



## Ligatures, typographie et informatique

Jacques André, Adolf Wild

► **To cite this version:**

Jacques André, Adolf Wild. Ligatures, typographie et informatique. [Rapport de recherche] RR-2429, INRIA. 1994. inria-00074246

**HAL Id: inria-00074246**

**<https://hal.inria.fr/inria-00074246>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

## *Ligatures, typographie et informatique*

Jacques André & Adolf Wild

**N° 2429**

5 décembre 1994

PROGRAMME 3



*rapport  
de recherche*



## Ligatures, typographie et informatique

Jacques André \* & Adolf Wild \*\*

Programme 3 — Intelligence artificielle, systèmes cognitifs  
et interaction homme-machine  
Projet Opéra

Rapport de recherche n° 2429 — 5 décembre 1994 — 58 pages

**Résumé :** Cette note comprend le texte de deux exposés sur les ligatures. Dans le premier, Adolf Wild regarde la *Bible* de GUTENBERG avec l'œil d'un typographe, d'abord au niveau de la page où ROSAVIRO a montré l'existence de constantes dans les grilles ; et ensuite au niveau des lignes qui étaient justifiées par l'emploi de ligatures et non par des blancs de taille variable. Dans la seconde partie, on étudie les ligatures du point de vue informatique : les normes d'échange de caractères (comme Ascii ou IsoLatin 1) ne contiennent pas de ligatures car celles-ci ne sont que des glyphes. Cependant elles existent dans les tables de codage des imprimantes. On montre alors comment les ligatures peuvent être reconnues et traitées par des formateurs comme T<sub>E</sub>X.

**Mots-clé :** Ligatures, normes d'échange de caractères, codage de caractères, PostScript, T<sub>E</sub>X

(Abstract: pto)

Ce travail a été partiellement financé dans le cadre du projet Didot, programme Comet de la CEE, numéro Cb/90/367/Cb. Ces deux notes ont été soumises aux *Cahiers GUTenberg* pour parution dans un numéro spécial sur les ligatures (à paraître en 1995).

\*. Irisa/Inria-Rennes, Campus de Beaulieu, 35042 Rennes cedex – [jandre@irisa.fr](mailto:jandre@irisa.fr)

\*\* Stadt Mainz Gutenberg Museum, Postfach 3820, D-6500 Mainz 1, Allemagne

## Ligatures, typography, and computer science

**Abstract:** This report contains two papers on ligatures.

First, the GUTENBERG's *Bible* is examined from a typographical point of view (both at the page level, with ROSAVIRO grid constants, and at the line level, where justification was made by using ligatures instead of today variable spaces).

Second, ligatures are seen with the computer scientist point of view: character exchange standards (such as Ascii or IsoLatin1) do not know ligatures because they are glyphs. In the other hand, encoding tables allow to print ligatures. Some formatters, such as T<sub>E</sub>X, are able to recognize and handle such ligatures.

**Key-words:** Ligatures, character exchange standards, encoding tables, PostScript, T<sub>E</sub>X

---

## La typographie de la *Bible* de GUTENBERG<sup>1</sup>

Adolf WILD

*Conservateur du Musée Gutenberg à Mayence  
Stadt Mainz Gutenberg Museum  
Postfach 3820, D-6500 Mainz 1  
Allemagne*

Traduction française par André PELTRE (Cicero-Systèmes)

---

**Résumé.** L'auteur regarde la *Bible* de GUTENBERG avec l'œil d'un typographe. D'abord au niveau de la mise en page où les travaux récents de ROSAVIRO ont montré l'existence de constantes dans les grilles ; puis, et surtout, au niveau des lignes qui étaient justifiées non pas, comme aujourd'hui, en jouant sur des espaces inter-mots de taille variable, mais sur des combinaisons de lettres : logotypes, ligatures, abréviations, etc.

Jamais l'art de l'imprimerie n'a été célébré comme dans le colophon – la conclusion – du *Catholicon* de Johannes BALBUS, qui est paru en 1460 à Mayence et qui a été, peut-être, imprimé par GUTENBERG lui-même :

C'est avec l'aide du Très-Haut qui délia la langue des enfants, et qui révèle souvent aux petits ce qu'il cache aux hommes de science, que fut terminé le *Catholicon*, ce livre admirable, l'an de l'incarnation du Sauveur M.CCCCLX, dans la mère patrie Mayence, insigne ville de l'Allemagne, que Dieu dans

---

1. Cette note correspond à un exposé fait, en allemand, lors des Rencontres de Lure en août 1992. Cet exposé avait alors fait l'objet d'une traduction française directe par André PELTRE (Cicero-Systèmes) qui en a aussi assuré la traduction écrite depuis. Qu'il en soit sincèrement remercié ici.

sa clémence a daigné rendre la plus illustre et la première des villes : et ce livre fut parfait sans le secours ordinaire de la plume ou du calame, mais par l'admirable enchaînement de formes et de caractères, grâce aux rapports et l'harmonie admirable.

La formule *mira patronarum formarum concordia, proportione et modulo* a été longtemps l'objet d'interrogation.

Il y a eu, très certainement des règles de proportion dans la mise en pages, règles auxquelles se soumettaient déjà les copistes pour leurs manuscrits, mais ces règles n'étaient pas codifiées par écrit.

Ce secret de fabrication, l'Argentin Raul Mario ROSARIVO croit l'avoir découvert au XX<sup>e</sup> siècle. Dans son ouvrage principal *DIVINA PROPORTIO TYPOGRAPHICA* (Krefeld 1961) et quelques autres articles, il réduit tout l'ordonnancement typographique à la proportion 1:1,5 ou bien 2:3 et prouve que le nombre de lignes par page était toujours divisible par 3. Il a aussi souligné la signification particulière magique et religieuse du nombre 3, déjà chez les anciens Égyptiens, mais surtout depuis le christianisme.

Sa thèse est vérifiée dans les faits car, aussi bien pour la *Bible à 42 lignes* de 1455 que celle à 36 lignes de 1458 et celle à 48 lignes de 1462 de FUST et SCHÖFFER et que pour également le « calendrier turc » à 21 lignes et le *Catholicon* à 66 lignes, il s'agit de multiples du nombre premier 3. Il arrive au module 1,5 en divisant 3 fois l'unité de mesure cicero (=12 points) par deux :  $(12/2=6)$ ,  $(6/2=3)$ ,  $(3/2=1,5)$ . Pour revenir de 1,5 à 12 il faut multiplier par 8.

Appliqué au format de la page (figure 1) imprimée, il en résulte une proportion de la largeur et de la hauteur de 8/12, c'est-à-dire 2/3. Ce qui se présente d'abord comme un jeu de numérogie ésotérique peut en fait se vérifier pour tous les premiers imprimés majeurs.

Prenons l'exemple de la *Bible à 42 lignes* de GUTENBERG.

Si on divise la page dans la proportion de 2 à 3 et si on trace à partir des points de partage une ligne horizontale et une ligne verticale, il en résulte une intersection. Une diagonale du coin supérieur gauche de la page au coin inférieur droit passe par cette intersection.

Dans la structure de la page classique, qui s'applique pour ROSARIVO à la *Bible* de GUTENBERG, on divise cette diagonale en 9 segments égaux. On enlève 1/9 en haut et 2/9 en bas. Par les points 1 et 7 de la diagonale on fait passer des verticales

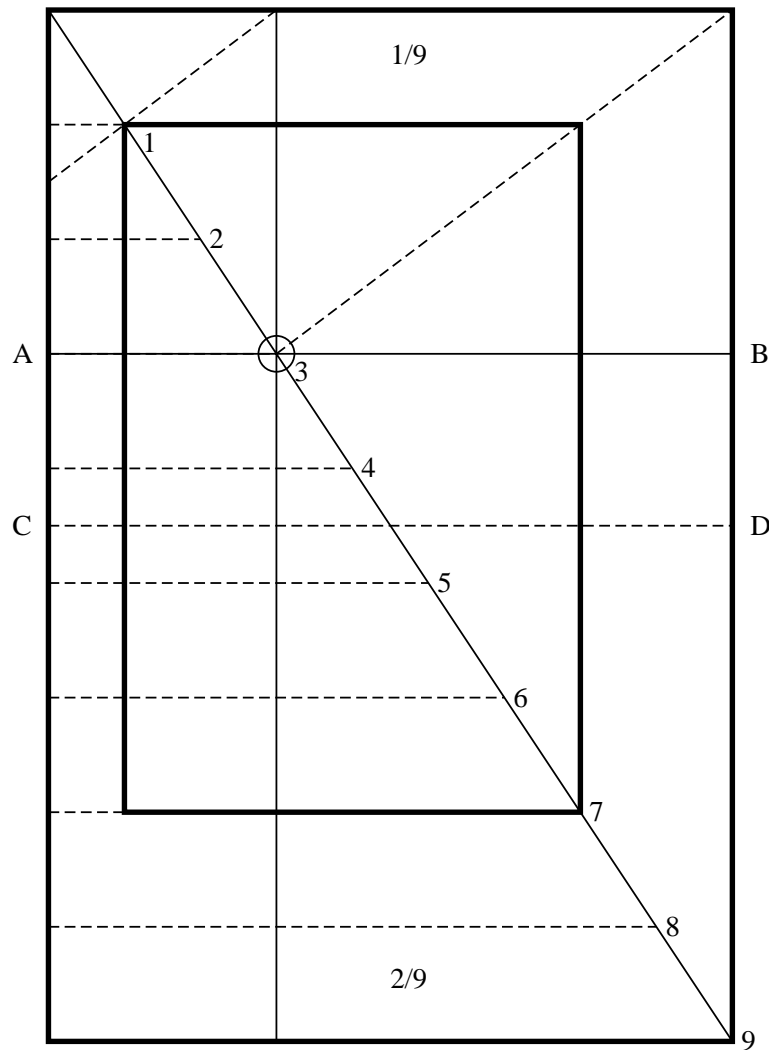


FIG. 1 - Répartition d'une page selon la proportion 1 : 1,5



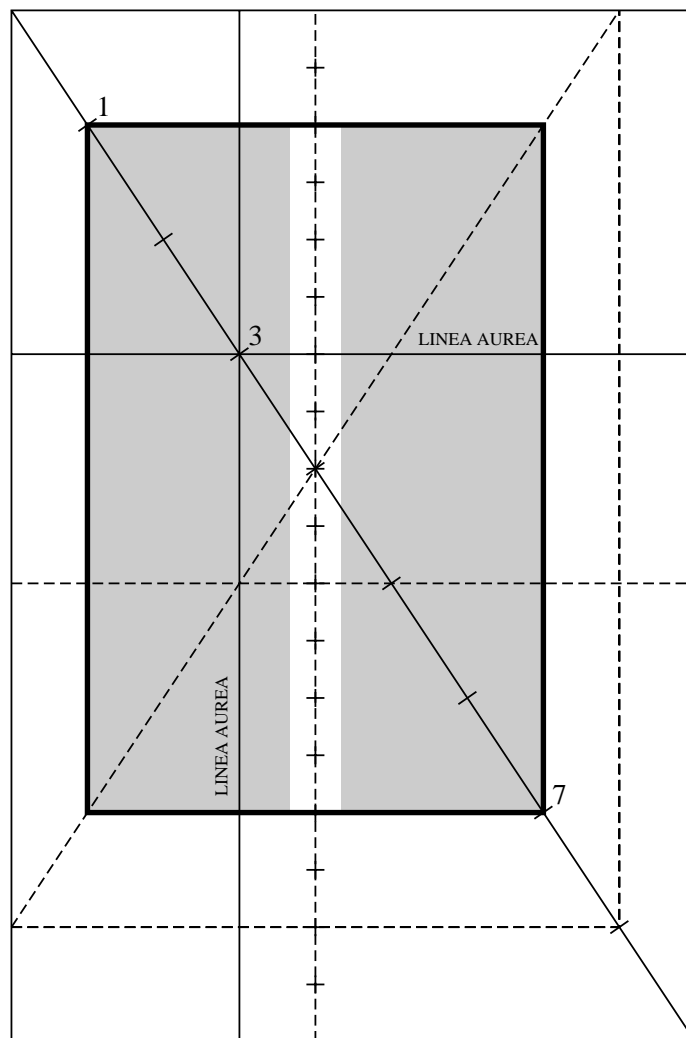


FIG. 2 - Répartition d'une page avec la partie imprimée de la Bible à 42 lignes de GUTENBERG

et des horizontales et on obtient un rectangle. Ceci est le rectangle de la partie imprimée qui est entouré des marges non imprimées.

La largeur des marges est donc, à partir de la pliure, dans une progression de 2/3/4/6.

L'intersection des lignes de partage 2:3 indiquées au début marque 3 neuvièmes de la diagonale. Les lignes ne partagent donc plus seulement la page, mais aussi la partie imprimée en hauteur et en largeur dans la proportion de 2 à 3. ROSARIVO appelle ces lignes les « lignes d'or » (*linea aurea*, figure 2).

ROSARIVO a aussi découvert autre chose à partir de ses mesures. « La hauteur de la partie imprimée est égale à la largeur de la page et aux deux tiers de sa hauteur. De même, la largeur de la partie imprimée correspond à deux tiers de la largeur de la page. La marge intérieure est égale à la moitié de la marge extérieure. La largeur de la marge intérieure multipliée par 9 donne la largeur de la page. Il en est de même si on multiplie la largeur de la marge extérieure par 4,5. Si on multiplie par contre la marge extérieure par 3 on obtient la largeur de la partie imprimée. La hauteur de la partie imprimée est obtenue en multipliant la marge inférieure par 3 et la marge supérieure par 6. La multiplication de la marge inférieure par 4,5 – ou de la marge supérieure par 9 – donnent la hauteur de page. »

Maintenant revenons aux lignes verticales et horizontales qui divisent la page dans la proportion 2/3. Traçons les parallèles à ces lignes qui séparent le deuxième du troisième tiers, on obtient alors sur la partie imprimée 9 rectangles égaux. Les rectangles aussi respectent la proportion 2/3. Ils contiennent 14 lignes imprimées. La marge entre les deux colonnes de texte est égale à un neuvième de la partie imprimée (figure 3).

La théorie de ROSARIVO part du principe qu'une mise en page harmonieuse doit obéir au module 1,5. Ce module est à utiliser systématiquement pour tous les éléments des rectangles, en particulier pour la proportion de la page par rapport à la partie imprimée d'où résulte la marge automatiquement. Seule une page structurée de cette manière peut avoir une typographie équilibrée. Elle refléterait une idée de base qui s'applique avec une précision scientifique à plusieurs étapes successives. La beauté s'exprime ici par la proportion, le rythme et l'harmonie. La perfection esthétique est le reflet de l'idéal.

Les réflexions et les calculs de ROSARIVO semblent se mêler intimement avec le calcul et la numérologie. Cela était parfaitement plausible dans le contexte des réalités du haut Moyen Âge. Si la règle des proportions contenue dans le colophon du

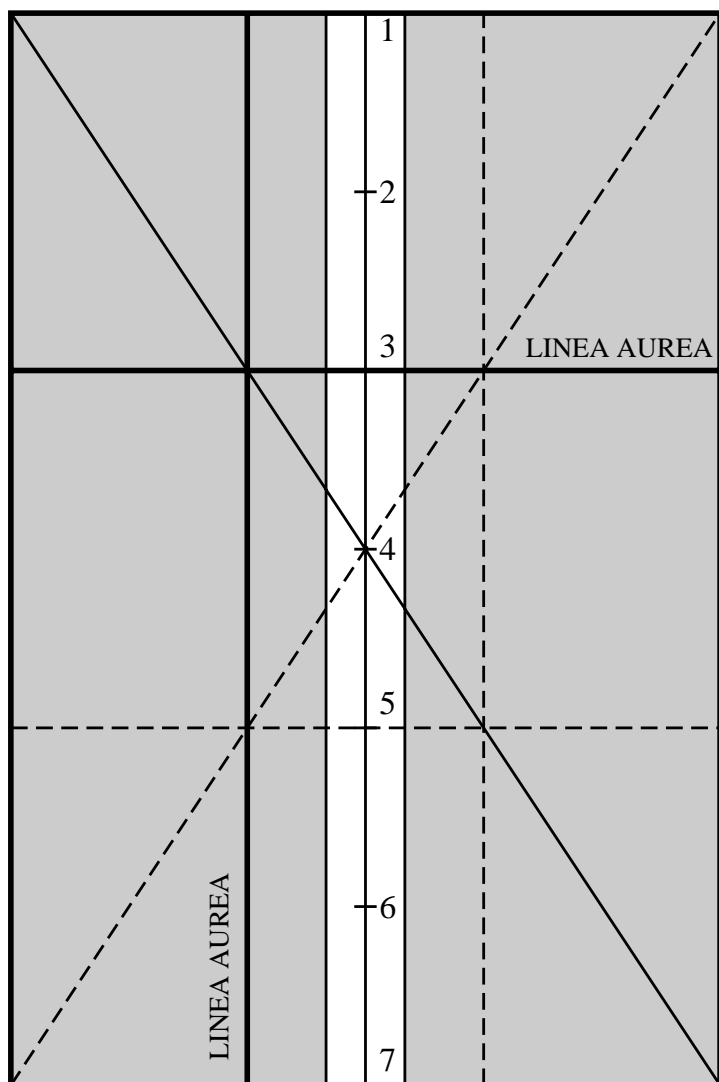


FIG. 3 - Répartition de la partie imprimée d'une page de la Bible à 42 lignes de GUTENBERG

*Catholicon* est ainsi fondée, la question reste ouverte. On ne peut faire des mesures objectives que sur les exemplaires conservés de la *Bible* de GUTENBERG. À cette occasion, il faut remarquer que ROSARIVO s'est mis à l'abri des critiques à l'avance, en faisant remarquer que l'impression a eu lieu sur des papiers humides dont les dimensions ont pu varier par la suite.

Les mesures exactes faites sur la *Bible* de GUTENBERG ne permettent ni de prouver, ni d'infirmer les formules de ROSARIVO. On a constaté des écarts importants qui peuvent toutefois s'expliquer par un manque de connaissance des règles de la part du compositeur. Les marges ont pu être massicotées par la suite. Même la plus forte objection n'est pas forcément pertinente. La *Bible à 42 lignes* a été composée – pour les premières pages – sur 40 lignes (I, 1-5r et I, 129-132r). La page I 5V<sup>o</sup> compte 41 lignes. Ces deux nombres ne sont pas divisibles par 3. Il est possible que l'on ait corrigé ces erreurs initiales, car à l'occasion de réimpressions pour augmenter le tirage, on composait ces pages également sur 42 lignes. On aurait alors composé le texte moins serré sans recourir à des abréviations. Il n'y a pas eu de problème de mise en page à deux exceptions près. Lors du passage de 40 à 42 lignes la surface de la page imprimée est restée inchangée. Les caractères, fondus avec des marges largement suffisantes, ont été limés d'environ 0,3 mm. Certaines lettres ont été fondues à nouveau. On peut aussi expliquer ce changement à partir d'arguments économiques car il se traduit par une économie de 30 pages. Au regard du prix du papier et de l'encre, cet argument est plausible.

À l'encontre des formules de ROSARIVO, on constate aussi les variations dans la largeur des colonnes et celle de l'intercolonnage. Les écarts dans la dimension des colonnes atteignent 85 et 88 mm. Pour les intercolonnes, on constate 2 groupes, l'un avec un écart de 22-23 mm, l'autre avec 20-21 mm.

On attribue ces écarts au comportement des divers compositeurs. La structure de la composition n'a donc pas toujours été respectée et GUTENBERG n'est pas intervenu avec autorité pour faire respecter la norme.

On comprend que la *Bible* de GUTENBERG se réfère au modèle des manuscrits du Moyen Âge. Mais on n'a pas trouvé un modèle précis. On trouve des analogies avec l'« autographe de THOMAS de KEMPTEN » et surtout « la Bible géante de Mayence », ouvrages écrits tous les deux en « écriture des missels » et conservés à la bibliothèque du Congrès à Washington. Dans le cadre de la tradition biblique, on a pu constater une influence des écritures manuelles des bibles parisiennes, les versions « Vulgates » de ces ouvrages servant de modèle pour les *scriptoria* des pays

rhénans. Il faut aussi se souvenir que Peter SCHÖFFER a travaillé comme copiste à Paris. Sa présence dans cette ville est prouvée, encore en 1449.

Remarquable aussi est le fait que GUTENBERG ait choisi pour sa *Bible* une écriture – la *Textura* ou « monocale » – qui à l'époque n'était plus utilisée que pour les écrits liturgiques, les missels surtout. C'est pourquoi on l'appelle aussi « écriture des missels ». Pour une bible, elle était étonnamment grande. Ce qui conduit à admettre que la *Bible* de GUTENBERG se situe dans la tradition de la lecture publique. Ceci se confirme aussi par le fait que tous les premiers possesseurs de la *Bible* étaient des institutions religieuses, un débouché sûr pour GUTENBERG. Elle était donc surtout utilisée pour la lecture à haute voix dans les réfectoires. Comme l'écriture choisie était très anguleuse, par opposition aux écritures courantes plus rondes, elle se prêtait bien à la constitution de lettres isolées. Bien que la *Bible* de GUTENBERG ressemblât beaucoup à un manuscrit, il y avait néanmoins une différence significative. Les fins de ligne étaient soigneusement alignées sur une ligne verticale, ce qui était impossible pour une écriture manuelle. Nous disons, en langage technique, que les lignes étaient « justifiées ». Dans la marge ne figuraient que les « divisions », le point et les « s » surélevés. Ce « s » n'était presque employé là que pour assurer un effet optique de netteté à la fin de la ligne. Les coupures de mots étaient, par ailleurs, strictement interdites en fin de colonne ou en fin de page. Dans ses efforts pour obtenir des lignes d'égale longueur, GUTENBERG renonçait à la méthode la plus simple – hier comme aujourd'hui – à savoir l'insertion de blancs variables entre les mots. À de rares exceptions près, le blanc utilisé est toujours égal à la chasse d'un « I ». Ceci procure un aspect très agréable et équilibré car les blancs si disgracieux dans la ligne sont évités. Mais cette restriction était en partie corrigée par les signes de ponctuation qui, à l'époque, n'étaient pas placés directement après le dernier mot mais à égale distance entre celui-ci et le prochain. Ils étaient fondus au milieu de ces deux lettres. Ceci procurait une certaine souplesse pour s'adapter à l'espace disponible.

Par ailleurs GUTENBERG s'était doté d'une série de variantes de caractères, toutes issues de la tradition de l'écriture manuelle, pour réaliser la justification. Il s'agissait :

1. de modifications dans le dessin de caractères isolés,
2. de « ligatures » c'est-à-dire de logotypes qui sont des groupes de lettres fondues ensemble,

3. d'abréviations, courantes mais disponibles sous différentes formes qui étaient utilisées selon la place disponible.

Les variantes dans les caractères résultent de la structure particulière de la *Textura*. Les lettres n'étaient plus arrondies, mais formées par l'assemblage de segments verticaux ou obliques, dans lesquels les traits verticaux dominaient. Le copiste devait positionner la plume pour le tracé de chaque segment. Il en résultait de petits tirets (ou crochets) au début et à la fin de ces segments, c'est-à-dire à gauche ou à droite, en tête ou en pied du caractère. Les petits tirets donnent à la composition une impression de « grille » caractéristique. GUTENBERG a encore accentué cette impression en faisant que tous les segments verticaux à l'intérieur d'un mot soient équidistants. Les petits tirets orientés à droite se rapprochent de la lettre suivante et esquissent une liaison. Des lettres comme le c, f, g, et t furent pourvues d'un appendice qui forme liaison avec le caractère suivant, à mi-hauteur.

Mais il y avait un problème dans ce système, là où les petits tirets venant de gauche et de droite étaient vis à vis. C'est pourquoi GUTENBERG a fondu pratiquement toutes les minuscules sous une seconde forme avec suppression du petit turet vers la gauche. Ces caractères s'appellent caractères de liaison. Ils sont positionnés à l'extrême gauche du cadratin, de telle manière à être en liaison avec le petit turet de droite de la lettre précédente. Il y avait des règles précises à leur emploi. De plus, pour tous les caractères de liaison commençant par un segment vertical du type « I » (c'est-à-dire i, m, n, p, r et u) il y avait une forme spéciale avec une tête coupée obliquement vers la droite. Pour certaines liaisons, l'effet obtenu était ainsi plus agréable à l'œil. Il y a aussi une forme spéciale et peu courante du X qui, en échange de ses tirets de gauche supprimés, est doté d'un turet supplémentaire vers la droite.

Les ligatures, dans la Bible comme dans les écritures manuelles, sont très nombreuses et ont constitué le deuxième moyen utilisé par GUTENBERG pour structurer la ligne. Elles correspondent à des habitudes anciennes des copistes. Deux lettres sont étroitement rapprochées et fondues ensemble en tant que logotypes. Il s'agit notamment de consonnes doublées, surtout de celles qui se superposent en haut ou en bas, de lettres dont les rondeurs sont redessinées ensemble ou de lettres s'imbriquant les unes dans les autres, du fait du décalage de leur tracé dans le sens vertical.

Pour économiser de la peine et de la place, les copistes utilisaient de très nombreuses abréviations, bien connues des lecteurs. On les retrouve chez GUTENBERG et elles enrichissent son répertoire de caractères qui se chiffre à 290 signes tous

FIG. 4 - *Les signes de la Bible à 42 lignes – d'après EMIL RUDER, Typographies, 1967*

ensemble (figure 4). Leur utilisation – ou leur non-utilisation – servait à l’optimisation de l’espace de la ligne. La plupart du temps, les abréviations sont marquées par des traits ou des crochets disposés au-dessus ou à côté des minuscules. Pour ces abréviations, GUTENBERG n’a pas créé de poinçons originaux, mais il a modifié les matrices des lettres grâce à des « poinçons d’accent ».

Les capitales n’offraient aucune possibilité pour contribuer à la justification. En fait elles n’étaient pas souvent utilisées, et seulement pour la première lettre d’un titre ou d’un verset. Seul le A capitale a deux formes. Toutes les autres capitales ont été légèrement limées et arrondies dans leur partie supérieure, lors du passage de 40 à 42 lignes.

Les capitales n’avaient pas un rôle déterminant parce que les lettres importantes étaient ajoutées à la main, et en couleur, sous forme de lettrines. L’espace prévu à cet effet était laissé libre lors de l’impression du texte en noir. C’est seulement au début, dans les quelques pages à 40 lignes, que l’on trouve des titres imprimés en rouge – les « rubriques ». L’impression se faisait lors d’un deuxième passage, après séchage des feuilles. Mais comme l’ajustement exact de cette deuxième impression et les mesures à prendre pour éviter des maculages étaient trop coûteuses en temps, on abandonna cette manière de faire. La difficulté résidait principalement dans le calage précis de la forme imprimante car le repositionnement de la feuille de papier était facile : il suffisait de replacer les petites aiguilles qui fixaient la feuille de papier dans les mêmes trous. Les impressions recto et verso coïncidaient exactement. Il s’agissait d’une composition « en registre ».

Pour les « rubriques » à insérer manuellement – titres, débuts et fins de texte et lettrines – GUTENBERG imprima une liste, la *Tabula rubricatoria*. Nous en avons deux exemplaires. L’acheteur des feuilles – non reliées – de la *Bible* les confiait à un rubriquaturiste et à un illumineur de son choix. Ceux-ci remplissaient les blancs, en suivant l’ordre de la liste. Les positions et la séquence des pages étaient marquées. Les cahiers étaient numérotés dans l’ordre sur les 5 premières feuilles. Avant le chiffre, une lettre spécifiait le cahier (par exemple : a1, a2, a3, etc.). Il y avait aussi des pages blanches et des parties vides à la fin des textes. On les laissait telles quelles, car elles avaient une fonction de séparation juste devant un texte important. On mettait en effet, volontiers, le début des textes importants en tête des cahiers. Le *Nouveau Testament* et l’*Apocalypse* sont précédés de pages blanches qui les distinguent. Pour arriver à ce résultat on acceptait même parfois d’avoir des cahiers



inégaux. Ce système de structuration visuelle est tout à fait dans la tradition du manuscrit.

Pour l'espace libre destiné aux lettrines, GUTENBERG proposait un modèle. On peut ainsi voir qu'il restait peu de place pour les dessins et enluminures. Les réserves n'autorisaient que des dessins relativement petits. Les images ne devaient donc avoir qu'un rôle très secondaire. Ceci est en opposition complète avec les bibles manuscrites de cérémonie qui comportaient des illustrations importantes. Ceci confirme encore une fois que la *Bible* de GUTENBERG était destinée à la lecture publique, ce qui limitait l'importance de l'illustration.

Malgré cette restriction, l'enlumineur devait encore surmonter une grande difficulté. Il fallait ajouter – à la main – 2509 lettrines sur 1 ligne, 1292 lettrines sur 2 lignes, 11 sur 3 lignes, 61 sur 4 lignes, 3 sur 5 lignes et 72 sur 6 lignes. Plus 3945 titres, 3 alphabets hébraïques ainsi que 1277 titres courants de page. Toutes les lettrines sont inscrites dans des rectangles. Les lettrines sur 5 ou 6 lignes étaient réservées pour le début des livres. Les lettrines plus petites pour le début des chapitres. Il y a aussi une certaine hiérarchie de la décoration qui facilite la recherche.

Faisant exception à ce schéma, le I capital était traité à part. Pour lui on ne réservait pas d'espace. Il était dessiné, à la main dans la marge, le long du bord du texte exactement comme dans la tradition du manuscrit. Il en résultait des formes graphiques très allongées pleines de fantaisie. Mais d'autres initiales pouvaient aussi servir de point d'accroche de motifs décoratifs floraux qui s'échappaient du contour de la lettre. . . avec ajout de figures animales le cas échéant. L'enlumineur n'avait de liberté totale que dans les marges et les colombelles. Les artistes ont, dans ce cadre, donné libre cours à leur créativité luxuriante. . . La hiérarchie des textes était alors pratiquement négligée. . . D'une manière générale, ce sont les premières pages du premier et du deuxième volume qui sont les plus décorées. Les frontispices étaient d'ailleurs aussi les pages les plus remarquablement décorées dans les manuscrits.

En ce qui concerne les lettrines, un chercheur a encore fait une constatation très intéressante, en examinant le *Psautier de Mayence* qui a été imprimé entre 1457 et 1459 dans l'atelier de FUST et SCHÖFFER. Les lettrines imprimées en couleur pour ce travail auraient pu aussi être insérées dans la *Bible* de GUTENBERG, d'après les mesures de ce chercheur. Ceci ouvre la voie à des spéculations multiples. . . que nous n'aborderons pas !

En résumé, on peut dire ceci : la *Bible* de GUTENBERG se situe dans la droite ligne de la tradition du manuscrit médiéval. Avec la régularité et l'unité confé-

rées par la typographie, celle-ci est portée à la perfection. Sur le plan esthétique, c'est l'aboutissement d'une longue évolution, sur le plan technique c'est un grand commencement.

Pour l'art de l'imprimerie, elle reste un modèle de soin, d'équilibre, de savoir-faire artisanal et d'exigence esthétique, modèle pour lequel seul le meilleur était acceptable !



---

## Codage, caractères spéciaux & ligatures<sup>2</sup>

Jacques ANDRÉ

Irisa/Inria-Rennes  
Campus universitaire de Beaulieu  
F-35042 Rennes cedex, France  
email: jandre@irisa.fr

---

**Résumé.** Pour pouvoir être traités par ordinateur, les caractères doivent être codés. Nous en donnons la raison et montrons les principales normes actuelles pour l'échange de caractères (Ascii, IsoLatin1, Unicode). Ces normes sont différentes des tables de codage des glyphes utilisées par les fontes numérisées.

Les claviers suivent en général les codages d'échange ce qui explique pourquoi on y trouve souvent des caractères « bizarres » comme « & @ £ \$ # ¤ » mais plus rarement « é ». Quant aux petites capitales (comme « É ») et ligatures (comme « fi » ou « ſt ») il faut en général utiliser au moins trois touches. Seuls certains logiciels, comme T<sub>E</sub>X, se servent des informations métriques des fontes pour produire automatiquement les glyphes correspondantes.

Il faut être de mauvaise foi pour prétendre encore aujourd'hui que le Macintosh (nous entendons par là la PAO, voire les fontes numérisées) ne produit que des caractères sans accent et de mauvaise qualité<sup>3</sup>. Mais il faut avouer en revanche que les « casses » informatisées n'ont pas la richesse qu'ont pu avoir certaines casses du temps du plomb. Ainsi, trouve-t-on, par exemple, peu de ligatures, et pas toujours celles que l'on aimerait. Par contre, les claviers sur lesquels on saisit aujourd'hui les textes (et où il n'est pas toujours facile de taper un É en petites capitales) ont presque toujours des touches offrant des caractères bizarres tels que « & @ £ \$ # ¤ ».

---

2. Cette note correspond à des exposés faits dans divers contextes, notamment lors de l'École Didot organisée à l'École Estienne à Paris en mai 1992.

3. Il est exact que les premières imprimantes à aiguilles n'avaient pas une très bonne définition, que certaines polices penchaient les caractères pour simuler l'italique, que toutes les lettres accentuées françaises – et plus généralement européennes – n'étaient pas disponibles, etc. Mais tout ceci relève d'un passé désuet ; hélas, les mauvaises réputations durent. . .

Nous les qualifions de bizarres car ils ne faisaient pas partie, avant l'informatique, de l'environnement du français moyen : on ne nous les a pas appris à l'école, on ne les trouvait pas sur les machines à écrire et on ne les voyait que rarement imprimés. Ils sont tellement bizarres qu'en général on ne sait même pas comment les appeler et bien souvent on est incapable de les dessiner du premier coup. Enfin, ils ne figurent pas dans les dictionnaires français importants (comme le Larousse, ou le Robert du moins dans l'édition de 1966) ni d'ailleurs dans les codes typographiques français d'aujourd'hui.

Nous allons donc essayer d'expliquer ici pourquoi l'informatique offre ces caractères et pourquoi les ligatures sont, aujourd'hui, limitées en nombre. Pour cela, il convient de commencer par quelques retours historiques.

## 1 Quand deux voies convergent. . .

Depuis plus de cinq cents ans, on distingue deux formes d'écriture : l'écriture manuscrite puis dactylographique et l'écriture imprimée. Chacune de ces formes a suivi son propre cheminement, mais aujourd'hui elles convergent. En dehors de l'aspect « qualité », voyons ce que contenaient leurs divers alphabets.

1. Les caractères imprimés ont suivi la voie « royale », et jusqu'à la fin du XIX<sup>e</sup>, les catalogues de fondeurs ne proposaient que des caractères de très grande qualité. Mais la « presse » (le secteur qui s'occupe de l'impression des journaux) et ses besoins de vitesse et de minimisation des coûts a entraîné le « labeur » (le secteur d'impression des livres) dans une continuelle dégradation en ce qui concerne notamment la richesse des alphabets. C'est en effet en 1885, avec la machine à graver les poinçons de Lynn Boid BENTON (basée sur l'emploi d'un pantographe), qu'apparaissent les premiers caractères homothétiques, un même dessin servant aussi bien pour un corps 6 qu'un corps 32. C'est l'époque où apparaissent d'abord la Monotype puis la Linotype. Ces machines utilisent alors des caractères homothétiques et remplacent casses et casseaux par des porte-matrices (figure 5-droite). Mais ces porte-matrices contiennent moins de caractères que les casses traditionnelles et ne comportent pas de façon naturelle d'emplacements pour ces caractères « exotiques », comme disent les Américains, que sont les « Ê », « st » ou autres « Å » et « ß ». Adieu donc la casse de GUTENBERG et ses 300 caractères environ

FIG. 5 - À gauche : casse française ; à droite, châssis porte matrice de la Monotype (d'après [Marshall93])

[Wild94][Karow94][Ponot93], adieu la casse parisienne au profit de quelques dizaines de caractères à peine.

L'invention de l'*offset* (New York, 1904) conduit naturellement à rechercher le moyen de composer des prêts-à-clicher et débouche sur l'invention de la photocomposition dans les années 1950 par les français HIGONNET ET MOYROUD [Marshall93]. Mais le développement de cette technique aux États-Unis et non en France fait que la limitation du nombre de caractères reste présente même dans les photocomposeuses de seconde génération<sup>4</sup>. Il faudra

---

4. Ainsi, par exemple, une police de la photocomposeuse *Graphic System* en 1975 ne contenait que 128 caractères, y compris toutes les espaces fines, demi-cadratins, etc., sans oublier les caractères dits de contrôle ; mais on n'y trouve que rarement les capitales accentuées, les petites capitales et

attendre les photocomposeuses de troisième génération avec des caractères numérisés (la première, *Digiset* de Hell, ne date que de 1965), pour trouver les fontes d'aujourd'hui avec des jeux de caractères plus riches.

2. Parallèlement à l'évolution des techniques d'impression s'est développée une seconde histoire, celles des chancelleries puis des secrétariats et de ce que l'on appelle encore aujourd'hui la bureautique<sup>5</sup> [Marshall94]. Longtemps restée en usage, la plume des scribes des chancelleries, s'est vue remplacée par la plume métallique – dont celle dite « sergent-major » – puis par la machine à écrire (la première Remington date de 1873 [Laufer80, Autrement94]). Là encore, des problèmes techniques (difficulté de mettre ensemble de très nombreuses barres) font qu'il faut limiter les caractères et comme cette invention est américaine, ce sont bien sûr les capitales accentuées, les ligatures et les abréviations<sup>6</sup> qui disparaissent les premières tandis que l'on remplace l'italique par le souligné. Quant à la technologie elle-même (impression par transfert d'encre après frappe au travers d'un ruban de tissu, utilisation de caractères à chasse fixe, etc.), elle n'a fait que creuser le fossé avec la typographie que la dactylographie ne prétendait toutefois pas imiter.

Mais les machines à écrire se sont améliorées sans cesse, l'aboutissement capital étant l'invention des machines à boule, de la *Varietyper* et de la *Selectric Composer* d'IBM qui a été la première tentative de concilier la qualité des chasses variables de la typographie avec la facilité d'emploi (pour un non-typographe) et le coût très bas des machines à écrire [Frutiger].

3. Enfin, depuis la seconde guerre mondiale, est arrivée l'informatique. Ses techniciens ont eu très vite à résoudre le problème de la communication homme-machine – saisir les données et imprimer les résultats – ce qui a conduit à toute une évolution du parc de matériels « d'entrées-sorties ». Pendant

---

les ligatures comme « ffi » ou « ffl ». Toutefois, les polices pouvant être composées à la demande, il était quand même possible de se faire faire des « casseaux » complémentaires. Signalons au passage le rôle important qu'a joué, pour la production de caractères gravés, la fonderie Micheau, à Beaune.

5. Les chancelleries, états-majors et secrétariats ont toujours eu de gros besoins de communication non seulement interne, mais aussi entre sites éloignés. D'où les « transmissions », qui ont beaucoup évolué de Chappe à Mors et dont un des apports fondamentaux a été le Telex.

6. Celles-ci existent encore couramment : quel étudiant n'en a pas utilisé pour prendre ses cours, par exemple « m̄ » pour « même », « tt » pour « tout » ou « tq » pour « tel que » ?

longtemps, les principales applications de l'informatique ont été comptables. On a donc tout naturellement développé en premier des matériels d'impression de factures, de feuilles de paye, etc. et les caractères ad hoc comme nous le verrons.

Les systèmes de traitement de textes ont d'abord été basés sur ces matériels lourds qu'étaient les imprimantes à barre mais l'emploi de terminaux similaires aux machines à écrire, fussent-elles à boules, marguerites ou aiguilles, n'a pas vraiment fait évoluer la qualité des fontes<sup>7</sup>. Ce n'est qu'avec l'apparition d'imprimantes de relativement bonne définition (comme celles utilisées par les premiers systèmes  $\text{T}_{\text{E}}\text{X}$ ) puis bien sûr, vers 1982, l'apparition des imprimantes à laser que se sont développées les fontes d'aujourd'hui et, avec l'aide de la micro-informatique, la PAO.

C'est donc une sorte de convergence vers la qualité typographique qui a marqué cette évolution, même si de nombreux exemples semblent montrer le contraire. En fait, ce ne sont pas les outils qui sont en cause, mais l'emploi qui en est fait aujourd'hui par des personnes qui n'ont pas eu la formation voulue.

## 2 Normes et tables de codage

### 2.1 Notion de codage

Autrefois, les caractères étaient mis dans des casses (comme celle de la figure 5-gauche) qui étaient « normalisées » : la casse parisienne avait toujours les mêmes lettres au même endroit et on peut dire que à chaque caractère était implicitement associé un certain code  $(x, y)$ ,  $x$  et  $y$  étant les coordonnées de la case correspondante dans la casse. Vers 1900, les porte-matrices de la Linotype (figure 5-droite), puis les films des photocomposeuses firent de même : chaque caractère (matrice ou lettre gravée sur le film) était assigné à une place précise. Mais au lieu de parler de  $x, y$ , on a simplement numéroté ces cases de 0 à  $n$ .

On fait de même actuellement avec les fontes numérisées : chaque caractère est numéroté de 0 à  $n$ . Ce que les ordinateurs manipulent, ce sont donc les numéros des caractères, pas les images des caractères qui, elles, seront produites par les

---

7. On trouvera dans [André89] une petite histoire de ces systèmes de traitement de textes.



imprimantes, écrans ou photocomposeuses. On retrouve donc la même différence qu'entre « caractère » (ou type) et sa marque imprimée (ou œil) du temps du plomb. On reviendra en 2.1.1 sur la notion sous-jacente de *glyphe*.

Pour coder les caractères, il suffit donc par exemple de remplacer « espace » par 0, « A » par 1, « B » par 2, . . . , « K » par 11, « L » par 12, . . . , « Z » par 26. Mais tel que, ce n'est pas évident de savoir si « 112 » correspond à « 1,12 » (c'est-à-dire « AL ») ou à « 11,2 » (c'est-à-dire « KB »). La solution est alors de dire : on code les caractères sur deux chiffres : « 01 » pour « A », « 02 » pour B, etc. Il n'y a plus d'ambiguïté mais on ne peut traiter que 100 caractères (numérotés de 0 à 99) ce qui est fort peu<sup>8</sup>. Si on en veut plus, il faut passer à un codage sur trois chiffres (on dit aussi à trois « moments »), c'est-à-dire qu'on a la possibilité d'avoir  $n = 1000$  caractères. C'est confortable, mais du coup les tables de codage (par exemple celles où on associe le dessin d'un caractère à un numéro de code), doivent avoir 1000 entrées (même si on a moins de 1000 caractères, il faut que l'on puisse dire « caractère inexistant » !). Il y a donc un choix à faire entre place et nombre de caractères possibles. C'est ce qui s'est passé avec les codes informatiques, sauf que ces codes sont en nombres binaires :  $n + 1$  n'est plus une puissance de dix ( $100 = 10 \times 10 = 10^2$ ,  $1000 = 10 \times 10 \times 10 = 10^3$ , etc.) mais une puissance de deux (par exemple  $32 = 2 \times 2 \times 2 \times 2 \times 2 = 2^5$ ).

Par ailleurs, chaque ordinateur a son propre système de codage interne. De même chaque imprimante ou photocomposeuse a ses propres codes pour les positions des caractères dans les fontes. Devant cette anarchie, il a fallu trouver des « normes » de codage. Mais les besoins étant différents selon les corps de métier (un typographe sera intéressé par la présence d'une ligature comme « *Œ* », alors qu'un employé de la poste envoyant un télégramme se contentera des deux caractères « c » et « t », voire « C » et « T »), divers types de normes co-existent, notamment :

- les normes d'échange de caractères,
- les normes de codage des caractères dans les fontes,
- les normes pour les claviers.

Le problème de la PAO est de les utiliser concurremment.

---

8. Le français nécessite par exemple 42 lettre accentuées ou non et ce en bas de casse, en petites capitales et en capitales, soit  $3 \times 42 = 126$  caractères. Sans oublier les chiffres, les parenthèses, la ponctuation, etc.

TAB. 1 - À un caractère peut correspondre plusieurs glyphes, et réciproquement.

Caractère(s)	Glyphe(s) possibles
Lettre majuscule A	A, A, A, A
Lettre majuscule ALPHA	A
Minuscule c suivie de minuscule t	ct, ct, c̄t
Esperluette	& c̄ &

### 2.1.1 Noms, œils et glyphes

La majorité de ces deux premiers types de normes définissent deux choses :

1. un numéro de code associé à un caractère, par exemple (en octal) 073 pour « ; », 101 pour « A », 102 pour « B » et 340 pour « à » etc.
2. un nom pour chaque caractère, par exemple *semicolon* pour « ; », *A* pour « A », *B* pour « B » et *grave* pour « à » – avec bien sûr des traductions officielles pour chaque langue (en France, c’est l’Afnor qui fait ce travail).

Par ailleurs, les normes font, souvent de façon implicite, la distinction très nette entre la notion de caractère et celle de leur représentation (ou trace imprimée ou affichée, c’est-à-dire ce que l’on appelle œil en français). Assez curieusement, cette notion simple prête souvent à confusion (sans doute parce que le mot « caractère » lui-même est ambigu ?). Avec les conventions de [Unicode91, page 12], on peut dire que les caractères n’ont d’existence qu’en machine, tandis que les glyphes correspondantes peuvent être visualisées sur papier ou sur écran. Mais l’important est de comprendre qu’il n’y a pas de relation biunivoque et surjective entre un caractère et une glyphe : à un caractère peut correspondre plusieurs glyphes et réciproquement à une glyphe unique peuvent correspondre plusieurs caractères comme le montre la table 1. On peut dire en gros que les normes d’échange manipulent des caractères et les normes de codage des fontes manipulent des glyphes.

On remarquera qu’avec cette définition, les ligatures *fi*, *ffi*, *c̄t*, etc. ne sont pas des caractères, mais simplement des glyphes possibles pour ces suites de caractères. En revanche, *œ* et *ij* sont des caractères (avec d’ailleurs des majuscules *Œ* et *IJ* correspondantes : on écrit « Œuvres de BALZAC » ; à la rigueur, faute de « Œ », on acceptera « OEuvres de BALZAC » mais sûrement pas « Oeuvres de BALZAC » !).

## 2.2 Normes d'échange de caractères

Les normes permettant l'échange de données alphabétiques ne sont pas récentes et expliquent le rôle prépondérant joué par les organismes de télécommunication des divers pays dans les instances de normalisation. En effet, les premières « normes » ont été les idéogrammes des codes du télégraphe de Chappe ou les fanions de la marine ; mais ce sont les alphabets Morse et ceux du Télex qui auront été les premiers « alphabets internationaux »<sup>9</sup>. Ces alphabets, à 6 moments (donc avec  $2^6 = 64$  caractères) ne comprenaient que les lettres majuscules, les chiffres, quelques signes de ponctuation et certains codes réservés (le Télex utilise par exemple un code pour « sonnerie »).

C'est donc dans l'esprit d'échange que plusieurs normes ont été définies, à 7 moments ( $2^7 = 128$  caractères), comme le codage EBCDIC (*Extended Binary Coded Decimal Interchange Code*) d'IBM, et l'ASCII (*American Standard Coded Information Interchange*). Ce dernier a été légèrement modifié et adopté comme « alphabet international N° 13 » IA5 en 1963 par l'ISO (*International Standard Organization*) et le CCITT (Comité Consultatif International du Télégraphe et Téléphone) sous le nom de norme ISO 646. Depuis, de nombreuses normes ont été redéfinies pour coder les caractères avec plus de bits ou moments ou pour s'adapter à diverses langues, dont les célèbres ISO LATINn (avec n de 1 à 7) sur 8 bits et UNICODE (ou ISO-10646) à 16 bits. Ces normes ont assuré toutes la compatibilité avec ISO 646, d'où la très grande importance de cette norme même si, aujourd'hui, elle a vieilli ! Ces normes ont les tailles du tableau ci-après.

Nom du code	nombre de bits ou moments	nombre de caractères
Telex	6	64
Ascii	7	128
Iso Latin	8	256
Unicode	16	65536

Voyons maintenant quel est le contenu de ces principales normes<sup>10</sup>.

9. Respectivement connus sous le nom de IA1 et IA2. Voir à ce sujet et de façon plus générale pour toutes ces normes [Marti90].

10. On trouvera dans [Marti90] la description de nombreuses autres normes, moins importantes ici.

### 2.2.1 ASCII ou ISO 646

Le principe de la norme ISO 646, reprenant la norme ASCII, est une structure à 7 bits (7 bits) permettant donc le codage de 128 caractères. Cette table comprend en fait 2 parties (figure 2) :

34 caractères<sup>11</sup> dits (à tort) « de contrôle » (tels que *Carriage Return*, *Line Feed*, *Bell*, etc.) et

94 caractères dits « graphiques » car on peut les afficher sur un écran ou les imprimer.

Ces 94 derniers caractères sont eux-mêmes répartis en 3 groupes :

- 82 caractères obligatoires :
  - 52 lettres : A–Z et a–z,
  - 10 chiffres : 0–9,
  - 20 signes de ponctuation ou autres : ! " % & ' ( ) \* + , - . / : ; < = > ? \_
- deux caractères « au choix » :
  - # ou £
  - \$ ou ₤ (symbole monétaire international *currency*)
- 10 positions réservées à des caractères d'usage national.

La norme ISO 646 comprend donc :

- des variantes nationales (parfois plusieurs pour un même pays – c'est le cas de la France)
- une version internationale de référence, IRV, où les positions optionnelles sont affectées d'un caractère précis.

Le tableau 3 montre la version de référence IRV et quelques exemples de ces versions nationales. Ces normes n'ont en fait pas été très suivies car mal adaptées ou flouées.

- La version internationale de référence n'était pas la version américaine ASCII : IRV contenait le symbole ₤ tandis que l'ASCII utilisait le dollar. On ne trouve donc le symbole ₤ sur pratiquement aucun matériel informatique<sup>12</sup>.

---

11. Les 32 premiers, le suivant qui est en fait l'espace et le dernier (de code binaire 1111111) pour DEL.

12. Sauf sur certains matériels à vocation internationale, notamment lorsqu'ils assurent la compatibilité avec ISO-Latin1. Dans le cas des claviers français des stations SUN, on a non seulement cette

octal	0	1	2	3	4	5	6	7
/0 x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL
/1 x	BS	HT	LF	VT	FF	CR	SO	SI
/2 x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB
/3 x	CAN	EM	SUB	ESC	GS	FS	RS	US
/4 x		!	"	#	\$	%	&	'
/5 x	(	)	*	+	,	-	.	/
/6 x	0	1	2	3	4	5	6	7
/7 x	8	9	:	;	<	=	>	?
/10x	@	A	B	C	D	E	F	G
/11x	H	I	J	K	L	M	N	O
/12x	P	Q	R	S	T	U	V	W
/13x	X	Y	Z	[	\	]	^	_
/14x	'	a	b	c	d	e	f	g
/15x	h	i	j	k	l	m	n	o
/16x	p	q	r	s	t	u	v	w
/17x	x	y	z	{		}	~	DEL

TAB. 2 - *Le codage ASCII.*

TAB. 3 - Les caractères optionnels de la version de référence de la norme ISO646 et de quelques variantes nationales – d’après [Marti90, page 243])

Version de référence	#	¤	@	[	\	]	^	‘	{		}	~
Allemagne (DIN66003)	#	\$	§	Ä	Ö	Ü	^	‘	ä	ö	ü	ß
Belgique	#	\$	à	°	ç	§	^	‘	é	ij	è	~
Espagne	#	\$	·	ı	Ñ	Ç	ı	‘	’	ñ	ç	"
France (NF Z62010/1982)	£	\$	à	°	ç	§	^	μ	é	ù	è	"
Grande Bretagne	£	\$	@	[	\	]	^	‘	{		}	~
Suisse romande			à		ç				é	ù	è	~
USA (norme ASCII)	#	\$	@	[	\	]	^	‘	{		}	~

- Les informaticiens américains se sont mis à utiliser nombre de leurs caractères optionnels (#, @, les accolades, etc.) ce qui a donné un poids anormalement fort à la version américaine Ascii de la norme 646.
- Il y avait une grande incohérence d’un pays francophone à l’autre à tel point d’ailleurs que la France a abandonné cette norme Z62010 au profit de l’Ascii en 1983.

C’est pourquoi, en 1988, ISO 646 a pris exactement le codage Ascii de la figure 2.

Par ailleurs, les caractères de commande correspondent à une technologie périmée : certaines commandes sont complètement dénuées de sens aujourd’hui<sup>13</sup>, d’autres ont été très mal définies<sup>14</sup>. Les fabricants de matériel ou de logiciel ont donc pris l’habitude d’y mettre des codes à eux<sup>15</sup>, mais avec la plus grande anarchie. Elle a donc été remplacée par de nouvelles normes d’échange.

touche *currency* mais aussi le  $\mu$  de la norme française (voir tableau 3) qui est restée dans IsoLatin1 (tableau 4).

13. Notamment à cause de changements de technologie, comme l’abandon des rubans papier pour lesquels DEL était prévu pour annuler un caractère erroné en perforant toutes ses positions dans la colonne.

14. Typiquement, ce codage permettait, pour certains écrans, d’afficher un caractère accentué en saisissant successivement le code ESC, le caractère, l’accent et enfin le code BS de retour en arrière. Mais ceci ne permettait pas de faire le moindre tri.

15. Ainsi,  $\TeX$  y place beaucoup de caractères comme fi, ffi, etc. comme on le voit table 5.

Néanmoins, compatibilité oblige, cette norme sert de base à toutes les autres normes et en particulier à IsoLatin1. C'est donc cette compatibilité à l'Ascii (et accessoirement à IsoLatin1) qui fait que l'on a sur nos claviers ces 94 caractères et tous ces caractères spéciaux « & @ £ \$ # { } » sur lesquels nous reviendrons en section 3

### 2.2.2 IsoLatin1

IsoLatin est un codage à 8 moments (8 bits) permettant de coder 256 caractères. En fait c'est un ensemble de normes regroupant les caractères de toutes les langues latines (européennes). Comme en Europe, on utilise plus de 256 caractères différents utilisés en Europe, il a été décidé de les regrouper par affinités . . . commerciales. C'est ainsi qu'il y a le premier groupe IsoLatin1 pour la zone occidentale (correspondant aux langues suivantes : allemand, anglais, danois, espagnol, féroïen, finnois, français, hollandais, irlandais, islandais, italien, norvégien, portugais et suédois), le second pour la zone orientale (avec l'albanais, le macédonien, le polonais, le roumain, etc.) etc. Pour des raisons politico-économiques, un codage spécial a du être ajouté pour la Turquie et ses partenaires !

La table 4 donne le contenu de la norme IsoLatin1. On y remarquera plusieurs choses :

1. Les 128 premiers caractères sont ceux de la norme Ascii. C'est vrai aussi pour IsoLatin $n$  quelque soit  $n$ .
2. On n'y trouve pas les deux caractères œ et Œ. On raconte que lorsqu'ISO a adopté cette norme, le représentant français était malade et qu'un autre pays en a profité pour imposer un caractère à lui (thorn?) prétextant que Œ pouvait s'obtenir par crénage de O et de E (ce qui est faux : voir section 16). Il est probable que ceci vient plutôt d'un certain désintérêt des Français en matière de normalisation au contraire des Allemands qui eux ont bien tous leurs *umlaut* et le ß. Mais œ et Œ figurent bien dans Unicode (section 2.2.3).
3. Les 32 premiers caractères sont, comme pour l'Ascii, des caractères « de contrôle », en général peu utilisés. D'où la tendance de nombreux organismes à récupérer ces positions pour y mettre, au moins dans les normes de codage des fontes, des caractères manquants. C'est ainsi que le codage de T<sub>E</sub>X dit de Cork y place les ij hollandais et nos œ français (voir table 5).

### 2.2.3 Unicode – une norme mondiale à 16 bits

De nombreuses langues rentrent difficilement dans la grille de ces normes occidentales. Aussi de nombreuses normes spécifiques ont-elles été définies<sup>16</sup>. Mais, depuis longtemps, les représentants de ces langues dans les instances de normalisation, rêvaient d'un codage unique pour toutes les langues du monde. Les travaux d'inventaire ont commencé vers 1988 et cette norme, ISO/IEC 10646, a été adoptée sous forme de DIS (*Draft International Standard*) en 1991. En parallèle, un consortium de fabricants de matériels et logiciels (et pas des moindres : Apple, Adobe, ATT, ITC, Microsoft, Xerox, etc.) définissait une norme similaire : Unicode. Finalement, en août 1991, les deux groupes de normalisation décidèrent de fusionner ces deux normes en une seule, à quelques détails près [Unicode91].

On trouvera divers exemples des dizaines de milliers de caractères ainsi définis dans [André94, Bettels93, Bigelow94]. Mais ce codage présente un inconvénient : tous les logiciels de manipulation de documents (qu'il s'agisse de T<sub>E</sub>X, Word, FrameMaker ou Grif, sans oublier HTML) tournent avec des mots de 8-bits. Or Unicode est codé sur 16 bits. Il faut donc réécrire la majorité de ces logiciels (certains sont déjà annoncés). Par ailleurs, si redessiner un grand nombre de caractères pour un tel codage ne pose pas trop de problèmes, en revanche faire en sorte qu'ils soient homogènes tout en respectant une certaine culture est encore une gageure aux yeux des bons typographes [Bigelow94].

## 2.3 Tables de codage des fontes

Comme nous l'avons dit, chaque imprimante ou chaque photocomposeuse avait sa propre façon de coder les caractères (et même plus généralement de piloter les sorties, ce qui explique que des auteurs ou des constructeurs aient défini des langages de description de page, comme DVI *DeVice Independant* de T<sub>E</sub>X puis PostScript d'Adobe). On a donc tout naturellement tenté de normaliser ces codes. On ne pouvait pas utiliser directement les normes d'échange qui n'ont pas la notion de glyphe et ne satisfont donc pas les besoins de la typographie : ni petites capitales, ni ligatures, ni les divers espaces ou tirets nécessaires, etc.

---

16. On trouvera dans [Marti90] des normes pour le grec, l'hébreu, l'arabe, mais aussi pour les 3000 caractères katakanas, voire des normes pour des caractères non alphabétiques comme le kanji.



TAB. 4 - *Le codage des caractères d'Europe occidentale : IsoLatin 1 (les cases blanches correspondent aux caractères de contrôle non imprimables).*

	!	"	#	\$	%	&	'
(	)	*	+	,	-	.	/
0	1	2	3	4	5	6	7
8	9	:	;	<	=	>	?
@	A	B	C	D	E	F	G
H	I	J	K	L	M	N	O
P	Q	R	S	T	U	V	W
X	Y	Z	[	\	]	^	_
'	a	b	c	d	e	f	g
h	i	j	k	l	m	n	o
p	q	r	s	t	u	v	w
x	y	z	{		}	~	
	`	´	^	~	-	˘	˙
¨		°	˘		˘	˘	˘
	ı	¢	£	¤	¥	¦	§
¨	©	ª	«	¬	-	®	-
°	±	²	³	´	µ	¶	·
¸	¹	º	»	¼	½	¾	¿
À	Á	Â	Ã	Ä	Å	Æ	Ç
È	É	Ê	Ë	Ì	Í	Î	Ï
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×
Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
à	á	â	ã	ä	å	æ	ç
è	é	ê	ë	ì	í	î	ï
ð	ñ	ò	ó	ô	õ	ö	÷
ø	ù	ú	û	ü	ý	þ	ÿ

TAB. 5 - Exemple de variante d'IsoLatin1 : le « standard de Cork » (utilisé par T<sub>E</sub>X pour les fontes) qui récupère les premières positions du codage Ascii – d'après [Ferguson90].

	00	10	20	30	40	50	60	70	80	90	A0	B0	C0	D0	E0	F0
0		ı		0	@	P	'	p	ž				À	—	à	
1			!	1	A	Q	a	q	š			–	Á	Ñ	á	ñ
2		`	"	2	B	R	b	r	Ž		¢	†	Â	Ò	â	ò
3		´	#	3	C	S	c	s	Ÿ		£	‡	Ã	Ó	ã	ó
4		˘	\$	4	D	T	d	t	Š		/	·	Ä	Ô	ä	ô
5		˙	%	5	E	U	e	u	ł		¥		Å	Õ	å	õ
6		-	&	6	F	V	f	v	°		f	¶	Æ	Ö	æ	ö
7		°	'	7	G	W	g	w	Ł		§	•	Ç		ç	
8		‚	(	8	H	X	h	x	ª		¤	,	È	Ø	è	ø
9		β	)	9	I	Y	i	y	ˆ			„	É	Û	é	ù
A		æ	*	:	J	Z	j	z	˜		“	”	Ê	Ú	ê	ú
B		œ	+	;	K	[	k	{	˘		«	»	Ë	Û	ë	û
C		ø	,	<	L	\	l		˘		<	...	Ï	Ü	ì	ü
D	'	Æ	-	=	M	]	m	}	^		>	‰	Í	Ý	í	ý
E	i	Œ	.	>	N	^	n	˘			fi		Î		î	
F	ı	Ø	/	?	O	_	o	˘			fl		Ï	ß	ï	ÿ

TAB. 6 - Exemple d'accès aux fontes très riches de  $\TeX$  par le seul code Ascii

Code (en Ascii)	glyphe
<code>\'e</code>	é
<code>\'E</code>	É
<code>{\sc \'e}</code>	É
<code>\oe</code>	œ
<code>\AA</code>	Å
<code>\dag</code>	†
<code>\alpha</code>	α
--	—
---	—

Mais il faut distinguer deux choses bien différentes :

1. les caractères réellement présents dans une fonte
2. la façon dont les formateurs s'en servent.

Par exemple, les ligatures « *fl* » et « *ffl* » sont présentes dans *Adobe-Garamond*, mais peu de formateurs sont capables, contrairement à  $\TeX$ , de les imprimer.

Le principe est de coder les caractères d'une fonte en utilisant un ou plusieurs codes d'une ou plusieurs normes d'échange.  $\TeX$ , par exemple, utilise des tables de codage très riches pour ses fontes. Mais on peut y accéder par les seuls 94 caractères du jeu de l'Ascii, un caractère, `\`, servant à préfixer les codes des autres (exemple table 6). Bien sûr, une saisie directe à 8 bits (voir table 5), est aussi possible tout comme divers systèmes de saisie directe dans une langue donnée ont été proposés (par exemple pour l'arabe – voir [Haralambous95a]).

$\TeX$  a probablement été le premier système à utiliser la notion de fonte telle qu'elle est répandue aujourd'hui (on verra à sujet [Rubinstein88, Karow92, Hersch93]), c'est-à-dire de base de donnée informatique avec une description de chaque caractère par ses contours, des possibilités de *hints* (adaptation à la grille) et des tables, dites TFM ( *$\TeX$  Font Metric*), fournissant les métriques aux utilisateurs (voir [CousquerPicheral92]). Mais le système adopté par PostScript étant plus répandu, c'est celui que nous décrivons d'abord.

### 2.3.1 Le codage des fontes PostScript

Le codage des fontes PostScript repose sur plusieurs principes [Adobe90, Adobe91, AndréBur91] :

1. Lors de l'impression, une fonte comprend 256 caractères (et ce dans n'importe quel corps ou n'importe quelle direction).
2. Chaque caractère est en fait une procédure de tracé de la glyphe correspondante ; cette procédure a un nom (par exemple A pour « A », semicolon pour « ; », eacute pour « é », etc).
3. Ces 256 procédures peuvent être choisies dans un ensemble beaucoup plus grand : le *Times romain* d'Adobe, par exemple, offre non seulement toutes les lettres accentuées du français, mais aussi celles du polonais, le symbole monétaire florin ou le « ž » (zcaron). Pour cela PostScript utilise un système de codage à deux temps<sup>17</sup> : une table de codage intermédiaire (dite *Encoding Vector*) comprenant 256 entrées permet d'indiquer quel (nom de) caractère correspond à tel code. Par exemple (table 7), il suffit, pour un formateur, d'utiliser le vecteur de codage correspondant à IsoLatin1 pour associer au code numérique 193 le nom Aacute qui dessinera donc la glyphe « À ». Certaines fontes, comme *Symbol* où se trouvent des caractères comme  $\alpha$ ,  $\Leftrightarrow$  ou  $\int$ , ont leur propre jeu de noms de procédures : au code 67 (table 7) on associera par exemple le nom Gamma pour dessiner la glyphe  $\Gamma$ .
4. IsoLatin1 occupant les 256 positions d'une table de codage, il est alors impossible d'y mettre les autres nom de glyphes (telles que « È » ou « ſ »), voire comme <sup>5/8</sup>. Le principe est alors d'utiliser plusieurs « fontes », c'est-à-dire de considérer qu'une fonte est formée de plusieurs morceaux, chacun avec son propre système de codage. Par exemple, le *Garamond romain* d'Adobe, employé ici, utilise les « fontes » suivantes :

---

17. Pour être complet, signalons qu'un opérateur, `glyphshow` permet d'imprimer une glyphe en ne connaissant que son nom.

TAB. 7 - Dans les fontes traitées par PostScript, le choix d'un vecteur de codage permet d'associer à un code numérique le nom d'une procédure de tracé d'un caractère et par là d'adapter une fonte à son propre codage.

Code d'entrée en décimal	Choix de vecteurs : nom (glyphe)			
	codage Adobe standard	codage Apple Quick Draw	codage EBCDIC	codage IsoLatin1
...	...	...	...	
65	A	A		A
66	B	B		B
67	C	C		C
...	...	...	...	
193	grave (‘)	exclamdown (¡)	A	Aacute (À)
194	acute (’)	logicalnot (¬)	B	Acircumflex (Â)
...	...	...	...	...

- AGaramond-Regular pour les caractères d'IsoLatin1,
- AGaramondExp-Regular pour les petites capitales<sup>18</sup> et quelques ligatures (« *ffl* » ...),
- AGaramondAlt-Regular pour des variantes, d'autres ligatures comme *Œ* ou des lettres finales.

Et de même pour les italiques, les gras, les capitales de titres, etc. La famille *Adobe-Garamond* est donc formée d'une vingtaine de fontes numériques.

Lors de la saisie d'un texte, il faut alors appeler une nouvelle fonte (en Word sur Macintosh par exemple, ceci revient à cliquer sur le nom de la fonte correspondante) chaque fois qu'une nouvelle table de codage doit être utilisée. C'est fastidieux, mais probablement pas plus que de changer de casse !

18. Ce sont effectivement de vraies petites capitales et non des capitales réduites optiquement à un corps plus petit.

### 2.3.2 Le codage et la sélection des fontes de (L<sup>A</sup>)T<sub>E</sub>X

(L<sup>A</sup>)T<sub>E</sub>X utilise, en gros, le même style de fontes que celles de PostScript (et éventuellement celles-la d'ailleurs). Mais, puisqu'il s'agit d'un formateur, leur emploi est mieux défini et se veut plus général que les autres systèmes surtout depuis deux ou trois ans (notion de fonte virtuelle, *New Font Selection Scheme*, etc.). On trouvera dans [CousquerPicheral92, Goossens94] les principaux textes sur ces mécanismes et y renvoyons le lecteur. En résumé, disons ici que L<sup>A</sup>T<sub>E</sub>X distingue nettement (même si c'est souvent transparent à l'utilisateur normal) :

1. le contenu physique des fontes (tables de métrique par exemple) ;
2. la façon, pour un utilisateur, d'appeler une fonte ; pour les non-T<sub>E</sub>Xistes, ceci se fait par l'intermédiaire de « style » ou classe de documents soit de façon automatique (mais programmable) – par exemple le seul fait d'écrire en tête de cet article

```
\title{Codage, caractères spéciaux & ligatures}
```

fait que le titre sera composé en *Times*, romain, corps 16 – soit de façon explicite ;

3. la façon de faire entrer les fontes physiques d'une famille commerciale (par exemple *Adobe-Garamond*) dans ce système ;
4. la façon de coder les caractères en fonction d'une fonte donnée et, par exemple, de la langue : ce peut être à 7 bits (table 6), à 8 bits (table 5), voire sur 16 bits [Plaice94].
5. etc.

## 3 Les caractères spéciaux

Revenons à présent sur ces caractères spéciaux qui constituent la majorité des caractères optionnels de l'Ascii et qui sont donc sur pratiquement tous nos claviers d'ordinateur : & @ £ \$ # ¤ et même { et }.

### 3.1 Une origine commune et une récupération informatique

Ces caractères ont tous en commun d'être des signes utilisés en comptabilité américaine. Rappelons que les ordinateurs ont longtemps servi essentiellement à des tâches de gestion (d'ailleurs BM dans IBM veut dire *Business Machine*). C'est donc là, et uniquement là, qu'il faut chercher l'origine de ces symboles.

Mais comme tous les informaticiens ne font pas de la gestion et comme ils ont toujours besoin de symboles « spéciaux », les développeurs de logiciels se sont appropriés ces symboles. Aussi, même si ces symboles n'ont vraiment raison aucune d'être sur tous les ordinateurs en tant que symboles comptables, on aurait du mal aujourd'hui à s'en passer pour les langages informatiques.

### 3.2 Signe dollar \$

C'est évidemment le plus connu de ces symboles. C'est le symbole monétaire des États-Unis d'Amérique ; il est donc très employé dans le reste du monde.

Le mot *dollar* vient lui-même du nom d'une monnaie mexicaine *dolera* (dont le vrai nom était *peso*, du latin *pensum*, poids) qui vient de l'allemand *Thaler* (en bas-allemand *Daler*) du nom d'une monnaie frappée (et rendue plus ou moins européenne en 1537, déjà !, par Charles Quint) dans la vallée (en allemand *Thal*) de Joachim : *Joachimthal* [Sedillot89].

Contrairement à ce que disent beaucoup d'américains, ce symbole n'est pas dessiné en surimposant le U et le S de US : cette explication n'est pas possible car le symbole \$ est attesté avant la création des États Unis d'Amérique ! ; il se retrouve d'ailleurs sur beaucoup d'autres pièces de monnaies (voir figure 9). D'autres hypothèses ont été émises. La première est qu'il s'agit d'un 8 déformé : cette monnaie espagnole *dolera* s'appelait aussi « la pièce de 8 » car elle valait 8 réales<sup>19</sup>. La seconde est que cette monnaie d'origine espagnole portait au revers le symbole de Gibraltar : deux<sup>20</sup> barres verticales pour symboliser le détroit, et un drapeau flottant comme un S. Enfin, ce symbole viendrait du « p » de *peso* avec une barre en biais (comme la deuxième série des pesetas de la figure 9).

19. La figure 9 montre une planche d'un catalogue de la fonderie Déberny et Peignot où l'on voit que le symbole de la monnaie portugaise Réis ressemble étrangement au signe dollar.

20. Il est intéressant de voir que l'on a tendance actuellement à ne plus mettre qu'une barre au signe dollar, voir figure 6, alors qu'on en mettait encore deux en 1926, voir figure 9.



FIG. 6 - *Quelques signes dollar (de gauche à droite : Times, Helvetica, Palatino, Bookman, ZapfChanchery et Courier)*

Ce caractère est utilisé en informatique comme délimiteur de chaînes (par exemple en  $\text{\TeX}$  pour marquer les formules mathématiques), comme pointeur dans les langages Ada, Pascal, etc.

### 3.3 Esperluette &

Il s'agit de la très vieille ligature « et » qui a fait l'objet d'études célèbres de Jan TSCHICHOLD [Tschichold] et de Gérard BLANCHARD [Blanchard92].

Ce caractère est très utilisé aux États-Unis (plus qu'en France) dans les noms de sociétés commerciales (par exemple *Bigelow & Holmes* et plus généralement sous la forme *& Co.*).

Son nom français est « esperluette » ; mais il y a beaucoup de variantes : « perluète » pour ISO Latin, « perluette » ou « eperluette » ; il est aussi appelé *et commercial* (voir ci-dessous le *à commercial*) ce qui confirme son origine comptable. L'origine de ce mot esperluette n'est pas non plus bien connue : le Grevisse dit que ce caractère s'appelait « ète » et qu'il était placé dans l'alphabet après le z, alphabet que les enfants chantaient « a, b, ..., z et puis le ète » ce qui aurait donc donné « éperluette » [Grevisse, art. 71]. Blanchard fait aussi le rapport avec une origine ludique (pirlolette). Son nom anglais est *ampersand* et est en fait un mélange de latin et d'anglais : *and per se and* (et<sup>21</sup> à lui tout seul « et »).

En informatique ce signe a des usages divers, mais *très rarement* liés à « et » au sens logique (pour lequel on emploie plutôt le caractère  $\cap$  ou  $\wedge$ , voire +). Toutefois certains éditeurs logiciels l'utilisent comme symbole de concaténation ;  $\text{\TeX}$  l'utilise essentiellement comme symbole de tabulation.

21. On raconte aussi la même histoire qu'en français d'alphabet chanté [Ullman63].





FIG. 7 - *Quelques esperluettes (de gauche à droite : Times, Helvetica, Palatino, Bookman, ZapfChanchery et Courier)*



FIG. 8 - *Quelques « à commercial » (de gauche à droite : Times, Helvetica, Palatino, Bookman, ZapfChanchery et Courier)*

### 3.4 À commercial @

Voici en tout cas un caractère qui était pratiquement inconnu en France il y a quelques années à peine.

Comme le &, ce caractère est aussi issu des chancelleries ; c'est la ligature latine *ad* (« à » en français) où le a et le d cursifs de l'onciale (*ad*) ont fini par se confondre.

Ce caractère n'est utilisé, aux USA, qu'en comptabilité pour indiquer les prix unitaires : ainsi « deux livres à 1 dollar pièce » s'écrit dans une facture « 2 books @ \$ 1 ».

Le nom français de ce caractère, selon la version française (Afnor) d'ISO Latin est « a commercial ». Cependant, le nom que lui donnent les informaticiens français tourne autour de sa forme : a-rabesque, a-rondi, a roulé, a-arrondi. Mais le nom le plus fréquemment employé, du moins dans les milieux universitaires, est « arobas ». Ce mot vient d'une confusion que nous nous expliquons de la façon suivante. Les traducteurs qui ont voulu faire imprimer des manuels techniques où apparaissait pour la première fois ce signe ont du s'adresser à un imprimeur qui leur aura sorti un catalogue de fondeurs français comme celui de la figure 9. On y voit effectivement un caractère qui a à peu près la même graphie, qui s'appelle arobas, mais qui correspond à quelque chose de complètement différent : c'est le symbole d'une unité de poids espagnol (*arroba*, poids de 25 livres espagnoles, soit 11, 502 kg, dont le vrai nom français d'ailleurs est « arobe » selon le Robert).

Ce caractère est très employé aujourd’hui avec le sens  $\@$  dans les adresses de courrier électronique, par exemple `gut@ens.fr` qui se lit « GUT(enberg) à École normale supérieure, France ». Mais son utilisation principale en informatique a été longtemps d’être un caractère « inutilisé normalement » ce qui permet notamment de coder (par exemple en  $\TeX$ ) des macros internes sans risque de confusion avec du texte ou une fonction définie par l’utilisateur (qui est donc censé ne pas utiliser ce caractère ainsi – un code, par exemple `\at` en  $\TeX$ , permettant néanmoins d’imprimer la glyphe correspondante).

### 3.5 Numéro #

Le symbol `#` est aussi issu d’une ligature latine : *numerus* (nombre), un « n » surmonté d’une barre, c’est-à-dire «  $\bar{n}$  » et dont la barre est descendue peu à peu au bas des jambes du « n ».

En comptabilité américaine, ce signe sert à indiquer des numéros (de pièce, de série, de compte bancaire, etc.) ; dans un facture, on écrira par exemple « 4 gonds numéro 78-9253 » sous la forme « *4 hinges # 78-9253* ».

Ce caractère est souvent appelé dièse en français à cause de sa ressemblance avec ce signe musical, mais « `#`  $\neq$   $\sharp$  ». Son nom américain est *number* (« numéro », c’est d’ailleurs la traduction officielle adoptée par l’Afnor pour ISO latin) ; mais le manuel du langage Ada l’appelle aussi *sharp* (dièse). En Grande-Bretagne, il est parfois appelé *hash* (hacher).

En informatique, il sert justement à indiquer des numéros (par exemple des paramètres en  $\TeX$ ) ou des bases de nombre, par exemple  $12_8$  peut s’écrire `12#8`.

### 3.6 Livre £

C’est le symbole monétaire *livre sterling* britannique. Il a été introduit dans le code ISO-646 (Ascii), à la demande des russes lors de la guerre froide en réaction à l’hégémonie américaine même temps que le symbole  $\text{r}$  international<sup>22</sup>.

Son nom vient d’une ancienne monnaie d’argent qui valait une livre (poids) d’argent. On retrouve donc la dualité poids/valeur monétaire pour *pounds* comme pour *peso* et nos livres (tournois ou de beurre). Le mot livre vient du latin *libra* qui

---

22. Le franc français dont l’abréviation est F (et non Fr) n’a pas besoin d’un symbole spécial, encore qu’il y en ait eu sur quelques machines à écrire.

FIG. 9 - *Extrait du catalogue de la fonderie Deberny & Peignot, 1926 (document communiqué par Alan MARSHALL, Musée de l'Imprimerie à Lyon).*

a aussi donné l'italien *lira*. Le mot *sterling*, nom adopté par IsoLatin1, est d'origine obscure ! Le dessin représente un « L », abréviation de *Libra*, et ressemble donc à celui de la lire italienne (voir figure 9).

En informatique, le symbole £ sert peu puisqu'il est en général présent, de façon optionnelle, uniquement à la place de # (voir table 3).

### 3.7 Symbole monétaire international ₭

Symbole très peu utilisé, créé de toutes pièces pour contrer le monopole du dollar. Il représente un cercle portant quatre rayons symbolisant une pièce d'or qui brille (voir figure 10).

En français on l'appelle « symbole monétaire international » mais beaucoup de nos collègues emploient son nom américain : *currency*. La norme ISO 7350 lui donne le nom « symbole monétaire général » tandis que ISO Latin (8859) se contente de « symbole monétaire ».

À noter que le Yen fait partie d'IsoLatin1 (code 245, table 4) mais pas l'Ecu (qui entre toutefois dans une norme, UER 3232, de codage de sous-titres [Marti90, page 258]) et que le codage Unicode comprend une dizaine de symboles monétaires [Bigelow94, figure 2] tandis qu'*Adobe-Garamond* comprend aussi plusieurs glyphes de symboles monétaires (figure 10).

### 3.8 Les accolades

Enfin, signalons que les accolades { et } sont, en France, des symboles de dimensions très variables servant à accoler plusieurs lignes de textes<sup>23</sup>. Elles n'ont pas du tout ce rôle de paires de parenthèses ou de crochets du code Ascii qu'on leur fait jouer actuellement en informatique ou en T<sub>E</sub>X.

### 3.9 Symboles américains et caractères français

Terminons par une note cocasse : tous ces caractères spéciaux d'origine américaine font donc obligatoirement partie des « fontes » numérisées d'aujourd'hui. Il n'est donc pas surprenant de trouver, des signes monétaires (dollar, yen, cents, etc.

---

23. Le *Dictionnaire de l'Académie* dit, en 1740 : « On appelle *accolade*, dans un compte, un trait de plume qui joint plusieurs articles pour n'en faire qu'un. » ; cité par [Frey].

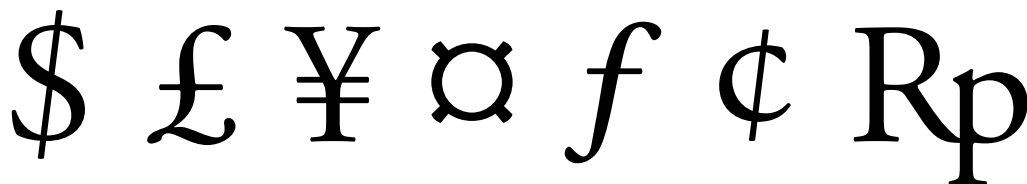


FIG. 10 - *Symboles monétaires proposés par Adobe-Garamond ; de gauche à droite : dollar, livre sterling, yen, symbole monétaire international, florin, cents, roupie.*

voir figure 10) et bien d'autres dans une fonte qui porte le nom de GARAMOND mais que celui-ci n'a sûrement pas gravés.

## 4 Ligatures et PAO

La notion de ligature a fait l'objet de nombreux écrits dont on trouvera une synthèse dans [Peignot9?]. Comme le dit Yannis HARALAMBOUS [Haralambous95], il existe plusieurs classes de ligatures typographiques. Nous nous limiterons ici à ce que René PONOT appelle plutôt « lettres à combinaison » [Ponot93] ; Haralambous [Haralambous95] parle de « ligatures esthétiques » mais on pourrait aussi les appeler « traditionnelles » puisqu'il convient de conserver le ductus des lettres liées par les copistes ou scribes. Si ce terme convient aux « *ff* » ou « *fi* » du *Garamond* bien connues des lecteurs de la Pléiade, une raison plus technique est souvent à l'origine des ligatures comme l'explique Émile LECLERC dans son *Traité de typographie* en 1947 (donc du temps du plomb) : « Les lettres doubles *fl*, *fi*, *ff*, *ffl*, *ffi*, sont fondues ensemble, parce que la bouclette supérieure en saillie de la lettre *f* rencontrant le point de la lettre *i* ou l'extrémité supérieure de la lettre *l* occasionnerait par pression latérale la rupture d'une de ces deux partie, peut-être des deux, et conséquemment, outre un aspect fâcheux, la perte de ces lettres mutilées » [Leclerc, page 92]. La figure 11 montre bien qu'en *Garamond* les deux lettres « *f* » et « *i* » se chevauchent et la ligature « *fi* » est donc indispensable. Ceci est moins vrai pour le *Times* ou d'autres caractères (voir aussi la figure 14).

Voyons alors, dans ce contexte des tables de codage pour ordinateurs et celles pour fontes, ce que nous offre la PAO en matière de ligatures. Toutefois, nous nous limitons ici au français, renvoyant le lecteur à [Haralambous95] pour les langues arabe, khmer, etc.

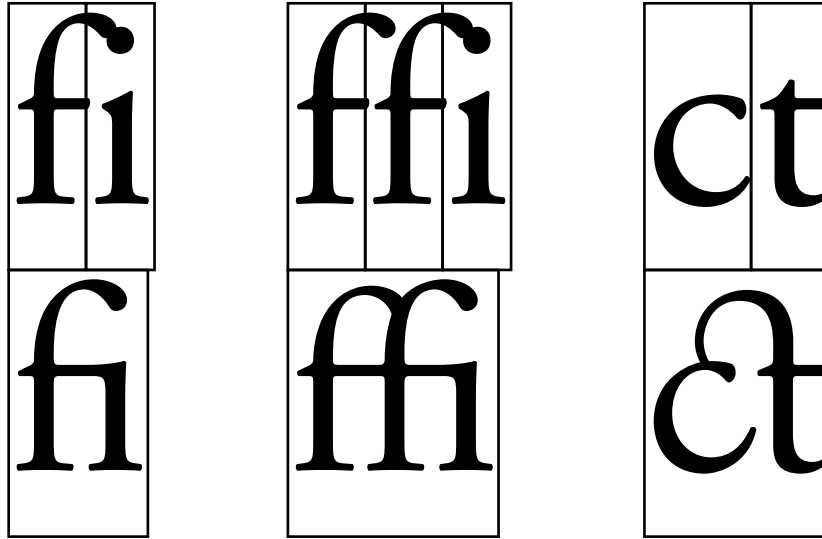


FIG. 11 - *En haut : séquence normale de caractères ; en dessous : les ligatures correspondantes.*

#### 4.1 Informatique et ligatures

Faire une ligature consiste à associer à une paire de caractères  $C_i C_j$  une glyphe unique  $C_k$ , étant entendu que l'un des premiers  $C_n$  peut être lui même le résultat d'une ligature. Les exemples typiques sont le remplacement de « ct » par « *ct* » ou ceux successifs de « ff » par « *ff* » puis de « *ffi* » par « *ffi* » (figure 11). On peut aussi considérer que les début et fin de mots sont des caractères vides. Ceci permet alors de traiter comme ligature les formes spéciales de certaines lettres, en début ou en fin de mot (voir figure 12). Mais rappelons que « œ » et « Œ » ne sont pas des ligatures mais des caractères à part entière.

Deux problèmes sont alors à résoudre point de vue informatique :

1. de quelles ligatures dispose-t-on ?
2. comment utiliser ces ligatures ?

## 4.2 Inventaire des ligatures

Les normes d'échange, on l'a vu plus haut, ne connaissent pas les ligatures puisque, pour elles, ce sont des glyphes et non des caractères. Quant aux fontes elles-mêmes, on peut distinguer plusieurs cas :

1. des fontes qui n'ont pas de ligatures ; c'est par exemple le cas des vieilles *Chicago*, *New York*, *Geneva* d'Apple ; on ne peut pas en dire grand chose, sauf qu'elles sont un peu à l'origine de la mauvaise presse qu'a pu avoir la PAO aux yeux des typographes !
2. des fontes qui n'ont que quelques ligatures notamment les deux plus courantes : « fi » et « fl » ; c'est pratiquement le cas de toutes les fontes numérisées (comme le *Times*) probablement car ces deux ligatures font partie des tables de codage standard d'Adobe ;
3. des fontes qui en disposent de plus ; par exemple *Lucida* de Bigelow & Holmes ; pour pouvoir les utiliser, il faut alors les faire entrer dans le vecteur de codage (voir table 7) ; soit à la place d'un autre caractère, soit à la place d'un code inutilisé par PostScript (c'est en gros ce que fait T<sub>E</sub>X dans le standard de Cork, table 5, où on voit que, par exemple, « ffi » est à la place du *trademark* d'IsoLatin1) ;
4. enfin, certaines familles de caractères, comme *Adobe-Garamond* sont ventilées dans plusieurs fontes (voir ci-dessus section 2.3.1). Par exemple, la figure 12 montre les ligatures (au sens large, en incluant les lettres finales que nous considérons comme des ligatures avec un caractère vide) disponibles dans cette famille. On trouve, d'autres exemples de ligatures ou de lettres initiales, notamment celles dessinées par François BOLTANA [Boltana95].

On trouve aussi des ligatures plus exceptionnelles comme « tt », « us » ou « sa » dans le *Delorme* [Delorme90]. L'intérêt y est surtout de montrer qu'il est possible d'intégrer des ligatures dans les fontes actuelles, et il est donc bien dommage que la fonte *Avant-Garde* d'Adobe n'ait pas gardé toutes ces ligatures qui en faisaient le charme (figure 13).

Enfin signalons, une fois de plus, que le marché des fontes n'étant pas du tout d'origine française, les « fondeurs » américains ne se posent pas toujours les bonnes

t a e n z v  
*ff fi fl ffi ffl et st*

FIG. 12 - *Lettres finales (en haut) et ligatures (en bas, en italique) proposées par Adobe-Garamond.*

FIG. 13 - *Ces ligatures d'Avant-Garde telle que ce caractère a été dessiné par Herb Lubalin, ne font malheureusement pas partie de la fonte numérisée par Adobe.*



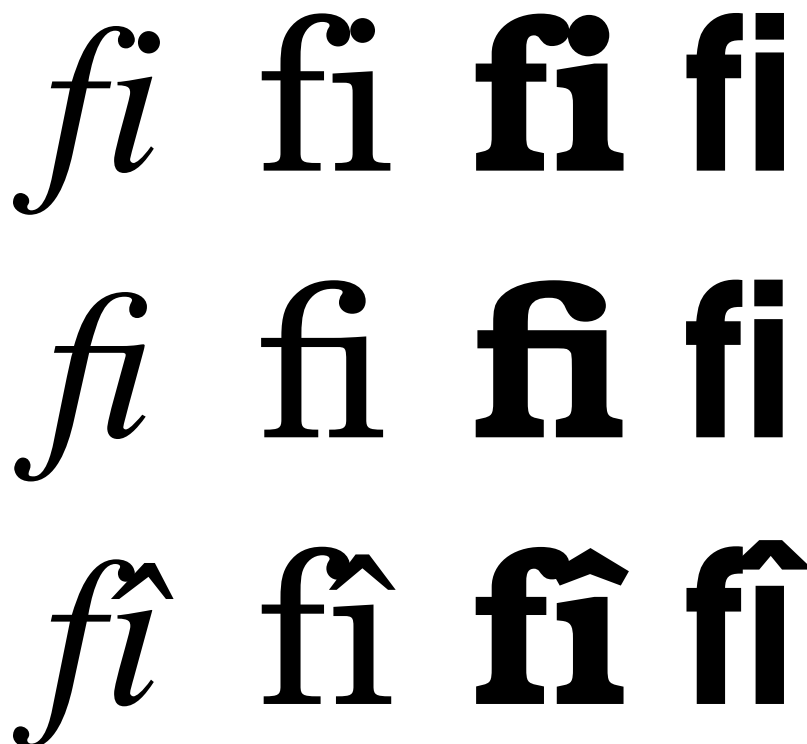


FIG. 14 - *En haut : quelques suites de lettres, sans ligatures ; au milieu les mêmes avec ligatures (on notera que toutes ne sont pas utiles) ; en dessous : quelques ligatures qui manquent !*

questions au sujet des ligatures : par exemple il est assez curieux de trouver dans des tables de fonte des ligatures pas toujours indispensables<sup>24</sup> alors qu'il y a des dessins intolérables pour d'autres combinaisons de lettres comme celles de la figure 14.

24. Elles ne sont souvent là que parce que les tables de codage les prévoient – c'est en particulier le cas de *Courier* alors qu'il n'y a aucun problème de chevauchement puisque tous les caractères ont une chasse unique.

### 4.3 Saisie des ligatures

Comment un utilisateur voyant dans un catalogue de fontes la présence d'une ligature va-t-il pouvoir s'en servir par le biais d'un formateur ? Tout dépend de ce dernier, et il y a plusieurs façons de faire (supposons que l'on veuille obtenir la ligature « fi » dans le mot « afin »).

1. Cas le plus simpliste : certains formateurs, Word par exemple – mais c'est aussi le cas de FrameMaker, ne connaissent pas spécialement les ligatures. Il faut alors taper le code interne de cette ligature, par exemple en Word/Macintosh on tapera « a commande option Q 222 n » après avoir trouvé (annexe D du manuel Word) que le code de « fi » est 222. Si de plus la ligature est dans une fonte spéciale, c'est le cas du « Æ » d'*Adobe-Garamond*, il faut alors non seulement donner ce numéro de code, mais en plus préciser quelle fonte on utilise. Même avec des remplacement automatiques, ce n'est pas une façon très naturelle de saisir un document et on comprend donc que les usagers de ces systèmes rechignent à employer les ligatures !
2. Pour de nombreux systèmes, utiliser des ligatures autres que « fi » et « fl » nécessite d'appeler une autre fonte ; par exemple, la ligature « ffl » se trouve dans *Adobe Garamond Expert*. Mais alors, il faut appeler cette glyphe soit avec son code (132), soit avec le caractère qui a le même code dans le vecteur de codage standard (ici il s'agit de « Z ») ;
3. Certains systèmes, comme QuarkXPress, permettent de déclarer, en début de travail, quelles sont les ligatures qui seront utilisées. Le remplacement se fait alors automatiquement par la suite. L'inconvénient de cette méthode est que si l'on change de fonte, il faut refaire cette initialisation.
4. Enfin, quelques rares systèmes, comme T<sub>E</sub>X, utilisent les informations données dans les tables de métrique, TFM [CousquerPicheral92] ou AFM [AndréBur91]. En effet, une fonte est fournie avec une « métrique » qui, outre les tables de chasse, contient aussi des informations comme la présence de ligatures. Par exemple, une fonte pourra avoir la partie suivante :

```

C 102 ; WX 410 ; N f ; B ... ; L f ff ; L i fi ; L l fl ;
...
C 174 ; WX 663 ; N fi ; B ... ;
C 175 ; WX 663 ; N fl ; B ... ;
...
C -1 ; WX 723 ; N ff ; B ... ; L i ffi ; L l ffl ;
C -1 ; WX 1000 ; N ffi ; B ... ;
C -1 ; WX 1000 ; N ffl ; B ... ;

```

Cette table peut se lire ainsi :

- le caractère de code (C) 102 a une chasse (Wx) de 410/1000 de cadratin, a pour nom (N) f ,... et il existe une ligature (L) avec « f » qui s'appelle ff , une avec « i » qui s'appelle fi et une avec « l » qui s'appelle fl ;
- le caractère de code (C) 174 a une chasse (Wx) de 663/1000 de cadratin et a pour nom (N) fi ;
- le caractère de code (C) 175 a une chasse (Wx) de 663/1000 de cadratin et a pour nom (N) fl ;
- le caractère de nom (N) ff a une chasse (Wx) de 723/1000 de cadratin et il existe une ligature (L) possible avec le caractère « i » qui s'appelle ffi et une avec « l » qui s'appelle ffl ; ce caractère n'a pas de code standard (C=-1) et doit être entré spécialement dans le vecteur de codage (voir figure 7) ;
- le caractère de nom (N) ffl a une chasse de 1000/1000 cadratin et n'a pas de code standard (C=-1) ;
- le caractère de nom (N) « ffi » a une chasse de 1000/1000 cadratin et n'a pas de code standard (C=-1) ;

En suivant ce chaînage, il est alors possible, pour un formateur, de décider des ligatures à utiliser sans que l'utilisateur ait à s'en soucier. C'est ce que nous pourrions faire ici : ce texte a été composé, pour respecter le style des rapports de recherche Inria-Rennes, en utilisant du *Times* où il n'y a pas les ligatures « ffi ». Sans le moindre changement à notre texte, il nous suffirait d'appeler *Adobe-Garamond* pour que les ligatures « ffi » soient, automatiquement, prises

en compte. Toutefois, l'auteur doit pouvoir interdire certaines ligatures : ainsi, par exemple en allemand, n'utilise-t-on pas de ligatures « fl » dans un mot composé d'un premier mot se terminant par « f » et d'un commençant par « l »<sup>25</sup>.

Cette méthode a toutefois l'inconvénient de nécessiter de nombreuses « fontes » dans le cas de langues très riches en ligatures. La tendance toute nouvelle est alors d'utiliser la notion de *cluster* comme dans le projet Omega [Plaice93, Plaice94].

## 4.4 Remarques

Si l'emploi de ligatures est donc assez courant désormais en PAO, certains problèmes restent souvent méconnus. En voici quelques-uns :

### 4.4.1 Ligatures et correcteurs orthographiques

Si l'on exclut les systèmes comme T<sub>E</sub>X, l'emploi des ligatures se fait par codage. Si on écrit le mot « effluve », il sera codé (voir ci-dessus) soit `e\l32uve` soit `eZuve` (en cachant le changement de fonte). Dans les deux cas, ce mot sera rejeté par un correcteur orthographique.

### 4.4.2 Ligatures et interlettrage

L'interlettrage, le crénage automatique (*track kerning* des anglo-saxons), les modifications d'approche, etc. sont des techniques peu appréciées<sup>26</sup> des bons typographes qui font confiance aux métriques données (ou imposées du temps du plomb) par les dessinateurs de caractères. Voici une bonne raison supplémentaire de ne pas en faire : lorsqu'on modifie la distance entre les lettres d'un mot, les ligatures qui sont alors assimilables à un seul caractère, voient leur glyphe non modifiée ce qui rompt l'esthétique du mot ; voir figure 15.

---

25. On écrira par exemple `auf{}lage` pour obtenir *auflageet* non *auflage*.

26. À quelques rares exceptions près, comme pour supprimer une veuve ou pour éviter de rejeter le dernier mot d'un titre à la ligne suivante.

inefficace  
inefficace  
inefficace

FIG. 15 - *Les modifications d'interlettrage ne touchent pas aux glyphes des ligatures.*

### 4.4.3 Ligatures et grands corps

Certains auteurs ont fait remarquer depuis longtemps que les ligatures et le crénage suivent des nécessités opposées : plus le corps est grand, plus le crénage doit être fort (au point de d'arriver à faire se toucher deux caractères consécutifs [Karow87, page 61]). Par contre, les ligatures deviennent de moins en moins utiles. Une fonte faisant correctement de l'ajustement optique (*optical scaling*) comme disent le faire les fonte *MultipleMaster* d'Adobe, devraient donc proposer des ligatures fortes pour les petits corps et disparaissant au delà d'une certaine force de corps.

### 4.4.4 Composition artificielle de ligatures

Comme nous l'avons vu, beaucoup de fontes n'offrent que peu de ligatures, par exemple « fi » mais pas « ffi ». Certains systèmes essaient alors de produire ces dernières par un crénage entre « f » et « fi » (c'est-à-dire par un déplacement arrière du point courant d'écriture). C'est notamment ce que fait ici T<sub>E</sub>X pour le *Times*. Mais les résultats ne sont pas toujours très valables (figure 16).

Même si nous considérons que « Œ » n'est pas une ligature, signalons que cette méthode est aussi utilisée pour imprimer les glyphes correspondantes (c'est d'ailleurs ce qui a justifié la non-présence de ce caractère dans IsoLatin1). Là aussi, les résultats peuvent être décevants (figure 16).

Mais dans tous les cas, cette méthode est à proscrire avec les outils travaillant avec des fonds d'écran (comme le récent Mosaic) : en effet la couleur du fond est en général perdue et le résultat est alors très piteux.

## 5 Conclusion

Nous espérons avoir montré dans cet article qu'il y a plusieurs partenaires dans l'emploi des caractères informatisés et en particulier des ligatures. On ne dira sans doute jamais assez qu'il ne faut pas confondre une « fonte » et l'usage que l'on en fait. On n'a pas le droit de dire par exemple que « le Macintosh » n'offre pas de ligatures « ffi ». C'est vrai avec certaines fontes, ou avec certains logiciels, ou les deux. C'est faux avec d'autres, à condition que l'auteur les utilise (quant à T<sub>E</sub>X, rappelons que ceci est fait automatiquement pour les fontes qui le permettent).



FIG. 16 - *La simulation de caractères par déplacement du point courant (ou par crénage) ne donne pas (ligne du haut) les résultats attendus (ligne du bas)*

C'est aux utilisateurs de se battre pour disposer des ligatures utiles, mais aussi des caractères accentués comme É ou de certains comme « œ ». Il ne faut surtout pas croire que c'est une question technique, elle n'est que politique !

### **Remerciements**

Nous tenons à remercier ici Alan Marshal (Musée de l'Imprimerie à Lyon) et Charles Bigelow (Stanford University) pour les nombreux renseignements qu'ils nous ont fournis, Philippe Louarn (Irisa-Rennes) pour son aide à composer ce document, ainsi que les personnes qui ont bien voulu corriger des versions successives de cette note, en particulier Pierre Legrand et Éric Picheral.

## Références

- [Adobe90] Adobe Systems Incorporated, *Adobe Type 1 Font Format*, version 1.1, Addison-Wesley Publishing Company, Reading, MA (USA), 1990.
- [Adobe91] Adobe Systems Incorporated, *PostScript Language Reference Manual*, Addison-Wesley Publishing Company, Reading, MA (USA), 2nd edition, 1991.
- [André89] Jacques ANDRÉ, « Langages de publication assistée par ordinateur », art. du *Traité d'informatique – Techniques de l'ingénieur*, H2440, 12 pages, 1989.
- [AndréBur91] Jacques ANDRÉ & Justin BUR, « Métrique de fontes PostScript », *Cahiers GUTenberg*, num. 8, mars 1991, 29–50.
- [Delorme90] Jacques ANDRÉ et Christian DELORME, « Le Delorme : un caractère modulaire et dépendant du contexte », *Communication et langage*, 86, 1990, 65–76.
- [André94] Jacques ANDRÉ, « Unicode – une casse de 38 000 signes », *Caractères*, num. 373, 1994, p. 32–36.
- [Autrement94] « Machines à écrire, des claviers et des puces : la traversée du siècle », numéro spécial de la revue *Autrement*, num. 146, juin 1994.
- [Bettels93] Jürgen BETTELS and E. Avery BISHOP, « Unicode: A Universal Character Code », *Digital Technical Journal*, vol. 5, num. 3, été 1993, 21–37.
- [Bigelow94] Chuck BIGELOW and Kris HOLMES, « The design of a Unicode font », *EPODD, Electronic Publishing, Origination, Dissemination and Design*, vol. 6(3), september 1993 (actes de RIDT'94 *Raster Imaging and Digital Typography*), 289–305.
- [Blanchard92] Gérard BLANCHARD, « Nœuds & esperluettes – actualités et pérennité d'un signe », *Communication et langages*, num. 92, 1992, p. 85–101.
- [Boltana95] François BOLTANA, « Ligatures et calligraphie assistée par ordinateur », *Cahiers GUTenberg*, 1995, (à paraître).



- 
- [CousquerPicheral92] Alain Cousquer & Éric Picheral, « Polices, T<sub>E</sub>X et Cie », *Cahiers GUTenberg*, num. 9, juillet 1992, 3–31.
- [Ferguson90] Michael J. FERGUSON, « Fontes latines européennes et T<sub>E</sub>X 3.0 », *Cahiers GUTenberg* num. 7, novembre 1990, p. 29–31.
- [Frey] A. FREY, *Nouveau manuel complet de typographie*, Manuels Roret, 1857 (réédition fac-similée par Laget, Paris, 1979).
- [Frutiger] Adrian FRUTIGER, « Typography with the IBM Selectric Composer », *Journal of Typographic Research*, vol. 1, num. 3, 285–292.
- [Goossens94] Michel GOOSSENS, Frank MITTELBACH et Alexander SAMARIN, *The LaTeX Companion*, Addison-Wesley, 1994.
- [Grevisse] Maurice GREVISSE, *Le bon usage*, Duculot, 1986.
- [Haralambous95] Yannis HARALAMBOUS, « Tour du monde des ligatures », *Cahiers GUTenberg*, 1995, à paraître.
- [Haralambous95a] *Cahiers GUTenberg*, numéro spécial « TeX et l’arabe », en préparation sous la direction de Yannis HARALAMBOUS, 1994.
- [Hersch93] Roger HERSCH (ed.), *Visual and Technical Aspects of Types*, Cambridge University Press, 1993.
- [Karow87] Peter KAROW, *Digital Formats for Typefaces*, URW Verlag, Hambourg, 1987.
- [Karow92] Peter KAROW, *Schrifttechnologie – Methoden un Werkzeuge*, Springer-Verlag, 1992. Version anglaise : *Font Technology*, URW Verlag, Hambourg, 1994.
- [Karow94] URW Company, *The hz-Program – Micro-typography for advanced typeetting* URW SOftware & Type GmbH, Hambourg, 1993. En français : Peter KAROW, « Le programme hz de microtypography », *Cahiers GUTenberg*, 1995, (à paraître).
- [Laufer80] *La machine à écrire*, actes réunis par Roger LAUFER, Sollin éd., 1980.

- [Leclerc] Émile LECLERC, *Nouveau manuel complet de typographie*, Encyclopédie Roret, SFELT, Paris, 1947.
- [LouarnAndré94] Jacques ANDRÉ et Philippe LOUARN, « La fonte du jour : Adobe-Garamond », *La Lettre GUTenberg*, num. 4, décembre 1994, p. 19–20.
- [Marshall93] Alan MARSHALL, *Ruptures et continuités dans un changement de système technique – le remplacement du plomb par la lumière dans la composition typographique*, thèse, Grenoble, 18 décembre 1991. Parue comme *Publication interne Irisa*, num. 638, mars 1992.
- [Marshall94] Alan MARSHALL, « Les origines de la PAO », *Caractères*, num. 382 et 383 (pages 36–38), juin 1994.
- [Marti90] Bernard MARTI et co-auteurs, *Télématique – techniques, normes, services*, Dunod, 1990.
- [Peignot9?] Jérôme PEIGNOT, « Petit traité de la ligature », *Communication et langages*, num. 73,
- [Plaice93] John PLAICE, « Language-Dependent Ligatures », *TUGBoat*, vol. 14, num. 3, 1993, 270–274.
- [Plaice94] John PLAICE, « Advances in the Omega Project », Congrès annuel TUG 1994, à paraître dans *TUGBoat*.
- [Ponot93] René PONOT, « Le *Didot* a-t-il besoin de ligatures? », *Cahiers GUTenberg*, 1995, (à paraître).
- [Rubinstein88] Richard RUBINSTEIN, *Digital Typography – An Introduction to Type and Composition for Computer System Design*, Addison-Wesley, 1988.
- [Sedillot89] René SÉDILLOT, *Histoire morale et immorale de la monnaie*, Bordas, 1989.
- [Tschichold] Jan TSCHICHOLD, *Formen Wandlungen der &-zeichen*, D. Stempel AG, Francfort. Traduction française de René Grasset à paraître à l'École Estienne.

- [Ullman63] B. L. ULLMAN, *Ancient writing and its influence*, Cooper Square Publishers, Inc., New York, 1963.
- [Unicode91] The Unicode Consortium, *The Unicode Standard – Worldwide Character Encoding*, Addison-Wesley, vol. 1 (1991) et 2 (1992).
- [Wild94] Adolf WILD, « La typographie de la *Bible* de GUTENBERG », ce rapport, pages 3–15 et *Cahiers GUTenberg*, 1995, (à paraître)..

Projet OPÉRA  
Récentes publications internes IRISA-PI

Les publications suivantes sont disponibles :

- sous forme papier ; s'adresser au *Service documentation, Irisa, campus de Beaulieu, F-35042 Rennes cedex, France ; FAX: (+33) 99 84 71 71 ;*
- pour les plus récentes, par ftp anonyme à *ftp.irisa.fr* (131.254.254.2), dans le répertoire */techreports*, sous forme compressée *PI-xxx.ps.Z* (où xxx=numéro de publication).

The following research reports from Opera team are available

- in paper form, on request to *Service documentation, Irisa, campus de Beaulieu, F-35042 Rennes cedex, France ; FAX: (+33) 99 38 38 32 ;*
- by anonymous ftp at *ftp.irisa.fr*, directory : */techreports*, as compressed files *PI-xxx.ps.Z* where xxx= research report number.

609. Hélène RICHY, Patrice FRISON et Éric PICHERAL, *Intégration d'un correcteur typographique dans l'éditeur structuré Grif*, Publication interne Irisa n° 609, 1991.
636. Jacques ANDRÉ et Roger HERSCH, *Un curriculum pour la typographie numérique*, Publication interne Irisa n° 636, 1992.
676. Jacques ANDRÉ, *Font metrics*, Publication interne Irisa n° 676, 1992.
677. Hélène RICHY, *Grif et les index électroniques*, Publication interne Irisa n° 677, 36 pages, septembre 1992.

715. Jacques ANDRÉ, Dominique DECOUCHANT Vincent QUINT et Hélène RICHY, *Vers un atelier éditorial pour les documents structurés*, Publication interne Irisa n° 715, 15 pages, mars 1993.
747. Jacques ANDRÉ and Irène VATTON, *Contextual Typesetting of Mathematical Formulae Taking Care of Optical Scaling*, Publication interne Irisa n° 747, 24 pages, juin 1993.
841. Jacques ANDRÉ and Hélène RICHY, *Utilisation des index d'un éditeur structuré dans le cadre d'actes médiévaux*, Publication interne Irisa n° 841, juin 1994.



---

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,  
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY  
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex  
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1  
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex  
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

---

Éditeur  
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)  
ISSN 0249-6399