (*Microsoft Windows*)
- ▶ Achim Blumensath, Andreas Böhmann et Michael Zedler ont préparé tous les fichiers nécessaires à l'utilisation de la famille *Minion* (`CTAN:fonts/minionpro`)
- ▶ pour les quatre fontes de base (disponibles gratuitement) utiliser le fichier `base-v2.zip` et l'installer dans l'arborescence TeX
- ▶ l'opération installera les fichiers `enc`, `fd`, `tfm` et `vf` nécessaires. Également installés sont `MinionPro.map` et les extensions `fontaxes.sty`, `MinionPro-FontDef.sty`, `MinionPro.sty`, `mt-MinionPro.cfg`, et `otfontdef.sty`

Λ_TE_X et la fonte OpenType *Adobe Minion* (2)

- ▶ pour utiliser les fontes OpenType avec Λ_TE_X standard une version PostScript Type 1 est générée à l'aide de scripts qui sont disponibles dans la distribution (`scripts.zip`)
- ▶ en particulier sur Linux cela devient :
 - > `cfftot1 -b MinionPro-Regular.otf -o MinionPro-Regular.pfb`
 - > `t1dotlessj MinionPro-Regular.pfb -o MinionPro-RegularLCDFJ.pfb`idem pour `MinionPro-Bold`, `MinionPro-BoldIt`, et `MinionPro-It`
- ▶ `t1dotlessj` lit un fichier PostScript Type 1 standard, qui ne contient pas le caractère « j sans point » et génère ce caractère en coupant le point du « j », puis le sauve dans un nouveau fichier PostScript Type 1 contenant uniquement le caractère `dotlessj`
- ▶ finalement il faut copier tous ces fichiers à l'endroit adéquat et mettre à jour la base de données T_EX
 - [1] > `texhash`
 - [2] > `updmap --enable Map MinionPro.map`

Un exemple complet de *Minion Pro* et *MnSymbol*

options par défaut (**MinionPro**)

1 Общие сведения

Заметьте разницу в стиле верстки выражений в абзацах и выключных: $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934$:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934. \quad (1)$$

Строчные греческие буквы вводятся как `\alpha`, `\beta`, `\gamma`, `\dots`, **прописные буквы** вводятся как `\Gamma`, `\Delta`, `\dots`: α , β , γ , λ , ξ , π , μ , Φ , Ω .

Оператор интеграла печатает команда `\int` (\int_0^π).

$$\int_0^\pi \sqrt{\pi^2 - x^2} dx = \frac{\pi^3}{4} \quad (2)$$

Окружение `array` можно также использовать для верстки выражений, имеющих один большой ограничитель, подставляя `\right.` в качестве невидимого правого ограничителя:

$$y = \begin{cases} +1 & \text{если } d > c, \\ -1 & \text{по утрам,} \\ 0 & \text{остальное время дня.} \end{cases}$$

2 Généralités

Remarquez le format différent des expressions mathématiques composées « en ligne » ou « hors texte » : $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934$,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934. \quad (3)$$

Les lettres **grecques minuscules** sont saisies de la manière suivante : `\alpha`, `\beta`, `\gamma`, etc. Les lettres **grecques majuscules** sont saisies ainsi : `\Gamma`, `\Delta`, etc. : α , β , γ , λ , ξ , π , μ , Φ , Ω .

La commande `\int` produit une **intégrale** (\int_0^π).

$$\int_0^\pi \sqrt{\pi^2 - x^2} dx = \frac{\pi^3}{4} \quad (4)$$

L'environnement `array` peut également être utilisé pour imprimer des expressions qui ont un délimiteur invisible obtenu par la commande `\right.` :

$$y = \begin{cases} +1 & \text{si } d > c, \\ -1 & \text{le matin,} \\ 0 & \text{la journée.} \end{cases}$$

Un exemple complet de *Minion Pro* et *MnSymbol*

Options sélectionnées : frenchmath, lf, minionint

1 Общие сведения

Заметьте разницу в стиле верстки выражений в абзацах и выключных: $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934$:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934. \quad (1)$$

Строчные греческие буквы вводятся как `\alpha`, `\beta`, `\gamma`, ..., **прописные буквы** вводятся как `\Gamma`, `\Delta`, ...: α , β , γ , λ , ξ , π , μ , Φ , Ω .

Оператор интеграла печатает команда `\int` (\int_0^π).

$$\int_0^\pi \sqrt{\pi^2 - x^2} dx = \frac{\pi^3}{4} \quad (2)$$

Окружение `array` можно также использовать для верстки выражений, имеющих один большой ограничитель, подставляя `\right.` в качестве невидимого правого ограничителя:

$$y = \begin{cases} +1 & \text{если } d > c, \\ -1 & \text{по утрам,} \\ 0 & \text{остальное время дня.} \end{cases}$$

2 Généralités

Remarquez le format différent des expressions mathématiques composées « en ligne » ou « hors texte » : $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934$,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} = 1.644934. \quad (3)$$

Les lettres **grecques minuscules** sont saisies de la manière suivante : `\alpha`, `\beta`, `\gamma`, etc. Les lettres **grecques majuscules** sont saisies ainsi : `\Gamma`, `\Delta`, etc. : α , β , γ , λ , ξ , π , μ , Φ , Ω .

La commande `\int` produit une **intégrale** (\int_0^π).

$$\int_0^\pi \sqrt{\pi^2 - x^2} dx = \frac{\pi^3}{4} \quad (4)$$

L'environnement `array` peut également être utilisé pour imprimer des expressions qui ont un délimiteur invisible obtenu par la commande `\right.` :

$$y = \begin{cases} +1 & \text{si } d > c, \\ -1 & \text{le matin,} \\ 0 & \text{la journée.} \end{cases}$$

Collection de fontes TeX Gyre

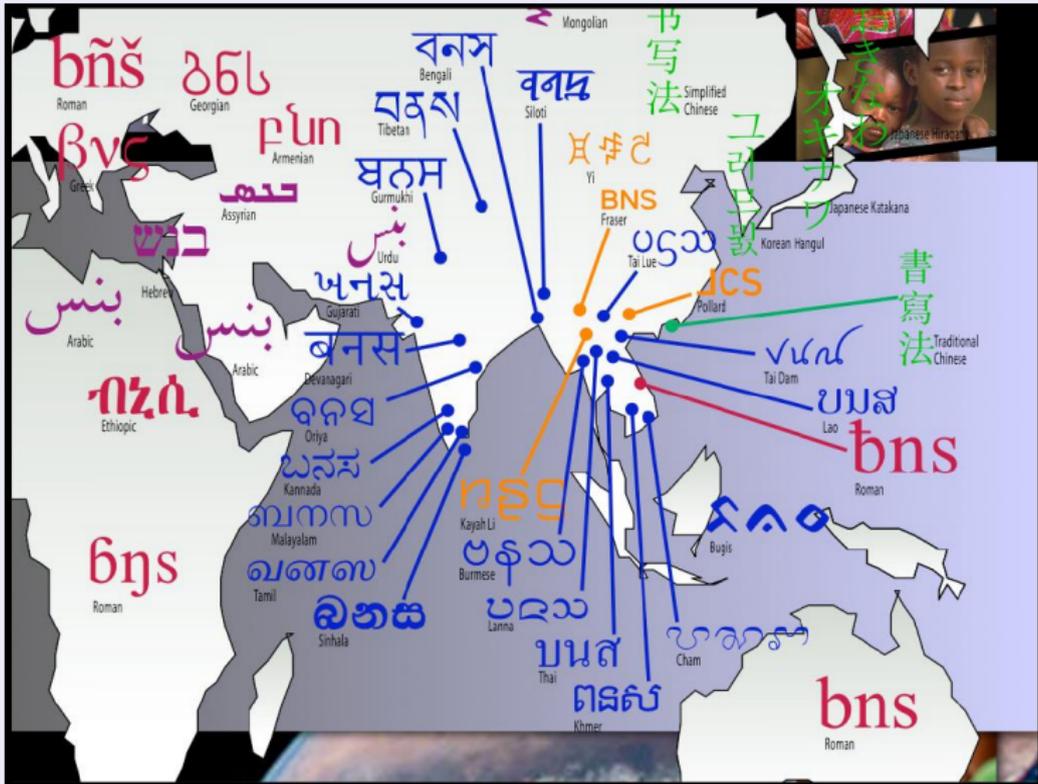
- ▶ projet « New Font Project: TeX Gyre » initié par GUST (co-financé par DANTE e.V., NTG, TUG, CS TUG, TUG India et GUST)
- ▶ but annoncé : améliorer la qualité et augmenter le nombre de caractères des familles de fontes distribuées librement avec Ghostscript
 - ▶ *TeX Gyre Adventor* famille sans empattement basée sur URW Gothic
 - ▶ *TeX Gyre Bonum* famille avec empattement basée sur URW BookmanL
 - ▶ *TeX Gyre Cursor* famille à chasse-fixe basée sur URW Nimbus Mono
 - ▶ *TeX Gyre Heros* famille sans empattement basée sur URW Nimbus Sans
 - ▶ *TeX Gyre Pagella* famille avec empattement basée sur URW Palladio
 - ▶ *TeX Gyre Schola* famille avec empattement basée sur URW Century Schoolbook
 - ▶ *TeX Gyre Termes* famille avec empattement basée sur Nimbus Roman No9
- ▶ ce développement d'une version OpenType est particulièrement intéressant pour Xe_ΛTeX
- ▶ jusqu'à présent peu de travail dans le domaines des symboles mathématiques, mais l'intention d'améliorer la situation existe
- ▶ ces fontes permettent la composition de textes techniques sans trop de formules mathématiques complexes

Xe_lTeX : une extension de T_EX pour Unicode et OpenType

- ▶ SIL (à l'origine *Summer Institute of Linguistics*, voir <http://www.sil.org>) est une organisation religieuse chrétienne fondée il y a 70 ans
- ▶ son but est d'améliorer la communication par la diffusion des documents dans un maximum de langues (écritures). SIL publie *Ethnologue*, *languages of the world* (<http://www.ethnologue.com/>)
Ce livre de 1272 pages décrit les 6912 langues parlées dans le monde
- ▶ ses services se disent indépendants des convictions religieuses et politiques, de la race, du sexe ou de l'ethnie des correspondants
- ▶ Jonathan Kew, dans le cadre de ses activités à SIL, a développé Xe_lTeX (scripts.sil.org/xetex)
- ▶ une version stable pour plusieurs plateformes informatiques était diffusée avec TeXlive 2007
- ▶ une version précompilée à jour pour *Microsoft Windows* ([xetex-w32.tar.bz2](http://www.fsci.fuk.kindai.ac.jp/kakuto/win32-ptex/web2c75-e.html)) est téléchargeable depuis l'URL :
<http://www.fsci.fuk.kindai.ac.jp/kakuto/win32-ptex/web2c75-e.html>



Les écritures d'Asie



Des langues plus ou moins complexes

è è̃ ï Ì̃ ل لَا بَقَ [6] [3]

ب ب م م م م

ക + റ = കറ

കാത കിത കോത
kat kit kot

Xe_{La}TeX : un brèf historique

- ▶ avril 2004 : Xe_{La}TeX 0.3 disponible (Mac OS X uniquement)
 - ▶ support intégré pour Unicode
 - ▶ accès direct à toutes les fontes installées sur l'ordinateur
 - ▶ AAT (Apple Advanced Typography) pour les caractéristiques typographiques
 - ▶ support graphique avec *Quicktime*
- ▶ février 2005 : version 0.9 de Xe_{La}TeX
 - ▶ support pour Opentype
 - ▶ cohabite correctement avec extensions importantes de L^AT_EX
- ▶ avril 2006 (Bacho_{TeX}) : Xe_{La}TeX pour Linux (première annonce publique de la disponibilité de Xe_{La}TeX)
- ▶ juin 2006 : Akira Kakuto annonce la disponibilité de Xe_{La}TeX pour MS Windows
- ▶ février 2007 : T_EX live 2007 contient Xe_{La}TeX 0.996 pour toutes les plateformes binaires où T_EX est distribué

Xe_ΛTeX : les bases

- ▶ moteur de composition basé sur eTeX
- ▶ extensions TeX--XeT (commandes `\beginL`, `\endL`, `\beginR` et `\endR` activées par `\TeXXeTstate=1`) pour composition bi-directionnelle (l'arabe, l'hébreu, etc.)
- ▶ utilise Unicode comme codage par défaut
- ▶ offre un environnement multiécriture qui simplifie considérablement le traitement multilingue
- ▶ la plupart des extensions L^AT_EX sont compatibles avec Xe_ΛTeX, p.ex. `graphics`, `xcolor`, `geometry`, `crop`, `hyperref`, `pgf` détectent automatiquement la présence du moteur Xe_ΛTeX
- ▶ utilisation directe de fontes OpenType, TrueType et PostScript sans installation de fichiers spécifiques pour T_EX (`tfm`, `vf`, `fd`, etc.)
- ▶ utilisation des caractéristiques OpenType (ligatures, swash, glyphes alternatifs, attachement dynamique d'accents, etc.)
- ▶ utilisation à travers Unicode de fontes standard contenant les caractères pour les alphabets (latin, cyrillique, grec, etc.) étendus, des alphabets non-latins et les écritures complexes

Xe_{La}TeX : césures

une couche supplémentaire `xu-hyphen` permet de charger les motifs de césures T_EX en Unicode, p.ex. : (répertoire `texmf-dist/tex/generic/xu-hyphen`)

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% xu-frhyph.tex (Wrapper for XeTeX to read frhyph.tex)
\begingroup
\expandafter\ifx\csname XeTeXrevision\endcsname\relax
\else
  % frhyph.tex uses ^^xx for T1 characters
  % redefine them to access the required Unicode characters
  % (only \oe{} actually matters here!)
  \input xu-t1.tex
\fi
\input frhyph.tex
\endgroup

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% xu-t1.tex %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% make T1 letters \active and map them to Unicode character codes
% (for use when loading hyphenation patterns that use ^^xx notation
% to represent characters in T1 font encoding, or literal 8-bit
% bytes if read using \XeTeXinputencoding "bytes")
\catcode`\=12 % ensure = isn't active or otherwise "weird"
\catcode`\^=7 % ensure ^ is the proper catcode for hex notation
%
\catcode"B0=\active \def^^b0{~~~~0159} % rcaron
...
\catcode"DF=\active \def^^df{SS} % SS
\catcode"F7=\active \def^^f7{~~~~0153} % oe
\catcode"D7=\active \def^^d7{~~~~0152} % OE
% we don't handle the non-letter codes in the control range
% but we'd better handle dotless-i (for Turkish)
\catcode"i9=\active \def^^i9{~~~~0131} % dotlessi
    
```

Composition multilingue avec T_EX

- ▶ la saisie des textes
 - ▶ commandes pour caractères non-ASCII
 - ▶ plusieurs codages 8-bits
 - ▶ préprocesseurs pour écritures complexes
- ▶ les fontes
 - ▶ limitées à 256 caractères adressables
 - ▶ codages spécifiques pour utiliser des glyphes spéciaux
- ▶ système complexe de macros imbriqués
 - ▶ difficile à comprendre et étendre
 - ▶ difficile à intégrer avec d'autres extensions
- ▶ exemples :
 - ▶ CJK de Werner Lemberg (`CTAN:language/chinese/CJK/`)
 - ▶ `arabtex` de Klaus Lagally (`CTAN:language/arabic/arabtex/`)

Conventions de saisie traditionnelles de T_EX

- ▶ Texte saisi est en ASCII (codage 8-bits)

<i>texte saisi</i>	<i>texte composé</i>	<i>notes</i>
<code>\'{a}</code>	á	commandes typiques « accents »
<code>\c{c}</code>	ç	
<code>\aa</code>	å	
<code>---</code>	—	ligature spécifique fontes T _E X
<code>\$_alpha\$</code>	α	symbole mathématique
<code>{\dn acchaa}</code>	अच्छा	préprocesseur <i>ad hoc</i>

Xe_ΛTeX propose une approche unifiée

- ▶ Unicode : caractères directement adressables
 - ▶ plus besoin de commandes spéciales pour adresser des caractères absents du codage (8-bit) courant
 - ▶ plus besoin de commuter entre codages différents en fonction de la langue/écriture
 - ▶ tous les caractères ont un code standard «unique »
 - ▶ exemples de caractères composés (entrée, résultat)
džabe džabe đak đak
džin džin Džin Džin
- ▶ utilisation d'un modèle caractère/glyphe et d'une technologie moderne pour rendre une fonte
 - ▶ écritures complexes traitées par un processeur externe travaillant sur le fichier texte à l'entrée

Composer les idéographes (CJK)

```
\font\han="STSong:color=660000" at 14pt
\font\rom="Gentium:color=006600" at 8pt
\newcommand\hc[2]{\begin{tabular}{l}
  \han #1\\[-1mm]\rom #2\end{tabular}}
\begin{tabular}{l}
  \hc{書<}{ka-ku}\\
  \hc{最も}{motto-mo}\\
  \hc{最後}{sai-go}\\
  \hc{働<}{hatara-ku}\\
  \hc{海}{umi}
\end{tabular}
```

書<

ka-ku

最も

motto-mo

最後

sai-go

働<

hatara-ku

海

umi

Composer des textes complexes : l'arabe

```
\color[rgb]{0,0,1}
\begin{arab}[utf]
\begin{center}
ش ئاڊي پ ڇ اڀنڊ \\\[3mm]
\end{center}
\begin{flushleft}
\fbbox{۱} نيمز ادخ ۾ تاعورش ۽
\fbbox{۲} بېترتڀي نيمز تقو نا
۽ ڊنم س ڀهنوا . ڀئ ه نارڀو ۽
وه لڀڪڀ ناس هڊنوا ورڇاٿم چو
ادخ ناٿم ڀڇ ڪڀڻاڀ ۽
\fbbox{۳} ڀنشور ”ات ونڌ مڪح ادخ نهڌت“
ڀئ ڀڀ ڀڻ ڀنشور وس .“ ڀئڻٿ
\end{flushleft}
\end{arab}
```

دنيا جي پيدائش

۱ شروعات ۾ خدا زمين ۽ آسمان کي پيدا ڪيو.

۲ ان وقت زمين بيترتيب ۽ ويران هئي .
اونهي سمنڊ جو مٿاڇرو اوندهه سان
ڍڪيل هو ۽ پاڻي جي مٿان خدا جي
روح ڦيرا پئي ڪي

۳ تڏهن خدا حڪم ڏنو ته ”روشني
ٿي.“ سو روشني ٿي پئي .

Spécifier les codes des caractères

- ▶ caractères sont codés sur 16 bits
 - ▶ codage UTF-16 d'Unicode utilisé
 - ▶ exception : quelques anciennes fontes codées différemment
- ▶ extension des primitives T_EX
 - ▶ `\char`, `\chardef` acceptent des nombres jusqu'à 65 536
 - ▶ notation à quatre chiffres avec la syntaxe `~~~~abcd`
`\char"5609~~~~6167 = 嘉慧`
- ▶ les caractères Unicode dans les plans supérieurs (> 0)
 - ▶ utilisation de « surrogates » (standard UTF-16)
 - ▶ convient pour la composition
 - ▶ ne convient pas pour manipuler le texte à l'entrée au niveau du caractère individuel
- ▶ élargissement des tables contenant les codes individuels des caractères `\catcode`, `\lccode`, `\uccode`, `\sfcode`
 - ▶ « Xe_lTeX plain » initialise ses tables avec les codes Unicode
 - ▶ `\lowercase{DŽIN}` džin
 - ▶ `\uppercase{Esi eyama klo míafe nuvowo da vo la}`
ESI EYAMA KLO MÍAFE NUVOWO ĐA VO LA
 - ▶ `\catcode`\王=\active \def王{...}`

Les codages à l'entrée et les césures

- ▶ par défaut le codage à l'entrée est Unicode (UTF-8 ou UTF-16)
 - ▶ le format du codage est détecté automatiquement
- ▶ textes non-Unicode :
 - ▶ codages historiques traités via des convertisseurs ICU
 - ▶ spécification du codage du fichier à l'entrée :
`\XeTeXinputencoding "nom du codage"`
 - ▶ spécification du codage des prochains fichiers lus :
`\XeTeXdefaultencoding "nom du codage"`
- ▶ motifs de césure
 - ▶ extension pour 16 bits
 - ▶ fichiers de césures T_EX standard dépendent du codage utilisé → modifiés pour fonctionner avec Xe_{La}TeX
 - ▶ les césures sont spécifiées comme des données simples, pas besoin de commandes, de caractères actifs, etc. → motifs simplifiés (p.ex. Devanagari)

`% césure avant et suivant une voyelle indépendante`

`1ॐ1`

`1ॐ1`

`1ॐ1`

`% césure suivant une voyelle indépendante, mais jamais avant`

`2ॐ1`

`2ॐ1`

Les fontes présentes sur la plateforme informatique

- ▶ X_YTeX permet l'utilisation de toutes les fontes (PostScript Type 1, TrueType, OpenType) installées sur l'ordinateur où X_YTeX tourne
- ▶ extension de la commande `\font` permettant de spécifier les « vrais » noms des fontes
 - ▶ `\font\rm="Adobe Caslon Pro" at 14pt \rm Bonjour GUT2007 !`
Bonjour GUT2007 !
 - ▶ `\font\it="Trebuchet MS" at 14pt \it Bonjour GUT2007 !`
Bonjour GUT2007 !
 - ▶ `\font\ch="Viva Std" at 14pt \ch Bonjour GUT2007 !`
Bonjour GUT2007 !
- ▶ plus besoin de `.tfm`, `.vf`, `.fd`, etc.

Les fontes traditionnelles T_EX

- ▶ Xe_ΛTeX permet toujours l'utilisation des fichiers `.tfm`
 - ▶ requis pour les fontes mathématiques
 - ▶ données non-Unicode : codes 0..255 seuls utilisés
- ▶ post-processeur PDF (`xdvipdfmx` par défaut sur Linux) peut utiliser les fontes PostScript Type 1
 - ▶ utilise les fontes `.pfb` présente dans l'arborescence `texmf` (à la `dvips`)
 - ▶ pas de support pour les fontes bitmap
 - ▶ actuellement pas de support direct pour les fichiers `.vf`
 - ▶ a accès aux mêmes fontes que le moteur de composition (Xe_ΛTeX)
 - ▶ génère du PDF par défaut (un format intermédiaire de type `.xdv`, « DVI étendu », est disponible)
 - ▶ seul le sous-ensemble des caractères d'une fonte actuellement référencés est inclus dans le fichier PDF

Correspondance des caractères

- ▶ conventions habituelles T_EX

```\TeX'---`un système de composition

génère : “T<sub>E</sub>X”---un système de composition

- ▶ il existe une option de *compatibilité*

`; TECKit mapping for TeX input conventions`

`U+002D U+002D <> U+2013 ; -- -> en dash`

`U+002D U+002D U+002D <> U+2014 ; --- -> em dash`

`U+0027 <> U+2019 ; ' -> right single quote`

`U+0027 U+0027 <> U+201D ; '' -> right double quote`

`U+0022 > U+201D ; " -> right double quote`

qui génère : “T<sub>E</sub>X”—un système de composition

# OpenType : support pour les langues et les écritures

- ▶ certains caractères ont une représentation qui varie avec la langue en question

```
\font\Doulos="Doulos SIL/ICU"
```

```
\font\DoulosViet="Doulos SIL/ICU:language=VIT"
```

Unicode cung cấp một con số  
duy nhất cho mỗi ký tự

Unicode cung cấp một con số  
duy nhất cho mỗi ký tự

```
\font\Minion="Minion Pro"
```

```
\font\MinionTrk="Minion Pro:language=TRK"
```

... gelen firmaları ...  
tarafından ...

... gelen firmaları ...  
tarafından ...

- ▶ certaines écritures asiatiques complexes nécessitent un moteur spécifique pour dessiner la forme des caractères

- ▶ `\font\x="Code2000:script=arab" \x` العربي `العربي`

- ▶ `\font\x="Code2000:script=deva" \x` हिन्दी `हिन्दी`

## OpenType : caractéristiques optionnels

- ▶ une définition de fonte peut inclure une ou plusieurs spécifiant des caractéristiques

- ▶ `\font\x="Minion Pro" \x Bonjour GUT2007 ! 0123456789`

Bonjour GUT2007 ! 0123456789

- ▶ `\font\x="Minion Pro:+smcp"`

BONJOUR GUT2007 ! 0123456789

- ▶ `\font\x="Minion Pro:+sups"`

Bonj<sup>o</sup>ur GUT<sup>2007</sup> ! 0123456789

- ▶ `\font\x="Minion Pro Italic:+onum"`

*Bonjour GUT2007 ! 0123456789*

- ▶ `\font\x="Minion Pro Italic:+swsh,+zero"`

*Bonjour GUT2007 ! 0123456789*

## OpenType : contrôler le corps des caractères

- ▶ certaines fontes OpenType se composent d'une série de corps ayant les corrections optiques adéquates. Un changement de taille charge automatiquement le corps correspondant.

- ▶ Minion Pro avec corps de 7, 10, 18, 24pt

sept dix dix-huit vingt-quatre

- ▶ ce choix peut être outrepassé à l'aide du paramètre `/S=` qui force le chargement du corps donné (ici montré à 16 pt)

Minion Pro/S=7

Minion Pro Caption

Minion Pro/S=10

Minion Pro Text

Minion Pro/S=18

Minion Pro Subhead

Minion Pro/S=24

Minion Pro Display

## Les points de coupure des lignes

- ▶ couper les lignes en l'absence d'espaces entre les mots
  - ▶ TeX coupe les lignes normalement à la « colle » associée aux espaces
  - ▶ Le chinois, le japonais, le Thai, etc. n'utilisent pas d'espaces entre les mots
  - ▶ โดยพื้นฐานแล้ว, คอมพิวเตอร์จะเกี่ยวข้องกับเรื่องของตัวเลข. คอมพิวเตอร์จัดการเก็บตัวอักษรและอักขระอื่นๆ โดยการกำหนดหมายเลขให้สำหรับแต่ละตัว. ก่อนหน้าที่Unicode จะถูกสร้างขึ้นมาได้มีระบบ encoding อยู่หลายร้อยระบบสำหรับการกำหนดหมายเลขเหล่านี้
- ▶ utiliser la méthode de coupure de la librairie ICU :  
`\XeTeXlinebreaklocale "th"`
  - ▶ โดยพื้นฐานแล้ว, คอมพิวเตอร์จะเกี่ยวข้องกับเรื่องของตัวเลข. คอมพิวเตอร์จัดการเก็บตัวอักษรและอักขระอื่นๆ โดยการกำหนดหมายเลขให้สำหรับแต่ละตัว. ก่อนหน้าที่Unicode จะถูกสร้างขึ้น, ได้มีระบบ encoding อยู่หลายร้อยระบบ สำหรับการกำหนด หมายเลข เหล่า นี้.

## Justification

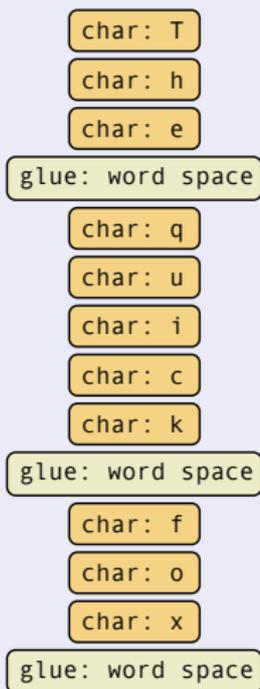
- ▶ la justification d'un texte sans espaces, y compris la coupure des lignes est une tâche non-triviale
- ▶ une solution est la composition en drapeau
  - ▶ 基本上，计算机只是处理数字。它们指定一个数字，来储存字母或其他字符。在创造Unicode之前，有数百种指定这些数字的编码系统。没有一个编码可以包含足够的字符：
- ▶ Alternativement l'utilisation de la commande `\XeTeXlinebreakskip` permet l'introduction de colle à chaque point de coupure potentiel
  - ▶ 基本上，计算机只是处理数字。它们指定一个数字，来储存字母或其他字符。在创造Unicode之前，有数百种指定这些数字的编码系统。没有一个编码可以包含足够的字符：

## Mise en œuvre du modèle caractère/glyphe (1)

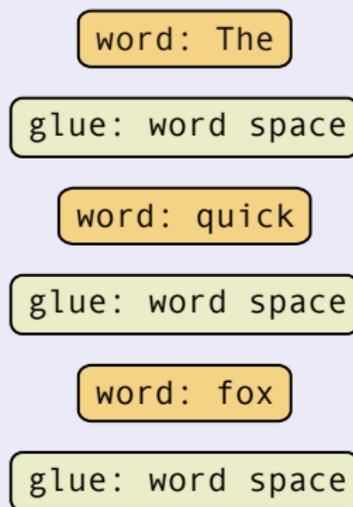
- ▶ exploiter optimalement les caractéristiques des fontes OpenType codées en Unicode
- ▶ changement radical du modèle T<sub>E</sub>X traditionnel
  - ▶ l'unité fondamentale de T<sub>E</sub>X est le « code spécifique d'un caractère dans une fonte donnée »
  - ▶ il est supposé qu'un tel caractère a des dimensions connues et invariables
  - ▶ les ligatures sont traitées par des substitutions de caractères
  - ▶ un paragraphe est construit à partir d'une séquence de nœuds de « caractères », à placer avec beaucoup de précision, et des nœuds de « colle »
- ▶ un caractère Unicode ne correspond pas nécessairement à un seul glyphe unique
  - ▶ plusieurs écritures nécessitent une sélection contextuelle des glyphes (p.ex. l'arabe)
  - ▶ les dimensions d'un caractères doivent être déterminés en tenant compte du contexte (pas en isolement)
- ▶ réalisation en Xe<sub>Λ</sub>T<sub>E</sub>X
  - ▶ le processus de composition collecte les suites de caractères (mots)
  - ▶ une interface de programmation du système (API) donne accès à leurs dimensions
  - ▶ un paragraphe Xe<sub>Λ</sub>T<sub>E</sub>X est une séquence de nœuds de « mots » séparés par de la « colle »
- ▶ le moteur de composition positionne les mots, pas les glyphes
  - ▶ ce qui incombe au moteur qui rend les fontes

## Mise en œuvre du modèle caractère/glyphe (2)

T<sub>E</sub>X : nœuds dans un paragraphe



Xe<sub>Λ</sub>T<sub>E</sub>X : nœuds dans le même paragraphe



## Mise en œuvre du modèle caractère/glyphe (3)

- ▶ support Unicode avec la bibliothèque ICU
  - ▶ ICU (*International Components for Unicode*, voir <http://www.icu-project.org/>) est une bibliothèque complète et hautement portable (des version C/C++ et Java existent) offrant un support avancé pour Unicode et l'internationalisation des logiciels qui est disponible sur pratiquement toutes les plateformes informatiques
  - ▶ traite la plupart des écritures gracieusement (langues latines accentuées ou celles du subcontinent indien et de l'Extrême-Orient)
- ▶ Xe<sub>ℒ</sub>TeX utilise ATSUI (l'équivalent de ICU sur Mac OS X) ou ICU d'après le format des tables présentes dans une fonte donnée et `fontconfig` pour la localiser
  - ▶ le processus de composition est indépendant de la technologie des fontes utilisées
  - ▶ la distinction se fait uniquement au niveau le plus bas où les dimensions des caractères dans une fonte doivent être déterminées
  - ▶ possibilité de mélanger les fontes AAT et OpenType
- ▶ `xdvipdfmx` utilise `freetype` ([www.freetype.org/](http://www.freetype.org/)), un moteur informatique pour rendre les images des glyphes avec grande précision

# Xe<sub>Λ</sub>TeX : traitement de la césure des mots (1)

- ▶ un paragraphe est construit à partir d'une liste de « boîtes contenant des mots »
  - ▶ ces boîtes sont considérées comme des unités indivisibles dans la liste de jetons (*tokens*)
  - ▶ TeX ne doit pas s'occuper des détails à bas niveau
- ▶ si une coupure de ligne acceptable ne peut être trouvée, on doit envisager la césure d'un mot
  - ▶ extraire les caractères des nœuds de « mots »
  - ▶ trouver les positions de césures en utilisant l'algorithme TeX
  - ▶ remettre en boîte les mots comme de fragments de mots et de nœuds de césures discrétionnaires

## X<sub>Y</sub>TeX : traitement de la césure des mots (2)

- ▶ modifier la liste de nœuds pour permettre la césure de mots

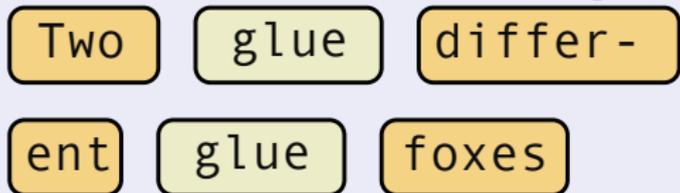


- ▶ problème : les points de césure non utilisés interrompent l'image de représentation visuelle



*Two differ-  
ent foxes*

- ▶ il faut refusionner les nœuds des mots après le choix des points de césures



*Two differ-  
ent foxes*

## Récapitulation Xe<sub>La</sub>TeX : spécifier les fontes

<i>Nom spécifié</i>	<i>fonte utilisée</i>
"Times New Roman/B"	<b>Times Bold</b>
"Times New Roman/I"	<i>Times Italic</i>
"Times New Roman/BI"	<b><i>Times BoldItalic</i></b>

<i>Nom spécifié</i>	<i>fonte utilisée</i>
"Warnock Pro" at 7pt	WarnockPro-Capt
"Warnock Pro" at 11pt	Warnock Pro
"Warnock Pro" at 16pt	WarnockPro-Subh

<i>Nom et corps optique spécifiés</i>	<i>fonte utilisée</i>
"Warnock Pro/S=7" at 12pt	WarnockPro-Capt
"Warnock Pro/S=10" at 12pt	WarnockPro-Regular
"Warnock Pro/S=18" at 12pt	WarnockPro-Subh
"Warnock Pro/S=36" at 12pt	WarnockPro-Disp
"Warnock Pro Subhead" at 12pt	WarnockPro-Capt
"Warnock Pro Subhead/S=0" at 12pt	WarnockPro-Subh

Le troisième ensemble d'exemples montre comment imposer un corps optique et aussi qu'une fonte peut avoir plusieurs noms.

## Extensions pour Xe<sub>Λ</sub>TeX

- ▶ l'extension `fontspec` de Will Robertson est une interface  $\text{\LaTeX}$  pour la sélection des fontes OpenType (`CTAN:macros/xetex/latex/fontspec.zip`)

- ▶ exemple d'utilisation

```
\usepackage{euler,fontspec,xltxtra,xunicode}
\defaultfontfeatures{Scale=MatchLowercase}
\setromanfont{Minion Pro}
\setsansfont{Myriad Pro}
\setmonofont[Color=\MYCOLOR]{Courier Std}
```

- ▶ l'extension `xunicode` de Ross Moore :
  - ▶ redéfinit les commandes  $\text{\LaTeX}$  pour un grand nombre de caractères spéciaux en séquences Unicode
  - ▶ ne traite pas des tirets et les quotes (spécifier l'option `tex-text` de `fontspec`)
  - ▶ permet ainsi de traiter un document  $\text{\LaTeX}$  non-Unicode avec Xe<sub>Λ</sub>TeX sans problèmes
- ▶ l'extension `xltxtra` de Will Robertson complète `fontspec` avec quelques commandes supplémentaires

# Xe<sub>l</sub>TeX et fontspec (1)

```
\fontspec{Myriad Pro}
normal : Myriad Pro 0123456789 \\
\defaultfontfeatures{Numbers=OldStyle, Colour=8888888}
\fontspec{Myriad Pro}
Now grey, with old-style figures: 0123456789
```

normal : Myriad Pro 0123456789  
Now grey, with old-style figures:  
0123456789

```
\fontspec[Numbers=OldStyle]{Myriad Pro}
`In 1842, 999 people sailed 97 miles in
13 boats. In 1923, 111 people sailed 54
miles in 56 boats.' \bigskip
```

`In 1842, 999 people sailed 97 miles in 13 boats. In  
1923, 111 people sailed 54 miles in 56 boats!'

```
\addfontfeatures{Numbers={Monospaced,Lining}}
\begin{tabular}[t]{t}{@{}cccc@{}} \hline
Year & People & Miles & Boats \\ \hline
1842 & 999 & 75 & 13 \\
1923 & 111 & 54 & 56 \\ \hline
\end{tabular}
```

Year	People	Miles	Boats
1842	999	75	13
1923	111	54	56

## Xe<sub>l</sub>TeX et fontspec (2)

```
\fontspec [
 UprightFeatures={Colour = 220022,
 SmallCapsFeatures = {Colour=115511}},
 ItalicFeatures={Colour = 2244FF,
 SmallCapsFeatures = {Colour=112299}},
 BoldFeatures={Colour = FF4422,
 SmallCapsFeatures = {Colour=992211}},
 BoldItalicFeatures={Colour = 888844,
 SmallCapsFeatures = {Colour=444422}},
] {Minion Pro}
Upright {\scshape Small Caps}\\
\itshape Italic {\scshape Italic Small Caps}\\
\upshape\bfseries Bold {\scshape Bold Small Caps}\\
\itshape Bold Italic {\scshape Bold Italic Small Caps}

\setromanfont{Adobe Garamond Pro}
\newfontinstance\lc[Scale=MatchLowercase]{Verdana}
The perfect match {\lc is hard to find.}\\
\newfontinstance\uc[Scale=MatchUppercase]{Arial}
L O G O \uc F O N T
```

Upright **SMALL CAPS**  
*Italic* **ITALIC SMALL CAPS**  
**BOLD BOLD SMALL CAPS**  
*Bold Italic* **BOLD ITALIC SMALL CAPS**

The perfect match is hard to find.  
L O G O F O N T

## Xe<sub>La</sub>TeX et fontspec (3)

```
\def\text
 {\parbox{5cm}{A MULTITUDE OF
 OBSTREPEROUSLY
 HYPHENATED ENTITIES}
 \par\vspace{1ex}}
\fontspec[HyphenChar=None]
 {Adobe Garamond Pro} \text
\fontspec[HyphenChar={-}]
 {Adobe Garamond Pro} \text
\fontspec[HyphenChar="F6BA]
 {Adobe Garamond Pro} \text
```

```
\fontspec[OpticalSize=0]{Warnock Pro Caption}
Warnock Pro optical sizes \\
\fontspec[OpticalSize=0]{Warnock Pro}
Warnock Pro optical sizes \\
\fontspec[OpticalSize=0]{Warnock Pro Subhead}
Warnock Pro optical sizes \\
\fontspec[OpticalSize=0]{Warnock Pro Display}
Warnock Pro optical sizes
```

```
\fontspec{Warnock Pro}Automatic optical size \\
\scalebox{0.4}{\Huge Automatic optical size}
```

A MULTITUDE OF  
OBSTREPEROUSLY  
HYPHENATED ENTITIES

A MULTITUDE OF OB-  
STREPEROUSLY HYPHEN-  
ATED ENTITIES

A MULTITUDE OF OB-  
STREPEROUSLY HYPHEN-  
ATED ENTITIES

Warnock Pro optical sizes  
Warnock Pro optical sizes  
Warnock Pro optical sizes  
Warnock Pro optical sizes

Automatic optical size  
Automatic optical size

## Xe<sub>l</sub>TeX et fontspec (4)

```
\fontspec[Letters=SmallCaps]
 {Warnock Pro} THIS SENTENCE no verb \\
\fontspec[Letters=UppercaseSmallCaps]
 {Warnock Pro} THIS SENTENCE no verb
```

THIS SENTENCE NO VERB  
THIS SENTENCE no verb

```
\fontspec[Numbers=Lining]{Warnock Pro}
0123456789
\fontspec[Numbers=SlashedZero]{Warnock Pro}
0123456789
```

0123456789      0123456789

```
\fontspec{Warnock Pro} \itshape
Without Contextual Swashes \\
\fontspec[Contextuals=Swash]{Warnock Pro}
With Contextual Swashes; cf. W C S
```

*Without Contextual Swashes*  
*With Contextual Swashes; cf. W C S*

```
\fontspec{Warnock Pro} Ta AV \\
\fontspec[Kerning=Off]{Warnock Pro} Ta AV
```

Ta AV  
Ta AV

```
\fontspec{Warnock Pro}
UPPER-CASE EXAMPLE \\
\addfontfeature{Kerning=Uppercase}
UPPER-CASE EXAMPLE
```

UPPER-CASE EXAMPLE  
UPPER-CASE EXAMPLE

## Xe<sub>L</sub>TeX et fontspec (5)

```
\fontspec[VerticalPosition=Superior]
 {Warnock Pro} Sup: abdehilmnorst (-12,345.67) \\
\fontspec[VerticalPosition=Numerator]
 {Warnock Pro} Numerator: 12345 \\
\fontspec[VerticalPosition=Denominator]
 {Warnock Pro} Denominator: 12345 \\
\fontspec[VerticalPosition=ScientificInferior]
 {Warnock Pro} Scientific Inferior: 12345 \\
\fontspec[VerticalPosition=Ordinal]{Warnock Pro}
`Ordinals': 1st 2nd 3rd 4th 0th
```

Sup: abdehilmnorst <sup>(-12,345.67)</sup>  
 Numerator: <sup>12345</sup>  
 Denominator: <sub>12345</sub>  
 Scientific Inferior: <sub>12345</sub>  
 `Ordinals': 1<sup>st</sup> 2<sup>nd</sup> 3<sup>rd</sup> 4<sup>th</sup> 0<sup>th</sup>

```
\fontspec{Warnock Pro} \vfrac{13579}{24680}
```

<sup>13579</sup>/<sub>24680</sub>

```
\fontspec{Warnock Pro} K Q R k v w y \\
\addfontfeature{Style=Alternate} K Q R k v w y
```

K Q R k v w y  
 K Q R k v w y

```
\fontspec{Adobe Jenson Pro} M Q Z \\
\addfontfeature{Style=Historic}M Q Z
```

M Q Z  
 M Q Z

# ArabXe<sub>L</sub>TeX : typographie arabe (1)

François Charette a développé une extension `arabxetex`, qui est basé sur `arabtex` de Klaus Lagally

```
\begin{arab}[novoc]
mi_tAl: \aemph{45} darajaT
\end{arab}
\begin{arab}[voc]
mi_tAl: \aemph{45} darajaT
\end{arab}
\begin{arab}[fullvoc]
mi_tAl: \aemph{45} darajaT
\end{arab}
```

مثال: ٤٥ درجة

مثال: ٤٥ دَرَجَة

مثال: ٤٥ دَرَجَة

## ArabXe<sub>L</sub>TeX : typographie arabe (2)

```
\begin{arab}[voc]
'amruN, 'ibiluN, 'u_htuN, '"u_ht"uN, '"Uql"Id"Is, ra'suN, 'ar'asu,
sa'ala, qara'a, bu'suN, 'ab'usuN, ra'ufa, ru'asA'u, bi'ruN, 'as'ilaTuN,
ka'iba, qA'imuN, ri'AsaTuN, su'ila, samA'uN, barI'uN, sU'uN, bad'uN,
^say'uN, ^say'iN, ^say'aN, sA'ala, mas'alaTuN, saw'aTuN, _ha.tI'aTuN,
jA'a, ridA'uN, ridA'aN, jI'a, radI'iN, sU'uN, .daw'uN, qay'iN, .zim'aN
, yatasA'alUna, 'a`dA'akum, 'a`dA'ikum, 'a`dA'ukum maqrU'aT, mU'ibAt,
taw'am, yas'alu, 'a.sdiq^A$\\;$'uh_u, ya^g^I'u, s^U'ila
\end{arab}
```

أَمْرٌ، إِبِلٌ، أُخْتٌ، أُخْتٌ، أُوقْلِيدِيسٌ، رَأْسٌ، أَرَأْسٌ، سَأَلَ، قَرَأَ، بُؤْسٌ، أَبُؤْسٌ، رُوْفٌ، رُوْسَاءٌ، بِيْرٌ، أَسْئَلَةٌ،  
 كَيْبٌ، قَائِمٌ، رِاسَةٌ، سُئِلَ، سَمَاءٌ، بَرِيءٌ، سُوءٌ، بَدءٌ، شَيْءٌ، شَيْءٌ، شَيْئاً، سَاءَلَ، مَسْأَلَةٌ، سَوَاءٌ، خَطِيئَةٌ،  
 جَاءَ، رِدَاءٌ، رِدَاءٌ، جِيئٌ، رَدِيءٌ، سُوءٌ، ضَوْءٌ، قِيءٌ، ظِمئاً، يَتَسَاءَلُونَ، أَعْدَاءُكُمْ، أَعْدَائِكُمْ، أَعْدَاؤُكُمْ  
 مَقْرُوَّةٌ، مُؤَيَّبَاتٌ، تَوَامٌ، يَسْأَلُ، أَصْدِقَاءُهُ، يَجِيءُ، سُؤِلَ

## ArabXe<sub>La</sub>TeX : le Coran

```
\newcommand{\hamzaB}{\char"200D\char"0640\raise-.95ex\hbox{\char"0654}\char"200D}
\begin{arab}[fullvoc]
mina 'l-qur'Ani 'l-karImi, sUraTu 'l-ssajdaTi 15--16:\\
'innamA yu'minu bi-\hamzaB al"Ay__atinA 'lla_dIna 'i_dA _dukkirUA bihA
_harrUA sujjadaN wa-sabba.hUA bi-.hamdi rabbihim wa-hum lA yastakbirUna
SAJDA [[15]] tatajAfY_a junUbuHum `ani 'l-ma.dAj`i yad`Una rabbahum
_hawfaN wa-.tama`aN wa-mimma razaqn_ahum yunfiqUna [[16]]\\
sUraTu 'l-baqaraTi 71--72:\\
qAla 'innahu, yaqUlu 'innahA baqaraTuN lla _dalUluN tu_tIru 'l-'ar.da wa-lA
tasq.I 'l-.har_ta musallamaTuN lla ^siyaTa fIhA|^JIM qAluW" 'l-\hamzaB a__ana
ji'ta bi-'l-.haqqi|^JIM fa_daba.hUha wa-mA kAdduW" yaf`alUna [[71]] wa-'i_d
qataltum nafsaN fa-udda$, $_ara|^|_i"tum fIhA|^SLY wa-al-ll_ahu mu_hrijuN mma
kun"tum taktumUna [[72]]
\end{arab}
```

مِنَ الْقُرْآنِ الْكَرِيمِ، سُورَةُ السَّجْدَةِ ١٥-١٦ :

إِنَّمَا يُؤْمِنُ بِآيَاتِنَا الَّذِينَ إِذَا ذُكِرُوا بِهَا حَرُّوا سُجَّدًا وَسَبَّحُوا بِحَمْدِ رَبِّهِمْ وَهُمْ لَا يَسْتَكْبِرُونَ ﴿١٥﴾ تَتَجَافَى جُنُوبُهُمْ عَنِ الْمَضَاجِعِ يَدْعُونَ رَبَّهُمْ خَوْفًا وَطَمَعًا وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ ﴿١٦﴾

سُورَةُ الْبَقَرَةِ ٧١-٧٢ :

قَالَ إِنَّهُ، يَقُولُ إِنَّهَا بَقَرَةٌ لَا ذَلُولٌ تُثِيرُ الْأَرْضَ وَلَا تَسْقِي الْحَرْثَ مُسَلَّمَةً لَا شِئَةَ فِيهَا قَالُوا أَلَكِن جِئْتَ بِالْحَقِّ ﴿٧١﴾ فَذَبْحُوهَا وَمَا كَادُوا يَفْعَلُونَ ﴿٧٢﴾ وَإِذْ قَتَلْتُمْ نَفْسًا فَادَّارَعْتُمْ فِيهَا وَاللَّهُ مُخْرِجٌ مَّا كُنْتُمْ تَكْتُمُونَ ﴿٧٣﴾

## ArabXe<sub>La</sub>TeX : typographie persane

```
\newfontfamily\farsifont[Script=Arabic,Scale=1.33]{Farsi Simple Bold}
\begin{farsi}[voc]
_hwAb, xwI^s, _hwod, ^ceH, naH, yal_aH, _hAneH, _hAneHhA, _hAneH-hA,
ketAb-e, U, rAh-e, t_U, nAmeH-i, man, bInI-e, An, mard, pA-i, In,
zan, bAzU-i, In, zan, dAr-_i, man, _hU-_i, t_U, nAmeH-_i, sormeH-_i,
gofteH-_i, ketAb-I, rAh-I, nAmeH-I, dAnA-I, pArU-I, dAnA-I-keH,
pArU-I-keH, rafteH-am, rafteH-Im, AnjA-st, U-st, t_U-st, ketAb-I-st,
be-man, be-t_U, be-An, be-In, be-insAn, beU, be-U, .sA.heb"|_hAneH,
pas"|andAz, naw"|AmUz
\end{farsi}
```

خواب، خویش، خود، چه، نه، یکه، خانِه، خانِه‌ها، خانِه‌ها، کتاب، او، راه، تو، نامه، من، بینی،  
 آن، مرد، پای، این، زن، باثوی، این، زن، دارِ من، خوی، تو، نامه، سرمه، گفته، کتابای،  
 راهای، نامه ای، داناای، پاروای، داناایکه، پاروایکه، رفته ام، رفته ایم، آنجاست  
 اوست، توست، کتابیست، بمن، بتو، بان، باین، بانسان، باو، باو صاحب خانِه،  
 پس ندان نو آموز

## ArabXe<sub>La</sub>TeX : typographie Urdu

```
\begin{urdu}[novoc]
,ham `i^sq kE mAr0.n kA itnA ,hI fasAna,h ,hae\\
rOnE kO na,hI.n kO'I ,ha.nsnE kO zamAna,h ,hae\par
ya,h kiskA ta.sawwur ,hae ya,h kiskA fasAna,h ,hae\\
j0 a^sk ,hae A.nkh0.n mE.n tasbI.h kA dAnA ,hae
\end{urdu}

\newfontfamily\urdufont [Script=Arabic,Scale=1.3]{Code2000}
```

ہم عشق کے ماروں کا اتنا ہی فسانہ ہے  
رونے کو نہیں کوئی ہنسنے کو زمانہ ہے  
یہ کسکا تصور ہے یہ کسکا فسانہ ہے  
جو اشک ہے آنکھوں میں تسبیح کا دانا ہے

```
\urdufont [Script=Arabic,Scale=1.2]{Nafees Pakistani Naskh} {Nafees Riqua}
```

ہم عشق کے ماروں کا اتنا ہی فسانہ ہے  
رونے کو نہیں کوئی ہنسنے کو زمانہ ہے  
یہ کسکا تصور ہے یہ کسکا فسانہ ہے  
جو اشک ہے آنکھوں میں تسبیح کا دانا ہے

ہم عشق کے ماروں کا اتنا ہی فسانہ ہے  
رونے کو نہیں کوئی ہنسنے کو زمانہ ہے  
یہ کسکا تصور ہے یہ کسکا فسانہ ہے  
جو اشک ہے آنکھوں میں تسبیح کا دانا ہے

## Typographie chinoise : zhspacing

- ▶ Xe<sub>l</sub>TeX gère très bien les différentes langues complexes, y compris celles utilisant les idéographes Han (chinois, japonais, koréen, etc.)
- ▶ la transition entre systèmes d'écriture (p.ex. alphabet latin  $\Leftrightarrow$  caractères Han) n'est pas pris en charge
- ▶ pas de traitement de la ponctuation pour les langues qui utilisent un alphabet non-latin
- ▶ pour le chinois YIN Dian a écrit une extension `zhspacing`, qui adresse certains de ces points (<http://code.google.com/p/zhspacing/>)

```
\documentclass{article}
```

```
\usepackage{zhspacing}
```

```
\begin{document}
```

这是中文测试。中文和English的混排。中文和 $E=mc^2$ 的混排。

```
\end{document}
```

这是中文测试。中文和English的混排。中文和 $E = mc^2$ 的混排。

## zhspacing : les espaces

Les espaces qui suivent un caractère chinois sont toujours ignorés. Par contre, un espace bien visible est introduit à la limite entre caractères chinois et latins et à une formule mathématique, comme le montre l'exemple suivant :

中Eng文, 中 Eng文, 中Eng 文 and 中 Eng 文.

qui donne la sortie correspondante :

中 Eng 文, 中 Eng 文, 中 Eng 文 and 中 Eng 文.

La 1<sup>ière</sup> et la 2<sup>ième</sup> génèrent une représentation identique, tout comme la 3<sup>ième</sup> et la 4<sup>ième</sup>. Les deux dernières ont un espace un peu plus large entre **eng** et l'idéographe qui suit, ce qui est dû à l'insertion d'un espace « standard » après la lettre **g**

## zhspacing : la ponctuation

La gestion de la ponctuation par `zhspacing` prend en charge l'introduction de l'espacement correct autour de la ponctuation, son alignement en fin de ligne et l'indivisibilité de certaines paires d'idéographes (p.ex. 禁则). Un exemple suit.

他强调，“三个代表”重要思想是在新的历史条件下运用马克思主义的立场、观点和方法的典范，是我们学习马克思主义的立场、观点和方法最现实、最生动的教材。“三个代表”重要思想是与时俱进的理论。

## zhspacing : la composition verticale

Un texte peut être composé verticalement en ajoutant la caractéristique `vertical` pour la fonte en question. Pour la ponctuation nous utilisons la fonte *AdobeSongStd-Light*.

```
\documentclass[12pt]{article}
\usepackage[dvipdfm]{graphicx}
\usepackage{zhspacing}
\zhspacing
\begin{document}
\newfontfamily\zhfont[RawFeature={vertical:}]{SimSun}
\newfontfamily\zhpunctfont[RawFeature={vertical:
+vert:+vhal}]{Adobe Song Std L}
\haltskipsscheme
\rotatebox{-90}{我是中国人，我爱自己的祖国。}
\end{document}
```

我是中国人，  
我爱自己的祖国。

## Les fontes et éditeurs Unicode

- ▶ **emacs** et **vi** avec les fontes adéquates
- ▶ **yudit**, un gratuiciel (<http://yudit.org>) sur Linux ou *Microsoft Windows*
- ▶ fontes Unicode, les disponibilités :
  - ▶ **Cyberbit** de Bitstream, voir <ftp://ftp.netscape.com/pub/communicator/extras/fonts/windows/cyberbit.zip>
  - ▶ une incarnation plus récente « Titus » (développé à l'université de Francfort, voir à l'URL <http://titus.uni-frankfurt.de/>)
  - ▶ fonte-partaguiciel **Code2000** pour le plan 0 (et *Code2001* et *Code2002* qui couvrent une grande partie des caractères des plans 1 et 2) voir <http://www.code2000.net/>
  - ▶ *Arial Unicode MS* qui vient avec le système *Microsoft Windows XP* ou *Vista*
- ▶ page *Unicode fonts* : <http://www.wazu.jp/index.html>
- ▶ page web de Luc Devroye : <http://www.cccg.ca/~luc/fonts.html>
- ▶ page d'Alan Wood : <http://www.alanwood.net/unicode/fonts.html>
- ▶ *Unicode tools and fonts* : <http://www.unifont.org/>
- ▶ pour lister les caractéristiques d'une fonte OpenType : [OpenType-info.tex](http://www.unifont.org/OpenType-info.tex)

## Xe<sub>L</sub>TeX : exemples

- ▶ d'abord quelques courts exemples
  - ▶ le bonjour habituel
  - ▶ le fichier standard  $\LaTeX$  `sample2e.tex`
  - ▶ les commandes Xe<sub>L</sub>TeX pour gérer les fontes
  - ▶ un processeur spécial pour les écritures complexes (*graphite*)
  - ▶ utiliser la fonte mathématique Cambria
- ▶ le Haut-Commissariat des Nations Unies aux droits de l'homme à Genève publie la *Déclaration universelle des droits de l'homme* (<http://www.ohchr.org/french>)
  - ▶ le site Unicode a repris ce texte traduit en 321 langues comme exemple des avantages d'Unicode (<http://www.unicode.org/udhr>)
  - ▶ le premier article en quelques écritures (sans traitement spécifique) (`udhrfirst.tex`)
- ▶ le site [www.sacred-texts.com](http://www.sacred-texts.com) contient des centaines de textes « sacrés », dont en UTF8 les textes d'Homère (`cla/homer/greek` en grec ancien), une bible polyglotte (`bib/poly` en anglais, grec, hébreu et latin), le Coran (`isl/uq` en arabe et anglais), Confucius (`cfu/cfu.htm` en chinois et anglais), le Rig Veda (`hin/rvsan` en sanscrit), etc.
- ▶ le site de l'institut des études indogermaniques de Francfort ([titus.uni-frankfurt.de](http://titus.uni-frankfurt.de)) contient de nombreux textes

## Unicode, les mathématiques, X<sub>Y</sub>TeX et les autres

- ▶ composer des structures mathématiques correctement
- ▶ raison d'être de T<sub>E</sub>X
- ▶ initiative de Murray Sargent (Microsoft), membre du groupe MathML de W3C
- ▶ *Unicode support for mathematics* par Barbara Beeton, Asmus Freytag et Murray Sargent ([www.unicode.org/reports/tr25/tr25-7.html](http://www.unicode.org/reports/tr25/tr25-7.html))
- ▶ *Unicode Nearly Plain-Text Encoding of Mathematics* ([unicode.org/notes/tn28/](http://unicode.org/notes/tn28/))
- ▶ intérêt inclure un processeur de math dans Office 2007
  - ▶ présentation *Math Editing and display in Office 2007* ([research.microsoft.com/workshops/fs2006/presentations/17\\_Sargent\\_071706.ppt](http://research.microsoft.com/workshops/fs2006/presentations/17_Sargent_071706.ppt))
  - ▶ discussion libre : <http://blogs.msdn.com/murrays>
  - ▶ processeur *ad hoc* basé sur le modèle de base de T<sub>E</sub>X (T<sub>E</sub>XBook appendice G)
  - ▶ développement d'une fonte mathématique *Cambria Math*
  - ▶ utilisation des caractères Unicode math et des tables spécifiques OpenType (**MATH**)
  - ▶ composant **MathFont.dll** pour communiquer avec les applications
- ▶ *Design Science Proposal to Microsoft to Help STEM (Scientific/Technical/Engineering/Mathematical) Publishers Work with Office 2007 Documents*  
[http://www.dessci.com/en/reference/white\\_papers/STMOffice2007Proposal.htm](http://www.dessci.com/en/reference/white_papers/STMOffice2007Proposal.htm)  
demande de partager la spécification de la technologie développée par Microsoft

## X<sub>Ǝ</sub>TeX : les mathématiques partout

- ▶ X<sub>Ǝ</sub>TeX utilise l'algorithme de l'appendice G du TeXBook pour composer les mathématiques
- ▶ pour une fonte « standard » X<sub>Ǝ</sub>TeX trouve l'information métrique caractérisant les caractères dans les fichiers `tfm`, puis `xdvipdfmx` se réfère aux fichiers `pfb` via les fontes virtuelles (si nécessaire) et le fichier `dvipdfm.map`
- ▶ pour une fonte OpenType (comme *Cambria Math*) X<sub>Ǝ</sub>TeX lit les paramètres métriques dans la table `MATH` en les mettant sous une forme que l'algorithme « appendice G » peut utiliser
- ▶ pour le moment X<sub>Ǝ</sub>TeX n'utilise pas de processeur spécifique pour les fontes math OpenType (intégration ultérieure à ICU ?)
- ▶ autres fontes math Unicode
  - ▶ *STIX* ([www.stixfonts.org/](http://www.stixfonts.org/)) par le projet *Scientific and Technical Information Exchange font*, un groupe d'éditeurs scientifiques qui ont financé la création d'une fonte Unicode qui conteint plus que 8000 glyphes mathématiques.
  - ▶ *Asana-Math* de Apostoulos Syropoulos ([asyropoulos@yahoo.com](mailto:asyropoulos@yahoo.com)) qui développe avec `fontforge` une fonte math incluant les tables `MATH`
- ▶ en guise d'interface LaTeX Will Robertson a commencé à travailler à une extension `unicode-math`, mais il attend la disponibilité des fontes *STIX* pour finaliser son travail

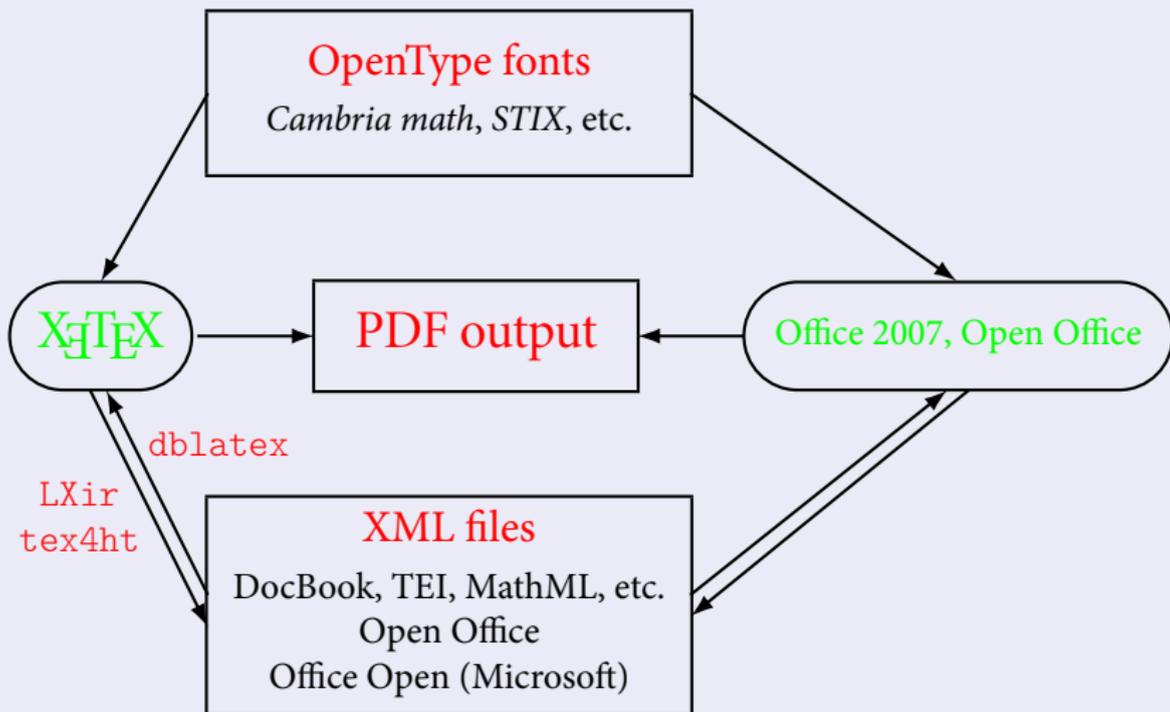
# Xe<sub>La</sub>TeX : développements à l'étude (Jonathan Kew à TUG2007)

- ▶ un remplacement modulaire de `babel` basé sur Unicode
  - ▶ la langue du document (termes fixes)
  - ▶ motifs de césure
  - ▶ conventions typographique (guillemets, ponctuation, espaces)
  - ▶ association implicite entre une écriture et interval de codes Unicode
- ▶ les options étendues de césure, p.ex. compléter `\hyphenchar` par `\prehyphenchar` et `\posthyphenchar` afin de permettre l'insertion d'un caractère Unicode précédent et suivant le point de césure
- ▶ extensions micro-typographiques à la pdfTeX (expansion de lettres, ponctuation dans les marges)
- ▶ extensions à la Oméga/Alpha (modèle de multi-directionnalité, boîte L/R)
- ▶ intégration avec XML, en particulier MathML et les fontes mathématiques

## X<sub>Y</sub>TeX : une ouverture décisive vers le monde extérieur

- ▶ T<sub>E</sub>X est entré avec X<sub>Y</sub>TeX pleinement dans le monde multilingue et multiculturel Unicode
- ▶ ses bases T<sub>E</sub>X garantissent une composition excellente en exploitant d'une façon optimale la richesse typographique des dizaines de milliers de fontes OpenType
- ▶ son intégration avec des interfaces système pour la gestion des fontes et l'internationalisation permettent de tirer profit des derniers développements dans ces domaines
- ▶ grâce à l'inclusion des paramètres math dans les fontes OpenType et d'un codage Unicode linéaire des maths dans le texte nous auront peut-être bientôt la possibilité d'échanger plus facilement des documents scientifiques ou techniques contenant des formules mathématiques

## J'ai un rêve...



## Questions ?

የኒኮድ ምንድን ነው? "ما هي الشفرة الموحدة "يونكود"؟ 什麼是 Unicode(統一碼/標準萬國碼)? Što je Unicode? ಸಾ ಸಾ ಸಾ ಸಾ ಫ್ರೆಂಚ್? Τί είναι τὸ Unicode; ? מה זה יוניקוד? यूनिकोड क्या है? Hvað er Unicode? ユニコードとは何か? 유니코드에 대해? یونی‌کد چیست؟ Что такое Unicode? Unicode คืออะไร? የኒኮድ ከንታይ ኢዩ? Qu'est-ce Unicode ?

*X<sub>Y</sub>T<sub>X</sub>, MathML et Unicode : le futur ensemble*