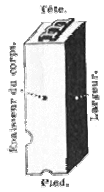
Unicode et les polices

Deux mondes

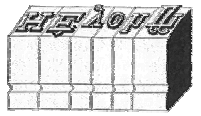
Rabat, le 24 septembre 2004

Patrick Andries
hapax@iquebec.com

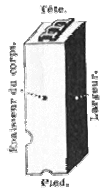
Unicode



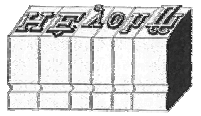
- Ne se préoccupe que des caractères.
- Caractères sont stockés en « ordre logiquea », c'est-à-dire ordre phonétique.
- Unicode suppose la présence d'un moteur de rendu qui pourra, le cas échéant, réordonner les caractères et pourra substituer des glyphes et placer de manière précise.



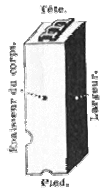
Pourquoi ce modèle ?



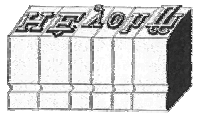
- Garantir que les opérations suivantes fonctionneront :
  - Correcteur grammatical, traduction automatique
  - Indexation
  - Recherche et remplacement de texte
  - Copier-coller
  - Exportation de texte
  - Changement de police
  - Changement de style de paragraphe ou de caractères



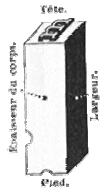
Composition



1. Découpe en passages
   - Même style, même police, même système d'écriture et même directionalité
2. Réordonnancement
   - Des passages pour passer de l'ordre logique à l'ordre visuel
3. Sélection des glyphes («façonnage»)
4. Justification
5. Rendu



1. Découpe



❖ En mémoire :

Il dit *grosso modo* « DE L'EAU ! » et expira.

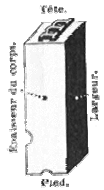
Il dit *grosso modo* «ءامد ي رأ!» et expira.

latin, romain

latin, italique

arabe, romain

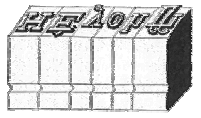
2. Réordonnancement

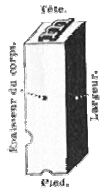


❖ Réordonné localement :

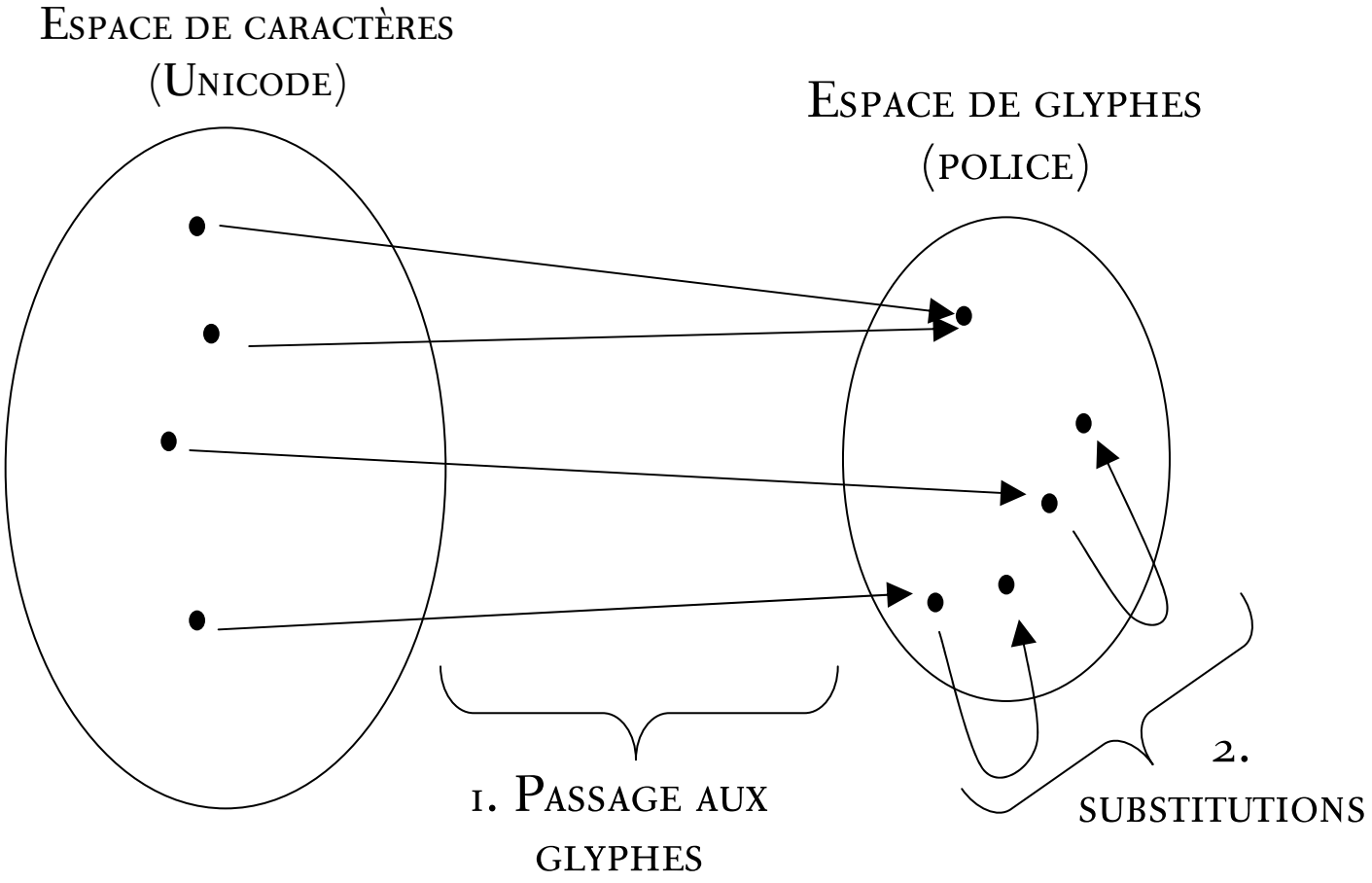
Il dit *grosso modo* « UAE'L ED » et expira.

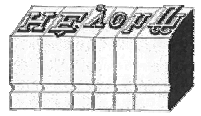
Il dit *grosso modo* « أريد ماءا ! » et expira.



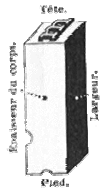


3. Sélection des glyphes





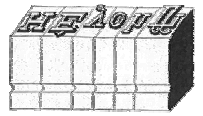
4. Justification



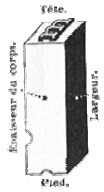
Les suites de glyphes sont ajustées

pour respecter les règles justification en vigueur :

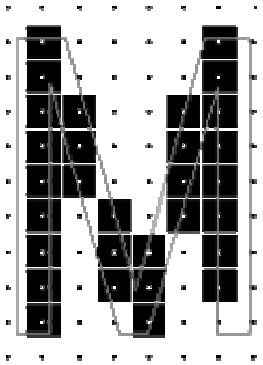
- Exemple de texte dit en drapeau, aligné à gauche ou encore fer à gauche.
- Exemple de texte dit justifié, en bloc ou en encore en beau paragraphe, avec au besoin la coupure de mots.



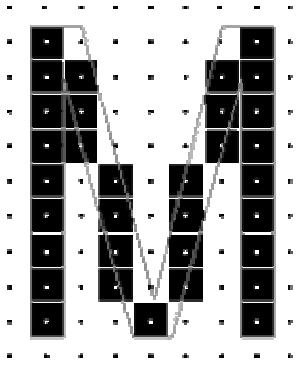
5. Rendu



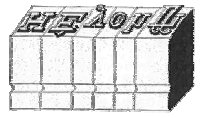
- Suite de glyphes affichées dans l'ordre final sur le dispositif



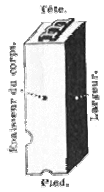
Tramage non optimisé



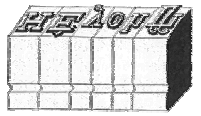
Ajusté à la grille



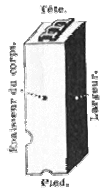
Polices



- PostScript Type 1 (Adobe, 1985)
- TrueType (Apple 1990, Microsoft, 1991)
- OpenType (gestation 1996-1999)
  - Nouveau format multi-plateforme, résultat travaux conjoints de Microsoft et Adobe
  - Extension de TTF
  - Typographie complexe

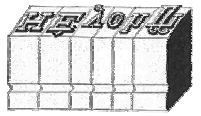


Police = tables

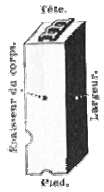


- Police composée de plusieurs tables
  - cmap (TrueType et OpenType)
  - glyf (TrueType et OpenType)
  - GSUB (OpenType)
  - GPOS (OpenType)
  - Autres (chasses des glyphes, type de chaque glyphes, quels blocs Unicode pris en charge par police?)
  - Privées

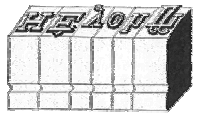
...



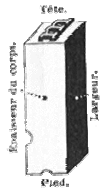
CMAP



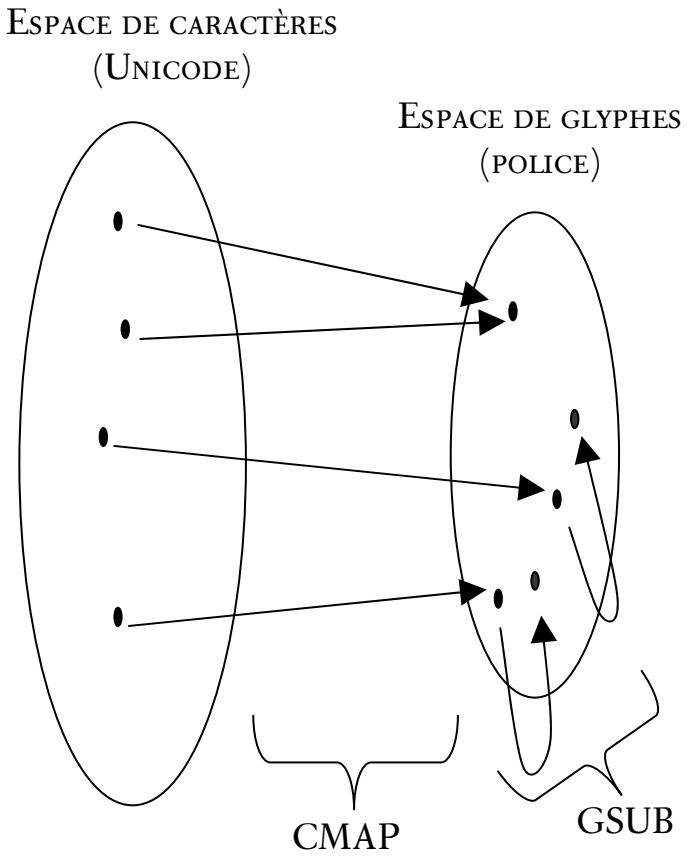
- Table obligatoire
- Permet de passer de l'espace des caractères à l'espace des glyphes
- Exemples :
  - U+0543 → glyphe n° 123
  - U+0178 → glyphe n° 124



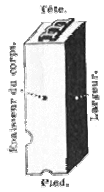
CMAP (SUITE)



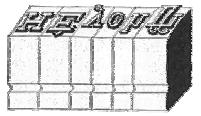
- Tous les glyphes ne correspondent pas à des caractères
- Tous les glyphes n'ont donc pas d'entrée dans la cmap



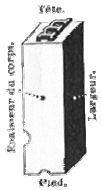
GLYF

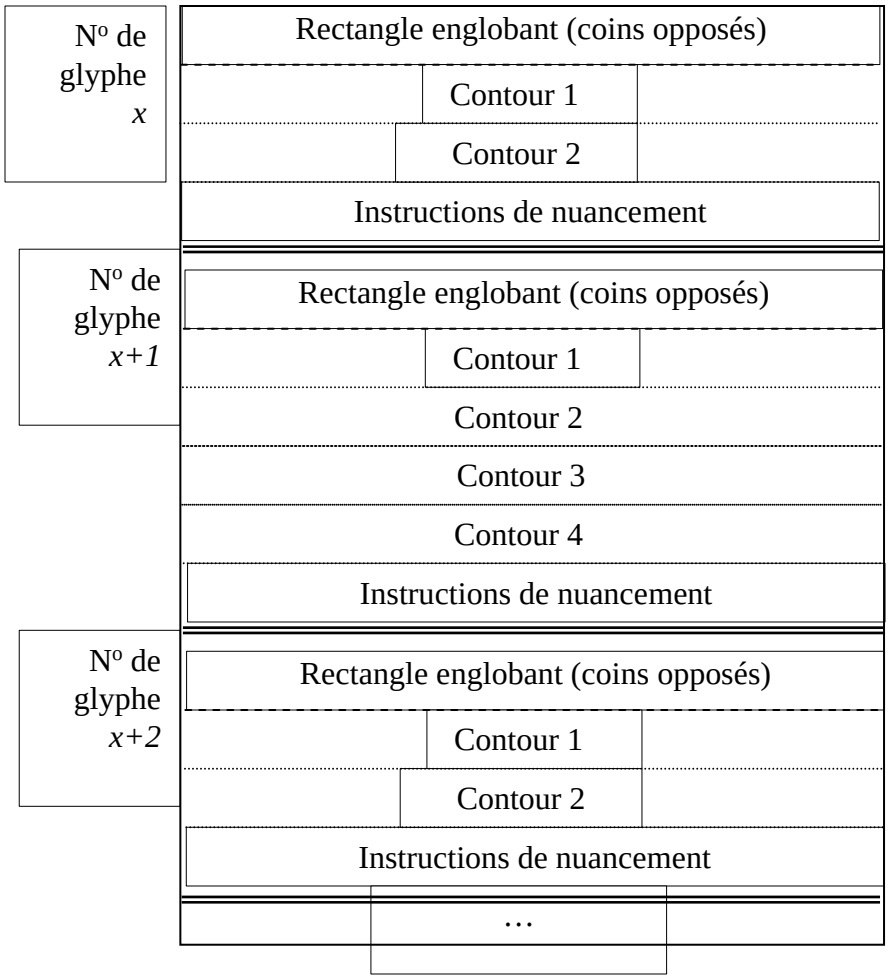


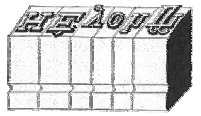
- Souvent la plus volumineuse table
- Contient les contours des glyphes
- Chaque entrée = un glyphe
  - Rectangle englobant
  - Contours fermés (série de points)
  - Instructions d'optimisation de rendu à faible résolution (« hinting », « nuancement »), du code assembleur



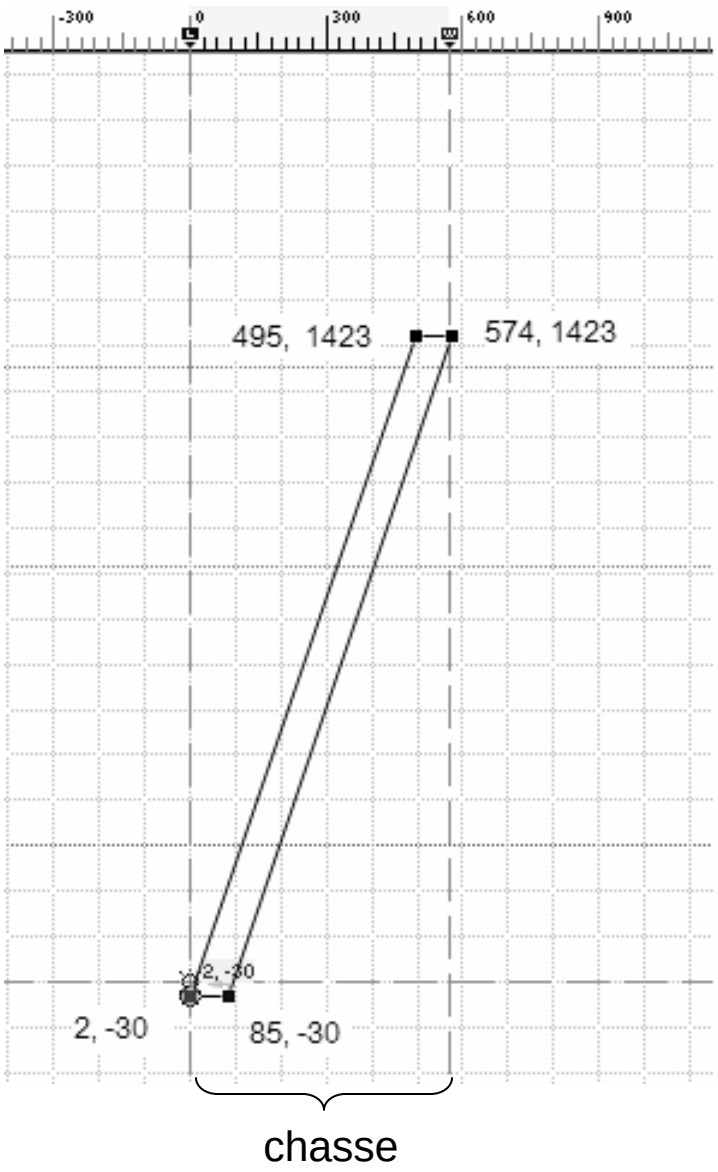
GLYF

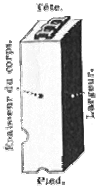



Nº de glyphe x
Rectangle englobant (coins opposés)
Contour 1
Contour 2
Instructions de nuancement
Nº de glyphe x+1
Rectangle englobant (coins opposés)
Contour 1
Contour 2
Contour 3
Contour 4
Instructions de nuancement
Nº de glyphe x+2
Rectangle englobant (coins opposés)
Contour 1
Contour 2
Instructions de nuancement
…

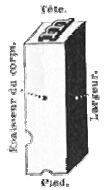


GLYF

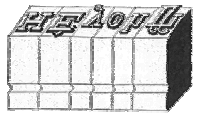




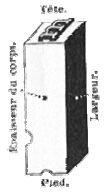
GLYF, ENTRÉE POUR « / »

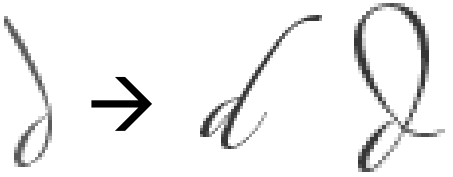


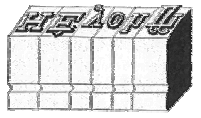

```
<Glyphe
      xMin="2" yMin="-30" xMax="574" yMax="1423">
   <contour>
            <pt x="85" y="-30" sur="1"/>
            <pt x="2" y="-30" sur="1"/>
            <pt x="495" y="1423" sur="1"/>
            <pt x="574" y="1423" sur="1"/>
   </contour>
   <instructions/> <!-- code compilé -->
</Glyphe>
```

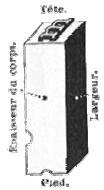
GSUB



- Substitution glyphe(s) → glyphe(s)
  - Simple a → A
  - Décomposition ffi → f f i
  - Sélection de variantes 



GSUB (suite)



- Ligature

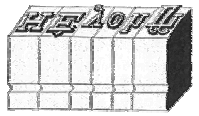
  f f i → ffi

  ل ا → ﻻ

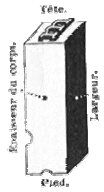
- Contextuelle

  dans → *dans*

- …



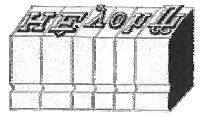
GPOS

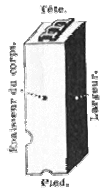


- Définit déplacement de glyphes par rapport à :
  - position habituelle

    (BACH) (BACH) (BACH)
  - d'autres glyphes (par ex. crénage)

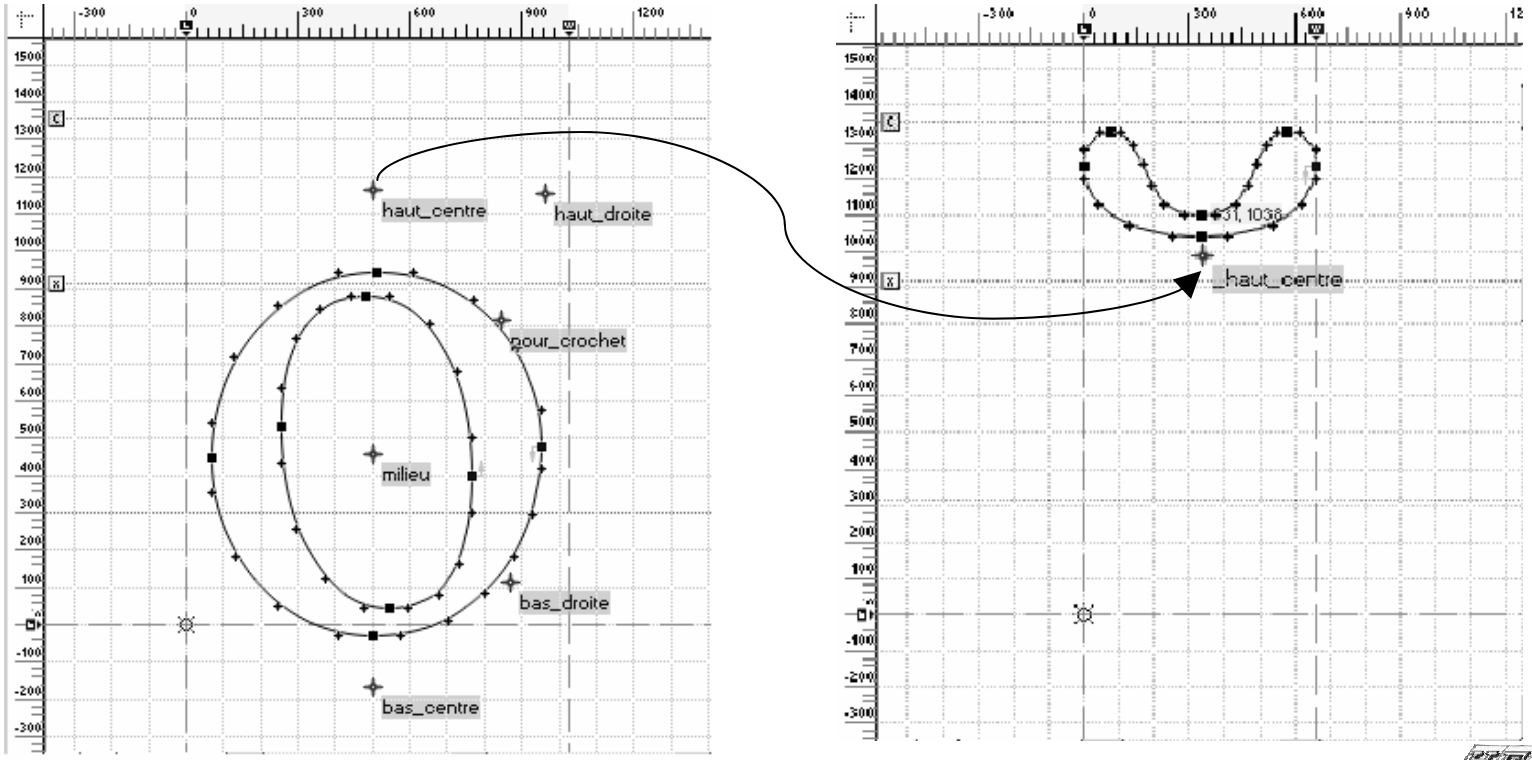
    AVION → AVION



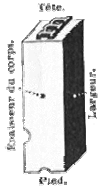


GPOS (suite)

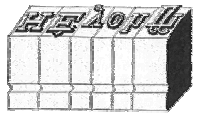
- Déplacement d'un diacritique par rapport à une base



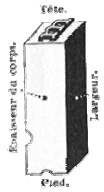
O tilde



Õ



GPOS

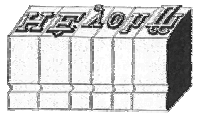


- Déplacement d'un diacritique par rapport à un autre
  - P. ex. pour le vietnamien

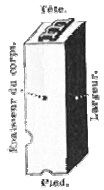
ẳ

Ẵ

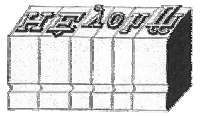
Ế

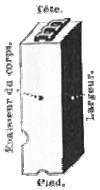


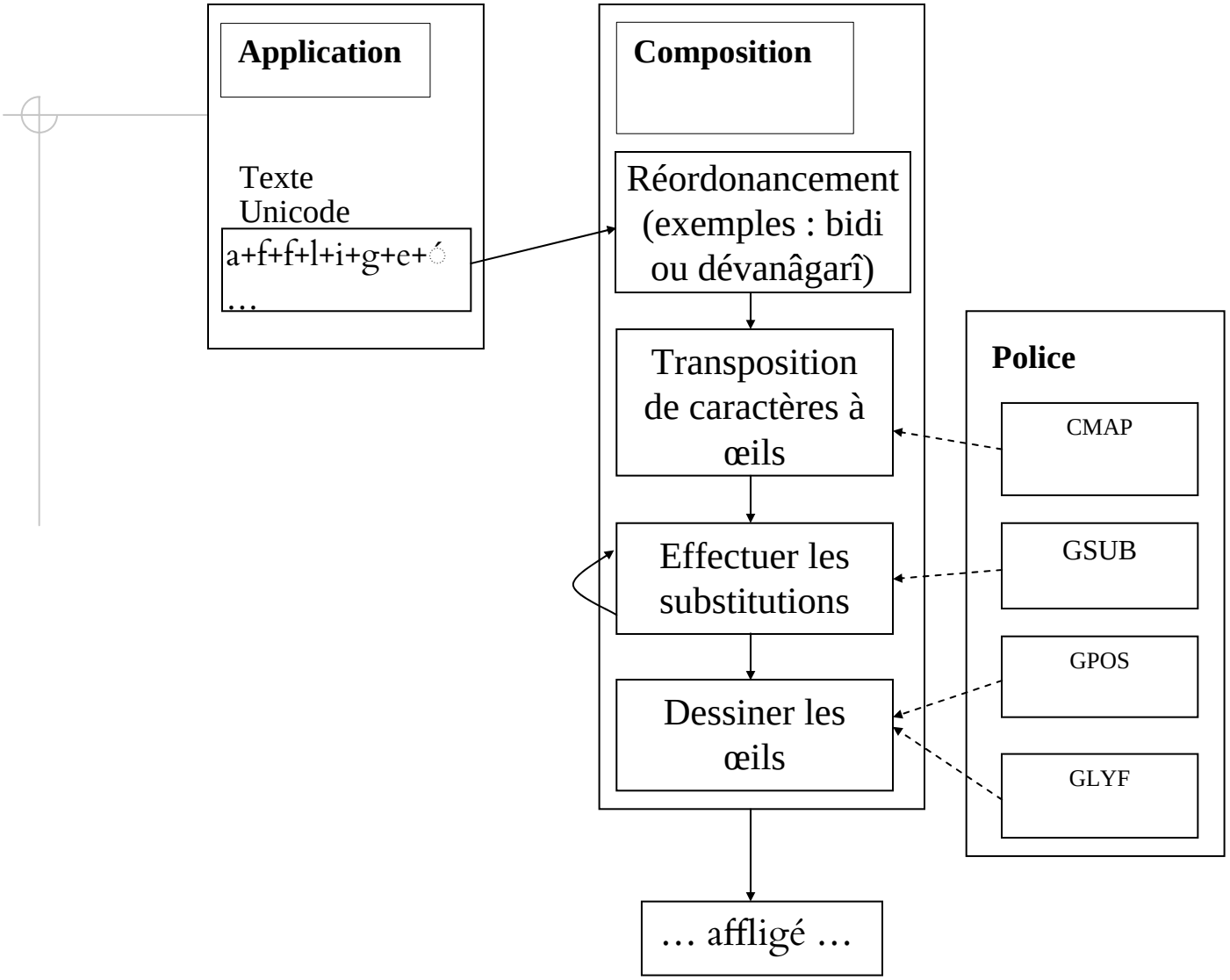
GSUB ou GPOS ?

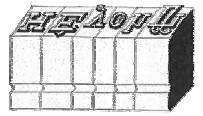


- Substitution plus rapide
- Et, bien sûr, moindre effort pour créer une police avec position (moins de glyphes)
- GPOS intéressant quand l'ensemble des combinaisons explose

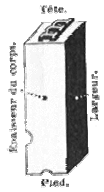




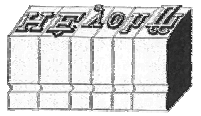

Application
Texte Unicode
a+f+f+l+i+g+e+◌́
…
Composition
Réordonancement (exemples : bidi ou dévanâgarî)
Transposition de caractères à œils
Effectuer les substitutions
Dessiner les œils
… affligé …
Police
CMAP
GSUB
GPOS
GLYF



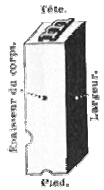
Concepts

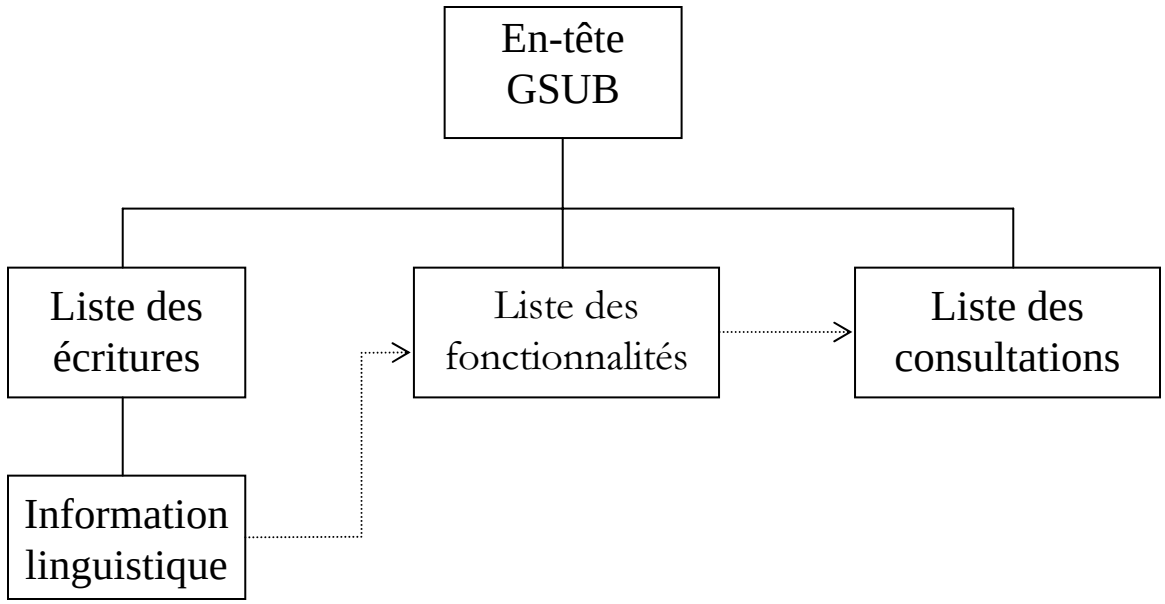


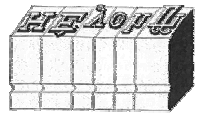
- Écriture
  - Système d'écriture (latin, arabe, cyrillique,...)
- Langue
  - Au sein d'une écriture (peut-être implicite ou p.ex. français, anglais pour le latin)
- Fonctionnalité (« feature»)
  - Regroupe plusieurs consultations
- Groupe de règles (« lookup »)
  - Regroupe un ensemble de règles GSUB ou GPOS



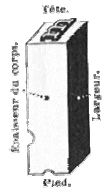
Schématiquement



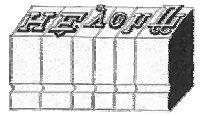




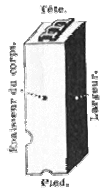
Exemple



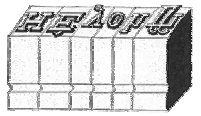
- Arabe (arab)
  - défaut (dflt)
    - ccmp (composition canonique)
      - décomposer
      - composer
    - init (formes initiales)
      - initiales
      - lam initial
    - rlig (ligatures obligatoires)
      - lam alif
    - …



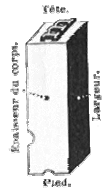
En pratique…



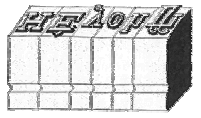
- Client logiciel /utilisateur humain choisit une écriture
  - Éventuellement une langue
- Client/utilisateur choisit fonctionnalités (certaines sont systématiques pour une écriture donnée, *rlig* par exemple)
  - = un ensemble de règles GSUB/GPOS



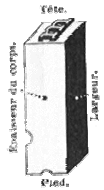
Groupe de règles « lam alif »



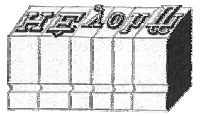
- lam.initial ل alif.madda ◌ٓ →
  lam.alif.madda.initial لآ
- lam.médial ـل alif.madda ◌ٓ →
  lam.alif.madda.médial ـلآ
- lam.initial ل alif.hamza أ → lam.alif.hamza.initial لأ
- …



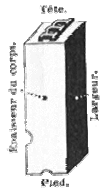
Fonctionalité *init*



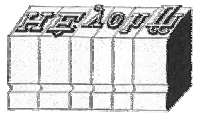
- Fournit la forme initiale de glyphes
  - uni0626 ج → djîm.init ﺟ
- Mais rien n'indique dans la police quand il faut appliquer cette fonctionnalité !
- C'est le rôle du moteur de rendu (parfois appelé moteur de façonnage pour cette activité spécifique)
  - Il doit pouvoir découper un texte en « mots » et identifier les positions initiales dans un mot



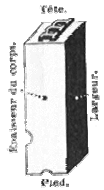
Conséquences



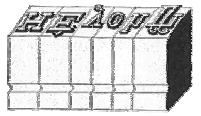
- Les fonctionnalités sont définies par les fabricants de moteur de rendu (p.ex. Uniscribe sur Windows)
- Le concepteur de police doit respecter les noms de ces fonctionnalités si on utilise ce moteur
- Il faut prendre en compte l'ordre défini de consultation de ces fonctionnalités pour une écriture donnée



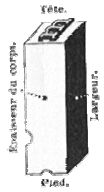
Ordre des opérations



- L'utilisateur ou le client logiciel choisit les fonctionnalités (parfois interactivement comme dlig)
- On rassemble toutes les opérations (GSUB ou GPOS) et on descend dans l'ordre dans lequel ils sont définis dans la table GSUB (pas nécessairement celui apparemment quand on regarde par le côté des fonctionnalités)



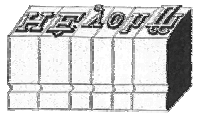
Ordre des substitutions



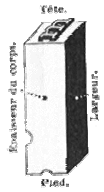
- Correspond à l'ordre de la définition
- Truc : d'abord effectuer les plus longues substitutions
- Soit une consultation :

  f f → ff

  f f i → ffi

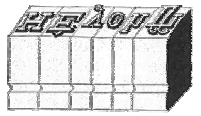


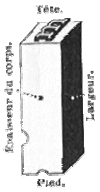
Ordre des substitutions



Efficace → Efficace

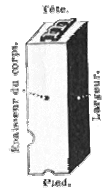
Efficace → Efficace





Merci

Ressources



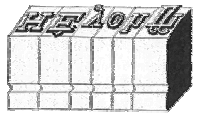
- Fontes et codage, par Yannis Yaralambous, O'Reilly, 2004, ISBN 2-84177-273-X
- Unicode en français: http://pages.infinit.net/hapax
- Listes de diffusion et forum de discussion

  (typo) https://www.irisa.fr/wws/info/typographie

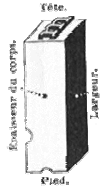
  (Unicode Afrique) http://fr.groups.yahoo.com/group/Unicode-Afrique/
  - (Unicode fr) nntp://fr.comp.normes.unicode
- FontForge (logiciel libre de création de polices)

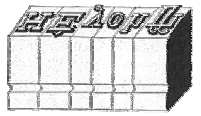
  http://fontforge.sourceforge.net/ (version française existe)
- BabelMap

  http://www.cooptel.qc.ca/ pandries/BabelMap fr.html

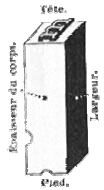


Supplément si jamais…

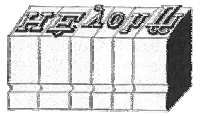
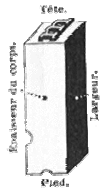
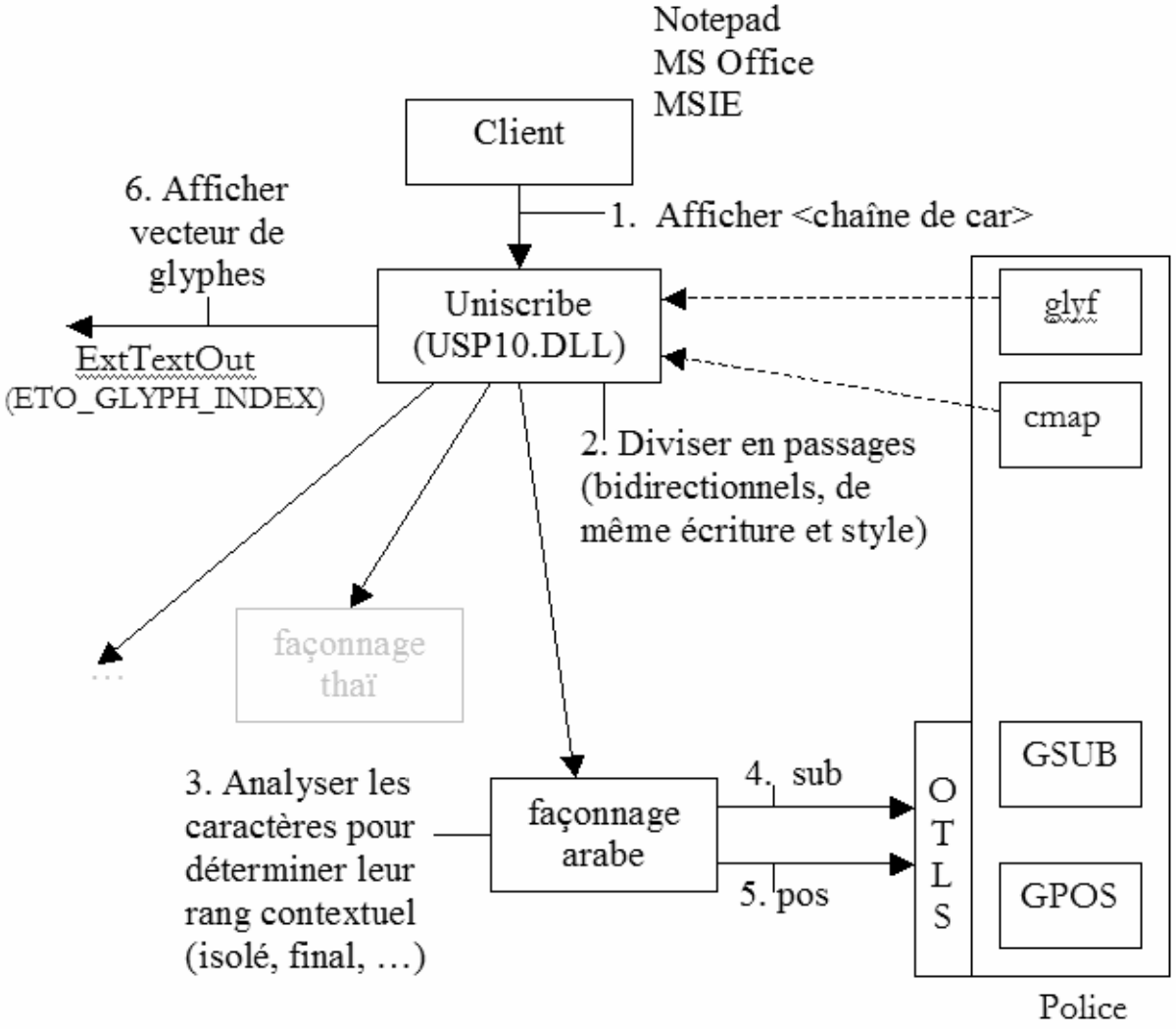
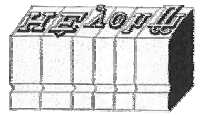




Uniscribe



- *Uniscribe,* moteur de « façonnage »
- Windows 2000, XP et des applications comme MSIE et Office 2002
- Un moteur de façonnage par écriture
  - Un sous-ensemble précis de fonctionnalités OpenType exécuté dans un ordre prédéfini pour chaque écriture.
  - arabe, dévanâgarî, tamoul, thaï, syriaque, etc...



Uniscribe, OTLS