Cahiers GUTenberg n°20 — mai 1995

# Création d'une police UNICODE*

Charles BIGELOW[a] et Kris HOLMES[b]
Traduction française de Héloïse TISSOT (École Estienne)†

[a] Irisa/Inria-Rennes
Department of Computer Science
Stanford, California 94305, USA
bigelow@cs.stanford.edu

[b] Division CN
Bigelow & Holmes Inc.
Menlo Park, California 94026, USA
D2782@applelink.apple.com

**Résumé**. L'internationalisation du traitement et de l'échange d'informations et de l'édition électronique a montré la nécessité de définir des normes pour le codage des caractères du monde entier. C'est l'objet de la norme UNICODE. Pour que l'homme puisse lire ces codes de caractères, il faut leur associer des polices produisant des images visuelles – des glyphes – correspondant aux codes. Nous décrivons et commentons ici la façon de dessiner une police traitant une partie de la norme UNICODE.

**Abstract**. *The international scope of computing, digital information interchange, and electronic publishing has created a need for world-wide character encoding standards.* UNICODE *is a comprehensive standard designed to meet such a need. To be readable by humans, character codes require fonts that provide visual images – glyphs – corresponding to the codes. The design of a font developed to provide a portion of the* UNICODE *standard is described and discussed.*

## 1. Introduction

UNICODE est une norme pour le codage des caractères du monde entier, définie par le Consortium UNICODE, un groupe de compagnies privées et d'institutions concernées par le codage des textes et leurs applications informatiques. La norme UNICODE[1] a été développée pour répondre au problème fondamental du traitement simultané de plusieurs langues par ordinateur dont notamment « la lourdeur des mécanismes de fontes lors du

*Cet article correspond à une communication faite par les auteurs lors d'une session commune aux deux congrès *RIDT'94 – Raster Imaging and Digital Typography* et *EP'94 – Electronic Publishing, Document Manipulation and Typography* – Darmstadt, avril 1994 – et reprise lors du congrès *TEX Users Group, Fifteenth Annual Meeting* – Santa Barbara, Californie, Août 1994. Cet article est paru sous le titre *The design of a* UNICODE *font* dans *EPODD – Electronic Publishing Origination Dissemination and Design*, vol. 6(3), septembre 1993, pages 289–305. Il est reproduit ici en français avec l'aimable autorisation des auteurs, des rédacteurs en chef de cette revue et des éditions John Wiley & Sons, que nous tenons à remercier {Ndlr}.

†Avec la collaboration de Jacques ANDRÉ.

[1] Désormais nous parlerons, en abrégé, d'UNICODE. Depuis la rédaction de cet article, en août 1993, UNICODE a été en fait reprise et légèrement modifiée dans le cadre de la norme Iso 10646. Voir à ce sujet l'article de Jacques ANDRÉ et Michel GOOSSENS dans ce *Cahier GUTenberg* [1] {Ndlr}.

codage des caractères et l'emploi de nombreuses normes, non cohérentes, à cause de conflits entre identités nationales » [2].

UNICODE fait la distinction entre *caractère* et *glyphe*[2] : « Les caractères ne résident qu'en machine, sous forme de chaînes en mémoire, sur disques ou en mémoire auxiliaire. UNICODE ne s'intéresse qu'au code des caractères. Par contre, contrairement aux caractères, les glyphes apparaissent sur les écrans ou sur le papier comme représentations particulières d'un ou plusieurs caractères se trouvant en mémoire auxiliaire. Un répertoire de glyphes forme une police. » [2].

Donc, en termes d'UNICODE, la relation entre glyphes et caractères n'est pas biunivoque. Ainsi, la capitale latine A, par exemple, est-elle codée comme caractère UNICODE 0041 (en hexadécimal), mais le glyphe visuel représentant ce caractère à un moment donné sur papier ou sur écran peut être un A *Times-Roman*, ou un A *Helvetica*, ou un A *Courier*, ou tout autre forme de A en fonction de la police choisie pour le texte. Par contre les A latin (0041), cyrillique (0410) et l'Alpha grec (0391) sont des caractères UNICODE distincts qui peuvent tous être représentés par le même glyphe. Même dans un alphabet unique, on peut associer plus d'un glyphe à un caractère donné. En arabe, par exemple, la forme d'une lettre écrite dépend de son contexte et le dessin du glyphe diffère selon que le caractère apparaît en position initiale, médiale ou finale du mot ou de façon isolée dans une chaîne de caractères. UNICODE 1.0 ne code pas ces variations contextuelles (voir cependant ci-dessous la section 4.6 « Révisions et mises à jour »).

Malgré l'utilité des précautions d'UNICODE pour distinguer caractère et glyphe, nous utiliserons dans cet article le mot *caractère* – de façon plutôt libre afin de tenir compte de la tradition et du langage courant – pour désigner un élément d'une écriture ou une unité d'un système d'écriture. Par écriture ou système d'écriture, nous entendons la représentation graphique d'un langage. Un caractère dans ce sens large est donc une unité quelconque d'écriture, telle qu'une lettre de l'alphabet latin, un idéogramme (ou plus proprement un logographe) du système d'écriture chinoise, un chiffre, un signe de ponctuation, un symbole, une unité d'espace, etc. Avant l'invention de la typographie, les caractères étaient dessinés « au vol » par la main du scribe et la distinction logique qu'UNICODE fait entre glyphe et caractère n'était pas aussi forte. L'écriture manuscrite contient habituellement des variations dépendant du contexte ; dans certains systèmes d'écriture, des règles explicites permettent de formaliser ces variations ; dans d'autres, ces variations sont inconscientes. La succession de variations contextuelles sur les glyphes est d'ailleurs à l'origine du passage des capitales aux minuscules ou du romain à l'italiquee, par exemple. L'invention de la typographie a abouti à des glyphes invariables, préfabriqués et produits en quantité, connus sous le nom de « types » (ou caractères en plomb). La séparation du processus de création de lettres de celui de la composition des textes, ainsi que les besoins de la bureaucratie d'indexer, de conserver et de retrouver des caractères typographiques ont favorisé la distinction entre le concept abstrait de caractère ou de lettre et la réalisation tangible de ce concept abstrait sous forme d'un type. Comme la majorité des caractères composés, affichés et imprimés par ordinateurs sont typogra-

[2] La notion de glyphe est connue en français sous le nom d'« œil » depuis des siècles. Mais force nous est de constater que le mot « glyphe » est déjà tellement employé qu'on l'utilise, à regret, ici. Voir ce *Cahier GUTenberg* page ? {Ndlr}.

phiques, la séparation implicite entre caractère abstrait et glyphe concrèt s'est retrouvée naturellement dans la formulation logique d'UNICODE.

UNICODE propose un schéma uniforme pour le codage et l'identification de caractères basé sur un code fixé à 16 bits, schéma couvrant la majorité des systèmes d'écriture en usage dans le monde aujourd'hui. La version 1.0 code environ 28000 caractères dont environ 3 000 lettres alphabétiques et marques diacritiques pour les langages européens, indiennes et asiatiques, 1000 symboles divers et éléments graphiques et 24000 caractères idéographiques (logographiques), syllabiques et phonétiques utilisé pour les écritures chinoise, japonaise et coréenne[2, 3].

UNICODE est une étape importante pour la portabilité des documents à travers les frontières des différentes nations, des langages et des systèmes informatiques. L'étendue des possibilités d'UNICODE constitue un défi aux développeurs de logiciels, mais UNICODE est déjà implanté sur trois systèmes récents, à savoir Microsoft Windows NT .1, AT&T Bell Laboratories Plan 9 [4] et Apple QuickDraw GX [5].

### 1.1. Grands jeux et petits jeux de caractères

UNICODE pose problème aux dessinateurs de caractères du monde occidental qui ne sont pas habitués à faire entrer des milliers de caractères dans un seul jeu. Pendant la grande majorité de ces cinq derniers siècles, les caractères latins ont été fondus et ont formés des casses (ou jeux de caractères) suffisants pour composer la majorité des textes écrits en langues européennes. Bien que, au début, certains caractères aient pu être dessinés avec beaucoup de variantes et de ligatures pour imiter la très riche variété de l'écriture manuscrite, le succès de l'imprimerie en Europe est partiellement dû à l'efficacité et au bas coût de la composition gâce à de petites séries de caractères alphabétiques. Bien que la *textura* – le caractère utilisé par GUTENBERG dans sa *Bible* à 42 lignes de 1455– 56 (le premier livre imprimé en Europe) – utilisait plus de 250 caractères et que l'*humanistica corsiva* gravée par Francesco GRIFFO pour le *Virgile* d'ALDE en 1501 – le tout premier livre imprimé en italique – comprenait plus de 200 caractères, les jeux de caractères furent de plus en plus petits afin de réduire les coûts de gravure, de fonderie, de composition, etc.

Les jeux classiques de caractères contenaient les lettres capitales et bas de casse (avec accents et signes diacritiques), les ligatures, les chiffres, la ponctuation et parfois les petites capitales. Dans le tout premier manuel européen d'imprimerie et de fonderie de caractères, *Mechanick Exercises on the Whole Art of Printing*, publié en 1683, Joseph MOXON montre une casse composée de capitales et de bas de casse contenant environ 140 signes [6]. Dans le premier volume de son *Manuel typographique*, paru en 1764, Pierre-Simon FOURNIER dresse une liste de 155 caractères (dont 32 petites capitales) pour une police romaine [7]. Dans leur abrégé classique, *Typographical Printing Surfaces*, LEGROS et GRANT comptent, en 1916, 154 caractères (dont 29 petites capitales) pour une police romaine anglaise [8].

Le passage de la composition manuelle à la composition mécanique à la fin du XIX$^{e}$siècle n'a pas modifié la taille moyenne des jeux de caractères, pas plus que ne l'a fait le passage de la composition mécanique à la composition photographique dans les années 60. La matrice Monotype comprenait 225 à 272 caractères, pour deux styles d'une même famille, par exemple le romain et l'italique ; les matrices doubles de la Linotype étaient de taille équivalente [9]. En photocomposition, le nombre de caractères sur les films typographiques ou dans les matrices changeait beaucoup d'une machine à l'autre ; cependant, le nombre de caractères d'un seul style variait à peu près de 88 à 128 [10]. La machine à écrire, invention du XIX$^{e}$siècle restée en usage pendant tout le XX$^{e}$siècle, offre elle aussi un jeu de caractères assez limité. La machine à écrire *IBM Selectric* de réputation mondiale des années 60 aux années 80, fournissait 88 caractères sur chaque « boule ».

Dans les années 80, les ordinateurs personnels et les imprimantes à laser, commercialisées dans le monde entier, ont favorisé la production de jeux de caractères plus grands, 256 caractères codés sur 8 bits, adaptés à diverses langues. Beaucoup de ces jeux de caractères, par exemple ceux du Macintosh Apple [11], de Windows de Microsoft [12] et de PostScript d'Adobe [13], sont basés sur la norme Iso Latin 1 avec en plus quelques lettres grecques et symboles divers [13]. Comme UNICODE, les normes de codage de caractères telles que Ascii ou Iso Latin 1 font correspondre aux caractères des codes numériques lisibles par ordinateur, de façon à fournir une base normalisée pour les échanges d'informations[3]. Ces normes fournissent aussi des codes pour les caractères dits « de contrôle » comme le `carriage return` (« retour chariot », Ascii décimal 13) qui n'est pas imprimable mais qui a un rôle important dans les fichiers de textes électroniques.

L'introduction de symboles non alphabétiques et de lettres non latines dans ces jeux de caractères à 8 bits a exigé que les créateurs de police décident si ces symboles et ces lettres devaient avoir un style spécifique ou être génériques. Comme la plupart des designers, nous préférons des glyphes possédant un style spécifique, car nous pensons que la création de nouvelles polices pour l'édition électronique est l'occasion de coordonner caractères picturaux, non latins et latins. Certains vendeurs de polices ont cependant favorisé les caractères picturaux génériques car ils sont relativement standardisés et qu'ils réduisent donc le coût de production des polices.

Le choix entre symboles de style spécifique ou générique semble mineur pour les jeux de caractères à 8 bits d'Iso Latin, car seule une douzaine de caractères est concernée. Mais si on sort du domaine limité de l'alphabet latin défini par Iso Latin pour entrer dans le vaste royaume des écritures du monde entier codé en UNICODE, il devient crucial de décider d'unifier ou non le dessin de tous ces alphabets, signes, symboles et caractères hétérogènes. Cette décision concerne le dessin de milliers de glyphes.

[3] On consultera les divers articles de ce *Cahier GUTenberg* sur ces normes {Ndlr}.

# 2. Objectifs de notre création

## 2.1. Système international

En élaborant une police UNICODE, notre première motivation a été de fournir un ensemble standardisé de glyphes qui puissent être utilisé comme fonte par défaut pour différents systèmes d'exploitation et langages. Nous voulions que, à l'intérieur d'une police donnée, les divers alphabets et symboles suivent le même style de dessin. Si la même police de base est utilisée dans différents systèmes acceptant UNICODE, la chasse et l'apparence des caractères seront prévisibles et stables. Tout document – qu'il s'agisse de brefs messages de courrier électronique ou de rapports techniques multilingues complexes – sera facilement portable sans déformation typographique.

UNICODE est une norme de codage de caractères et non de glyphes. Rien n'oblige que le glyphe correspondant au code pécis d'un caractère UNICODE ait un dessin particulier ni que les glyphes de différents sous-ensembles d'UNICODE partagent des caractéristiques communes. Par conséquent, pour concevoir facilement une police UNICODE, on pourrait assembler de façon arbitraire des polices disparates : une police latine d'une certain style, une grecque d'un style différent, une cyrillique d'un troisième, des opérateurs mathématiques d'un quatrième, et ainsi de suite, toutes avec des graisses, chasses, proportions et formes différentes. Deux polices UNICODE différentes élaborées ainsi au hasard auraient des jeux de caractères UNICODE identiques mais produiraient des glyphes très différents.

Mais ce genre de conception aléatoire poserait des problèmes : les caractéristiques typographiques des documents, des fenêtres de programmation, des écrans et des autres images basées sur du texte ne seraient pas conservées lors du portage entre systèmes utilisant des polices différentes. Chaque élément typographique pourrait changer : style, fins de lignes, coupures de pages, longueur du document, taille des fenêtres, etc.

Notre but était d'éliminer ce genre de problèmes en créant une police par défaut qui offre stabilité et fiabilité, quelque soit le système utilisé.

## 2.2. Dessin harmonisé de caractères

Notre seconde motivation a été de tester le concept de dessin harmonisé *(harmonized design)* pour des jeux internationaux de caractères. Par « harmonisation », nous entendons la régularisation et l'accord des graisses et des alignements de base des alphabets disparates pour qu'ils fonctionnent ensemble, leurs différences inutiles étant minimisées mais leurs différences essentielles et significatives étant préservées. Ainsi filtrons-nous le « bruit » des artifices de dessin insignifiants et des accidents historiques pour amplifier au contraire le « signal » des caractéristiques significatives du caractère. Avec une police harmonisée, quand on passe, dans un texte, du latin au cyrillique, ou du grec à l'hébreu, ou quand on y introduit des expressions mathématiques ou autres symboles, alors la taille, la graisse et le rythme des caractères ne devraient pas changer d'apparence ni choquer ou distraire le lecteur ; les formes de base devraient néanmoins être distinctes et immédiatement reconnaissables.

Si l'harmonisation d'une police semble souhaitable, il n'est pas possible en pratique de tester l'efficacité typographique de ce style de conception unifiée sans utiliser une vraie police de symboles et caractères intégrés et harmonisés. Les idées de design formel doivent être conçues avant d'être évaluées. UNICODE fournit des codes, des noms et des représentations graphiques minimales des caractères, mais n'offre aucune ligne directrice pour harmoniser des jeux de caractères disparates. D'ailleurs, il n'existe aucune règle générale d'harmonisation de caractères ; seules quelques familles de caractères existantes peuvent nous servir de modèle.

Parmi celles-ci, citons (bien qu'elle ne soit jamais parue dans son intégralité) la curieuse famille *Romulus* de Yan VAN KRIMPEN qui comprenait à l'origine le romain, le grec, la cursive chancelière et des caractères sans empattements [14]. La famille *Encyclopedia* de Nikolaï KUDRYANSHOV qui incluait des alphabets latin, grec et cyrillique, avec ou sans empattements, ainsi que des signes et symboles spéciaux, est un autre exemple intéressant. Elle fut élaborée à l'origine pour une encyclopédie soviétique, mais elle est restée peu connue hors de Russie [15]. Le *Times New Roman* de Stanley MORISON [16, 17], l'*Helvetica* de Max MIEDINGER [18], l'*Univers* d'Adrian FRUTIGER [18] et le *Gill Sans* d'Eric GILL [17] sont autant d'exemples carctéristiques d'harmonisation de caractères initialement créés uniquement pour l'alphabet latin mais qui, à la suite de leur diffusion internationale, ont été étendus aux caractères grecs et cyrilliques. L'élaboration d'une police UNICODE nous mène encore plus loin.

### 2.3. Lisibilité comparée

La conception d'écritures non latines est aussi l'occasion de rechercher les « universels » de la lisibilité. La plupart des recherches sur la lisibilité en Europe ou en Amérique se sont concentrées sur les caractères latins, évaluant et comparant le rôle de facteurs tels que la graisse, la présence d'empattements, les capitales et le style [19, 20]. Les typographes et les lecteurs des langues latines sont prêts à donner leur avis sur la lisibilité comparée des caractères avec ou sans empattements, ou du romain et de l'italique ; pourtant il ne s'agit que de détails quand on le compare aux différences profondes qui existent, au niveau de la forme ou du détail, entre l'alphabet latin et des alphabets tels que l'hébreu, l'arabe, l'arménien ou le sanscrit et qui offrent pourtant des niveaux de lisibilité à peu près équivalents à ceux du latin. En tant que dessinateurs de caractères, nous aimerions savoir dans quelle mesure les « règles » qui gouvernent la lisibilité des langues latines s'appliquent aux non latines. SHIMRON et NAVON, par exemple, notent une différence significative dans la distribution spatiale des caractéristiques distincives entre alphabets romains et hébreux [21]. Nous aimerions voir publiées de telles études inter-culturelles pour toutes les écritures non-latines. Mais, en attendant, l'expérimentation reste une autre façon d'étudier ces notions même si ce n'a qu'un sens intuitif ou anecdotique.

### 2.4. Histoire et culture

Enfin, une de nos motivations relevait d'un plaisir artistique : dessiner les nombreux glyphes d'UNICODE. C'est l'occasion de contempler le passé tout en participant au futur en développant des milliers de formes graphiques abstraites élaborées depuis plusieurs millénaires par des générations de scribes de nombreuses civilisations. Concevoir une telle police est le moyen d'étudier et d'apprécier, à une échelle microscopique, la richesse de la culure et de l'histoire de la lettre. Pour nous, étudier et reproduire le dessin d'une lettre grecque ou copte dérivée d'un hiéroglyphe égyptien, ou n'importe quelle autre de ces milliers de formes écrites, revient à essayer de comprendre ce que pensait notre collègue il y a des milliers d'années lorsqu'il faisait cette même forme. C'est comme travailler dans l'esprit du théorème de PYTHAGORE avec le mode de pensée d'un géomètre d'alors. La forme n'est plus une ombre sans vie sur une page, mais une idée vivante. Nous ne pouvions pas laisser passer une telle occasion !

## 3. Conception de notre police

### 3.1. Choix d'une linéale

Nous avons choisi de baser notre première police UNICODE sur des caractères sans empattement. Il y a plusieurs raisons à celà.

1. Dans la plupart des cultures typographiques, les caractères sans empattements sont de création plus récente et sont par conséquent plus distants du contexte historico-culturel que les caractères avec empattements. Par ailleurs, l'utilisation courante de linéales pour les logos des compagnies aériennes, pour les échanges et les communications internationales, etc., les a rendues familières dans le monde entier. Les linéales tendent donc à avoir une apparence neutre mais rassurante, sans connotation spécifique à une nation, une société, une langue, une culture ou une religion particulière.

2. Notre police UNICODE est d'abord destinée à être utilisée dans un environnement technique : préparation, échange et impression de documents électroniques. Depuis les années 20, les linéales ont été promues pour leur aspect moderne, technique et industriel ; elles semblent donc appropriées à tout traitement de documents électroniques.

3. Les linéales permettent de bien voir les propriétés graphiques essentielles des lettres. Bien que l'histoire ait montré que les linéales ne sont jamais vraiment dépourvues de caractéristiques stylistiques, elles ont une apparence plus simplifiée que leurs homologues avec empattements. Elles permettent ainsi de montrer les traits fondamentaux d'un alphabet sans se noyer dans des détails. Le dessinateur peut ainsi analyser plus facilement les différences distinctives entre caractères et par là de mieux comprendre les différences entre alphabets.

4. Les linéales offrent de suite une solution simple au problème de l'italique. La création d'un caractère romain (le terme caractère *droit*, moins culturel, serait peut-être plus approprié puisque UNICODE est supposé inclure le grec, le cyrillique, l'hébreu et d'autres alphabets) nécessite celle d'un italique (de même, les termes *oblique*, *en biais*, *penché* ou *incliné* seraient peut-être meilleurs). Les caractères avec empattements sont habituellement accompagnés de vraies cursives dont les formes ne sont pas seulement inclinées mais aussi différentes (comparez le « a » du *Times romain* au « *a* » du *Times italique*), alors que les italiques linéales ne sont souvent que des versions penchées du romain.

   Beaucoup de typographes puristes pensent que la qualité de simples obliques est inférieure à celle des vraies cursives. Stanley MORISON affirma cependant, en 1926, que l'italique idéale pour un romain devait être une version inclinée de ce romain [22]. Bien que la théorie de MORISON ne marche pas pour les caractères avec empattements (il ne la suivit d'ailleurs pas lui-même quand il créa son *Times*), la norme pour les linéales est de pencher les romains pour obtenir l'italique correspondante.

   Étant donné qu'un caractère oblique peut être obtenu par transformation d'un dessin droit, la première version de *Lucida Sans* UNICODE pourraît être munie d'une italique par simple appel à quelque algorithme de transformation géométrique du systèmes d'affichage. Il ne serait donc pas nécessaire, dans un premier temps, de créer une police spéciale. Si on implémente UNICODE comme une grande police unique, on économise la place mémoire (centrale ou dans les imprimantes) des italiques en utilisant un programme de calcul des obliques. De plus, une oblique est beaucoup plus facile à concevoir qu'une vraie italique cursive même si les créateurs consciencieux n'utilisent pas qu'un simple algorithme pour pencher, mais modifient les formes obliques avec finesse.

   En plus de ces facteurs économiques, il y a aussi un autre argument en faveur d'un caractère d'accompagnement oblique pour une police UNICODE : la notion de « vraie italique » n'est pas un concept universel. La distinction européenne entre caractère droit et caractère cursif ne fait pas partie de la tradition de certaines écritures non-latines, par exemple l'hébreu (qui utilise par contre d'autres règles) ; une simple oblique est donc une distinction graphique plus universelle, bien que minimale, qui peut être appliquée de façon identique à toute écriture non latine.

Nous avons basé notre police UNICODE sur *Lucida Sans* qui est la version sans empattement d'une grande famille de caractères comprenant aussi des polices avec empattements, à chasse fixe, en script et pour les mathématiques [23, 24]. Tout comme d'autres caractères sans empattements « humanistes » tels que *Gill Sans* ou *Syntax* de Hans Ed. MEIER [25], *Lucida Sans* est plus proche des caractères de la Rennaissance, avec leur ductus manuscrit, que des linéales industrielles des XIX et XX$^{e}$siècles. Étant donné que notre étude de chaque écriture, latine ou non, débute par celle de l'écriture manuelle et de la calligraphie, nous pensons qu'il est opportun de baser la police UNICODE sur une création moderne qui conserve une sensibilité manuscrite. Bien sûr, *Lucida Sans* étant notre propre création, nous comprenons sa « logique » fondamentale et comment l'étendre à d'autres écritures. Nous avons aussi le droit légal de le faire[4] !

À propos de nos arguments en faveur d'une « italique » oblique pour UNICODE, rappelons ici que *Lucida Sans Italique* originale est une vraie cursive, basée sur les formes de chancellerie de la Renaissance et non un romain penché ; la version inclinée d'UNICODE sera donc appelée *Lucida Sans (*UNICODE*) 'Oblique'* pour la distinguer de vraies cursives.

### 3.2. Mise au point des signes diacritiques

Quand nous avons vu le très grand nombre d'accents et de signes diacritiques qui apparaissent dans UNICODE, nous avons conclu que, pour améliorer la lisibilité dans la composition de textes internationaux, les accents et les diacritiques devaient être conçus sensiblement différemment de la version standard de *Lucida Sans*. Dans un texte monolingue d'orthographe classique, il y a relativement peu d'accents différents et la plupart des bons lecteurs ont le sens de l'« image des mots » susceptibles d'être rencontrés : ils peuvent ainsi distinguer des mots se ressemblants. Les accents doivent donc être clairement différenciés mais sans exagération : trop grand ou trop épais, ils sont plus gênants qu'utiles au lecteur. Il n'en est pas de même pour les textes multi-lingues où il peut y avoir beaucoup d'accents différents selon l'orthographe des langues utilisées, lesquelles langues sont d'ailleurs souvent mal connues du lecteur. Pour faciliter la lisibilité, ou du moins pour faciliter le dechiffrage, les signes diacritiques nécessitent une plus grande différenciation. Nous avons donc créé des signes diacritiques bas de casse pour *Lucida Sans* UNICODE qui sont un peu plus grands et d'une modulation légèrement différente de ceux de *Lucida Sans* original. Suivant une pratique courante, nous avons aussi utilisé les accents des bas de casse pour les capitales.

Les sous-ensemble Latin 1 et Latin européen (voir figure 1) d'UNICODE contiennent plusieurs caractères composés de lettres et de signe diacritiques, mais cela ne limite pas les combinaisons possibles « lettre + diacritique ». UNICODE inclut en effet un jeu de signes diacritiques flottants (voir figure 2) qui peuvent se combiner avec n'importe quelle lettre. Ils ne varient pas en fonction du contexte où ils sont utilisés, c'est-à-dire qu'il n'y a qu'un seul jeu pour les capitales et les bas de casse. Dans notre première version de *Lucida Sans* UNICODE, nous les avons implémentés comme des signes diacritiques pour bas de casse et nous avons calculé leurs positions par défaut pour qu'ils soient centrés au dessus d'un « o » bas de casse. Dans l'idéal, il devrait y avoir au moins deux jeux de glyphes, un pour les bas de casses et un pour les capitales qui seraient automatiquement choisis par le gestionnaire de composition en ligne du système d'exploitation ou de l'application. Il faudrait aussi un jeu de tables de crénage qui optimiseraient l'allure de chaque paire « lettre + diacritique ». En raison de nombreuses questions et problèmes concernant la position idéale des signes diacritiques, nous remettons à une version future de la police la mise au point de doubles jeux de glyphes et de tables de crénage.

[4] Charles BIGELOW est un spécialiste, et vigoureux défenseur, du droit de copie des fontes. Voir par exemple « Du piratage des fontes », *TSI – techniques et science informatique*, 6(3), 1987, 255–259 et de nombreuses interventions dans `comp.font`. {Ndlr}

### 3.3. Chasse variable et chasse fixe

Dans une police à chasse variable (comme celle que vous lisez en ce moment), la chasse d'un caractère est proportionnelle à sa forme. Un « m » à chasse variable (d'une fréquence spatiale de trois cycles) est plus large qu'un « n » (fréquence de deux cycles) qui a son tour est plus large qu'un « i » (un seul cycle). Le rythme d'un mot comme « minimum » est une alternance régulière de formes positives et négatives qui donnent un rythme régulier au déroulement du texte. Pour conserver l'aspect traditionnel des différents alphabets et symboles d'UNICODE, nous avons conçu notre première police UNICODE avec une chasse variable.

Dans une police à chasse fixe, comme *Courier*, tous les caractères ont la même largeur de telle sorte que le « `m` » est resserré et le « `i` » étendu ; « `minimum` » a un rythme irrégulier puisque la fréquence spatiale des lettres change sans cesse dans la largeur fixe des cellules. La première norme à utiliser des polices à chasse fixe a été celle du Telex, suivie par celles pour les terminaux d'ordinateurs et les imprimantes jusqu'au milieu des années 80. Malgré leurs défauts esthétiques, les polices à chasse fixe offrent simplicité et rigueur dans la métrique et sont donc toujours utilisées par défaut, même dans les systèmes de fenêtrage récents : émulateurs de terminal TTY, « consoles », interfaces de programmation, émulateurs d'imprimante ligne à ligne et autres logiciels survivants d'une époque révolue.

Les polices à chasse fixe ont une utilité indéniable ; c'est pourquoi on nous a demandé, peu après la sortie de notre *Lucida Sans* UNICODE à chasse variable, d'en faire une version à chasse fixe compatible avec les anciens logiciels « rétros » encore largement utilisés dans le monde informatique. Nous avons basé cette seconde police compatible UNICODE sur notre *Lucida Sans Typewriter*. Celle ci est souvent utilisée pour émuler les polices de terminal et d'imprimante à lignes, mais aussi pour distinguer les programmes du texte dans les documents techniques (à propos de cette utilisation, voir les actes du colloque annuel de 1993 du *TEX User Group* [26]). Les polices à chasse fixe sont habituellement ignorées en typographie car on les associe à la machine à écrire, technique considérée de qualité inférieure à la « vraie » typographie. Mais les connaisseurs savent que les caractères à chasse fixe sont très vieux : les inscriptions grecques *stoichèdon* datent du Ve siècle avant notre ère. L'élaboration d'une police à chasse fixe est loin d'être évidente et requiert beaucoup d'ingéniosité et d'adresse, même si le résultat n'est pas apprécié à sa juste valeur. Parmi les jeux de caractères alphabétiques d'UNICODE, le cyrillique pose un problème intéressant en chasse fixe : comparé au latin, il contient un plus grand nombre de caractères à grande fréquence spatiale (trois cycles ou plus par lettre) sur l'axe horizontal. L'hébreu, par contre, est plus facilement adaptable, les lettres à grande fréquence spatiale étant rares.

### 3.4. Crénage vertical des signes diacritiques

Nous avons aussi été génés, lors de l'élaboration de notre police UNICODE à chasse fixe, par la hauteur des signes diacritiques. Le logiciel de la plupart des télétypes et des systèmes de fenêtrage supposent, à cause de leur origine ancienne, que les caractères sont disjoints : aucun élément ne peut s'introduire dans les cellules de caractères, ni dessus, ni dessous, ni de côté. Quand le texte est déroulé, manipulé, collé ou effacé, des morceaux de caractères s'introduisent dans l'espace de cellules voisines et peuvent poser plusieurs types de problèmes : pixels dispersés sur l'écran, morceaux de lettres coupées, etc. Faire l'hypothèse qu'un caractère est enfermé dans une cellule pouvait se concevoir du temps des premiers terminaux et télétypes ; mais ce n'est plus vrai avec les polices typographiques. Dans beaucoup de polices numériques de PostScript ou de TrueType, les signes diacritiques sur les capitales dépassent la hauteur du corps. En effet, afin de réduire la taille mémoire des fontes numériques, les caractères accentués, capitales ou bas de casse, sont composés à l'aide de deux programmes indépendants, l'un pour la lettre et l'autre pour le diacritique. De plus, une seule forme de diacritiques est utilisée, celle pour les bas de casse. Cette forme étant souvent plus grande que celle qu'il faudrait pour les capitales, les signes diacritiques placés sur les capitales se trouvent ainsi plus hauts que le sommet du corps proprement dit.

Le « crénage » vertical n'est en général pas un problème dans les textes imprimés. Si on compose à l'aide de bas de casse et de capitales, les capitales accentuées sont peu nombreuses et coïncident rarement avec les descendantes de la ligne supérieure ; si on ne compose qu'en capitales, les capitales accentuées peuvent être fréquentes, mais les descendantes sont rares (à part quelques Q et J). Si les accents rencontrent trop souvent les descendantes, on peut toujours soit interligner d'avantage, soit réduire le corps sans modifier l'interlignage. Mais ces deux solutions ne sont pas applicables aux écrans de terminaux à taille fixe. Dans le premier cas, on réduirait le nombre de lignes dans la fenêtre ; dans le second, c'est la résolution effective qui diminuerait au détriment des détails et de la lisibilité.

Nous avons résolu ce problème en créant une version spéciale, de Procustre, de *Lucida Sans Typewriter*. Les capitales et les signes diacritiques de cette nouvelle police, *Lucida Console*, sont spécialement raccourcis pour le latin, le cyrillique, le grec et les autres alphabets qui utilisent des capitales. On obtient ainsi un caractère à chasse fixe, qui respecte le corps, qui a des capitales accentuées distinctes, un bonne différentiation de taille entre capitales et bas de casse, ces dernières étant de grand œil. Grâce à la possibilité de combiner *Lucida Sans* UNICODE et *Lucida Sans Typewriter* et comme ces deux polices existent déjà, nous avons gagné en rapidité et en efficacité lors du processus de création.

## 4. Incertitudes et problèmes

Après avoir donné des arguments en faveur de la création d'une police UNICODE, nous devons aussi exposer les arguments contre une telle entreprise et les différents problèmes que nous avons rencontrés.

### 4.1. *Ars longa, vita brevis*

La première objection pourrait être tirée du célèbre proverbe d'HIPPOCRATE : *Ars longa, vita brevis* (« l'art est long, la vie est brève »). Une police qui contiendrait tout UNI-

CODE 1.0 aurait une taille décourageante, la norme continuant de grossir au fur et à mesure que le Consortium UNICODE lui ajoute des caractères. Même sans le jeu chinois-japonais-coréen, les alphabets et symboles forment environ 4000 caractères distincts, soit 16 fois plus que dans les jeux habituels de caractères codés sur 8 bits. Pour que tous ces caractères soient harmonisés, il faut qu'ils soient conçus par un seul créateur ou par quelques uns en étroite collaboration. Dans la plupart des cas, il faut donc des années pour qu'une police UNICODE soit achevée, et les premières versions seront inévitablement incomplètes. Nous pensons élaborer la police comme une série de logiciels, où chaque nouvelle version ajoutera des « fonctionnalités » sous forme de jeux de caractères complémentaires.

Une police incomplète ne devrait pas porter le nom UNICODE car les utilisateurs pourraient croire par erreur qu'elle contient l'ensemble complet des 28 000 caractères. Cependant, il est difficile de nommer une police formée de nombreux sous-ensembles. Notre permière version de *Lucida Sans* UNICODE contient 1725 caractères, ceux des blocs UNICODE suivants: Ascii, Latin 1, Latin européen, latin étendu, phonétique, modificateurs de lettres, signes diacritiqeus flottants, grec, cyrillique, hébreu, ponctuation générale, exposants et indices, pictogrammes des caractères de contrôle, éléments de formulaires et graphiques, filets et formes géométriques. D'autres jeux, arménien, arabe, sanscrit et les variantes d'indien ainsi que divers jeux de symboles (formes de nombres, technique, dingbats, etc.) seront ajoutés dans les prochaines versions. Nous n'avons pas encore trouvé à ce jour une méthode de dénomination qui puise décrire avec concision tous les sous-ensembles contenus dans une police UNICODE, d'autant plus que beaucoup de systèmes d'exploitation imposent de sévères limitations à la longueur des noms de fichiers.

Ces 4 000 lettres et symboles ne sont qu'une fraction d'UNICODE complet : il y a environ 24 000 caractères *idéographiques Han*[5] [3]. Il s'agit des logographes utilisés en chinois, japonais et coréen. UNICODE « unifie » les caractères communs à ces écriture puisque les idéogrammes sont pour la plupart dérivés de sources communes ; mais, selon les pays et les cultures, les styles graphiques se différencient. En conséquence, une interprétation unique du jeu *idéographique Han* n'est pas possible pour les trois pays. Nous devons donc admettre que l'ampleur du travail sur la représentation de milliers de glyphes du *idéographique Han* dépasse la portée de notre seule participation, malgré notre émerveillement pour ces caractères : « l'art est long, la vie est brève » ! Une solution pratique à ce problème serait d'adjoindre à la partie « alphabétique » de notre police UNICODE des polices idéographiques de même style. Pour le japonais, par exemple, nous pourrions associer à *Lucida Sans* UNICODE une police Kanji de style « gothique »[6] avec graisse et proportions en harmonie. de même pour le chinois et le coréen. Notre rôle consisterait alors à collaborer avec des collègues chinois, japonais et coréens.

[5] *Han* veux dire tout simplement « chinois », au sens ethnique du terme : le peuple chinois, la civilisation chinoise, etc. Han-zi, c'est l'écriture chinoise (« idéogrammes chinois », ou logographes chinois, mais idéogrammes, quoique éthymologiquement incorrect, est le plus utilisé et le mieux compris). En japonais, Hanzi devient Kanji, mais ce sont exactement les mêmes idéogrammes ; dans le cas du coréen, il y a un peu plus de différences. L'évolution historique des écritures a fait que les mêmes caractères d'origine ont à l'heure actuelle des graphies variables. C'est ce que nous voulons indiquer, tout en évitant des conflits de susceptibilité nationale, en conservant le mot Han de préférence à chinois. {Ndlr}

[6] Le terme japonais « gothique » est équivalent à « sans empattement » ; il est également utilisé en anglais pour les linéales, en particulier pour les créations américaines du XIX[e] siècle comme *Franklin Gothic* ou *News Gothic*.

### 4.2. Frontières culturelles et création

La seconde objection relève du problème des « frontières culturelles » : les créateurs de caractères compétents dans une certaine tradition le sont-ils dans une autre ? Nous avons déjà mentionné les différences culturelles entre les caractères *idéographiques Han* utilisés dans les différents pays asiatiques, mais on retrouve des problèmes similaires, à plus petite échelle il est vrai, pour les écritures latines. Le *Times romain* adapté par Stanley MORISON en 1931 d'une création française du XVI^e^siècle (avec sans doute quelques influences flamandes) est un exemple instructif. Bien que MORISON fit autorité en typographie au XX^e^siècle et que le *Times romain* fût un des caractères les plus employés, des typographes reprochèrent au *Times New Roman* son style trop anglo-saxon. Pour répondre à ces critiques françaises et permettre au *Times* de mieux attaquer le marché français, la société Monotype grava une version spéciale de ce caractère en tenant compte des préceptes de Maximilien VOX, célèbre typographe français[7] [16, 27]. La re-création par Vox rendit le *Times* sensiblement plus proche du style recherché du *Romain du Roi* gravé par Philippe GRANDJEAN vers 1693. Si Anglais et Français – peuples engagés dans des échanges littéraires et culturels depuis la fin du VIII^e^siècle (lorsque Alcuin de York, sur invitation de Charlemagne, devint le maître des écoles du Palais de l'Empire franc) – peuvent ne pas avoir la même notion de goût typographique, que dire pour des cultures radicalement opposéess !

Les Français n'ont pas été les seuls à vouloir modifier le *Times*. Monotype en grava une version pour l'Allemagne avec des capitales plus maigres, plus adpatée à l'allemand où chaque substantif prend une majuscule [16]. À notre connaissance, cette version n'est pas disponible pour les imprimantes à laser. Le typographe suisse Hans MEIER a lui aussi dessiné un *Times* expérimental pour le département d'informatique de l'ETH à Zurich [25].

On retrouve, de façon encore plus flagrante, un tel gouffre de différence culturelle avec un caractère dessiné pour l'hébreu par Eric GILL en 1937, caractère que Henri FRIEDLAENDER considère comme une « curieuse tentative hybride d'imposer le système de la capitale romaine dans la Merooba alors que celle-ci est gouvernée par des lois complètement différentes » [28]. Le *Hadassah* de FRIEDLAENDER est toujours utilisé en Israël, mais plus les caractères de GILL.

Face au dilemme culture/design, nous pouvons nous référer aux sciences sociales telles que l'ethnologie et la linguistique, qui supposent que la culture humaine est compréhensible et transmissible au delà des frontières d'une société donnée. Une telle démarche peut aussi s'appliquer à la typographie. Pour quelquun de tradition latine, dessiner un caractère non-latin est l'occasion d'apprendre de nouvelles règles de dessin, de nouveaux concepts et de nouveaux modes de lisibilité. En effet, comme le montre l'histoire de la typographie européenne, la majorité des bons dessinateurs de caractères a regardé les caractères non-latins avec enthousiasme et respect. Les *Grecs Aldins* de Francesco GRIFFO [29], les *Grecs du roi* de Claude GARAMOND [30], les *Hébreux* de Guillaume LE BÉ [31] et les *Grecs*, *Hébreux*, *Syriens* et *Arabes* de FOURNIER [7, vol. 2] témoignent du

[7] Vox, de son vrai nom Samuel Monod, 1894–1974, a notamment créé la revue *Caractères*, la classification qui porte son nom, les *Rencontres de Lure* et l'*ATypI, Association de Typographie Internationale* {Ndlr}.

savoir-faire, de la compréhension et de l'attachement que les créateurs ont montrés pour les écritures étrangères.

Si nous devions nominer un patron ou un modèle historique de la création de caractères pluri-culturel pour son apport à la typographie moderne, ce serait Robert GRANDJON, graveur français de poinçons du XVI$^{e}$siècle. Artiste renomé pour son habileté, graveur de romains et italiques élégants, il a été particulièrement génial dans la gravure de splendides polices pour l'arménien, l'arabe, le cyrillique, le syrien et proablement aussi pour l'hébreu et le grec. À Rome, ses mécènes écclésiastiques l'appelaient « l'excellent», le « plus extraordinaire » ou « le meilleur » des graveurs de lettres [30].

### 4.3. Homogénéisation

On pourait penser que c'est l'homogénéité et non l'harmonisation qui régularise les dessins de lettres. Si l'on efface les différences distinctives, on augmente les possibilités de confusion. Aussi, filtrer le bruit superflu risque au contraire d'éliminer les parties signifiantes d'un signe et donc de l'altérer, l'affaiblir et le parasiter.

Pour éviter ce problème, nous avons essayé d'apprendre les règles de base inhérantes à chaque langue non latine avant d'en commencer le dessin pour UNICODE. Nous avons étudié des manuels pour enfants, des livres de calligraphie pour adultes, des photos d'écritures manuscrites historiques, des caractères créés par des typographes spécialistes de ces écritures et toutes sortes d'exemples de textes typographiés et d'images. Notre approche n'est pas infaillible, mais du moins est-elle basée sur l'écriture elle-même et non sur les notions latines de création de lettres « correctes ». Nous n'avons pas essayé d'adapter ces écritures non latines à un modèle latin précis. Pour *Lucida Sans Hébreu*, par exemple, nous avons mis au point et testé trois versions. Dans l'une, les lettres hébraïques étaient alignées sur les capitales latines ; dans une autre, elles l'étaient sur les bas de casse ; dans la troisième, à mi-chemin entre les deux. C'est cette dernière qui offrait le meilleur résultat quand de l'hébreu était utilisé en même temps que des capitales et des bas de casse latines. De même, nous avons respecté la modulation traditionnelle de la graisse pour l'hébreu : les éléments horizontaux sont plus gras que les verticaux, ce qui est contraire aux habitudes latines. En les graissant ainsi, nous avons donné aux caractères hébreux une « présence » équivalente à celles des caractères latins.

### 4.4. Norme de caractères ou de glyphes ?

Ainsi que nous l'avons dit plus haut, UNICODE est une norme de codage de caractères et non de glyphes. Elle ne prend pas en compte les diverses versions de caractères si ils sont seulement « allographiques », c'est-à-dire si les différences n'ont pas de sens orthographiquement parlant, ou si elles ne changent pas la signification d'un mot. Par exemple, UNICODE considère la capitale latine « B » comme un caractère et le bas de casse « b » comme un autre caractère car ils sont différents pour l'orthographe latine. En anglais, *Bill* est le diminutif de William alors que *bill* signifie bec d'oiseau, bord de casquette ou facture[8]. En revanche, UNICODE considère qu'un « *b* » italique ou qu'un « **b** » gras ne sont pas différents d'un « b » romain parce que ces lettres sont simplement des allographes ou, comme disent les linguistes, parce que les différences graphiques ne sont pas orthographiquement signifiantes.

La correspondance biunivoque d'isomorphisme entre caractère et glyphe dans une police n'est pas toujours sans inconvénients. Dans bien des cas, les polices doivent contenir plus de glyphes que de caractères. La belle typographie latine nécessite souvent des variantes de lettres ou de symboles graphiquement distincts mais ne correspondant pas à des caractères différents. Les livres et la presse de qualité, par exemple, utilisent souvent des petites capitales, des formes anciennes, des lettres ornées et des ligatures de fantaisie. UNICODE ne propose pas de code pour ces distinctions alors que de nombreux graphistes aimeraient en disposer dans leurs polices. Pour certaines écritures non latines, la différence entre caractère et glyphe devient encore plus grande, plus fondamentale. On sait que les écritures arabes utilisent des variations contextuelles, allographes, pour la plupart des lettres. Mais ces variations (nécessaires pour les formes initiales, médiales et finales ou pour les lettres isolées) sont graphiques et non sémantiques. Elles ne sont donc pas codées en UNICODE 1.0[9]. Un caractère peut donc correspondre à plusieurs glyphes selon le contexte où il est utilisé.

Un problème un peu différent peut se produire avec des langues qui utilisent de légères variantes de même caractères. Ainsi, S cédille et T cédille sont utilisés en turc et en roumain ; mais selon le groupe d'utilisateurs on utilise soit la cédille française, soit une sorte de virgule en-dessous de la lettre. Mais UNICODE ne distingue pas ces variantes.

Notre première version de *Lucida Sans* UNICODE considère simplement le codage des caractère comme un codage de glyphes et n'offre pas de variantes, à l'exception de quelques cas où l'utilisateur peut redéfinir les codes. Nous avons simplifié à l'extrême ce problème subtil et complexe mais, grâce à la correspondance biunivoque entre caractère et glyphe, la police est plus facile à utiliser et plus portable puisque n'importe quel système peut accèder à tous les glyphes directement par le code du caractère. Nous considérons la dualité glyphe/caractère comme une limite temporaire de la technique et non comme une énigme philosophique pour les dessinateurs. Des systèmes, comme par exemple Quickdraw GX de Macintosh [5], autorisent de plus en plus la manipulation de polices avec des variantes de glyphes. UNICODE devrait être capable d'inclure autant de variantes de glyphes que nécessaire pour atteindre la qualité et les besoins exigés par les créateurs et les utilisateurs.

### 4.5. Une seule grande ou plusieurs petites polices ?

Une fois les glyphes dessinés et numérisés pour des sous-ensembles d'UNICODE (Ascii, latin 1 ou cyrillique par exemple), il y a deux façons de les mettre en œuvre : soit comme un ensemble de petites polices indépendantes, correspondant chacune à une classe d'UNICODE ; soit comme une grande police unique dans laquelle toutes les sous-

[8] En français, on peut comparer « Marc », prénom, et « marc », eau de vie ou résidu de café {Ndlr}.

[9] Mais elles le sont maintenant en UNICODE 1.1. Voir ci-dessous section 4.6 et [1].

polices sont réunies. Le choix de la méthode est, bien sûr, moins important pour les dessinateurs que pour les auteurs de systèmes. PIKE et THOMPSON [4] ont étudié le gain de mémoire et la vitesse de chargement d'une police UNICODE complète construite comme un ensemble de sous-polices à partir desquelles on peut accéder indépendamment aux caractères ; c'est la méthode utlisée pour les polices bitmap et les polices définies par contours de *Lucida Sans* UNICODE du sytème Plan 9. Windows NT 3.1 de Microsoft utilise l'autre méthode : la première version de *Lucida Sans* UNICODE y est mise en œuvre comme une seule police TrueType de 1740 glyphes ; le mécanisme de manipulation des polices est plus simple et l'harmonisation des sous-polices plus facile puisque tous les caractères peuvent être analysés et coordonés en une seule fois par le logiciel de traitement des *hints*.

Nous pouvons fournir *Lucida Sans* UNICODE indifféremment sous forme de plusieurs petites polices, comme pour Plan 9, ou sous forme d'une seule grande police, comme pour Windows NT3.1.

### 4.6. Revision

Toute norme est susceptible d'évoluer. Nous avions basé la première version de notre police UNICODE sur la version 1.01 de la norme UNICODE. Mais avant même que nous ayons fini, nous avons dû apporter quelques changements pour que la police soit compatible avec la version 1.01. En 1992–1993, UNICODE fusionna avec la norme Iso/IEC 10646 pour devenir UNICODE version 1.1. Celle-ci contient 5 000 caractères supplémentaires dont des ligatures pour les formes des variantes contextuelles de glyphes dans les sous-ensembles latin, arménien, hébreu et arabe [1, 32]. L'augmentation continuelle de la taille et de la complexité de ces grands jeux de caractères nécessite une mise à jour constante en fonction de l'évolution des normes et de l'addition de nouveaux jeux de caractères. Les futures conventions de dénomination devront donc inclure des numéros de version et préciser le contenu du jeu de caractères.

## 5. Conclusion

Dans un ensemble de 1700 dessins de caractères, beaucoup de détails peuvent être remis en question. Mais nous pensons avoir abordé nombre des problèmes parmi les plus généraux et les plus intéressants.

Pour clore notre article, nous donnons ci-après un spécimen de la première version de *Lucida Sans* UNICODE en classant les caractères en fonction des sous-ensembles d'UNICODE.

### Remerciements

Nous sommes reconnaissants à Steve SHAIMAN et Georges MOORE de Microsoft de nous avoir commandé la première version de *Lucida Sans* UNICODE ; à David MCBRIDE, Dave Ohara, Asmus Freytag, Michel Suignard et bien d'autres personnes de chez Microsoft qui ont contribué à tester et mettre au point cette police en vue de son utilisation sous Windows NT3.1. Nous remercions Rob Pike, John Hobby et toutes les autres personnes des laboratoires d'AT&T Bell d'avoir préparé et rodé une version de *Lucida Sans* Unicode sur Plan 9. Merci à Glenn Adams du Consortium Unicode d'avoir attiré notre attention sur l'intérêt et la possibilité de dessiner une police Unicode. Nous voudrions aussi remercier Pierre MacKay, William Bright, Michael Sheridan, Rick Cusik et Kathy Shinhofen qui nous ont confié d'utiles spécimens d'alphabets et des conseils efficaces ; ils l'ont peut-être oublié, mais pas nous et leur en sommes reconnaissants.

## Bibliographie


[1] Michel Goossens, 'Les codages Unicode et ISO/IEC 10646', *Cahiers GUTenberg* n° 20 (ce cahier), mars 1995 {Ndlr}.

[2] The Unicode Consortium, *The Unicode Standard: Worldwide Character Encoding,* Version 1.0, volume 1, Addison-Wesley, Reading, MA, 1991.

[3] The Unicode Consortium, *The Unicode Standard: Worldwide Character Encoding,* Version 1.0, volume 2, Addison-Wesley, Reading, MA, 1992.

[4] Rob Pike and Ken Thompson, 'Hello world . . .', in *Proceedings of the Winter 1993 USENIX Conference*, pp. 43–50, San Diego, (1993).

[5] Apple Computer, *Inside Macintosh: QuickDraw GX Typography*, Addison-Wesley, Reading, MA, 1994.

[6] Joseph Moxon, *Mechanick Exercises on the Whole Art of Printing*, Herbert Davis and Harry Carter, eds., Oxford University Press, Oxford, 1962.

[7] Pierre-Simon Fournier, *Manuel Typographique*, volume 1, chez l'Auteur, rue des Postes, chez Barbou, rue St. Jacques, Paris, 1764. volume 2, *idem*, 1766.

[8] Lucien Legros and John Grant, *Typographical Printing-Surfaces*, Longmans, Green, London, 1916. (reprinted, Garland Publishing, New York, 1980).

[9] Arthur H. Phillips, *Computer Peripherals & Typesetting*, HMSO, London, 1968.

[10] Arthur H. Phillips, *Handbook of Computer-Aided Composition*, Marcel Dekker, Inc., New York, 1980.

[11] Apple Computer, *Inside Macintosh*, volume VI, Addison-Wesley, Reading, MA, 1991.

[12] Microsoft Corporation, Redmond, Washington, *TrueType 1.0 Font Files: Technical Specification*, Revision 1.02, May 1993.

[13] Adobe Systems, *PostScript Language Reference Manual*, Addison-Wesley, Reading, MA, 1985.

[14] John Dreyfus, *The Work of Jan van Krimpen*, Sylvan Press, London, 1952.

[15] Evgenia Butorina, *The lettering art: Works by Moscow book designers*, Kniga, Moscow, 1977.

[16] John Dreyfus, 'The evolution of Times New Roman', *Penrose Annual*, **66**, (1973).

[17] Monotype Typography, Salfords, Redhill, *Library of Non-Latin Typefaces*, 1992.

[18] Linotype Corporation, Eschborn, *LinoType Collection*, 1988.

[19] Miles A. Tinker, *Legibility of Print*, Iowa State University Press, Ames, 1963.

[20] Bror Zachrisson, *Studies in the Legibility of Printed Text*, Almqvist & Wiksell, Stockholm, 1965.

[21] Joseph Shimron and David Navon, 'The distribution of visual information in the vertical dimension of roman and hebrew letters', *Visible Language*, **XIV**(1), (1980).

[22] Stanley Morison, 'Towards an ideal italic', *The Fleuron*, **V**, (1926).

[23] Charles Bigelow and Kris Holmes, 'The design of Lucida: an integrated family of types for electronic literacy', in *Text processing and document manipulation*, ed. J. C. Van Vliet, Cambridge University Press, (1986).

[24] Microsoft Corporation, Redmond, Washington, *Microsoft TrueType Font Pack for Windows User's Guide*, 1992.

[25] Hans Ed. Meier, 'Schriftgestaltung mit Hilfe des Computers – Typographische Grundregeln mit Gestaltungbeispielen', Technical report, Institute for Computer Systems, ETH, Zurich, (1990). {Une partie de cette note est parue en français : « Règles fondamentales de mise en page », *Cahiers GUTenberg*, n° 12, juin 1992, 5–38. Ndlr }.

[26] 'TEX Users Group annual meeting proceedings', *TUGboat*, **14**(3), (1993).

[27] John Dreyfus, communication personnelle.

[28] Henri Friedlaender, 'Modern Hebrew typefaces', *Typographica*, **16**, (sans date).

[29] Nicolas Barker, *Aldus Manutius and the Development of Greek Script and Type in the Fifteenth Century*, Chiswick Book Shop, Sandy Hook, Connecticut, 1985.

[30] *Cabinet des poinçons de l'Imprimerie nationale de France*, Imprimerie nationale, Paris, 1948. Raymond Blanchot étant directeur et Georges Arnoult inspecteur de la typographie.

[31] Hendrik D. L. Vervliet and Harry Carter, *Type Specimen Facsimiles II*, University of Toronto Press, Toronto, 1972.

[32] The Unicode Consortium, 'The Unicode Standard, version 1.1', Unicode Technical Report 4, Unicode, Inc., Menlo Park, California, (1993).

<table>
<tr><td>ASCII</td><td>0000→</td><td>! " # $ % & ' ( ) * + , - . / 0 1 2 3 4<br>5 6 7 8 9 : ; < = > ? @ A B C D E F G H<br>I J K L M N O P Q R S T U V W X Y Z [ \<br>] ^ _ ` a b c d e f g h i j k l m n o p<br>q r s t u v w x y z { | } ~</td></tr>
<tr><td>Latin1</td><td>00A0→</td><td>¡ ¢ £ ¤ ¥ ¦ § ¨ © ª « ¬ - ® ¯ ° ± ² ³ ´<br>µ ¶ · ¸ ¹ º » ¼ ½ ¾ ¿ À Á Â Ã Ä Å Æ Ç È<br>É Ê Ë Ì Í Î Ï Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü<br>Ý Þ ß à á â ã ä å æ ç è é ê ë ì í î ï ð<br>ñ ò ó ô õ ö ÷ ø ù ú û ü ý þ ÿ</td></tr>
<tr><td>Latin européen</td><td>0100→</td><td>Ā ā Ă ă Ą ą Ć ć Ĉ ĉ Ċ ċ Č č Ď ď Đ đ Ē ē<br>Ĕ ĕ Ė ė Ę ę Ě ě Ĝ ĝ Ğ ğ Ġ ġ Ģ ģ Ĥ ĥ Ħ ħ<br>Ĩ ĩ Ī ī Ĭ ĭ Į į İ ı Ĳ ĳ Ĵ ĵ Ķ ķ ĸ Ĺ ĺ Ļ<br>ļ Ľ ľ Ŀ ŀ Ł ł Ń ń Ņ ņ Ň ň ŉ Ŋ ŋ Ō ō Ŏ ŏ<br>Ő ő Œ œ Ŕ ŕ Ŗ ŗ Ř ř Ś ś Ŝ ŝ Ş ş Š š Ţ ţ<br>Ť ť Ŧ ŧ Ũ ũ Ū ū Ŭ ŭ Ů ů Ű ű Ų ų Ŵ ŵ Ŷ ŷ<br>Ÿ Ź ź Ż ż Ž ž</td></tr>
<tr><td>Latin étendu</td><td>0180→</td><td>ƀ Ɓ Ƃ ƃ Ƅ ƅ Ɔ Ƈ ƈ Ɖ Ɗ Ƌ ƌ ƍ Ǝ Ə Ɛ Ƒ ƒ Ɠ<br>Ɣ ƕ Ɩ Ɨ Ƙ ƙ ƚ ƛ Ɯ Ɲ ƞ Ɵ Ơ ơ Ƣ ƣ Ƥ ƥ Ʀ Ƨ<br>ƨ Ʃ ƪ ƫ Ƭ ƭ Ʈ Ư ư Ʊ Ʋ Ƴ ƴ Ƶ ƶ Ʒ Ƹ ƹ ƺ ƻ<br>Ƽ ƽ ƾ ƿ ǀ ǁ ǂ ǃ Ǆ ǅ ǆ Ǉ ǈ ǉ Ǌ ǋ ǌ Ǎ ǎ Ǐ<br>ǐ Ǒ ǒ Ǔ ǔ Ǖ ǖ Ǘ ǘ Ǚ ǚ Ǜ ǜ ǝ Ǟ ǟ Ǡ ǡ Ǣ ǣ<br>Ǥ ǥ Ǧ ǧ Ǩ ǩ Ǫ ǫ Ǭ ǭ Ǯ ǯ ǰ</td></tr>
<tr><td>Phonétique</td><td>0250→</td><td>ɐ ɑ ɒ ɓ ɔ ɕ ɖ ɗ ɘ ə ɚ ɛ ɜ ɝ ɞ ɟ ɠ ɡ ɢ ɣ<br>ɤ ɥ ɦ ɧ ɨ ɩ ɪ ɫ ɬ ɭ ɮ ɯ ɰ ɱ ɲ ɳ ɴ ɵ ɶ ɷ<br>ɸ ɹ ɺ ɻ ɼ ɽ ɾ ɿ ʀ ʁ ʂ ʃ ʄ ʅ ʆ ʇ ʈ ʉ ʊ ʋ<br>ʌ ʍ ʎ ʏ ʐ ʑ ʒ ʓ ʔ ʕ ʖ ʗ ʘ ʙ ʚ ʛ ʜ ʝ ʞ ʟ<br>ʠ ʡ ʢ ʣ ʤ ʥ ʦ ʧ ʨ</td></tr>
</table>

Figure 1 – Palette de Lucida Sans UNICODE (première des trois parties)

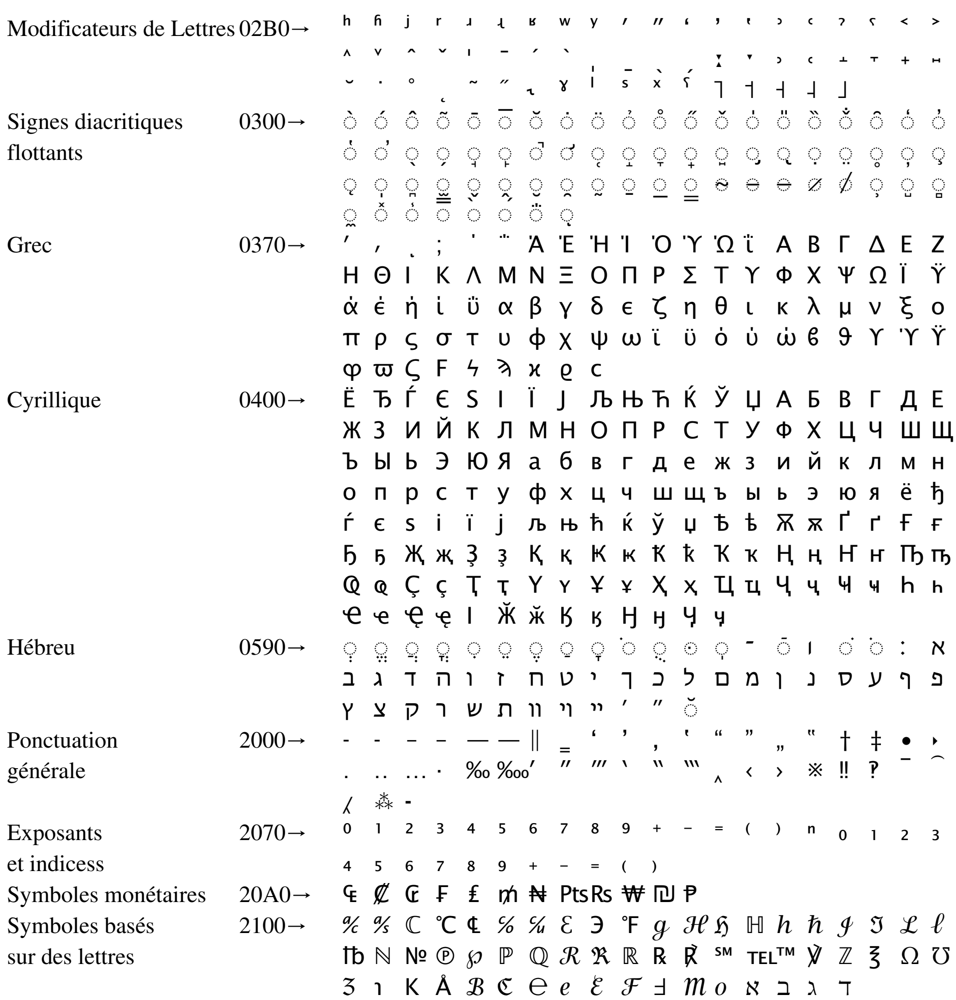


Figure 2 – Palette de Lucida Sans UNICODE (seconde partie)

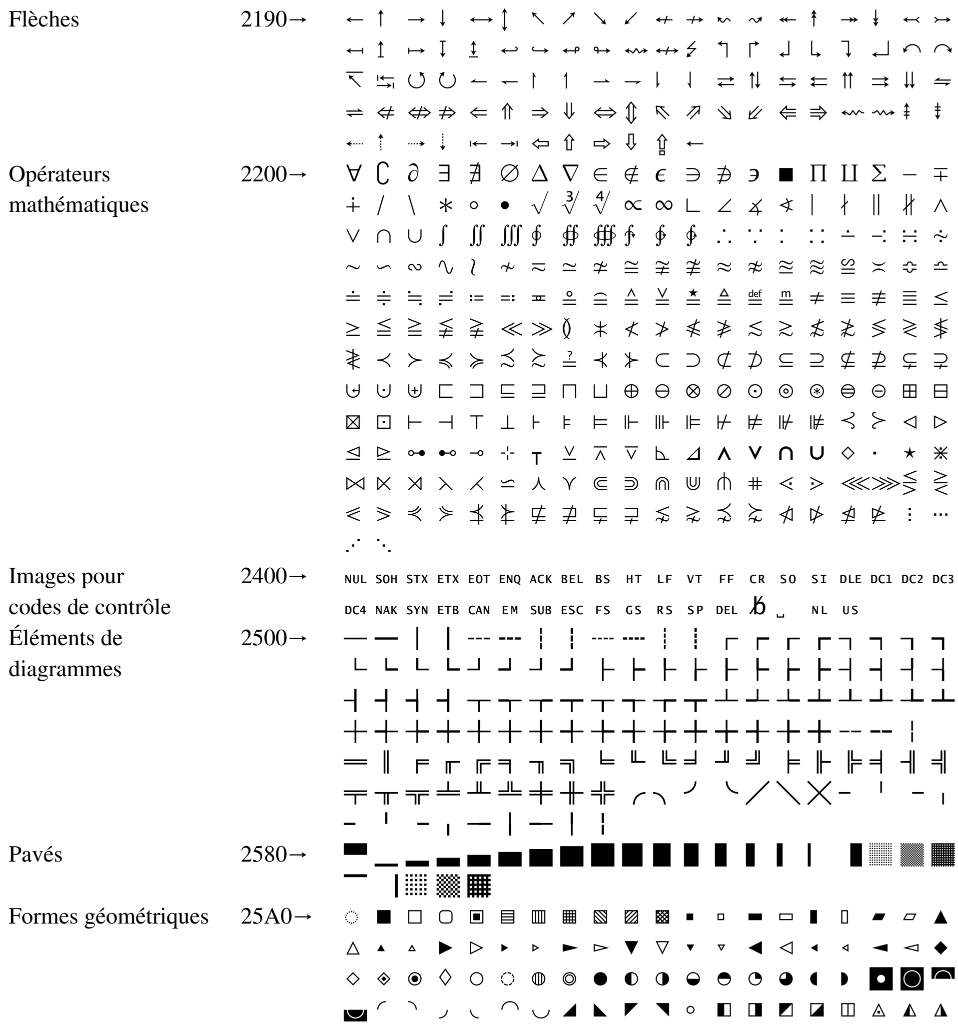


Figure 3 – Palette de Lucida Sans UNICODE (troisième et dernière partie)