

La cohabitation des langues et des écritures électroniques dans l'édition télématique

Jean-Michel Borde

DANS **ÉTUDES ET DOCUMENTS BERBÈRES** 2008/1 N° 27 , PAGES 185 À 191
ÉDITIONS **LA BOITE À DOCUMENTS**

ISSN 0295-5245

DOI 10.3917/edb.027.0185

Date de mise en ligne : 02/02/2020

Article disponible en ligne à l'adresse

<https://shs.cairn.info/revue-etudes-et-documents-berberes-2008-1-page-185?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour La Boite à Documents.

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur [cairn.info/copyright](https://shs.cairn.info/copyright).

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

La cohabitation des langues et des écritures électroniques dans l'édition télématique¹

L'enjeu

L'ouverture des systèmes informatiques se traduit par une explosion du volume de connaissance échangée et partagée. Des obstacles de poids viennent pondérer la perspective a priori favorable à un développement autant durable que global que l'on pourrait attendre du sens de ces échanges. Parmi ceux-ci la contrainte de l'internationalisation multiculturelle se dresse comme une barrière, qui constitue dans bien des cas un obstacle incontournable. Sauf à y consacrer de singuliers efforts, nous sommes encore loin de donner une garantie équitable de citoyenneté mondiale de ce point de vue particulier. Le phénomène, peu ou prou identifié a donné naissance au concept de fracture numérique. Celui-ci se manifeste partout de manière équitable et synchrone, au gré des générations ou des localisations².

Cette contrainte socio économique se double d'une deuxième évolution, qui en majore les effets, il s'agit des métas descriptions, intermédiaires indispensables au commerce³ international. Il s'agit des métadonnées autour desquelles sont en train de s'agréger de manière durable les paramètres de description des objets échangeables⁴. Avant d'en poser les principes d'évolution en cours, nous souhaitons dans les lignes qui suivent revisiter les fondations de l'édifice pour observer que, là non plus, les travaux sont encore loin d'être terminés.

Les caractères

On parle de familles de caractères⁵ lorsqu'ils sont regroupés pour écrire dans une langue donnée. On distingue, par exemple, les familles de caractères latins des cyrilliques, pourtant ces deux familles vont permettre d'écrire les mots de plusieurs langues chacune⁶. Le latin permet d'exprimer la plupart des langues d'Europe occidentale par écrit, alors que le cyrillique concerne plus de soixante langues différentes.

L'anglais même, par exemple, emploie les caractères de l'écriture latine.

Des variantes à ne pas négliger :

Les diacritiques

Entre autres choses, au plan de l'écriture l'anglais et le français pourtant tout deux utilisateurs du Latin se distinguent par l'emploi des diacritiques, accents et autres accessoires pour le français. Ainsi,

1. Le terme Télématique est pris ici comme générique de l'échange électronique hâtivement qualifié d'Internet. Nous rappelons que l'Internet n'est en fait qu'une famille de protocoles d'échanges ouverts et standardisés.

2. Jeunes et vieux, nord et sud.

3. Le mot est pris ici au sens littéral et générique de la relation humaine interpersonnelle.

4. Matériel ou conceptuel.

5. Attention à ne pas confondre les notions de caractère, de glyphe qui en est la représentation dessinée, comme de police qui qualifie une variante de représentation graphique, plus ou moins stylisée. Le caractère est avant tout le nom d'une lettre dans une écriture et dans une langue donnée. Ces derniers peuvent avoir eux-mêmes des variantes linguistiques, comme les noms de langue qui varient internationalement, selon la langue d'expression. Les accords de concordance sont « plus ou moins établis », c'est en cet parfois difficile à faire. Pensons aux caractères liés Ae ou Œ.

6. Cette communauté de problème est ressentie au point de formaliser la question au plan diplomatique. Il existe ainsi une « Union Latine » qui est une organisation représentant les intérêts des communautés linguistiques lusophones et hispanophones.

nous le verrons, des précautions d'emploi devront être prises dans la réalité des écritures interopérables pour procéder à des échanges garantis et transparents.

Les symboles

Ce sont des caractères translinguistiques qui vont permettre d'illustrer les écritures avec des représentations de grandeurs par exemple pour les poids et mesures. On s'aperçoit cependant que la représentation symbolique n'a pas forcément été conduite très en profondeur et que des variantes, là aussi, peuvent causer des troubles d'emploi, nombreux et parfois difficilement solubles.

Les sigles, acronymes

Nécessaires raccourcis dans la pratique quotidienne, ils compliquent en retour la perception du sens. UER = EBU soit Union européenne des radios-télévisions devient European Broadcasters union...

On gagne ainsi apparemment ce que l'on perd d'une autre manière, parfois en induisant de la confusion. Ceci peut prendre des proportions non négligeables dans l'indexation des termes d'un lexique multilingue, dans le cas ; comme ici, des dénominations composées.

Dans certaines langues des acronymes peuvent devenir des mots courants troublant d'autant le traducteur, qui risque de chercher vainement le sens de quelques uns.

En bref, on ne saurait ainsi limiter l'expression de l'écriture en français à l'usage des 26 lettres d'un alphabet latin. La représentation de l'écriture mobilise deux caractères pour représenter la même lettre, en minuscule et en majuscule. De plus, le recours à une ponctuation simple, (.,;:!?{}~) nécessite d'autres caractères additionnels. Au total on s'achemine vers un constat sensiblement alourdi, qui demande de vastes connaissances, autant culturelles que techniques, pour être correctement maîtrisé. De cette maîtrise correcte dépend une internationalisation équitable, respectueuse des cultures et pratiques, notamment dans le cas de certaines variantes.

Les Claviers

Pour manipuler les caractères, la machine à écrire fut une invention très pratique de 1873. On note, parce qu'on tend à l'oublier, qu'Émile Baudot inventa l'année qui suivit, en 1874, le téléscripteur avec l'ambition de transmettre les caractères saisis que l'on appela ensuite le Télex. L'écriture internationale télex fit l'objet, en son temps, d'une normalisation, dite ITA2⁷, qui est l'ancêtre historique de la normalisation des caractères des écritures du monde.

Pendant 100 ans ce concept ne fut pas significativement remis en cause. Les premières mises en application de l'informatique commerciale⁸ furent l'objet de développements propriétaires, donc antagonistes entre elles, des représentations écrites, ce qui en interdisait l'interopérabilité de façon simple. C'est à cette époque que l'IBM impose l'EBCDIC⁹ qui conduira à une réflexion américaine pour maîtriser une forme publique de représentation, qui sera nommée ASCII¹⁰. En 1980, le Pc qui arriva et rendit quasi instantanément la communication interpersonnelle internationale, sans limite et donc mondiale. Cette machine à communiquer nouvelle se heurte très tôt aux limites de l'internationalisa-

7. International Telegraph Alphabet No 2

8. Dans les années soixante.

9. *Extended Binary Coded Decimal Interchange Code* - Mode de codage des caractères sur 8 bits créé par IBM à l'époque des cartes perforées. Il existe au moins 6 versions différentes bien documentées (et de nombreuses variantes parfois créées par des concurrents d'IBM), incompatibles entre elles.

10. *American Standard Code for Information Interchange*, a donné ensuite lieu à la normalisation ISO 646, une norme ISO depuis 1972. « Alphabet international de référence n° 5 » (AIR5), « Alphabet international n° 5 » (AI5) ou encore (en anglais) *International Alphabet No. 5* » (IA5).

tion. Il lui faut supporter des codages, dits étendus, passant des 7 éléments de représentations, soit 170¹¹ emplacements disponibles dans les tables¹², à 8 éléments pour disposer d'un peu plus de marges, soit 256 emplacements dans les tables. Cette ouverture leur permet, enfin, de supporter une représentation presque comparable à celle de l'imprimerie courante, utilisée pour la presse où les livres. A cette époque, les imprimeurs résolvaient localement et manuellement leurs exceptions de représentations individuelles. Ce n'était pas vraiment un problème car les délais étaient couramment plus longs, et, surtout, il n'était pas question de réutiliser les saisies, qui étaient réalisées à fonds perdus. Ce n'est que plus tard, que ce que l'on appelle le prépresse¹³, par la PAO, va devenir totalement numérique avec le souci d'apporter de nouveaux services éditoriaux assortis de garanties nouvelles, accessibilité multimodale et accès aux archives en ligne, indexation globale de fonds pour faciliter les recherches¹⁴ etc.

Cette logique d'extension de type hypermédia diachronique a accompagné une logique d'extension multimédia synchrone assurant la banalisation des représentations de toute la connaissance, texte et illustrations comprises.

La famille AT 102 put se déployer en plusieurs dizaines de versions scripturales mais la plupart du temps sur une trame QWERTY. Seul le français de France restera attaché à la version dite AZERTY des claviers.

Les autres claviers

Il s'agit des variantes des autres types de micro ordinateurs dans les quelles nous ne rentrerons pas ici. Nous pensons plutôt aux claviers des autres machines de plus en plus omniprésents dans notre entourage. A commencer par celui, numérique, de notre téléphone portable pourtant largement utilisé pour transmettre des messages écrits avec le SMS. On imagine déjà avec cette simple prolifération de moyens et de méthodes des difficultés pour harmoniser des dispositions simples répondants aux problèmes courants de tout un chacun.

Le stylo numérique

Une alternative ?

Il y a quelques mois, une innovation est apparue, apparemment prometteuse qui offre la possibilité de nous passer d'ordinateur en permanence et, surtout de s'affranchir de l'usage du clavier et donc de son apprentissage ! Le stylo intégrant un enregistreur de mouvement pouvant restituer à une machine d'accueil USB, la séquence de mouvement et de déplacements de la pointe pour le transcrire en signe de type caractères ou dessin. Cette solution n'est qu'apparemment séduisante car dans la réalité, il s'agit d'un système un peu plus lourd. Il faut en effet adjoindre un support micro repéré qui est indispensable pour que les mouvements du stylo soient enregistrés selon les références d'un repère précis.

Les aides à la traduction

La généralisation de l'échange de textes électroniques puis la banalisation de leurs représentations codées ouvrirent la facilité à l'approche de la traduction automatique. Le texte codé devient un

11. Valeurs globales incluant ce que l'on appelle les caractères de commande.

12. On parle de table, pour désigner le tableau de classement des caractères qui les associe à un code de représentation hexadécimal ainsi qu'à une désignation univoque, normalisée.

13. Désigne tout ce qui se passe en amont de la presse. Cela va de la saisie de textes à la mise en pages (la publication assistée par ordinateur) (ou PAO) en passant par la photogravure, la numérisation des photographies, l'Imposition (bien que celle-ci soit de plus en plus souvent effectuée à l'imprimerie même) et la sortie de films ou de plaques.

14. Les systèmes de composition avancés utilisent souvent des extensions propriétaires ou idiosyncratiques au-delà des standards ASCII et ISO 8859, ou utilisent plutôt Unicode.

médiateur de la connaissance internationale. Pour cela il faut que certains outils évoluent dans la même direction pour permettre de raffiner les interprétations de sens alternatives qui émaillent fréquemment ce genre de travail. Parmi ceux-ci : dictionnaires, lexiques, thésaurus, vocabulaires, glossaires, répertoire, jargon, grammatologie...

Dictée vocale

Une nouvelle interface de commande pour les machines

L'énonciation des phonèmes par reconnaissance dans les textes concaténés est une technique bien maîtrisée. Elle a permis la mise au point de nombreux services, comme les annonces publiques dans les gares ou les aéroports. La dictée correspond à la technique inverse, il s'agit cette fois de faire reconnaître les phonèmes vocalisés par la machine. Cet exercice est plus compliqué car aléatoire et sensible aux variantes des locuteurs : accents, débits, ponctuation etc.

Encore aujourd'hui la dictée vocale, pour fonctionner, requiert une phase d'apprentissage minimale de la machine pour conduire correctement et rapidement le dialogue. Au fil des années, une seule compagnie a fini par monopoliser un marché de toute façon limité, pour l'instant. Nous sommes encore loin de donner des instructions de cette façon, sauf dans le cas de certaines applications de serveurs téléphoniques vocaux. Peut-être que l'augmentation de la miniaturisation et du nomadisme nous y conduiront avec obligation.

La correction orthographique

Vaste débat aussi vieux que celui de l'orthographe lui-même !!! Là encore la variabilité, fonction de l'actualité influencée par la mode est très difficile à suivre, des mots nouveaux entrent dans notre vocabulaire tous les jours !

Langue française

Ce n'est certes qu'une langue parmi d'autres, avec cependant quelques spécificités liées à son rôle historique d'établissement de certains droits internationaux. Parce qu'elle langue de référence, la terminologie établie en français est une référence reconnue dans de nombreux domaines de l'activité.

En ce sens, la langue française s'est posée (opposée ?) comme partenaire de l'expression anglophone, permettant d'éviter la désastreuse menace d'une domination mono linguistique.

On conçoit dans le monde tel que nous le connaissons aujourd'hui que ce rôle est essentiel au maintien de certains équilibres socio-économiques. La terminologie est essentielle au support des actes du commerce parce qu'elle définit les produits par leurs noms, leurs marques, leurs caractéristiques et performances comme les outils de distribution et de consommation.

Définitions *Source WikiPédia*

Un **glyphe** est une représentation graphique (parmi une infinité possible) d'un signe typographique, autrement dit d'un caractère (glyphe de caractère) ou d'un accent (glyphe d'accent). Un caractère particulier peut ainsi être créé en ajoutant un glyphe d'accent à un glyphe de caractère. Les logiciels informatiques ont accès au dessin de ce glyphe par l'intermédiaire d'une police de caractères : le tracé du glyphe y est le plus souvent défini par un ensemble de points ou de courbes de Bézier.

Le **caractère** (d'imprimerie), à l'origine, en typographie, était une pièce métallique ayant le dessin en miroir d'une lettre ou d'un signe un relief, permettant l'impression. Les caractères étaient rangés dans une boîte appelée casse. Pour composer un texte, on les prenait et les plaçait sur un composeur qui permettait de les aligner, et notamment de respecter la largeur définie pour le texte

(justification). Une fois les lignes composées, elles étaient transférées sur le cadre pour former la forme (matrice d'impression).

Chaque signe imprimable est donc en plusieurs exemplaires dans la casse. On a plusieurs caractères – pièces métalliques – pour un signe imprimé donné, et même plusieurs jeux de caractères (romain, italique, gras,...). On parlait aussi de **caractère mobile**, par opposition à l'impression par une plaque gravée dans la masse (comme en xylographie). De manière générale, Un caractère représente une variante d'un signe typographique destiné à l'impression. Les caractères sont regroupés en fonte de caractères dont ils partagent les caractéristiques (corps, empattements, graisse, etc.). On distingue les caractères du *haut de casse* (capitales ou majuscules), rangés en haut de la caisse », des caractères dits *bas-de-casse* (minuscules). Usuellement, en imprimerie on parle de capitales (du latin *caput, capitis* la tête) et bas-de-casse. Avec l'arrivée de l'informatique on distingue le concept de glyphe, qui est une représentation graphique d'un signe typographique ou caractère. Le caractère est donc un archétype qui se décline selon la police, par exemple, pour le caractère « A » (a capitale), on a les glyphes

A, A, **A**, **A**, A, A, **A**, **A**, A, A, **A**, **A**,...

Chaque caractère informatique est référencé par un nombre, la définition de la représentation binaire des caractères constituant un codage de caractères.

Par abus de langage, on appelle souvent ASCII » des normes qui étendent ASCII, mais qui ne sont pas compatibles entre elles. En particulier, les standards Windows-1252 (couramment utilisé sur Microsoft Windows dans les pays occidentaux), ISO 8859-1 (couramment utilisé sur Internet et Unix) et la page de code 850 (couramment utilisée sur DOS) ne sont pas la norme ASCII. Cet abus de langage ne va pas sans causer des confusions causant des incompatibilités, souvent rendues visibles par le fait que les caractères non ASCII comme les lettres accentuées » (èÉç) s'affichent mal. On écrit parfois **ASCII de base** pour bien identifier ASCII, et pas un standard plus étendu.

ISO/CEI 8859 est constitué à ce jour des parties suivantes :

- ISO 8859-1 (*latin-1 ou européen occidental*) – probablement la partie la plus largement utilisée de ISO 8859, couvrant la plupart des langues européennes occidentales : l'allemand, l'anglais, le basque, le catalan, le danois, l'écosse, l'espagnol, le féringien, le finnois (partiellement), le français (partiellement), l'islandais, l'irlandais, l'italien, le néerlandais (partiellement), le norvégien, le portugais, le rhéto-roman et le suédois, certaines langues européennes sud-orientales (l'albanais), ainsi que des langues africaines (l'afrikaans et le swahili). Le symbole de l'euro et la capitale Ī, qui manquaient, sont dans la version révisée ISO 8859-15 (latin-9). Le jeu de caractères correspondant ISO-8859-1, approuvé par l'IANA, est le codage par défaut des anciens documents HTML ou des documents transmis par messages MIME, tels que les réponses HTTP quand le type de média du document est text » (par exemple les documents text/html »).

- ISO 8859-2 (*latin-2 ou européen central*) – prend en charge celles des langues d'Europe centrale ou de l'Est basées sur un alphabet romain. Ceci inclut le bosniaque, le croate, le polonais, le tchèque, le slovaque, le slovène et le hongrois. Le symbole de l'euro manquant est présent dans la version ISO 8859-16.

- ISO 8859-3 (*latin-3 ou européen du Sud*) – le turc, le maltais, et l'espéranto ; largement supplanté par ISO 8859-9 pour le turc, et par Unicode pour l'espéranto.

- ISO 8859-4 (*latin-4 ou européen du Nord*) – l'estonien, le letton, le lituanien, le groenlandais, et le sami.

- ISO 8859-5 (*cyrillique*) – Couvre la plupart des langues slaves utilisant un alphabet cyrillique, y compris le biélorusse, le bulgare, le macédonien, le russe, le serbe et l'ukrainien (partiellement).

- ISO 8859-6 (*arabe*) – Couvre les caractères les plus courants de l'arabe. Ne prend pas en charge

d'autres langues à alphabet arabe. Nécessite un moteur de rendu qui prend en charge l'affichage bidirectionnel et l'analyse contextuelle.

- ISO 8859-7 (*grec*) – Couvre la langue grecque moderne (orthographe monotonique). Peut être utilisé aussi pour le grec ancien écrit sans accents ou dans l'orthographe monotonique, mais il manque les signes diacritiques pour l'orthographe polytonique. Ceux-ci ont été introduits avec Unicode.

- ISO 8859-8 (*hébreu*) – Couvre l'alphabet hébraïque moderne tel qu'il est utilisé en Israël. En pratique, deux codes différents existent : ordre logique (nécessite un moteur de rendu bidirectionnel pour l'affichage) et ordre visuel (gauche à droite).

- ISO 8859-9 (*latin-5* ou *turc*) – Grosso modo le même que l'ISO 8859-1, où les lettres islandaises peu utilisées sont remplacées par des lettres turques. Il est aussi utilisé pour le kurde.

- ISO 8859-10 (*latin-6* ou *nordique*) – Un réarrangement du latin-4. Considéré plus utile pour les langues nordiques. Les langues baltes utilisent plus souvent le latin-4.

- ISO 8859-11 (*thaï*) – Contient la plupart des glyphes requis pour la langue thaï.

- ISO 8859-12 – Était supposé couvrir l'alphabet devanāgarī, mais ce projet a été abandonné en 1997. ASCII et Unicode ISO/CEI 10646 couvrent le devanāgarī.

- ISO 8859-13 (*latin-7* ou *balte*) – Ajoute quelques caractères supplémentaires pour les langues baltes qui manquaient en latin-4 et latin-6.

- ISO 8859-14 (*latin-8* ou *celtique*) – Couvre des langues celtiques telles que l'irlandais (orthographe traditionnelle), le gaélique écossais, le mannois (langue disparue) et le breton (certaines anciennes orthographes).

- ISO 8859-15 (*latin-9* ou parfois de façon impropre latin-0) – une révision de 8859-1 qui abandonne quelques symboles peu utilisés, les remplaçant avec le symbole de l'euro et les lettres Š, š, Ž, ž, Œ, œ, et Ÿ, ce qui complète la couverture du français, du finnois et de l'estonien.

- ISO 8859-16 (*latin-10* ou *européen du Sud-Est*) – Prévu pour l'albanais, le croate, le hongrois, l'italien, le polonais, le roumain et le slovène, mais aussi le finnois, le français, l'allemand et l'irlandais (en nouvelle orthographe). Cette police mise plus sur les lettres que les symboles. Le signe de monnaie est remplacé par le symbole de l'Euro.

Webographie sommaire

normalisation	http://www.iso.ch http://www.iso.org/iso/fr/support/faqs/faqs_widely_used_standards.htm http://www.itu.int
Prépresse	http://generationig.free.fr
Imprimerie	http://fr.wikipedia.org/wiki/Imprimerie http://www.typographie.org/gutenberg/garamond/garamond_1.html
Télématique	http://fr.wikipedia.org/wiki/T%C3%A9l%C3%A9matique http://www.ofil.refer.org/tribune/n22/utilisat1.htm
Caractères	http://fr.wikipedia.org/wiki/Caract%C3%A8re_(typographie) http://fr.wikipedia.org/wiki/Lettre_(typographie) http://www.table-ascii.com/
Langues	http://fr.wikipedia.org/wiki/ISO_639 http://publications.europa.eu/code/fr/fr-370200.htm

- Pays** http://www.iso.org/iso/fr/country_codes.htm
http://www.iso.org/iso/fr/country_codes/iso_3166_code_lists.htm
- Applications** <http://www.issn.org/>
<http://www.isan.org/>

Ouvrages et articles

- Simon Nora, Alain Minc, *L'informatisation de la Société*, La Documentation Française, Paris, 1978 (ISBN 2020049740)
- Bernard Marti, *Télématique, techniques, normes, services*, éd. Dunod, Paris, 1990
- Publications de l'IREM de Rennes en histoire et épistémologie des mathématiques Faire des mathématiques à travers leur histoire T4 Caractères d'imprimerie et courbes de Bézier Loïc Le Corre – 2002
- Minitel : histoire du réseau télématique français Antonio Gonzalez Emmanuelle Jouve FLUX Cahiers scientifiques internationaux Réseaux et territoires International Scientific Quarterly on Networks and Territories ISSN 1154-2721 version imprimée ISSN 1958-9557 version électronique Revue publiée avec le concours du CNRS, de l'ENPC et de l'Institut de la Gestion Déléguée
- ISO/IEC 8859-6 :1999 Technologies de l'information – Jeux de caractères graphiques codés sur un seul octet – Partie 6 : Alphabet latin/arabe
- ISO/IEC 10646:2003 Technologies de l'information – Jeu universel de caractères codés sur plusieurs octets (JUC)

Jean-Michel BORDE

AILF Informatique et langue française