

Les caractères spéciaux

Si, lors de la rédaction de votre thèse, vous êtes amenés à utiliser des caractères autres que latins (symboles mathématiques, formes géométriques, caractères grecs, chinois, arabes ou phonétiques), vous devez faire attention au choix de la police de caractères et du système d'encodage que vous allez utiliser. En effet, il faut veiller à ce que ces caractères puissent être affichés correctement sur différents ordinateurs et systèmes d'exploitation, qui ne posséderont peut-être pas la police de caractères particulière que vous avez employée.

I. Problèmes d'encodage :

Lorsque vous saisissez du texte dans un logiciel de traitement de texte à partir d'un clavier, le logiciel encode les lettres que vous voyez apparaître à l'écran sous forme de nombres. **T**, par exemple, se voit attribuer le code 0054, que l'interface visuelle affiche comme **T**.

Il existe de nombreux systèmes d'encodage de caractères, qui pour la plupart ne couvrent pas tous les caractères et symboles en usage dans les langues du monde. De plus, les systèmes peuvent employer des nombres différents pour coder le même caractère ; et les polices ne possèdent en général qu'un nombre limité de caractères.

II. Unicode :

Pour garantir la stabilité des données, un système d'encodage standardisé a été développé : la norme Unicode, qui assigne à chaque caractère une valeur numérique unique et garantit un affichage correct sur toutes les plates-formes, dans tous les logiciels et toutes les langues.

Unicode permet de représenter la totalité des caractères employés dans les langues du monde, ainsi que les signes de ponctuation et les symboles (la version 4.0 propose ainsi des codes pour 96 447 caractères).

Aussi, lorsque vous créez des documents de texte, vous devez vous servir de polices de caractères utilisant Unicode, particulièrement si vous manipulez des caractères spéciaux. Les polices de base (Arial, Times New Roman) fournies sur les systèmes Windows et Macintosh sont compatibles avec Unicode (notez que la police « Symbol » ne l'est pas et est donc à éviter), mais ne proposent pas toute la gamme des caractères, d'où la nécessité de recourir à d'autres polices.

Voici quelques conseils et pistes à suivre selon les types de caractères dont vous avez l'utilité :

- Caractères phonétiques :
SIL International propose plusieurs polices de caractères compatibles avec Unicode (mais notez que les versions répandues « SILDoulos IPA » et « SIL Doulos IPA 93 » ne l'étaient pas). Consultez le site en suivant ce lien : <http://www.sil.org/linguistics/resources.html>
- Transcription du grec :
L'Institute of Biblical Greek propose une information détaillée sur la saisie de grec ancien, ainsi que le téléchargement de polices Unicode. Consultez le site (en anglais) en suivant ce lien : <http://www.biblicalgreek.org/links/fonts.html>
- Transcription de l'arabe :
Alan Wood propose une liste de polices arabes, dont de nombreuses pouvant être téléchargées gratuitement. Consultez son site (en anglais) en suivant le lien : <http://www.alanwood.net/unicode/fonts.html#arabic>
- Transcription de l'hébreu :
Voyez la liste de polices proposées sur le site suivant : http://www.wazu.jp/gallery/Fonts_Hebrew.html
- Transcription du chinois :
Voyez la liste de polices proposées sur le site suivant : <http://www.wazu.jp/index.html#Chinese>.
- Autres :
Ce site propose des exemples de polices : <http://www.wazu.jp/>

NB : Microsoft propose un utilitaire (téléchargeable à l'adresse <http://www.microsoft.com/typography/property/property.htm>) qui permet, lorsqu'on effectue un clic droit sur un fichier de police ayant une extension .TTF, de savoir si cette police est compatible avec Unicode et quelle est l'étendue de caractères qu'elle propose :

III. Ressources :

[Unicode Home Page.](#)

[Traduction française](#) de *What is Unicode ?*

[Une liste de polices multilingues Unicode](#) téléchargeable sur le net.

Les liaisons entre [Iso 10646 et Unicode](#).

[Qu'est ce que : Unicode](#) réalisé par AJLSM

Ci-dessous d'autres polices de caractères pour coder des caractères spécifiques :

- **Polices arabes**
Police [TITUS Cyberbit Basic](#)
- **Polices grecques**
Pour coder des [caractères grecs](#)
- **Exemples : Unicode et XML**
Transcription des caractères arabes : [thèse de Ryad Zaafrani - chap 2.3.2.](#)
Transcription des caractères grecs : [thèse de Valérie Bonnet - chap 2.1.](#)