L'appropriation des technologies multiscripturales

Nouveau système technique, nouveaux usages, vers une culture planétaire translinguistique

Henri Hudrisier

Université de Paris 8, Département Documentation
Maison des Sciences de l'Homme Paris Nord, 4 rue de la Croix Faron
F-93210 La Plaine Saint-Denis
h.hudrisier@libertysurf.fr

RÉSUMÉ. *Unicode existe, mais les usages que nous pourrions en attendre émergent à peine. Étudier l'appropriation d'une technique impose de distinguer deux niveaux d'usage, celui des professionnels de l'information et de la communication, au-delà la création d'une multitude d'usages dans toute la largeur du spectre social (professionnels, académiques, culturels ou grands publics). Les systèmes techniques, notion empruntée aux historiens des techniques constitue un concept très opératoire pour comprendre les générations techniques successives de la technoculture et nombre de mécanismes de l'appropriation des technologies. Cependant la multiscripturalité (liée à la langue maternelle et à l'écriture d'usage) plonge au plus profond de l'inconscient collectif des sociétés et des individus. La grammatologie nous aide à penser le développement d'une technoculture adaptée à un futur de la multiscripturalité. On conclura en proposant quelques scénarios prospectifs.*

ABSTRACT. *Unicode is now a reality, but its complete uses are just coming out. To study the appropriation of a technology requires distinguishing between two levels of usage : that of computer and communication professionals, and that of a larger social spectrum (businesses, academia, cultural institutions, the general public). Technological systems, a notion borrowed from historians on technology, provides a concept which is extremely helpful in understanding successive generations of technology and technocultures and studying technological appropriation. Meanwhile, multiscripturality (linked to our native languages and writing systems) goes to the heart of our societal and individual collective unconsciousness. Grammarology helps us comprehend the development of a technoculture favorizing the future of multi-writing. In conclusion, several prospective scenarios are proposed.*

MOTS-CLÉS : *appropriation technique, écriture, langue, typographie, Unicode, ISO/CEI 10 646, ASCII, ingénierie linguistique, informatique, multilinguisme, multiscripturalité, technoculture, systèmes techniques, grammatologie, prospective, histoire technique, idéogramme, alphabet.*

KEYWORDS: *technical appropriation, writing systems, language, typography, Unicode, ISO/CEI 10 646, ASCII, language engineering, computer systems, multilingualistic, multiwriting, technoculture, technical systems, grammarology, prospective, history of technologies, ideogram, alphabet.*

« Un système d'écriture composé de vingt à quarante lettres peut à la rigueur être remplacé par un autre. Il en serait de même pour la langue si elle renfermait un nombre limité d'éléments; mais les signes linguistiques sont innombrables. »

Saussure (Saussure, 1916)

« Chaque trait horizontal est une masse de nuages en formation guerrière, chaque crochet un arc bandé d'une force rare, chaque point un rocher tombant d'un sommet élevé, chaque bec un crochet de cuivre, chaque prolongement de ligne un sarment vénérable et chaque trait libre et délié un couteau prêt à bondir. »

Wang-Hsichin, célèbre calligraphe chinois (321-379 ap. J.C.)

« Ce qui réside dans les machines, c'est de la réalité humaine, du geste humain fixé et cristallisé en structures qui fonctionnent. [...] Pour redonner à la culture le caractère véritablement général qu'elle a perdu, il faut pouvoir réintroduire en elle la conscience de la nature des machines, de leurs relations avec l'homme et des valeurs impliquées dans ces relations. »

Gilbert Simondon (Simondon, 1958)

La communication planétaire informatique et multilingue est désormais un fait avéré. L'internet pénètre dans les territoires apparemment les plus inaccessibles de la planète et on peut théoriquement consulter et réaliser des pages électroniques dans la plupart des langues et écritures du monde.

Cette culture planétaire multilinguistique est une nécessité du XXI^e siècle. Les industriels linguistiques des grandes langues dominantes commercialisent, d'année en année, des logiciels de plus en plus efficaces et à des prix abordables dans des secteurs qui avaient longtemps été considérés comme des enjeux mythiques : traduction automatique, dictée en flux continu, bases de données documentaires et agents de recherche intelligents. Ces progrès viennent à point dans une planète qui devient d'année en année plus petite, ce qui contraint de plus en plus les peuples à l'absolue nécessité de s'entendre tout en préservant leur diversité culturelle et linguistique.

Cependant ni l'évidence du progrès technique, ni l'indiscutable intérêt d'usage d'une technologie ne suffisent pour qu'une innovation se développe à une large échelle dans tous les domaines où elle pourrait être utile. C'est là toute la question de l'indispensable appropriation des technologies. Il devient dès lors utile d'en comprendre la logique pour pouvoir ensuite proposer des pistes de développement de cette appropriation multiscripturale.

1. Les modalités d'appropriation d'une technique

La plupart des théories sur l'appropriation et l'usage des techniques sont nées et ont été utilisées pour étudier en aval de leur industrialisation l'usage des produits de

large diffusion, entre autres ceux de la communication. Avec la mondialisation des réseaux, on assiste à un véritable court-circuit de la conception, du développement et de la diffusion des technologies informatiques. Les phases de conception et même de définition préindustrielle sont portées sur la place publique, augmentant le nombre des développeurs indépendants potentiels, ce qui brouille et modifie les stratégies marketing traditionnelles des produits informatiques et a pour conséquence de complexifier l'étude de l'appropriation. Autre conséquence de *L'âge de l'accès* (Rifkin, 2000) : l'accélération de la convergence multimédia augmente considérablement les exigences d'interopérabilité, ce qui a pour effet de porter au grand jour le débat jusqu'ici technique et très confidentiel de la normalisation des modèles et codes des systèmes d'information (Solaris, oct. 1999). Ces normes deviennent de ce fait une nouvelle catégorie très paradoxale d'objets d'information que des non spécialistes devraient s'approprier (Hudrisier, 2000).

Dans ces conditions, nous proposons pour étudier les modalités d'appropriation d'une technique de chercher à comprendre comment une innovation (ou un nouveau système technique), peut pénétrer dans les catégories professionnelles directement concernées pour y susciter le plus grand nombre possible de nouveaux usages, se structurer ensuite en tant que nouvelle activité, voire nouveau métier, et finalement induire une nouvelle niche de prospérité. Pour qu'il y ait véritablement appropriation sociale, il faut ensuite que cette appropriation de premier niveau permette à de très nombreuses autres professions et disciplines, appartenant à tout éventail des occupations humaines, d'augmenter, parfois considérablement, leur champ d'action grâce à ces nouveaux produits et services. Il s'agit là d'une appropriation de deuxième niveau qui justifie véritablement le premier niveau puisque c'est seulement lorsque se réalisent cette infinité de nouveaux usages induits sur toute la largeur du spectre des activités sociales que peuvent enfin s'amortir les investissements qui ont permis le développement d'activités induites chez les professionnels directement concernés.

Il apparaît donc que l'appropriation d'une nouvelle technique n'est pas, loin s'en faut, un phénomène simple. La population mondiale se structure en de très nombreux sous-ensembles non exclusifs les uns des autres qui se distinguent par leurs caractéristiques sociales, culturelles, linguistiques, nationales, professionnelles...

Dans un cas comme celui qui nous occupe, c'est le monde entier et toutes les cultures linguistiques qui sont touchés¹. Par ailleurs, chacune des facettes de la socioculture linguistique (nationale, professionnelle, artistique, académique, religieuse...) induiront évidemment des prises en compte différentes de la technique.

– Catégories linguistiques d'abord, toutes les langues ne sont pas touchées également par les situations transscripturales. Pour aller vite, les communautés

1. Même les cultures orales sont concernées si tant est que des ethnolinguistes s'intéressent à ces civilisations.

n'appartenant pas à l'écriture latine sont beaucoup plus concernées parce qu'elles sont perpétuellement confrontées à la nécessité de faire coexister plusieurs écritures.

– Catégories professionnelles bien sûr, les métiers de l'informatique de la communication sont évidemment sollicités pour concevoir déjà des ingénieries adaptées aux nouveaux besoins d'internationalisation, mais toutes les autres activités sociales peuvent être aussi concernées à des degrés divers tant pour les métiers que pour les disciplines académiques. Bien sûr le degré d'implication pour des professions et activités académiques dépendra de leur degré de sensibilité à la communication interlinguistique.

– Les métiers du commerce international, du tourisme, de l'édition, des musées et bibliothèques, les professionnels de la traduction et les enseignants de langue étrangère, mais aussi quantité d'autres activités comme le cinéma, la médecine, la police mais aussi par exemple les travailleurs sociaux, toutes ces activités seront peu ou prou touchées par la question de l'échange d'informations multiscriturales.

2. Système technique et histoire du progrès technique

Il est bien connu que le progrès technique n'évolue pas de façon continue ni selon la simple logique de la réponse à de nouveaux besoins qu'il pourrait satisfaire. Toute appropriation d'une innovation connaît plusieurs phases historiques, correspondant au passage par ruptures successives d'un *système technique global* à celui qui lui fait suite. Ces *générations technologiques* peuvent nous aider à comprendre la logique de l'appropriation.

On connaît le concept de *système technique* popularisé notamment par Bertrand Gille (Gille, 1978). La notion de système technique suppose que les différentes composantes d'un *ensemble* (ou système) *technique* soient homogènes et appartiennent à un même état d'avancement technologique, ce qui leur permet de fonctionner en harmonie. Si les historiens des sciences s'intéressent aux systèmes techniques, c'est parce qu'ils trouvent dans ce concept une logique de la périodisation du progrès technique. On constate, en effet, que le progrès technique avance par mutation d'un système technique moins sophistiqué et moins cohérent vers un nouveau système technique plus intégré et plus universel. Les historiens des techniques distinguent ainsi des époques successives dans lesquelles l'environnement technique et technoculturel forme système. Ces systèmes techniques globaux sont constitués selon le degré de granularité auquel on les observe par quantité de systèmes techniques plus circonscrits. Une des caractéristiques essentielles de l'avancée technologique dans cette théorie des systèmes techniques tient précisément à ce que l'ensemble fait système. En d'autres termes, il est fondamental de prendre en compte l'interdépendance de tous les paradigmes techniques, tant culturels que technologiques, d'un système ou d'un ensemble.

En fait, une innovation quelle qu'elle soit, ne pourra véritablement s'imposer qu'autant qu'une nouvelle théorie du nouveau système technique se sera fait jour. Cela permettra à tous les acteurs tant les techniciens que les usagers, mais aussi les entrepreneurs et leurs financiers, de comprendre les nouveaux avantages, les nouveaux usages donc les nouveaux marchés permettant de concevoir, de développer, d'industrialiser et d'utiliser cette innovation.

Il s'agit donc de recomposer une nouvelle technoculture. Cette théorie des ensembles techniques se succédant par générations successives, de par la nécessité préalable de leur appropriation technoculturelle autorisant leur développement, explique assez complètement l'évolution des innovations dans des domaines très techniques et industriels. L'exemple de l'ensemble technique des débuts du machinisme et de son paradigme fondamental que constitue la machine à vapeur est de notre point de vue très démonstratif et nous avons essayé de l'exposer en parallèle avec le sujet qui nous occupe. Mais, précisément dans le cas de la numérisation de l'écriture les interactions de la technique sont beaucoup plus complexes parce que l'écriture est elle-même technique sans que nous en ayons une conscience très claire parce que les langues, qui sont, elles aussi des techniques de communication sont comme l'écriture profondément ancrée au plus profond de notre inconscient individuel et collectif. C'est dire que nous touchons là à l'univers autant technique que culturel qui participe de la construction de nos liens sociaux et de notre condition humaine. Dès lors nous voyons avec plus de clarté la nature des enjeux et les difficultés d'analyse de cette appropriation du codage universel des écritures.

En complémentarité avec les hypothèses d'historien des techniques, nous devons aussi faire appel à la théorie *grammatologique* qui a précisément pour objectif de chercher à appréhender à la fois la technique, la sémiotique, l'histoire, la philosophie de l'information et l'analyse culturelle comparée de toutes les traces d'information et de communication (dont l'écriture de la langue est un cas particulier). Nous cherchons ainsi à pénétrer cette complexité par itérations successives. Il nous faudra poser le cadre d'un ensemble technique multiscriptural dont nous chercherons par référence à notre ignorance technique personnelle à préciser les points qui nous ont paru difficilement compréhensibles par le *grand public cultivé*. À partir de ce référentiel commun d'appropriation nous observerons plusieurs segments de cet ensemble technique, tant dans leur dimension d'appropriation endogène (l'appropriation technique mais aussi culturelle par ceux qui en sont les premiers aménageurs et bénéficiaires : le premier niveau) que dans une approche prospective pointilliste destinée à ouvrir des pistes d'usages beaucoup plus exogènes : les effets attendus de la mise en place d'une numérisation cohérente et globale de toutes les écritures et pratiques scripturales associées (musique, chorégraphie, notation mathématique) d'abord sur un premier cercle d'usage, celui de l'intercommunication linguistique... ensuite sur la société tout entière dans ses besoins d'intercompréhension planétaire quelques soient les catégories et occupations sociales (l'appropriation de deuxième niveau).

Remarquons enfin, pour préciser le rapport entre multilinguisme et multiscripturalité, que les enjeux d'une future culture mondiale multilinguistique passent certainement par le développement de compétences transcripturales, c'est-à-dire propres au fonctionnement harmonieux, cohérent et si possible synergique de toutes les écritures.

Le non spécialiste pourrait se demander : quel est le problème ? Où est le problème ? Des traitements textes existent dans toutes les langues du monde et internet permet d'échanger des pages d'information à travers toute la planète. Que vient faire Unicode ? Pourquoi un codage numérique commun à toutes les écritures du monde ? Supposons d'abord que ce non spécialiste ait déjà compris la nécessaire articulation du code informatique entre le bit et l'octet, en quelque sorte sa logique de double articulation du langage (Hudrisier, oct. 1999). Au-delà il prendra sans doute un certain temps pour comprendre que la difficulté n'a pas été d'informatiser chacune des langues du monde (quoique ce fut plus difficile pour certaines) mais de les faire interagir toutes ensemble dans un même univers. Nous touchons là précisément aux propriétés nécessaires d'un système technique multiscriptural.

Il faudra ensuite prendre conscience que la langue, mais aussi l'écriture, sont liées à des réflexes très archaïques ancrés dans les racines de nos appartenances sociétales et culturelles, qui font que les progrès en ces domaines sont souvent très lents. L'histoire de l'écriture revue à travers les concepts de la grammatologie peut nous aider à bâtir quelques scénarios prospectifs d'appropriation des technologies multiscripturales.

3. Nouveaux potentiels et situations d'usage

Tout nouveau système technique et les outils qui lui sont associés génèrent potentiellement de nouvelles logiques d'usage (Perriault, 1989) qui doivent, pour émerger s'insérer dans une nouvelle technoculture prenant en compte le nouveau système technique. Certains nouveaux usages apparaissent évidents et, dans ce cas, les progrès techniques sont attendus avec impatience et l'appropriation sociale n'est pas en question. Nous adoptons ainsi assez rapidement des progrès techniques facilitant des tâches très matérielles ainsi que des progrès thérapeutiques (tout au moins tant que nos sociétés sociolibérales acceptent de nous aider à en rétribuer l'usage). D'autres innovations, au contraire, ne sont aucunement attendues mais sont trouvées par hasard, par détournement d'outils techniques initialement prévus pour autre chose et elles apportent ainsi, naturellement, une réponse à un besoin d'usage imprévu, mais opportun, même s'il n'a pas été repéré auparavant. Enfin, et c'est très souvent le cas, d'autres dispositifs techniques, d'autres systèmes techniques émergent lentement et difficilement. Cette dernière situation est caractéristique des systèmes techniques liés à des enjeux d'usage souvent considérables mais difficiles à mettre en évidence. Il est alors nécessaire, pour que se développent ces innovations, que puisse se légitimer un nouveau contexte technoculturel qui rentre

dès lors en concurrence et finit par remplacer l'ancien contexte technoculturel qui répondait au système technique antérieur et aux anciennes logiques d'usage. Ainsi, pour que puissent émerger de nouvelles logiques d'usage ouvertes par de nouveaux systèmes techniques, il faut que s'engage une difficile dialectique entre l'appropriation technique, l'économie et la culture et il est rare que ces évolutions puissent faire l'économie de la lente succession des générations technologiques mais aussi humaines².

Le cas de figure qui nous occupe est très paradoxal. Le progrès technique devrait être attendu avec impatience mais les *habitus* liés à l'écriture sont ancrés au plus profond de nos réflexes inconscients et nous interdisent sans doute d'admettre que sur le plan informatique nous ne pouvons plus continuer de fonctionner avec le seul catalogue des caractères de notre langue (l'alphabet français par exemple) mais avec un catalogue commun élargi à l'ensemble des langues. Certes, nous sommes perpétuellement agacés par l'absence de lettre grecque sur le clavier, par les mille et un inconvénients de l'échange de courrier avec des territoires d'écriture étrangère au notre, mais sommes-nous prêts à voir notre espace quotidien d'écriture perdre la marque tangible de notre écriture d'usage³ ? Je crois qu'il faut d'abord examiner les questions posées par l'appropriation concrète de ces nouveaux possibles, ce qui impose aussi d'examiner le contexte technoculturel dont dépend le développement potentiel de la communication panscripturale pouvant déclencher une nouvelle offre technique et une prospérité économique de l'échange interscriptural. Pour évaluer nos capacités de progrès en la matière, nous nous devons d'examiner plusieurs facettes qui sont simultanément indispensables mais que je ne peux qu'exposer successivement :

- comment fonctionne notre culture technique et quels sont les systèmes techniques successifs qui caractérisent ce point particulier de l'informatique linguistique ?

- comment cette culture techno-informatique, et plus particulièrement les systèmes techniques particuliers que constituent l'ASCII et maintenant Unicode et le codage des caractères sur 4 octets, interagissent avec les cultures d'écriture (qui, ne l'oublions pas, sont aussi des techniques) ?

- comment peut-on accéder à une nouvelle technoculture universelle d'écriture qui susciterait des usages intéressants et élargis ?

2. Dans certains cas ce sont des générations démographiques, dans d'autres cas il suffit d'attendre que se fassent jour de nouvelles promotions de professionnels ou de nouvelles catégories d'usagers éduqués selon les nouveaux principes techniques.

3. C'est bien d'écriture d'usage et non pas maternelle qu'il s'agit. On peut constater, par exemple, que la plupart des africains de langue *bantoue* (par exemple) mais écrivant en français ou en anglais ne semblent pas désirer outre mesure un système permettant d'écrire tous les diacritiques de ces langues (voir l'article de Sylvie Baste et celui de Andrei Popescu-Belis).

Dans chacun de ces thèmes, j'essaierai le plus souvent de situer les enjeux occidentaux (et plus précisément ceux des usagers des écritures latines) face à d'autres sociétés différemment impliquées dans un futur multiscritural. En effet, les situations plurilinguistiques mais aussi pluriscriturales correspondent à des situations d'échanges très inégales.

Le multilinguisme et la multiscrituralité sont-ils des handicaps ou au contraire une richesse et une chance données à telle ou telle communauté ou territoire ? La grande langue dominante et l'évidente simplicité du système d'écriture latine non accentuée continueront-ils de favoriser la seule communauté linguistique anglophone avec tous les risques de dépendance sociale et culturelle que cela implique, ou, au contraire, se dessine-t-il une nouvelle donne planétaire ?

4. Le codage à multiples octets de l'écriture comme nouveau système technique

L'ASCII, que le grand public considère encore très fréquemment comme la composante fondamentale de l'informatique alphanumérique, constitue le paradigme central d'un système technique ancien (celui de l'informatique à 7, puis 8 bits). Nous savons, bien sûr, qu'il s'agit là d'un abus de langage confirmé par l'usage. Le véritable ASCII et la norme ISO/CEI 646 qui lui succède (dont on perçoit l'évidence de généricité du sigle avec la norme ISO/CEI 10 646) sont codés sur 7 bits, ce qui ne permet pas de conserver les accents de l'écriture latine et cantonne ce système de codage à ne pouvoir coder que des textes en anglais ou en latin (sauf mécanisme complexe) ! Cependant, nombreux sont ceux qui ont continué à désigner sous le terme d'ASCII, ou parfois sous le terme moins ambigu d'*ASCII étendu*, la famille de codage à 8 bits ISO/CEI 8859-n. Cette famille de norme de codage se décline sur une grosse dizaine de langues (ou plutôt d'écritures) à partir de la plus connue d'entre elles, celle qui est utilisée en Europe occidentale ou en Amérique, ISO/CEI 8859-1, connue aussi sous le nom de *ISO latin* ou plutôt d'*ISO latin 1*⁴. Un mécanisme d'échappement permet dans ce codage à 8 bits, de passer d'une de ces normes à une autre. Cet ensemble de normes est devenu aujourd'hui très compliqué. Les risques d'ambiguïté sont nombreux puisque ce sont les mêmes positions de codage numérique qui sont utilisées pour des caractères complètement différents dans des écritures qui peuvent être aussi éloignées que l'hébreu, le latin, l'arabe ou le thaï. Cette complication même ainsi que l'impossibilité sur cette famille de norme de codifier des langues idéographiques sont les causes principales de la nécessité de passer à un système de codage fondé sur une *base de numérisation* beaucoup plus large : 4 octets.

Ainsi, parallèlement à l'obsolescence de ce codage alphanumérique à 7 ou 8 bits, émerge un nouveau système technique de prise en compte numérique de l'écriture,

4. Ce qui a le mérite d'être plus précis puisque certaines langues d'écriture latine doivent recourir à d'autres variantes de codage de ISO/CEI 8859 (Europe du Nord, de l'Est, Turquie... sans évoquer des situations plus complexes en Afrique et surtout au Vietnam).

beaucoup plus large, construit selon les principes de l'information structurée, des langages à balises, de la convergence des médias et pour lequel la codification de l'écriture sur deux puis quatre octets ne constitue, en fait, qu'un sous-ensemble technique. Cette réalité est particulièrement mise en évidence lorsqu'on constate que XML exige, pour fonctionner, que les modes de codage de l'écriture soient parfaitement définis, explicitement déclarés et soient si possible universels⁵. Il va aussi de soi que, face à l'augmentation générale des performances des machines informatiques (qui *tournent* maintenant en général sur 16 ou 32 bits), le maintien d'un système de représentation des caractères sur une norme à 1 octet ne se justifie plus et devient même contre-productif.

C'est ainsi que les nouvelles générations de systèmes d'exploitation et de bureautique intègrent maintenant de façon standard Unicode qui est définitivement adopté par beaucoup de constructeurs. En toute logique, ce nouveau sous-système technique de représentation de l'écriture⁶ devrait générer son environnement technique d'usage, des logiciels d'application et de typographie adaptés comme des polices universelles de caractères, des *moteurs de rendu d'écriture* eux aussi universels, des interfaces de saisie qui devraient être des claviers virtuels, mais aussi des environnements de traitement de texte permettant aisément l'affichage de textes multilingues, sans omettre de multiples applications et outils dépendant de l'informatique qui devraient pouvoir devenir multiécritures comme les tableaux d'affichage des aéroports, par exemple.

C'est d'ailleurs un classique de l'histoire des techniques. Par définition, un nouveau système technique succède par mutation au système technique qui le précède : ici Unicode supprime l'ASCII. Les industriels qui interviennent au cœur du processus technique (en général les multinationales) ont évidemment intérêt à ce que le parc de matériel se renouvelle. Ces mêmes entreprises connaissent parfaitement les raisons techniques du changement en cours. Elles ont collaboré au processus de normalisation et elles abandonnent aussi vite qu'elles le peuvent l'ancien système, surtout (comme c'est le cas) si elles peuvent livrer un ordinateur fondé sur Unicode tout en continuant à faire *tourner* l'ASCII.

5. Similarités entre les systèmes techniques successifs de la machine à vapeur et ceux des technologies multiscripturales

Cependant, même dans la communauté industrielle, le passage d'un système technique aussi fondamental que l'ASCII à un système technique de la codification des caractères sur 2 ou 4 octets n'est pas si simple.

5. On sait qu'en XML Unicode est implicite sauf déclaration contraire explicite.

6. La représentation informatique de l'écriture constitue, en effet, en soi un système technique mais on comprend que celui-ci s'insère dans un système technique plus vaste, celui de l'informatique puis, au-delà, dans le système technique plus vaste de l'information vidéo numérique, des réseaux, de la robotique, etc.

Dans le cas d'un progrès technique matériel⁷ comme celui, par exemple, du machinisme et de son paradigme fondamental la machine à vapeur, le passage d'un système technique à un autre est souvent long, complexe et nécessite des ajustements itératifs entre l'idéalité du progrès proposé par la technique et la science et son appropriation effective. À l'évidence, une machine à vapeur que l'on aurait installée sans discernement pour faire tourner un ancien *moulin à foulon* aurait posé une multitude de problèmes. D'abord, celui de l'adéquation des mécanismes de l'ancienne fabrique souvent en bois ou de toute manière mal adaptés aux nouvelles cadences de la machine. En quelques heures, ces anciens mécanismes auraient été brisés. De ce point de vue, l'analogie est criante quand on voit le contexte majoritairement à 8 bits et fondé sur l'ASCII qui caractérise encore nombre de logiciels et d'environnements informatiques. Tant que cet environnement à 8 bits subsistera, les nouveaux usages auront du mal à émerger, parce que même si des mécanismes d'adaptation ont été prévus (UTF8) un environnement aussi ancien et mal dimensionné provoquera inmanquablement d'innombrables dysfonctionnements.

Observons les similitudes que l'on constate dans la mise en cohérence du système technique machiniste, d'une part et du système technique informatique et plus largement numérique, d'autre part. Dans les deux cas une des grandes difficultés consiste à accorder sur un même niveau de progrès technique un environnement technique mais aussi culturel avec les paradigmes-clés dudit système technique.

Dans le cas du machinisme on perçoit bien la machine à vapeur comme paradigme central. Autour d'elle, il a fallu adapter très vite tout ce qui concernait son nécessaire approvisionnement en énergie. Tout cela a nécessité une adéquation du nouveau système technique aux infrastructures de transport, d'extraction du charbon ou même de forestage du bois de chauffe. Mais il a fallu aussi développer de nouvelles méthodes de fabrication métallurgique indispensables pour construire les machines à vapeur mais aussi tout l'environnement des machines les plus diverses entraîné par cette nouvelle source d'énergie. Dans le même temps qu'il a fallu résoudre ces questions, il était nécessaire de trouver des solutions pour acheminer les matières premières et expédier sur une beaucoup plus large échelle des produits industriels bien plus abondants.

Dans le cas du système technique de communication numérique et planétaire, on perçoit bien que, comme la machine à vapeur, la numérisation de l'information scripturale constitue un des paradigmes fondamentaux⁸. Comme dans le cas du machinisme la numérisation n'a aucun sens si elle n'est pas liée à un environnement d'approvisionnement, de traitement, de stockage et réseaux⁹. Pour ce qui concerne

7. Les technologies informatiques par leur aspect virtuel, et surtout parce qu'elles sont étroitement intriquées dans l'écriture, la communication humaine et la cognition ajoutent un étage de complexité.

8. Qualité de *centralité* partagé avec la numérisation du multimédia et les réseaux banalisés.

9. Voir l'article de Jacques André et celui de Yannis Haralambous.

une informatique multiscriturale, il semble que les recommandations du W3C et de l'IETF (RFC2277)¹⁰ devraient assez rapidement induire la mise en place d'infrastructures de réseaux correspondant au nouveau système technique. Mais, au-delà, qu'en est-il de la prise de conscience effective des utilisateurs ou même des administrateurs de réseaux (notamment les réseaux privés) ? Comprennent-ils toujours les raisons d'être de ces nouvelles dispositions et leur nécessité absolue afin d'assurer une continuité rigoureuse de l'échange international des données ?

Reprenons encore une fois notre parallèle avec les débuts du machinisme. Pour ce qui est du sous-ensemble technique que constitue la machine à vapeur, celles de la première génération dépourvues de tout automatisme étaient seulement capables d'effectuer des tâches matérielles simples : par exemple, soulever la lourde masse d'un marteau pilon. Cela permettait, certes, de rendre toute une catégorie de services. On pouvait, par exemple enfoncer plus facilement des pieux, ce qui est fondamental pour construire des ponts, creuser des puits d'extraction minière, bref amorcer les infrastructures qui vont permettre au machinisme de se développer. Mais, sans les automatismes permettant d'enchaîner automatiquement les cycles moteurs¹¹, on ne pouvait pas véritablement intégrer ces premières machines comme des moteurs pour faire tourner des fabriques. En fait, ce n'est que lorsqu'elles ont été optimisées comme des machines à double effet qu'elles ont pu faire rouler une locomotive.

De la même façon, une nouvelle informatique multiscriturale doit, elle aussi, devenir totalement transparente, universelle et entièrement automatique dans ses processus d'adaptation et de reconfiguration, tant au niveau des interfaces et périphériques (imprimantes, claviers) qu'au niveau des processus plus complexes (automatismes d'adaptation des logiciels systèmes, moteurs de recherche, moteurs de rendu d'écriture, mise en parallèle d'écriture, interfaces vers la TAO). L'article d'Emmanuel Giguët et Nadine Lucas est, de ce point de vue tout à fait éclairant sur les enjeux de recherche et de développement indispensables.

La machine à vapeur et le nouveau système technique du machinisme n'a enfin pu se développer dans nos sociétés que lorsqu'une nouvelle technoculture a pu largement se diffuser tant chez les élites techniciennes que chez les décideurs économiques, financiers, politiques mais aussi dans l'ensemble de la société pour que puisse exister une classe ouvrière apte à faire fonctionner le système technique industriel et que se développe une nouvelle société consumériste intégrant notamment le salarié comme principal consommateur¹² à même de bénéficier de l'extraordinaire profusion de biens produits par le machinisme industriel.

10. Internet Engineer Task Force, Request For Comment 2277, Dürst (Martin) & Freytag (Asmus), *Unicode in XML and other languages*, www.rfc-editor.org

11. On détourne une faible partie de l'énergie produite par la machine pour entraîner des systèmes de cames et tirettes permettant d'ouvrir et de fermer les soupapes alimentant le piston.

12. Le fordisme.

Là encore, l'analogie avec l'informatique multilingue et multiécriture fonctionne. Unicode, et au-delà, une communauté multiscriturale et multilinguistique ne se développeront pas seulement parce que des informaticiens et des experts en normalisation en comprendront les mécanismes, les contraintes et les protocoles. Cela ne se développera pas non plus, si une élite plus élargie de professionnels, de développeurs de logiciels, de techniciens des réseaux, d'industriels, de vendeurs en comprennent parfaitement les aspects techniques et les enjeux linguistiques. C'est une véritable révolution de la technoculture linguistique qui doit s'amorcer dans la société tout entière et qui ne se fera que si de très nombreux acteurs sociaux trouvent intérêt à utiliser et à voir fonctionner ces flux multilingues et multiécritures. L'internet peut être un champ concret d'applications puisqu'il peut permettre de chercher¹³, partout dans le monde, une adresse, un produit, un fournisseur, un film à visionner, une référence culturelle ou scientifique, cela bien sûr malgré les barrières de langue et d'écriture. Précisément, le web sémantique et l'information structurée devraient faciliter ce type d'application générique.

L'enjeu implicite du machinisme consistait en une mécanisation omniprésente de la machine capable à terme de se substituer à l'homme pour toutes les tâches pénibles ou démesurées. Cela a nécessité l'adoption obligatoire de normes uniformes notamment pour le calcul, la métrologie et la définition des matériaux. C'est dans ce contexte que s'est développée l'adoption généralisée du système métrique mais aussi, prioritairement, l'adoption de normes pour les poutrelles métalliques.

6. Les enjeux d'une communication planétaire multiscriturale

Une informatique mondialisée et interconnectée nous oblige évidemment à disposer d'un système technique d'échange et de traitement omnilingue parfaitement au point. Unicode constitue un des piliers de cet enjeu. Mais le web et Unicode, même s'ils sont devenus des réalités industrielles, ne suffisent pas seuls à répondre à tous les enjeux d'une communication planétaire multiscriturale.

La conquête industrielle, coloniale et économique du monde par l'Occident n'aurait pu se faire sans vaisseaux, sans boussole, sans compas et sans gouvernail (voilà pour les aspects techniques) ; mais elle n'aurait pas pu se faire, non plus, sans une vulgarisation universelle de la géographie physique, géologique, humaine, politique rapportée à une cartographie fiable, mais aussi sans d'innombrables études botaniques, ethnologiques... et même linguistiques. Cela s'est fait à l'échelle scientifique des temps modernes puis de l'époque contemporaine. L'imprimerie, notamment pour les cartes, y a apporté son concours fondamental. Enfin, les besoins

13. Chercher, mais aussi inscrire, structurer et baliser dans sa propre langue pour décrire une offre de produits ou de services.

exprimés pragmatiquement par les marchands, les soldats et les missionnaires ont marqué les limites socialement acceptables de cette aventure humaine.

Actuellement, l'aventure de la rencontre planétaire des hommes¹⁴ est devenue beaucoup plus symbolique, communicationnelle et linguistique. Pour ne pas perdre pied dans cette nouvelle ouverture du monde, nous devons pouvoir disposer, sinon des théories qui existent déjà, mais de manuels scolaires, de littérature de vulgarisation mettant en valeur ce nouveau contexte universel des écritures du monde. Il faudra aussi que nous puissions facilement appréhender cette richesse du patrimoine *panscriptural* en liaison avec le territoire. Notons, en incise, qu'il faudrait notamment inventer un tourisme linguistique et surtout scriptural comme il existe un tourisme des paysages, de l'architecture, des folklores, et des arts. De ce point de vue, la connaissance à gros trait des familles d'écriture, une sorte de cartographie des territoires pas seulement linguistique (cette cartographie est assez commune) mais scripturale est un premier préalable et une invitation sociale à l'échange transculturel.

Au niveau d'une culture technique professionnelle, l'acquisition d'un savoir-faire typographique universel, limité au scriptural n'est pas une utopie. Pendant deux cents ans, il a existé à l'Imprimerie nationale un corps de typographes spécialisés dans la composition des textes en langues orientales. Le besoin s'en était fait sentir lors du retour d'Égypte et Bonaparte lui-même créa ce *corps de typographes des textes orientaux*. Ce n'étaient pas de savants lettrés, connaissant une multitude de langues, cependant ils avaient été formés à l'École des Langues Orientales. Ils y avaient acquis quelques notions de chinois, d'arabe, de sanskrit, de sumérien, d'égyptien, de copte... mais surtout ils connaissaient, dans la totalité de ces écritures des règles de composition et de disposition des textes. Il est dommage qu'il ne reste plus, à ce jour aucun représentant de cette élite d'ouvriers typographes parce que c'est certainement à leur école qu'on aurait dû former les graphistes, mais aussi les typographes, les informaticiens des réseaux internationaux et les utilisateurs de données multicripturales des décennies futures¹⁵.

7. L'Extrême-Orient idéographique à l'origine du changement de paradigme du système technique d'écriture

Puisque l'informatique avait été inventée simultanément en Angleterre et aux États-Unis, il était logique que l'ASCII, et l'ASCII étendu adapté aux autres écritures phonétiques reste pour un temps, le paradigme central de la technologie informatique. Cependant, à partir du moment où les pays extrême-orientaux comme

14. Il ne serait pas politiquement correct de dire conquête même si cet aspect négatif subsiste encore aujourd'hui, notamment au niveau économique.

15. Sur cette importante question nous conseillons : [Lexique des règles typographiques en usage à l'Imprimerie nationale, 1990]. Nous renvoyons aussi le lecteur aux articles de Yannis Haralambous et de Olivier Randier dans ce même volume.

le Japon ont exigé et produit pour eux-mêmes des codes d'écriture à 16 bits¹⁶, il était déraisonnable que les informaticiens et des industriels occidentaux continuent de s'entêter aussi longtemps à nier l'évidence et à ne pas voir l'évolution logique de l'informatique textuelle.

La codification universelle de l'écriture était industriellement démontrée par le Japon avec le JIS, restait le principal obstacle pour l'Occident, entendre ce fait, le comprendre et se l'approprier industriellement, culturellement et socialement. Malgré de brillants précurseurs visionnaires, notamment tous ceux qui se sont impliqués dans Unicode et ISO/IEC10646, ce fait objectif a trop longtemps été nié.

7.1. La grammatologie comme révélateur de nos cultures alphabétiques éthnocentriques

Selon la thèse de l'évolution grammatologique développée par Gelb (Gelb, 1973) puis Derrida (Derrida, 1967), les évolutions de toutes les techniques d'*engrammation*¹⁷ et surtout les révolutions de l'écriture accompagnent et induisent les progrès technocognitifs de l'humanité. La conclusion finale de l'ouvrage de Gelb se résumait malheureusement avec une évidente simplicité : son histoire de l'écriture concluait à l'universalité indépassable et, donc, à la supériorité de l'écriture alphabétique. Ce jugement lapidaire ne fait plus aujourd'hui l'unanimité, mais c'était dans les années soixante et même soixante-dix une évidence intellectuelle partagée par l'inconscient collectif de la quasi-totalité des intellectuels occidentaux et, par là même, évidemment de bien des chercheurs et industriels de l'informatique.

Cependant, nous pouvons avec vingt ans de recul par rapport à la fin des années soixante-dix, conclure différemment de Gelb sur ce point¹⁸ et comprendre que ce qui fut le vrai pendant de longues périodes historiques (du miracle grec [V^e siècle av. J.-C.] jusqu'à nos jours) pourrait très vite s'avérer faux dans le futur proche de l'ingénierie linguistique multilingue et multiécriture. De ce point de vue, il serait

16. Dans son article Nadine Lucas fait référence aux sources IRG (Ideographic Rapporteur Group pour ISO/IEC) où sont rassemblées toutes les normes de codage préexistantes : préfixe G pour la Chine, T pour Taiwan, H pour Hong Kong, J pour Japon, K pour la Corée du Sud, KP pour la Corée du Nord, V pour Vietnam et U pour divers dont Singapour. On peut aussi se reporter utilement au livre de Bernard Marti [Marti, 1990].

17. Dans une vision grammatologique de la communication, il me paraît utile de pouvoir décliner *engrammer* ou *engrammation* à partir du terme *engramme* (en psychologie trace mnésique), pour désigner toute sorte de marquage de l'information qu'ils soient permanents ou volatiles. Le mot enregistrement réfère trop, de mon point de vue à quelque chose de permanent.

18. Sur ce point seulement parce que cette erreur manifeste n'invalide pas (comme le fait un peu brutalement remarquer Derrida) toutes les avancées théoriques et grammatologiques de Gelb.

déraisonnable de continuer à vouloir faire vivre une informatique mondialisée en réseaux mais dans laquelle la norme ISO/CEI 8859-1 (Latin-1) constituerait la norme fondamentale et de référence pour le codage des caractères.

Les langues ayant adopté une écriture idéographique disposent d'un potentiel de médiation cognitive sans doute bien supérieur à celui des langues ayant adopté des écritures alphabétiques. Ce fait, déjà constaté par Liebnitz et noté aussi par Saussure, apparaît d'évidence si on pense que la notation numérique des nombres par les chiffres arabe nous renvoie, toutes proportions gardées, à une sorte d'expérience cognitive idéographique sectorisée. On constate aisément que cette notation est réputée bien supérieure, pour ce qui est de l'efficacité cognitive, à leur notation en toutes lettres¹⁹. Cette notation numérique (et idéographique) est évidemment translinguistique puisque l'Italien, l'Allemand, l'Anglais qui lisent *uno, due, tre - eins, zwei, drei - one, two, three...* écrivent et comprennent universellement 1, 2, 3.

Ainsi Liebnitz évoquant son projet d'écriture algébrique dans *Nouveaux essais*, III, I, §1 : « ...une espèce d'écriture universelle, qui auroit l'avantage de celle des Chinois, parce que chacun l'entendrait dans sa propre langue, mais qui surpasserait infiniment la chinoise, en ce qu'on la pourroit apprendre en peu de semaines, ayant les caractères bien liés selon l'ordre et la connexion des choses... » ; chez Saussure (*opus cité* p. 48), « Pour le Chinois, l'idéogramme et le mot parlé sont au même titre des signes de l'idée ; [...] Les mots chinois des différents dialectes qui correspondent à la même idée s'incorporent également bien au même signe graphique. Nous bornerons notre étude au système phonétique, et tout spécialement à celui qui est en usage aujourd'hui et dont le prototype est l'alphabet grec. » On notera dans cette citation combien Saussure avait pris de précaution pour focaliser les limites de sa linguistique générale et relativiser son apport en laissant ouverte toute la problématique d'une linguistique des langues idéographiques.

Cependant, cette évidente supériorité cognitive des écritures idéographiques est aussi liée à des difficultés non moins évidentes d'apprentissage, ainsi qu'à des situations d'incapacité cognitive dans lesquelles, même un lettré, peut être incapable de lire et d'énoncer des idéogrammes dans des champs sémantiques étrangers à sa discipline²⁰. Cette disparité des situations cognitives d'apprentissage, cette congruence de l'écriture à la sémantique et à la pensée est très souvent difficile à

19. Disant cela, nous ne nions pas les aspects positifs de la notation en toutes lettres, par exemple, pour des raisons de redondance voulue pour assurer une certaine sécurité (sur un chèque ou dans un acte notarié) [suite de la note suggérée par Patrick Andries]. Gardons cependant à l'esprit pour ce qui concerne les idéogrammes chinois signifiant des nombres que ceux-ci sont aussi sûrs que des notations en toutes lettres car il en existe un différent pour chaque position décimale importante. Cependant, certains idéogrammes facilement falsifiables *un* (= un trait qu'on peut changer en dix facilement = deux traits en croix, par exemple) ont des glyphes différents dans le contexte bancaire.

20. La même difficulté se rencontre quelquefois pour la lecture de toponymes ou de patronymes peu connus.

admettre pour un occidental quand elle ne lui est pas parfaitement inconnue sinon niée.

Il faudra bien des transformations de notre inconscient collectif et de nos *a priori* culturels pour que nous admettions que les langues et écritures autres que les nôtres offrent des potentiels culturels, cognitifs, industriels, logiques, ergonomiques, politiques²¹ médiatiques²², translinguistiques²³ différents de nos langues et de nos écritures.

8. Dès l'imprimerie, ou même dès la formalisation d'un mode d'écriture, on observe des diversités d'appropriation selon les communautés d'écriture

S'il est indéniable que les Chinois ont été les premiers à maîtriser l'imprimerie, on prétend quelquefois que les Coréens, alliant dans une même écriture *hangul* phonétiques et idéogrammes chinois²⁴, auraient su profiter de cet avantage, puisqu'ils avaient la possibilité d'aller plus loin que les Chinois en déployant l'imprimerie dans une de ses futures dimensions fondamentales : une véritable réutilisation des caractères. Quoi qu'il en soit, l'Occident chrétien a réinventé sept siècles plus tard l'imprimerie.

Là encore, nous pouvons recourir au concept de système technique et de sa nécessaire mise en cohérence avec tout un environnement technoculturel. Les Chinois avait découvert l'imprimerie mais elle était restée un média presque exclusivement réservé à l'administration impériale. De ce fait, la mise en cohérence technoculturelle n'a pas fonctionné avec la même radicalité qu'en Occident. Il faut bien constater, en effet, que l'apparition de l'imprimerie en Europe déclenche la transformation d'un *univers technique* extrêmement large. Notons par exemple, sur le plan technique la mise au point d'encre permettant un séchage rapide donc une impression *recto-verso*, l'invention de nouveaux alliages pour la fonte des caractères avec le mélange plomb antimoine, mais aussi une quantité d'innovations sur le plan de l'organisation même de la composition, de l'impression et de la fabrication et de la distribution du livre sur une large échelle. Sur le plan social, culturel et cognitif les progrès aussi sont énormes et correspondent à cette mise en cohérence du

21. Pour les Chinois qui n'approfondissent l'apprentissage des idéogrammes que dans le champ lexical de leurs métiers et disciplines, cet état de fait induit des stratifications non seulement hiérarchiques mais par matière disciplinaire des lettrés, ce qui entraîne une rigidité sociale qui joue sur le politique, mais aussi sur l'histoire même de la Chine.

22. L'arabe favorise à l'évidence la lecture silencieuse mais augmente considérablement les difficultés d'énonciation à voix haute, ce qui complique, par exemple, la tâche d'un présentateur de télévision.

23. Ainsi par exemple l'écriture idéographique chinoise permet de communiquer par écrit pour des locuteurs de langues s'écrivant en chinois mais non intercompréhensibles à l'oral.

24. Voir plus bas.

nouveau système technique que représente l'imprimerie avec l'ensemble de la technoculture occidentale.

Mais ceci est une autre histoire largement débattue : citons le grand classique que constitue *La Galaxie Gutenberg* (Mac Luhan, 1967) ou encore le livre d'Élisabeth Esenstein (Esenstein, 1991). Quoiqu'il en soit de cette importante question, il reste que cela a permis à l'Occident d'écriture latine de normaliser très tôt son(ses) écriture(s)²⁵ en chaînes de caractères isolés regroupés en mots séparés par des espaces et syntaxiquement balisés par le système de ponctuation²⁶. Cette invention fondamentale, précisément parce qu'elle était liée aux techniques de l'intelligence et de la culture a induit en Occident une supériorité politique et économique par rapport aux autres écritures du monde avec lequel nous vivons encore.

9. Quelques exemples de la diversité des cultures d'écriture

Cependant d'autres langues avaient, en d'autres temps et lieux, bénéficié culturellement et au niveau cognitif d'autres caractéristiques d'écriture. La culture indienne, par exemple, ne trace pas de frontière entre le calcul et le poétique. Ces deux capacités culturelles sont fondamentales pour la religion hindoue qui demande aux fidèles de se livrer inlassablement à de complexes calculs de *karma* et d'âge du monde liés aux âges des êtres dans leurs réincarnations successives. Il s'ensuit que ce sont les Indiens qui ont su inventer nos chiffres et le système décimal. (sur ce sujet on pourra lire le livre de Georges Ifrah : Ifrah, 1994). C'est sans doute, en partie grâce à cela que la communauté indienne fournit parmi les meilleurs cerveaux de l'informatique dont le talent est universellement reconnu.

Les Arabes dont la langue appartient au rameau sémitique ont, comme les Hébreux, inventé une écriture essentiellement consonantique, ce qui leur a donné l'avantage de disposer d'une écriture beaucoup plus universelle :

- universelle dans l'espace puisque l'écriture de l'arabe classique (ou standard) fédère des parlers dialectaux quelquefois assez différents les uns des autres ;
- universelle dans le temps puisque l'écriture étant dissociée de la langue parlée, la dérive dans le temps de l'écriture est beaucoup plus faible que dans les langues européennes alphabétiques qui connaissent une dérive phonétique au fil des siècles qui influence, à son tour la langue écrite²⁷.

Certes, embarrassés par l'obligation religieuse de préserver une écriture indissolublement liée au Coran et à la révélation, les Arabes n'ont pas su

25. L'écriture latine mais aussi le grec et le cyrillique.

26. Certes, l'invention des espaces datait d'avant Jésus-Christ mais la ponctuation fut codifiée à la Renaissance.

27. Nos langues et nos orthographes ont évolué par rapport au latin et nous ne disposons pas, en Europe, d'une langue écrite vivante qui pourrait être commune à toutes les langues latines.

abandonner l'obligatoire cursivité de leur écriture (les lettres se sont figées dans leur forme manuscrite même dans la typographie) comme ils n'ont pas su standardiser la forme des lettres différenciées selon le contexte. Nous y reviendrons.

Nous ne nous attarderons pas sur les avantages culturels et cognitifs qui ont permis, malgré l'insistance de l'Occident, que l'Extrême-Orient conserve son particularisme idéographique. Le maoïsme a failli avoir raison des idéogrammes de la Chine populaire. En Corée du Nord, le dogmatisme marxiste a, par contre, réussi à éliminer les idéogrammes sino-coréens qui sont une des composantes de l'écriture coréenne et qui ont pu se maintenir en Corée du Sud. Au Japon, les Américains dans l'enthousiasme de leur victoire de 1945, ont presque réussi à *éradiquer* ce qu'ils considéraient comme une survivance passéiste. Au Vietnam, les missionnaires jésuites, la colonisation française et les Vietnamiens eux-mêmes ont presque fait disparaître toute trace de l'idéographie chinoise²⁸.

Si, malgré ces très nombreuses sollicitations d'avoir à se moderniser en devenant alphabétiques, ces communautés idéographiques ont su globalement résister, c'est grâce à la qualité extrêmement synthétique de ces écritures qui procurent à ceux qui ont pris le temps de les apprendre une évidente efficacité cognitive. Le livre de Pascal Griolet (Grirolet, 1985) relate l'histoire de la modernisation du japonais, évolution qui a su se faire sans jamais se laisser dominer par l'évidence alphabétique.

10. La trompeuse évidence de l'ordre alphabétique

L'argument de l'évidente logique alphabétique tombe, dès lors qu'on considère que l'ensemble des idéogrammes étymographiquement²⁹ signifiants sont ordonnable à partir des clés et des nombres de traits associés à la clé d'un caractère. Ainsi se construit une logique d'ordonnement des idéogrammes bien largement supérieure à notre ordre alphabétique latin qui n'a d'autre intérêt que son apparence

28. L'histoire linguistique et scripturale du Vietnam est complexe. Avant le XIII^e siècle ne sont attestés que des textes en *Chu hân* (des caractères chinois archaïques) à partir du XIII^e siècle apparaissent des caractères idéographiques spécifiques les *Chu nôm*, (terme qui signifie écriture populaire ; une écriture dans laquelle se surajoutent des éléments idéographiques qui précisent la prononciation vietnamienne. Au XIII^e siècle Alexandre de Rodes entreprend une romanisation de l'écriture qui est celle qui domine aujourd'hui : le Quô ngu (voir l'article de Michel Bottin, cf. aussi Glossaire <http://iquebec.ifrance.com/hapax/glossaire.htm> et Hudrisier, 1997).

29. Nous proposons ce terme pour faire passer l'idée que la graphie des idéogrammes est bien sûr liée au sens et comme l'étymologie organise les mots en familles. Cependant si les clefs organisent l'ensemble du dictionnaire selon une première logique (qui peut être ténue dans certains cas), la présence de plusieurs caractères élémentaires dans un seul caractère rend cette première logique beaucoup plus complexe.

universalité conventionnelle nous permettant de classer sans *a priori* la totalité d'une encyclopédie.

Dans les anciens systèmes de codification alphanumérique on cherchait souvent à ordonner au mieux les caractères d'une écriture dans les tables informatiques. On s'efforçait pour que les premières lettres de l'alphabet aient des codes numériquement inférieurs à ceux des lettres qui les suivaient pour pouvoir facilement générer un ordre alphabétique³⁰. Ce principe n'existe plus pour Unicode et l'ISO/CEI 10 646 qui renvoie vers d'autres normes pour activer des tris alphabétiques³¹.

11. La prégnance d'une culture d'écriture

L'abécédaire est souvent le premier livre d'écriture offert à l'enfant. Mais qu'est-ce que l'alphabet ? Quel est son degré d'universalité ? Qu'est-ce qui peut équivaloir à notre abécédaire latin pour un petit Arabe, un petit Japonais, un petit Chinois, un petit Indien ?

Pour le petit Arabe, et en cela il partage un point commun avec le petit Japonais, la langue (ou plutôt l'écriture) va même se plier jusqu'à lui pour lui faciliter l'apprentissage. Les grammairiens qui vivaient à la cour de Bagdad, aux VII^e et VIII^e siècle, ont prévu cette facilité qui consiste à noter toutes les voyelles qui, d'habitude ne se notent pas en arabe pour aider à l'apprentissage de la lecture à travers le Coran³².

Le petit Japonais va bénéficier de la même sollicitude. Contrairement au petit Chinois qui est confronté dès le début de son apprentissage au catalogue de caractères le plus important du monde (jusqu'à 60 000 pour des dictionnaires relativement exhaustifs) le petit Japonais bénéficie d'une langue dont la notation est relativement flexible liant deux syllabaires phonétiquement équivalents mais sous deux formes distinctes (*hiragana* et *katakana*) et un ensemble d'idéogramme chinois (10 à 12 000). Dérogeant aux règles de la notation canonique qui associe, selon des règles complexes, les idéogrammes (*kanji*) et les deux syllabaires : les *kana*³³, l'écriture japonaise se simplifie pour faciliter l'appropriation de l'écriture par l'enfant ; et même par celui qui n'a pas choisi d'être un lettré. Il devra d'abord apprendre la cinquantaine de sons de sa langue et les associer aux deux versus des syllabaires avec lesquels il va pouvoir noter phonétiquement toute la langue japonaise. Ce n'est qu'ensuite, progressivement, qu'il va affronter les idéogrammes

30. En fait ce principe n'est absolu que pour le vrai ASCII à 7 bits. À partir de Latin-1 ce n'est plus vrai (par exemple le code de F est inférieur à celui de É).

31. Voir sur ce point l'article de Patrick Andries.

32. Voir l'article de Rachid Zghibi.

33. Mot raccourci pour désigner le double syllabaire *hiragana* et *katakana* (voir l'article de Nadine Lucas).

(les *kanji*) selon un programme culturellement codifié (900 idéogrammes minimum à la fin du primaire et 1 800 minimum à la fin du secondaire). Les bandes dessinées ou même certaines publicités vont, elles aussi, jouer ce rôle *d'aménagement abécédaire* en notant la phonétique des caractères supposés mal connus en petits caractères phonétiques *sur-titrés* au-dessus des *kanji* : les *furigana*³⁴.

Voilà pour les facilités, mais qu'en est-il des violences ? Certaines communautés linguistiques peuvent être conditionnées, dès l'enfance, par la prégnance culturelle d'une écriture. Comment ne pas s'étonner que des Serbes et des Croates³⁵ qui parlent pratiquement la même langue doivent l'écrire, ici en alphabet latin et, de l'autre côté, en cyrillique ? Il en va de même dans certaines régions du continent indien où la même langue l'hindoustani va se noter, ici, en ourdou (des caractères arabes) et, là, en écriture hindi. Dès le stade de l'abécédaire, ces communautés sont ainsi culturellement confrontées à la construction d'un mur mental, d'une frontière que ne surpassera jamais aucun mirador.

Cette remarque n'est pas polémique. Ces communautés ont été marquées dans leur inconscient collectif au long des générations, conforté par des différences religieuses à leur tour relayées par le nationalisme. Ces enfants ne sont pas responsables de l'abécédaire maternel qui leur est ainsi imposé dès le plus jeune âge. Que dire encore de certaines minorités des marges de l'ex-URSS qui pendant le XX^e siècle ont parfois changé trois ou quatre fois d'écriture ? Leur langue se notait en écriture arabe, certains linguistes russes leur ont proposé l'écriture latine *pour se moderniser*. D'autres linguistes *plus soviétiques* leur imposèrent le cyrillique et la fin de la suprématie de l'Armée rouge ramena les mollahs et l'écriture arabe. Pour certaines communautés linguistiques, cela leur redonne une continuité scripturale et linguistique avec l'Iran ou le Pakistan où on utilise l'écriture arabe, mais certains sont turcophones or les Turcs écrivent en écriture latine depuis Atatürk ! Les Berbères ont été et sont encore pareillement malmenés pour ce qui est de leur écriture. Ils disposent traditionnellement d'un alphabet, le *tiffinag*, plusieurs fois millénaire. Mais les colonisateurs (d'abord les Arabes ensuite les Français) ont trouvé plus facile (pour eux) de leur imposer d'écrire en écriture arabe ou même, mais ce fut une transcription plus théorique que véritablement usitée, en caractères latins. Le renouveau du tiffinag en Kabilie ou chez les Targui³⁶ n'est pas toujours du goût du pouvoir algérien ni des fondamentalistes musulmans, mais il faut noter que le Maroc soutient le développement de classes d'apprentissage du berbère en tiffinag.

34. Voir l'article de Nadine Lucas.

35. Rappelons que la latinisation du croate est un choix récent qui s'explique de par l'appartenance de la Croatie à l'Empire austro-hongrois.

36. Il faut souligner que le tiffinag n'est toujours pas rentré dans Unicode (voir l'article de Patrick Andries). NOTA : Targui est le pluriel de Touareg.

12. Des-ordres alphabétiques

L'utilité de l'ordre alphabétique est une notion qui est discrètement mais fermement inculquée à l'enfant dès ses premiers abécédaires. En regroupant sous une même lettre la girafe, le gnou, la grand-mère et le gyrophare on conduit graduellement l'enfant à un apprentissage systématique des composants de son écriture, mais aussi vers le dictionnaire, l'encyclopédie, l'annuaire et les bases de données. Cependant, même si les fichiers informatiques, les encyclopédies, les dictionnaires, de nombreux catalogues semblent suivre *naturellement* l'ordre alphabétique, la complexité des situations réelle fait que la diversité est de règle. Ainsi, l'ordre alphabétique de l'écriture latine connaît des variations de détails selon les finalités de ses usages : une bibliothèque n'utilise pas traditionnellement le même ordre que les lexicographes ou les éditeurs d'un annuaire. L'écriture latine connaît plusieurs ordres alphabétiques dans différentes langues disposant de lettres ou de signes diacritiques spécifiques : les Scandinaves, par exemple, renvoient le « E dans l'A » (Æ) à la fin de l'alphabet, les Espagnols regroupaient à la fin de la lettre N toutes les occurrences du N avec *tilde* (Ñ), mais ils ont changé de convention et intercalent maintenant tous les N avec ou sans tilde. Ils agissent encore de même pour les « doubles L » et les « CH ».

Cependant, l'ordre des lettres qui existe dans certaines écritures peut avoir plusieurs facettes fonctionnelles (ordre d'apprentissage, ordre phonétique, classement par similarité de forme, ordre des lettres numérales). Par exemple, tous ces ordres sont utilisés concurremment en arabe.

Pour leur part, les Japonais sont confrontés à trois ordres au moins qui ne recouvrent pas le même matériel d'écriture : pour les syllabaires, les *kana*, un ordre phonétique logiquement articulé (consonne et voyelle) mais aussi un ordre poétique celui de l'ordre d'apparition de tous les sons contenus dans un seul poème. Mais pour classer les *kanji*, les Japonais utilisent, comme les Chinois, la classification par clefs puis par nombre de traits ajoutés à la clef.

13. Les dernières années de l'alphanumérique

La norme universelle de codification des caractères sur quatre octets ISO/IEC 10 646 et du standard industriel qui lui correspond se limite encore pragmatiquement à 2 octets : Unicode a déjà remplacé l'ASCII et depuis déjà quelques années la plupart des ordinateurs et tous les nouveaux logiciels systèmes ont adopté Unicode, même si nous ne le voyons pas et même si nous ne disposons pas encore des outils périphériques (claviers virtuels multilingues par exemple) qui nous permettraient de mettre en valeur ce potentiel multiécriture totalement révolutionnaire. En fait, les questions les plus cruciales aujourd'hui concernant le développement de ces techniques ne sont plus des questions industrielles importantes (les mécanismes, les principes et une grande partie des tables sont

maintenant totalement développés), mais des questions d'appropriation culturelle tant chez les informaticiens non directement spécialistes que chez l'honnête homme confronté à une omniprésence de l'altérité du monde qu'il ne sait pas dominer.

Revenons à l'informatique : il y a une vingtaine d'années, l'informatique se contentait d'être presque exclusivement aménagée par rapport à la langue anglaise et à l'écriture latine non accentuée. Elle fut assez vite adaptée aux écritures alphabétiques des grandes langues européennes occidentales. Cette exclusivité a cessé. Les grandes langues appartenant à d'autres alphabets sont devenues disponibles informatiquement sur des variantes du code ASCII (cyrillique, grec, arabe, hébreu, etc.) et les langues partiellement et entièrement idéographiques (chinois, coréen, japonais) ont amorcé un processus obligatoirement plus lourd de codification des caractères : deux octets au lieu d'un (16 bits au lieu de 8). C'est cette évolution qui a déclenché la mise en œuvre d'une codification universelle sur 4 octets³⁷ (32 bits).

Ce phénomène est très important en termes d'évolution informatique de l'écriture. Toute typographie suppose l'usage d'une casse (le casier en bois du typographe) permettant de ranger tous les caractères (lettres majuscules et minuscules, accentuées, chiffres, ponctuation...) d'une police de caractères. Ainsi la casse parisienne comprend 104 cases³⁸. Le premier système de codification informatique des caractères permet de gérer les caractères sur 7 bits (128 positions théoriques, mais en réalité 64 codes réels) : c'est le code ASCII qui ne dispose pas de lettre accentuée. Puis apparaissent des codifications sur 8 bits (1 octet) qui disposent de 256 positions théoriques, en réalité 96 [0xA0-0xFF] cases supplémentaires [il reste le fonds commun ASCII, 0x20-0x7F] réelles pour les caractères proprement dits : ce sont les normes ISO8859-n qui permettent de gérer alternativement mais jamais simultanément les autres grandes langues alphabétiques (écriture cyrillique, grecque, arabe, hébraïque, et le thaï...).

Ce jeu alterné des codes d'écriture présente de nombreux inconvénients : cela rend difficile la réalisation de fichiers de bibliothèque multi-écritures et, surtout, cela entrave la circulation d'informations multilingues sur internet. Par exemple, un document en langue française transitant par certains serveurs anglophones perdra ses accents et cédilles, quant aux documents en écriture non latine, ils peuvent perdre toute lisibilité dès qu'ils quittent leur aire linguistique³⁹.

37. Il s'agit de puissances de 2, et il est logique de considérer une évolution technologique allant de 8 à 16, puis à 32 bits (1, 2, 4 octets ou groupes de 8 bits).

38. La plupart des questions qui suivront sont abondamment développées dans l'article de Jacques André.

39. La réalité de ces questions est complexe. Il est important de bien faire la distinction entre la technologie du codage direct des caractères et les systèmes de présentation de l'information multilingue (comme Mime) dans lesquels l'information est systématiquement transformée dans le format ASCII à 7 bits. Sur ces sujets lire l'article de Sylvie Baste ou celui de Jacques André.

La première phase de cette norme universelle, le PMB (plan multilingue de base) déjà codé sur 4 octets, mais ne mobilisant que 16 bits est le résultat déjà en partie opérationnel issu du travail coordonné des experts de l'ISO et du consortium Unicode (qui au début prévoyaient 2 octets par sagesse technique et ont adopté 4 octets par la suite). Ce codage qui offre 65 536 codes théoriques possibles, pourrait résoudre tous les problèmes posés par la majorité des langues modernes déjà informatisées. Mais cette solution reste insuffisante pour coder de façon univoque les caractères de toutes les écritures du monde notamment les écritures anciennes ou les codes non linguistiques (musique, chorégraphie, langue des signes, glyphes mathématiques, codes ésotériques, pictogrammes...) Par ailleurs, la mise en place d'une informatique et de réseaux à 16 bits suppose des aménagements fondamentaux par rapport aux systèmes à 8 bits (réécriture de logiciels systèmes, des protocoles de télécommunication, évolution technologique des machines) qui plaide pour que la norme qui succède à l'ASCII soit directement prévue sur 32 bits (4 octets soit 4 294 967 296 [4 milliards] de codes théoriques⁴⁰) même si sa mise en œuvre informatique doit être progressive.

Il faut noter que la totalité des systèmes de glyphes, de schémas et de signalétique déjà développés par des civilisations humaines (du code de la route aux schémas codés pour l'électronique...) ne doit pas dépasser quelques millions. Le système à 4 octets mis ainsi en place pour coder les caractères génère ainsi une quantité phénoménale de *cassetins*⁴¹ dans l'immense casse virtuelle. A ce jour, les experts d'Unicode et de l'ISO/CEI 10 646 ne revendiquent qu'une vingtaine de bits (voir note ci-dessus). Le très grand nombre de codes potentiels restant pourrait servir à coder directement des toponymes, des patronymes, des taxons scientifiques, des copyrights, des listes d'autorité, des codes multimédias et bien sûr des logos picturaux qui seraient un prolongement direct de la culture typographique. La jachère informationnelle existe virtuellement. Pour que cela devienne une ressource concrètement disponible il faudra d'abord que se soit généralisée une informatique systématiquement à 2 puis 4 octets et non plus une informatique à 8 bits comme celle à laquelle nous sommes quotidiennement confrontés. A ce jour, un certain nombre de domaines de la normalisation des systèmes d'information ont de plus en plus besoin de normaliser des listes de vocabulaires, des listes de métadonnées (par exemple les *Learning Resources Metadata*⁴²). Aujourd'hui, émerge un énorme besoin pour identifier de façon absolue, non ambiguë et directement référençable un très grand nombre de codes pouvant être considérés comme des universaux

40. Ceci est le potentiel théorique des combinaisons possibles sur 4 octets. Il conviendrait d'abord de supprimer un bit, ce qui ramène le potentiel de codage à 2 milliards. Cependant soucieux de rester dans le champ normatif défini *a priori* : la codification de caractères, Unicode et l'ISO/CEI 10 646 vont de 0x00 à 0x10 FFFF (environ 20 bits), sur lesquels il reste environ 880 000 codes disponibles (137 000 sont réservés pour usage privé, 95 156 sont attribués par Unicode 3.2) [Les précisions ci-dessus m'ont été données par Patrick Andries].

41. Un petit casier élémentaire d'une casse (de l'italien *cassetino* = petit tiroir).

42. Anciennement LOM (*Learning Object Metadata*).

internationalement partagés. Ces codes pourraient facilement trouver place dans l'espace laissé libre par les 12 bits restant.

L'espace entre 20 et 32 octets n'étant pas à ce jour formellement revendiqué, on peut toujours supputer sur son usage potentiel. Cependant, il est utile de rappeler que c'est de toute façon un espace fini (qui serait rapidement saturé avec des données référents à des individus puisque nous sommes déjà 5 milliards d'humains). Notons enfin que les ressources codées sur ces tables numériques seraient situées sur un mécanisme informatique qui sert de fondement à notre culture informatique donc un espace de donnée très rapidement disponible dans tout système de traitement.

Là encore, une appropriation harmonieuse de la culture informationnelle numérique exige que l'on problématise la succession des générations technique selon la temporalité des futurs à court moyen et long terme. Aujourd'hui les normalisateurs de données (commerce, LRM, vocabulaires divers...) tendent à vouloir ramener l'énorme quantité de référents de base propres à chaque spécialité à des identifiants uniques. Mais ces normalisateurs de données et autres documentalistes envisagent déjà un monde toujours numérique, encore plus complexe et qui aurait peut-être besoin de revenir à une approche plus linguistique que par identifiants numériques. (pour ce qui est de la gestion des universaux normalisés, voir par exemple les derniers travaux de l'ISO-TC47/SC3 & SC4).

Quoi qu'il en soit, l'usage des parties du registre de code généré par un codage des caractères sur 4 octets et laissés libre par Unicode et ISO/CEI 10 646 est un enjeu informatique, culturel et cognitif fascinant, certes en partie lié à ces techniques mais qui dépasse largement notre sujet et qui sort de toute façon du domaine que se sont fixé les experts de la normalisation des caractères.

14. Nécessité d'une culture globale de l'écriture préalable à une large appropriation d'Unicode

Cependant, passer d'une centaine de cases de caractères à plusieurs dizaines de mille, puis à plusieurs centaines de mille (pour ne parler que des glyphes d'écritures), suppose des savoirs-faire qui ne sont pas exclusivement techniques. Les langues ont développé des systèmes d'écriture intimement liés à la culture et qui présentent une diversité réelle de laquelle la typographie alphabétique occidentale donne une bien faible idée.

Elles peuvent être alphabétiques, syllabiques, idéographiques, idéographique-syllabiques (japonais) ou idéographiques et alpha-syllabiques par composition (coréen). Elles peuvent s'inscrire dans toutes les directions du plan d'écriture : haut en bas, gauche à droite, droite à gauche, être cruciverbées et surtout elles peuvent présenter tous les hybrides possibles de déroulement⁴³. S'inscrire de gauche à droite,

43. Voir les articles de Rachid Zghibi et Yannis Haralambous.

puis de droite à gauche alternativement (boustrophédon⁴⁴), ou simultanément dans la même ligne (en arabe par exemple pour la notation des nombres). Les écritures peuvent présenter de nombreuses ligatures ou au contraire des scissions de lettre en deux parties venant encadrer une ou plusieurs lettres. Dans certaines écritures, comme en arabe, les lettres peuvent changer de forme suivant leur position dans le mot (isolée, initiale, médiane ou finale). Certaines lettres peuvent s'empiler (*digrammes* ou *trigrammes* des écritures indiennes), ce qui nécessite de gérer simultanément un déroulement horizontal et vertical ; mieux, les lettres peuvent s'agencer en matrices syllabiques (coréen).

Ce très bref tour d'horizon donne à penser que les questions de stockage ou de placement des caractères dans une mise en page ne sont pas simples dans le contexte d'un système informatique multiscritpural. Et pourtant, on peut comprendre l'enjeu des futures évolutions à travers trois exemples :

- celui des noms de domaine sur l'internet (DNS : *Domain Name System*) qui n'est aujourd'hui possible qu'en ASCII-7bits. A la demande des Asiatiques un groupe de travail de l'IETF s'est constitué pour imposer l'emploi d'Unicode à ce niveau (IDN, *International Domain Name*) ;

- autre exemple, les systèmes d'identification des œuvres (littéraires et cinématographiques) qui sont déjà disponibles et qui se généraliseront dans deux ou trois ans coderont le copyright et les droits d'auteur sous une forme intégrée aux fichiers informatiques. Ce marquage ne pourra pas se réaliser en code ASCII ; en effet les Chinois, les Japonais, les Indiens, les Arabes font aussi du cinéma et de la littérature et il n'y a aucune raison pour qu'ils translitèrent leurs noms ou leur copyright en caractères latins ;

- les bibliothèques constituent encore un autre métier qui devra adopter Unicode assez rapidement. En effet, la translittération latine est longtemps restée le seul moyen réaliste de pouvoir disposer d'un mode d'échange des bibliographies internationales. Ce code commun d'appréhension mondiale des références (comme pour la toponymie), sur un seul mode de lecture et d'écriture (l'alphabet latin), doit évidemment être maintenu parce qu'il correspond à une nécessité concrète d'internationalisation de l'échange. Il est donc indispensable et réaliste de continuer de produire des banques de données bibliographiques par translittération des titres et des noms d'auteurs mais il serait aussi beaucoup plus commode de pouvoir les inscrire aussi dans leurs langues respectives.

Ce dernier exemple (qui pourrait être décliné à l'identique dans le domaine de la cartographie, des annuaires internationaux ou des banques de données industrielles ou commerciales), présente l'avantage de préciser les nouveaux besoins en matière d'échange international de documents écrits⁴⁵.

44. Du grec *bous*, bœuf, et *strophein*, tourner. Une écriture qui s'inscrit comme les sillons d'un labour.

45. Voir l'article d'Emmanuel Giguet et Nadine Lucas.

A l'évidence ce n'est pas la coprésence de toutes les écritures du monde sur le réseau qui nous donnera le don immédiat des langues. Par contre, un réel potentiel de présentation multilingue de l'information aura à la fois des effets de commodité et des conséquences inéluctables d'apprentissage réel. Nous ne devons en aucun cas considérer que la translittération est une étape dépassée. Celui par exemple qui consulte une carte routière dans un pays dont il ne connaît pas l'écriture a le besoin impérieux de disposer des toponymes sous forme translittérée (voire aussi traduite⁴⁶ quand c'est opportun), mais il devrait aussi pouvoir apparier le nom d'une ville, par exemple, avec le même nom en écriture locale. Ce qui se comprend aisément pour cet exemple trouve sa correspondance dans la recherche d'une adresse, l'acquisition d'un produit, mais aussi d'un livre. Par contre, la généralisation de ces écrits de référence sous des formes multilingues mais surtout multiécritures réellement parallèles (et si possible interactives dans le cas de document électronique) ne sera réellement possible à des coûts non prohibitifs⁴⁷ que si Unicode et l'environnement associé nécessaire se généralisent. Mais pour ce faire, il faut que se développe une véritable culture, mais aussi sans doute un métier de l'échange multiscritptural.

14.1. Les langues sémites : des écritures sans presque aucune voyelle

« *Lis... je ne suis point de ceux qui lisent, **Lis** je ne suis point de ceux qui lisent, répliquais-je encore. Pour la troisième fois, l'ange me saisit, me pressa au point de m'enlever toute force, puis me lâcha en disant : **lis, au nom de ton Seigneur qui t'a créé...** » (El Bokari, 1964)*

Mahomet savait-il lire avant la date de sa révélation ? Cette question se rattache, en fait aux particularités de l'écriture arabe (et de l'hébreu) lorsqu'elle n'est pas vocalisée, c'est-à-dire accentuée pour faciliter son énonciation. L'arabe se lit plus facilement à voix basse mais il est, par contre, difficile à lire en public, à *proclamer*. Mahomet, qui était d'une famille de riches marchands de La Mecque, était certainement suffisamment lettré pour pouvoir lire les contrats, les livres de comptes, les messages commerciaux apportés par les caravaniers. Mais sans doute veut-il, par ses deux premiers refus à l'archange Gabriel, signifier à Allah qu'il n'est pas de ceux qui peuvent proclamer.

Si la voyelle grecque a accouché de la suprématie de l'Antiquité gréco-latine, la résistance à vocaliser la langue arabe aura été porteuse d'autres avantages :

- la non-accentuation de politesse (ne pas mettre les points diacritiques des voyelles : ne pas vocaliser) permet à des peuples parlant des dialectes divers de se comprendre à la lecture grâce à un arabe consonantique standard. Cela permet à une

46. En restant sur le territoire du *Latin I* l'exemple de la Belgique démontre que des cartes au moins bilingues peuvent être indispensables.

47. C'est un effet attendu de toute norme que de favoriser la large dissémination de processus industriel, donc d'induire la baisse des coûts des composants et des logiciels.

fédération culturelle et linguistique diversifiée de partager une même langue écrite sur la longue durée historique ;

– les écritures sémites consonantiques, arabe ou hébraïque, ont aussi permis, beaucoup plus tôt qu'en Occident, la pratique de la *lecture silencieuse*⁴⁸. Encore aujourd'hui, l'hébreu et l'arabe ouvrent à une plus grande performance de lecture et d'écriture ;

– cela pourrait être une des explications possibles de la fastueuse et féconde Renaissance arabe mais aussi de la tradition intellectualiste juive. Dans le continent logique de l'écriture arabe cohabitent des dialectes arabes pluriels.

Ceci a pour conséquence *une moindre transparence phonétique*, générant de plus grandes difficultés à apprendre la langue écrite et de plus grandes difficultés encore à la lire à haute voix. Mais, en contrepartie, ceux qui ont ainsi appris à lire et écrire disposent d'un accès facile à la lecture silencieuse rapide, et ceci très tôt dans l'Histoire.

14.2. D'autres particularismes scripturaux potentiellement dopés par Unicode

Que l'on pense, très près de nous, à l'usage germanique qui veut que les substantifs commencent à l'initiale par une capitale (dans un bon usage typographique cette capitale est d'un tiers moins haute que les capitales de noms propres ou de début de phrase⁴⁹) ; il y a là un avantage indéniable pour le traitement linguistique automatique de l'allemand. En effet un simple tri de ces petites capitales permet de trouver rapidement tous les substantifs d'un texte allemand.

Les très nombreux particularismes du chinois, du coréen, du japonais, des différentes écritures de la famille indienne, sont autant d'aménagements culturels historiques des écritures mais aussi de difficultés pour le traitement et la mise en œuvre informatique des textes. Par exemple, nombre d'écritures comme les écritures indiennes ou l'arabe associent les lettres, le plus souvent en les transformant (2 lettres associées : digrammes ; 3 trigrammes et en sanscrit on peut associer 4 ou 5 lettres), ce qui complique le jeu régulier d'une disposition successive des lettres en principe parallèle à l'écriture et à la lecture.

Le coréen lui est un cas limite de l'écriture qui déroge à ce que nous considérons (avec nos réflexes ethnocentriques d'Européens d'écriture latine) comme un

48. La lecture silencieuse, c'est-à-dire la compréhension du texte sans qu'il soit besoin de le dire à haute voix (ou même *mezzo-voce*), pour que le lecteur accède à la compréhension du texte par l'écoute de sa propre voix est une évolution historique de l'appropriation culturelle de l'écriture. On cite souvent Saint-Augustin qui s'étonnait que son ami Saint-Ambroise de Milan sache lire à voix basse. « *Quand il lisait, ses yeux parcouraient la page et son cœur examinait la signification, mais sa voix restait muette et sa langue immobile.* » Saint-Augustin, *Confessions*, VI, 3, Paris, 1959, (Manguel, 1998).

49. Facilité que la plupart des traitements de texte n'offrent pas.

principe premier de l'écriture : la chaîne de caractères. Les caractères *hangul* qui constituent la composante phonétique de l'écriture coréenne, s'associent par composition à partir d'éléments alphabétiques qui viennent se disposer dans un rectangle idéal de la même façon que se composent les différents formants d'un caractère chinois. Dans ce dernier cas, on est très loin des suites successives de lettres que nous considérons comme la logique de composition normale d'une écriture. Dans la norme Unicode, la composition/décomposition des *hangul* a été très systématiquement organisée, ce qui présente des avantages indéniables en termes de traitement linguistique.

Si on a déjà évoqué l'apparente difficulté du catalogue des caractères idéographiques (qui se confond avec un premier niveau du lexique) et la difficulté de saisie des textes en chinois, on ne peut ignorer en revanche les avantages translinguistiques des idéogrammes.

Le CJC (chinois, japonais, coréen) constitue précisément une sous-partie conséquente du PMB. C'est une série de quelques milliers de caractères qui ont été normalisés et définis en commun par les Chinois, les Japonais et les Coréens⁵⁰. Ce catalogue a pu être défini même s'il existe des variantes de forme entre, par exemple, les caractères chinois traditionnels utilisés par la diaspora mais aussi à Taiwan et à Hong Kong et les caractères simplifiés de la République populaire chinoise.

Cependant, comme le montre l'article de Nadine Lucas, le CJC n'est pas la panacée. Chaque partenaire linguistique a le sentiment, comme dans tout consensus, d'avoir perdu un peu de sa spécificité culturelle. Il reste que même avant la mise en chantier de ce catalogue commun les Chinois, les Coréens, les Japonais pouvaient déjà accéder par l'écriture à une compréhension interlinguistique même si elle se limite à ce seul niveau lexical. Sans avoir la qualité d'une véritable traduction, cette capacité, d'échange entre des peuples qui parlent des langues très différentes, constitue un atout connu de tout temps dans cette ère géographique. L'informatisation des réseaux n'a fait qu'optimiser cette capacité extrêmement efficace d'intercommunication translinguistique. La norme ISO/CEI 10 646, quant à elle, a contribué à optimiser et sans doute à rapprocher en les normalisant les caractères de ce catalogue qui, rappelons-le encore, constitue également un lexique.

L'écriture japonaise, quant à elle, rassemble un certain nombre de qualités qui se conjuguent opportunément avec les NTIC. À la suite d'une longue histoire qui est celle de l'évolution de l'écriture japonaise et de l'adaptation de l'écriture chinoise aux spécificités linguistiques et culturelles du Japon, on peut dire, en simplifiant beaucoup, que les idéogrammes notent le lexique ainsi que les patronymes et les toponymes, tandis que les *kana* transcrivent ce qui correspond à l'aspect syntaxique

50. Il faut cependant noter que les experts américains sont pour beaucoup dans l'aboutissement de la mise en chantier du CJC, ce qui n'est pas sans poser quelques problèmes (voir l'article de Nadine Lucas). Notons aussi qu'une petite zone est réservée pour les *Chu hán* et les *Chu nôm* vietnamiens.

et aux flexions ainsi qu'un certain nombre de transcriptions étrangères. Quels merveilleux atouts pour mettre en œuvre une ingénierie linguistique qu'une langue qui peut ainsi potentiellement *débrayer* le lexique de son contexte syntaxique ! Ce particularisme mixte de l'écriture japonaise a été et reste encore un sérieux avantage pour résoudre, au Japon, les inconvénients résultant de l'idéographie intégrale que subissent les Chinois avec beaucoup plus de difficultés. On a déjà dit que les petits Japonais apprennent d'abord à *phonétiser* intégralement l'écriture. Cette réalité pédagogique et culturelle a permis de réaliser des systèmes de saisie d'écriture. Après une saisie phonétique obligatoirement réductrice, le traitement de textes lance en temps réel une phase de traitement linguistique (*post-saisie*) qui permet de se voir proposer des idéogrammes (*kanji*) par équivalence phonétique. Cette nécessité de devoir disposer de fonctions aussi perfectionnées dans un simple traitement de textes fut ainsi un acquis très précoce de l'industrie japonaise résultant des difficultés d'ingénierie linguistique qu'elle avait dû surmonter.

Les Chinois, eux, pour saisir un texte, utilisent le plus souvent le pinyin (méthode efficace et populaire) mais ils disposent aussi du *bopomofo* (syllabaire autochtone) et ils inventent souvent de nouvelles méthodes (clefs, nombres de traits, par les 4 coins, etc.). Cependant avec les *kana* les Japonais utilisent une partie réellement constituante de leur propre écriture⁵¹. C'est ainsi que l'industrie informatique japonaise a pu commencer, depuis plus de vingt ans, à développer des systèmes de codification d'écriture sur deux octets (JIS)⁵² prenant ainsi une avance considérable sur ses voisins mais aussi sur le monde entier.

14.3. *Maîtriser la diversité des situations de multiscrituralité*

Les situations de multiscrituralité sont très inégales. Il est aujourd'hui très difficile de préjuger des avantages et des désavantages que pourront tirer les communautés linguistiques qui se trouvent en situation soit de domination quasi absolue (c'est le cas de l'écriture latine non accentuée) soit dans des cas de domination intermédiaire sur un vaste territoire et en interaction avec une famille de langue (c'est le cas typique du cyrillique et de l'arabe), soit en situation très spécifique, ce qui entraîne l'usage obligatoire d'un autre système d'écriture dès que l'on s'échappe de sa propre langue. C'est le cas de langues comme l'arménien, le grec, le cambodgien par exemple. Il faut aussi citer le cas des communautés linguistiques extrême-orientales qui utilisent des écritures idéographiques, que nous

51. Et surtout ils utilisent des réflexes de translittération intégrale en *kana* qui sont ceux de leur premier apprentissage de l'écriture japonaise.

52. En fait le codage JIS de base qui correspond à l'ISO-2022-JP utilise un nombre variable d'octets (l'ESC permet de passer d'un codage sur un octet à celui sur deux). JIS X 0221-1995 est en fait la norme nationale japonaise qui correspond à l'ISO 10 646 (traduite officiellement en japonais) donc 2 ou 4 octets. [Les précisions ci-dessus m'ont été données par Patrick Andries].

avons plusieurs fois évoquées. Ces populations disposent en effet d'un excellent potentiel de multiscritpuralité.

Les experts chinois, coréens et surtout japonais sont déjà depuis une décennie en situation de maîtrise globale potentielle de l'échange mondial des écritures. Autre avantage, ces premières générations de codage de l'écriture sur 2 octets se sont déjà amorties sur leur marché intérieur et ont vulgarisé l'usage obligatoire d'un codage 16 bits. De plus, contraints d'apprendre les idéogrammes dès leur jeune âge, contraints aussi de connaître l'alphabet latin qu'ils maîtrisent quasi obligatoirement, il leur est beaucoup plus facile que pour des Occidentaux d'ajouter l'apprentissage ne serait-ce que formel⁵³ du cyrillique, de l'arabe ou, par voisinage, de quelques écritures de la famille indienne. Nécessité absolue faisant loi, ils n'ont évidemment pas de résistance pour s'appropriier Unicode et ils fondent ainsi précocement la plus importante communauté d'utilisateurs de l'encodage à 2 ou 4 octets.

On ne saurait s'arrêter dans ce tour d'horizon des situations de multiscritpuralité. Ce qui nous paraît important c'est que cette diversité qui constitue une des richesses intellectuelles du monde soit connue, reconnue dans ses spécificités, avantages ou handicaps. Il est fondamental que les intervenants socioéconomiques des TIC, les techniciens et tous les autres utilisateurs de l'informatique planétaire acquièrent une culture générale solide de cette réalité. Cela semble fondamental parce que l'uniscritpuralisme, quand ce n'est pas l'unilinguisme anglophone, pourrait peut-être devenir un handicap dans quelques années (ou décennies) et non plus la rente de situation qu'il constitue aujourd'hui.

N'oublions pas, nous aurions trop tendance à le faire, que nous sommes devenus, en tant qu'europeens, les citoyens d'un même espace politique multilingue et pluriscritpural. Les billets de banque en Euro sont là pour nous rappeler cette situation de double écriture latine/grecque ; mais dans un avenir relativement proche (cinq à dix ans) l'Europe devrait s'ouvrir ensuite à l'écriture cyrillique (Bosnie, Serbie, Bulgarie, Ukraine, Russie dont l'intégration à l'ensemble européen ne semble pas programmée pour l'immédiat).

15. Le mythe du grec comme langue officielle de la jeune nation américaine

Nombre de Grecs, notamment des Américains d'origine grecque racontent l'anecdote historique suivante dont je n'ai pas aujourd'hui pu vérifier les sources et dont tout semble indiquer qu'elle ne repose sur aucun fondement sérieux :

« Au moment de l'indépendance des États-Unis, il y aurait eu un projet de vote (ou il y aurait eu un vote non abouti) pour désigner la langue officielle du jeune État (on sait que les USA n'ont pas, à ce jour, de langue officielle *de jure*).

53. Savoir reconnaître les lettres et connaître les grandes règles d'écriture : le même savoir-faire que les typographes orientaux de l'Imprimerie nationale.

Le grec aurait, en première approche, recueilli de nombreux suffrages sous le prétexte qu'il était la langue de la première démocratie et qu'il n'était pas une des langues des anciens oppresseurs (l'anglais, l'espagnol, le français). Il est à noter que la Grèce était alors une colonie ottomane et que les émigrants grecs aux USA sont une réalité plus tardive. En deuxième approche, l'anglais aurait été définitivement retenu. »

D'où vient cette histoire ? Est-elle fondée sur une proposition qui n'aurait jamais été véritablement discutée dans une quelconque instance constituante de l'État américain ? Existait-il un *lobby* pour la langue grecque à cette époque ?

Notons qu'il existe une variante allemande du même mythe :

« En 1793, le 3^e congrès refusa de traduire toutes les lois en allemand, demande présentée par des colons germanophones de Virginie, un colon allemand de Pennsylvanie, Muhlenberg, aurait pu faire la différence en votant contre... cependant il ne s'agissait pas de déclarer l'allemand comme seule langue officielle, seulement de traduire les textes. Par contre, les États-Unis ont fait interdire l'usage officiel de langues : le français à l'ouest du Mississippi dès 1803 et en Californie l'espagnol 30 ans après son *rattachement*. »⁵⁴

Ce qui est intéressant dans ce mythe c'est la fiction d'équilibre scriptural que cela aurait entraîné. Les GI seraient-ils devenus des Gamma Iota ?... Mais surtout l'informatique et l'ASCII se seraient développés sur la base d'un autre alphabet en même temps que les équilibres linguistiques ne seraient peut-être pas anglophones. Comme francophones (sinon comme Grecs) nous pouvons toujours rêver !

16. Des pistes d'appropriation pour un système universel d'informatique scripturale

Dès lors que nous avons fait dans nos pays d'écriture latine le constat négatif de nombreuses difficultés techniques, psychologiques et culturelles pour nous approprier le fantastique potentiel que représentent Unicode et ISO/IEC10646, comment pouvons-nous proposer des pistes d'appropriation ? Elles sont nombreuses, diverses, font appel à une approche plurielle multidisciplinaire. En faire le catalogue ressemblera obligatoirement à une liste à la Prévert, mais nous ne devons pas refuser ce désordre constructif.

Une des premières pistes serait que nous puissions culturellement favoriser la circulation entre les langues et les écritures du monde présent et passé. Cela implique que nous légitimions plus qu'elle ne l'est aujourd'hui la partie de la linguistique qui s'intéresse aux sciences de l'écriture et que nous fassions admettre par la communauté scientifique et académique des linguistes, des enseignants et spécialistes de l'étude des langues et écritures (y compris anciennes) et par les

54. Précisions de Patrick Andries.

praticiens de la typographie, de la traductique et de la localisation des logiciels qu'Unicode, ISO/IEC10646 et d'autres outils associés puissent devenir le laboratoire partagé et universel⁵⁵ de ces sciences, disciplines académiques et métiers. Toute nouvelle science exige en effet la construction d'atlas, de corpus balisés reconnus et universellement partagés. C'est le but du site *Écritures du monde* proposé par Michel Bottin.

Les avantages d'une telle attitude seraient nombreux :

- la communauté des experts qui travaillent à l'établissement des tables d'écriture s'en trouverait élargie et confortée et les écritures qu'ils continuent de codifier, d'instrumentaliser informatiquement et d'industrialiser seraient prises en compte de façon beaucoup plus complète ;

- la communauté des linguistes, des historiens de l'écriture, des spécialistes de la calligraphie, des épigraphistes, des paléographes gagnerait énormément à s'impliquer dans cette aventure techno-industrielle.

Le premier gain serait que, au-delà d'une simple approche historique des familles de langues, d'écritures et des associations langue-écriture cet univers serait perçu de façon beaucoup plus large et transversale faisant avancer notre savoir sur ces sujets en s'appuyant sur ce qui est déjà un apport considérable : les classifications pragmatiques et opératoires qui ont déjà été faites par les commissions d'experts et qui sont d'ores et déjà instrumentalisées informatiquement dans un contexte industriel fiable, universel et provisoirement exhaustif.

Ce savoir sur les écritures du monde, leur histoire et leur géographie ainsi que leurs liens réciproques d'emprunt, de généricité, d'abandon au profit d'autres écritures aurait intérêt, lui aussi, à être instrumentalisé informatiquement en liaison concertée avec la prise en compte d'Unicode et de ISO/IEC10646 comme laboratoire académique partagé et universel. Il nous semble, bien sûr, que les communautés linguistiques déjà confrontées au pluriscripturalisme⁵⁶ seront évidemment mieux armées pour développer ce type de savoir pan-scriptural. Ce n'est cependant pas une raison pour abandonner tout espoir de maîtrise occidentale⁵⁷ et européenne sur ces sujets.

55. Nous entendons bien que ce laboratoire n'est pas le catalogue de la norme. Cela sortirait trop des limites de son *scope*. Cependant ce laboratoire s'appuierait sur les travaux de la norme et en retour pourrait apporter de l'information utile aux experts en normalisation.

56. Et mieux au *pluriscripturalisme interne* par exemple les Japonais qui maîtrisent obligatoirement l'idéographie chinoise, leurs syllabaires et l'écriture alphabétique.

57. Rappelons encore que Unicode est né aux États-Unis mais sous l'influence de quelques personnes pour la plupart de souche anglo-saxonne ayant été exposées à ces écritures durant leurs études, comme missionnaires ou de par leur profession. Situation paradoxale, la péninsule indienne (pourtant une mosaïque de langues) n'a pas participé à ces travaux, ce qui les entraîne à se plaindre trop tardivement [précisions de Patrick Andries].

L'habileté que nous pourrions tirer d'expériences pratiques et de l'étude théorique et encyclopédique de ces sujets est indispensable au développement d'usages beaucoup plus simples, ceux auxquels on pense immédiatement :

- un élargissement du commerce mondial et des capacités de pouvoir communiquer dans la langue et dans l'écriture de chacun ;
- une bonne maîtrise des systèmes de documentation translinguistique d'abord en termes d'échange de corpus d'écriture sans les dénaturer, avec pour but second de savoir les exploiter ensuite, les échanger, les retrouver, les traduire, les translitérer et, éventuellement, en donner dans la langue chacune des traductions ou équivalents sémantiques ;
- une capacité de localiser les différents logiciels et systèmes d'information en pensant dès leur conception aux écueils et impossibilités d'équivalence que nous trouverons dans des écritures étrangères.

En effet, de nombreux logiciels fondent l'ergonomie de leurs interfaces en exploitant les primitives morphologiques de notre écriture latine : ainsi celle-ci se compose de chaînes régulières de caractères isolés mais regroupés en mots entre lesquels sont ménagés des espaces. Les lettres peuvent être accentuées et comportent le plus souvent un faible nombre de signes diacritiques, enfin les ligatures sont exceptionnelles. Les styles typographiques sont nombreux, comme dans toutes les écritures mais la communauté informatico-typographique a retenu et popularisé trois styles principaux : gras, italique et romain.

Par contre, la plupart des écritures du monde disposent d'un catalogue de styles différents souvent connotés d'une fonction d'usage spécifique (calligraphie, presse, diplomatie, commerce, écrits intimes, usage religieux, inscriptions lapidaires...) ou d'une périodicité historique. En informatique, ces styles peuvent évidemment être adaptés à condition de ne pas créer de contresens ou d'ambiguïtés culturelles majeures pour distinguer un mot ou une lettre parmi d'autres, mettre une partie de texte *en emphase*⁵⁸, attester une sélection etc. Là encore, une culture même presque uniquement formelle de la pluralité d'écriture sera indispensable non seulement pour les typographes (et plus seulement une élite de l'Imprimerie nationale), mais aussi pour les gestionnaires de sites internationaux, les concepteurs de logiciels, etc.

Notons cependant, pour qu'il n'y ait pas de confusion et de malentendu que, dans leur sagesse les experts de l'Unicode et ceux de l'ISO/CEI 10 646 ont bien circonscrit leur domaine d'intervention et l'on limité au niveau du caractère en laissant à d'autres experts le soin de modéliser et éventuellement de normaliser la production concrète de différents glyphes à partir du caractère abstrait. De ce fait, les questions touchant aux différents styles d'écriture et plus encore aux mécanismes et modes de mise en emphase appartiennent formellement à d'autres domaines techniques qu'il faudra bien développer assez rapidement et sur lesquels le besoin de

58. Ce terme est un anglicisme bâti sur un mot français, mais nous pouvons constater qu'il est passé dans l'usage en tant que vocabulaire technique.

normalisation se fera sentir très rapidement. On doit en profiter pour rappeler que les besoins de mise en parallèle de documents multilingues sont aussi des mécanismes qui auront besoin d'être commodément mis en œuvre et sur lesquels il faudrait aussi disposer bien vite de normes d'interopérabilité.

16.1. *S'inspirer des cultures d'écritures pour inventer des pistes d'appropriation*

Pour inventer un futur dans lequel la circulation multilingue et surtout pluri-écriture prendrait de l'importance il est fondamental d'élaborer des scénarios. Dans le cadre modeste de cet article je me contenterai d'en proposer un seul en prenant la précaution de relativiser mon propos et d'insister sur le statut d'hypothèses des lignes qui vont suivre.

Il importe d'abord de rappeler ce que nous soulignons plus haut, le développement d'un futur *omniécriture*⁵⁹ à large échelle suppose non seulement (bien que ce soit très important) la mise en place et la généralisation d'un code commun à tous les caractères des écritures du monde entier, cela suppose aussi la mise en place de tout un environnement technique, industriel économique et culturel. Pour raisonner en termes d'outils et de fonctionnalités, ce qui est plus simple, rappelons ainsi que nous devons disposer non seulement de traitement de textes réellement multilingues donnant accès à des fonctions simples à mettre en œuvre pour paralléliser deux ou plusieurs textes mais aussi le développement de très nombreux catalogues de référence, d'outils de nature encyclopédique permettant de renseigner les écritures du monde entier, leur logique d'ordonnancement des caractères, leurs règles d'écritures et de mise en page, etc. Cet environnement multiécriture est indispensable pour que se développe une circulation multilingue, mais il va de soi que ces deux marchés et ces deux environnements techniques sont complètement imbriqués. De ce fait les outils de traduction, les dictionnaires électroniques multilingues, mais aussi les vocabulaires multilingues techniques, scientifiques ou commerciaux participent aussi du foisonnement de toute une pléiade de logiciels et de documents susceptibles d'augmenter dans une très forte proportion le nombre des sollicitations de communication multilingue et multiécriture. Cet environnement notamment le dernier point ne se fera pas tout seul et suppose évidemment que le marché de la circulation multilingue, et pour ce qui importe ici multiécriture, s'accroisse⁶⁰.

Par exemple, un opérateur d'import-export qui ne traite qu'en anglais et en translittération l'importation ou l'expédition de ses transactions à l'autre bout du monde, sera dans une position fort différente s'il dispose à chaque fois de la

59. En employant ce terme je tiens à marquer la nécessité à terme d'une prise en compte exhaustivement universelle des écritures, y compris les écritures non linguistiques (chorégraphie...) ou paralinguistiques (langue des signes...).

60. Nous comprenons bien, évidemment que nous sommes ici face au paradoxe bien connu : est-ce la poule qui fait l'œuf ou l'œuf qui génère la poule ?

visualisation des équivalents en écriture locale. On peut parier sans crainte de trop se tromper, qu'un transitaire spécialisé sur la Chine, par exemple, finira, assez vite, par mémoriser les idéogrammes des principales villes chinoises et ceux correspondants aux noms de ses principaux correspondants. Il mémorisera aussi vraisemblablement un certain nombre de concepts (douane, embarquement, facture, etc.).

Ce que nous avons esquissé pour le métier de l'import-export, nous pourrions le décliner dans des domaines professionnels très divers : les bibliothèques, la police, l'enseignement, la santé, la recherche scientifique, le tourisme. A chaque fois la spécificité des métiers et le caractère spécialisé de ceux qui auront intérêt à échanger des données multilingues induiront obligatoirement des curiosités pour apprendre non pas forcément la totalité de la langue des autres, mais pour le moins savoir reconnaître à l'écrit quelques concepts-clés.

En effet, si nous nous projetons dans un futur où il est facile de traiter, de stocker et de faire circuler des documents multilingues et multiécritures, il deviendra inadmissible que l'on puisse être compétent dans tel ou tel domaine (commercial, technique, culturel) touchant une aire linguistique non d'écriture latine et être incapable d'en reconnaître le moindre mot. Projeté à 50 ans dans le futur il est vraisemblable que cela sera vu comme une insuffisance notoire de notre époque qui a vu émerger les premiers balbutiements de la circulation planétaire globale (tant pour les personnes que pour les marchandises et l'information), mais qui n'aura pas su s'adapter assez rapidement à la situation multilingue générée par les réseaux. Cette très faible culture de la langue et de l'écriture des autres est un point faible de la culture de notre époque qui est sans doute amenée à évoluer. Comment cela se fera-t-il concrètement dans le futur ? C'est sans doute sur ce point que nous aurons beaucoup de surprises mais que nous ne pouvons aujourd'hui pousser plus avant notre scénario d'évolution.

De notre point de vue, toute une série de caractéristiques fonctionnelles des écritures doivent être considérées comme des stimulants du raisonnement dans la mesure où ils participent du dispositif de mise en évidence de la pensée dans une culture et une langue donnée. De ce point de vue, les dispositifs des écritures idéographiques et des écritures mixtes (idéo-syllabiques) doivent être considérés avec beaucoup d'attention. Dans notre scénario propositionnel pour accélérer un processus de croissance des communications multilingues nous voudrions qu'une importante réflexion soit faite (non pas pour faire apprendre le chinois et le japonais à tous, c'est parfaitement irréaliste), mais pour extraire de la culture d'écriture chinoise un certain nombre de principes fondateurs et de données de base et les organiser pour constituer un aperçu global des grandes écritures du monde aussi important, dans le contexte émergent de communication planétaire, que l'aperçu géographique et cartographique qui nous est obligatoirement inculqué au cours de nos études.

Nous pouvons néanmoins ouvrir quelques autres pistes de détail simplement pour stimuler l'imaginaire de ce futur multiscritural.

16.2. Les balises du Coran et les DTD du solfège

Il existe, en arabe, plusieurs symboles religieux musulmans qui ont été codifiés dans le PMB. Ceux-ci ont essentiellement une fonction de balisage sémantique et sémantico-ordinal en ce sens qu'ils déterminent non seulement le début et la fin des versets à l'intérieur des sourates mais aussi leur numéro d'ordre. Un autre caractère note aussi l'arrêt de lecture pour introduire un instant de prière⁶¹. À l'évidence, une telle pratique scripturale peut être vue comme une préfiguration de ce qui est actuellement réalisé avec des langages à balises. Les calligraphes qui ont inventé ces caractères ont en effet perçu qu'il fallait mettre en évidence certains endroits stratégiques du texte sans pour autant polluer la lecture en introduisant un énoncé technique, un ordre exprimé par un texte écrit. En effet, ils auraient pu écrire : « ici nous repartons sur un autre chapitre du Coran et on peut prévoir de s'arrêter ici pour reprendre la lecture un jour prochain », ou encore : « ici s'arrêter de lire et prier un instant ». Ils pensaient, à juste titre que s'ils s'étaient tenus à ces énoncés écrits, ils auraient introduit un niveau second du texte parasitant le texte sacré. Ils ont donc inventé le concept de balise.

Il n'y a pas si longtemps, les chrétiens aussi introduisaient des caractères typographiques en forme de croix (généralement rouges) qui punctuaient les missels et les livres de prière. On proposait ainsi aux prêtres et aux fidèles d'interrompre la lecture par un signe de croix, une gémflexion ou un instant de prière silencieuse. Quand on considère ainsi avec un certain recul les différentes *cultures d'écriture* on s'aperçoit aussi que l'écriture musicale est de ce point de vue beaucoup plus riche parce que beaucoup moins linéaire qu'un texte alphabétique.

Toute proportion gardée, la clé, et les premières indications qui précèdent la première portée d'un texte musical (par exemple un bémol à la clé), constituent une sorte de DTD de la partition qui suit. Mais l'écriture musicale, pour s'en tenir à la notation que nous connaissons, présente bien des qualités préfigurant une culture d'écriture déployée sur un large éventail de fonctionnalités. Ainsi les règles du solfège permettent : le parallélisme synchronique de plusieurs lignes d'écriture correspondant à des voix différentes (partition d'orchestre), à des niveaux différents de modalité (le texte d'un chant et sa ligne mélodique) mais laisse encore la place à un balisage local (*andante, moderato cantabile, etc.*). On est proche de la nécessité de mettre en parallèle plusieurs langues dans le même texte !

Il faut aussi noter que les notations mathématiques ou chorégraphiques sont sans doute encore beaucoup plus riches en innovations de la scripturalité.

61. Voir l'article de Rachid Zghibi.

16.3. *L'idéographie ouvre des pistes*

Si les écritures idéographiques présentent un certain nombre d'inconvénients (par exemple pour ce qui est de leur apprentissage), en contrepartie elles offrent nombre d'avantages notamment sur le plan cognitif. Nous sommes là, en effet, face à un matériel d'écriture propre à exciter l'imagination et, sans doute, à élargir le catalogue des modèles offerts aux concepteurs informatiques et à nombreux autres utilisateurs de modèles formels (les documentalistes, les enseignants par exemple, puisque dans une certaine mesure l'ordre des caractères chinois est aussi un ordre logique du lexique).

Cette connaissance même superficielle de l'organisation, de la généricité et de la sémantique étymographique des caractères chinois devrait être enseignée systématiquement à ceux qui sont chargés de produire des idéogrammes dans des interfaces informatiques et, plus largement, dans toute signalétique. Malheureusement, nombre de graphistes ne sont absolument pas formés ou informés de ces logiques. Ils n'ont trop souvent cure de ces principes structurels et sémantiques de composition idéographique et, de ce fait, ils produisent des pictogrammes sans structure ou dont la structure est mal articulée. Notons que, de ce point de vue, la signalisation routière s'est construite collectivement et par adjonction et sélection naturelle sur la longue durée. Elle constitue un bon exemple d'*idéographie compositionnelle* articulée sur quelques principes généraux du code traduit par des éléments ou formants graphiques ou colorés qui viennent signifier des énoncés simples mais eux aussi, linguistiquement articulés : par exemple, dans la famille générique des interdictions caractérisée par la couleur rouge organisée selon une sous-typologie en bordures rouges de forme ronde ou triangulaire, viendra, par exemple, s'articuler la catégorie des injonctions chiffrées qui, selon leur forme et leur contexte d'insertion, seront une limitation de vitesse, une fin de limitation, une interdiction de dépasser un certain tonnage, une largeur ou une hauteur limite de la voie routière, etc.

Par contre, si comme c'est trop souvent le cas ils continuent de se développer anarchiquement, selon la seule imagination graphique des différents concepteurs de signalétique, il deviendra socialement impossible de maîtriser cette profusion de codes et de logos.

L'existence du code de signalétique routière et l'évidence qu'il s'est développé majoritairement sous l'influence occidentale devrait nous rassurer sur notre capacité sociétale et culturelle à pouvoir, même en Occident uni-alphabétique, disposer d'aptitudes cognitives idéographiques. Nous pouvons apprendre, dès notre plus jeune âge, (mais aussi vraisemblablement à l'âge adulte), des milliers, voire des dizaines de milliers de codes à condition qu'ils soient intelligemment articulés dans leur symbolique iconique et idéographique.

16.4. *Quelle capacité avons-nous pour organiser et symboliser des univers abstraits ou concrets que nous voulons documenter et parcourir à loisir ?*

Même si les commissions de normalisation du codage des caractères (Unicode ou ISO/CEI 10 646) ont limité par construction leurs champs d'action aux caractères stricto sensu, il reste que nous pouvons nous interroger. Que ferons-nous des milliards de codes possibles ? Sans doute y construire des catalogues univoques socialement utiles ? Comment pourrions nous nous y prendre ? Saurons-nous associer l'Extrême-Orient idéographique et l'Occident alphabétique dans la construction commune de tels codes ? Ce serait certainement, si cela se fait, un véritable cadeau interculturel et interlinguistique qu'offriraient en se mariant les deux grandes principales familles d'écriture aux générations futures.

La question technophilosophique qui consiste à chercher à savoir si cela n'a rien à voir avec l'écriture ou aurait au contraire toute pertinence à y être associée est importante. Avant de répondre, nous devons nous assurer que nous raisonnons sans préjugé alphabético-centrique.

Pourtant, à ce jour, nous n'avons pas démontré que l'on pouvait organiser avec cohérence quelques milliers de codes de commande ou de balisage multimédia qui pourraient être symbolisés par des codes univoques à la fois iconiques et numériques. Et pourtant, nous reconnaissons l'extrême utilité d'un code comme « @ » qui, à ce jour, a commencé à trouver une place réellement spécifique à la fois comme articulation et comme balise signifiante d'une adresse de courrier électronique dont cet ancien caractère est devenu le symbole presque univoque.

Les musiciens, les mathématiciens, les chimistes, les biologistes... et les usagers du code de la route démontrent, tous les jours, que les besoins sont importants et qu'il est sans doute urgent de trouver des solutions internationalement non ambiguës⁶².

Bien sûr nous savons qu'au prix de beaucoup de virtuosité dans la conception du format, mais aussi de presque autant de virtuosité pour l'utilisateur final, les habitués de TeX savent ainsi utiliser des multicoles ou combiner une infinité de cassetins avec les seuls 128 caractères ASCII pour coder tous les caractères du monde. Par exemple \e permet d'avoir le é, \mu permet d'avoir le μ, \ae d'avoir æ et on peut même coder des hiéroglyphes⁶³.

62. Nous pouvons même remarquer que dans tous ces domaines, notamment le code de la route, les usagers arrivent très vite à s'adapter et décoder des codes qui connaissent pourtant des variations nationales. On peut d'ailleurs remarquer que pour le code de la route il serait sans doute urgent de parvenir à pouvoir symboliser tout ou presque tout de façon iconique parce qu'il faut bien constater que dans certains cas des informations absolument vitales ne sont annoncées que par un texte en langue nationale (et pourtant des chauffeurs de poids lourds iraniens, finlandais ou bulgares traversent nos continents).

63. Ces précisions m'ont été suggérées par Jacques André.

On comprend que dans l'étroitesse de la norme ASCII, les informaticiens n'aient pas réussi à s'entendre et n'aient pas vu l'utilité de donner une forme affichable aux caractères de commande. Cela laissait la place libre pour d'autres visualisations, pour attribuer d'autres fonctions aux mêmes codes... Mais déjà sur un code à 2 octets, les informaticiens et les spécialistes multimédias devraient pouvoir s'entendre pour concevoir et organiser une idéographie articulée internationalement reconnue. Cela aurait, d'ailleurs, comme effet second d'organiser en un véritable solfège, dans une véritable écriture signifiante de leurs principaux axes théoriques, tous leurs concepts et les grandes catégories fonctionnelles, sémantiques, structurelles et référentielles que la discipline informatique nous permet de rendre réels dans notre technoculture.

Du point de vue du grand public cultivé qui s'intéresse au développement de la technoculture informatique mais pas forcément au détail de ses *habitus* méthodologiques, cette question des codes de commande peut faire débat. Je comprends parfaitement que pour un puriste de l'informatique il n'est pas admissible qu'un registre qui sert à coder des caractères serve aussi à coder des commandes de base. Pourtant, que ce soit sur ce registre, ou sur un autre, que ce soit confié, bien sûr, à d'autres commissions de normalisation il apparaît intéressant que l'on puisse disposer sur un code largement partagé de toute une série de commandes, de balises de bases et pourquoi pas de codes essentiels ou d'universaux fondamentaux (les éléments chimiques par exemple). De ce point de vue, je repose avec insistance la question : que ferez-vous des 12 bits restants sur 32 non réservés par Unicode et l'ISO/CEI 10 646 ?

Sans doute cette vision par trop concrète et iconolâtre (plutôt qu'iconoclaste !) ne correspond pas précisément à la culture habituelle du chercheur en informatique. Il lui paraît logique que ce qui a été défini selon une sémantique univoque, puis qui a été structurellement organisé et référencé sera, de toute façon, réutilisable en machine. Toute codification parallèle est *a priori* redondante et inutile.

Et pourtant, les grandes communautés professionnelles qui se rencontrent dans les NTIC sont à la fois les informaticiens, les gens d'images, les professionnels de la documentation et les linguistes. Nous avons sans doute là les premiers éléments de l'association interdisciplinaire susceptible de nous construire non seulement une interécriture planétaire mais, reprenons le terme, comme une image, un solfège iconique universel de nos primitives interculturelles et de nos listes d'autorités⁶⁴.

Nous devrions pouvoir être capables de construire des consensus après nous être entendus sur les véritables principes efficients de l'idéographie, de la schématique et de la signalétique ; nous pourrions sans doute entreprendre, dès lors, la construction

64. On peut être en effet fasciné que les musiciens aient pu inventer des codes de notation apparemment complexes (mais très correctement articulés) qu'ils savent parcourir avec *maestria*. Ils ont aussi mis au point des interfaces homme/machine hautement improbables si on sait qu'un simple piano offre un nombre de touches supérieur à la plupart des claviers informatiques sans aucune autre indication explicite que leur organisation structurelle et graphique (touches blanches et noires).

d'une écriture à l'échelle de notre technoculture certes multiculturelle mais aussi planétaire.

Ce type de raisonnement existait déjà en germe dans le texte même de la norme ISO/IEC 10 646 puisqu'il y est signalé que nous disposons actuellement d'un catalogue déjà conséquent de schémas, logotypes, codes divers pour l'électricité, la physique, l'astronomie...

Ce débat achoppe, en effet, sur des questions de principe édictées par les informaticiens qui ont défini avec peut-être trop de rigueur et d'économie⁶⁵ le catalogue des caractères du PMB⁶⁶ ainsi que les premiers plans complémentaires activables facilement avec un mécanisme de sériation UTF16 ou UTF32⁶⁷. Mais la même rigueur d'économie de catalogue doit-elle être appliquée dès lors qu'on définit des plans de codifications sur quatre octets ?

Ce qui me paraît certainement le plus nécessaire, c'est de définir les différents modes de production de codes graphiques, schématiques et idéographiques. Si on devait se lancer dans l'enregistrement codifié et normalisé des logos et de la signalétique, il faudrait cartographier ces différents domaines avec autant de rigueur que cela l'a été pour les codes d'écriture. Il faudra construire leur typologie et préciser, catégorie par catégorie, leur potentiel d'usage, leurs champs d'application, leur mode éventuel d'articulation graphique. Cela revient en quelque sorte à organiser une théorie du domaine dans laquelle la forme concrète du logo, aurait statut de glyphe et dans laquelle sa structure abstraite correspondrait au caractère. C'est d'ailleurs la voie dans laquelle s'est déjà lancé un chercheur de l'UQAM (Université du Québec à Montréal) comme Jacques Cartier mais qui pourrait être en partie revue sous l'éclairage particulier des concepts d'Unicode. Les informaticiens lorsqu'on leur expose ces problèmes (ainsi d'ailleurs que les typographes informaticiens) ont trop souvent tendance à nier l'utilité en soi de la création de tels catalogues graphiques internationaux pour schématiser des primitives fonctionnelles ou intellectuelles. On peut comprendre les raisons de leurs réticences à ce que s'ouvre une telle boîte de Pandore. Deux milliards de codes c'est beaucoup à l'échelle des primitives et catalogues d'autorité de la technoculture de l'humanité si

65. Je crois en effet après avoir entendu les professionnels qu'il y a là un véritable débat dont les principes sont importants pour la cohérence globale du catalogue mais en gardant à l'esprit que chacun des principes appliqués, si j'ose ce mauvais jeu de mot *à la lettre*, risque d'avoir des conséquences culturelles importantes. Qu'on se souvienne par exemple du débat décrit par Nadine Lucas à propos de la codification des caractères chinois : fallait-il comme l'auraient voulu certains experts extrême-orientaux coder les idéogrammes à partir de leurs composants ou, comme ce fut décidé, les coder de façon globale sans en donner les séquences de composition ?

66. Les professionnels de la typographie se plaignent d'ailleurs de ne pas avoir été assez consultés (voir article de Olivier Randier par exemple).

67. Voir l'article de Patrick Andries.

l'on choisit avec mesure les domaines que l'on voudrait coder de façon unitaire⁶⁸, univoque et universelle. Par contre, ce même espace pourrait être rapidement gaspillé et saturé si l'on décidait d'attribuer ces codes à des usages individuels ou à la libre initiative de l'achat d'espace.

Ce dernier sujet abordé me paraît à la fois marginal (codifier ce qui n'est pas de l'ordre des écritures attestées et des codes graphiques faisant autorité est évidemment hors sujet des commissions de normalisation dont il est question ici) mais au cœur de l'évolution technoculturelle indispensable pour survivre et se développer harmonieusement dans une culture fondée principalement sur la communication.

La mise en ordre et en cohérence de nos différents codes et symboles devient en effet de jour en jour plus cruciale. D'une part il est évident, normal et sain que nous ayons de part le monde de grandes différences culturelles dans ce que nous considérons cependant comme des codes de base, d'autre part, il est par contre indispensable que nous ayons catalogué et décrit dans le détail toutes ces différences et que nous nous en soyons nourris pour éventuellement améliorer notre capacité à nous entendre par-dessus les frontières linguistiques et culturelles. De ce point de vue, la frontière alphabétique/idéographique est une marque d'altérité importante et pleine de promesses. Il est en effet indéniable qu'une part importante de notre culture et de notre technoculture, procède historiquement de ce choix fait à l'âge des Lumières : la neutralité du regard telle que nous la concevons en Occident, est induite par l'encyclopédie alphabétique avant d'être méthodique. Cependant dans notre culture nous savons difficilement articuler une *logographie*, mais il serait abusif de déduire que c'est exclusivement la conséquence de notre prédisposition *alphabético-centriste*. La future convivialité avec la culture idéographique qu'induirait obligatoirement la codification universelle de l'écriture devrait favoriser l'émergence d'une culture métissant les aptitudes propres à chacune d'elles.

Mais d'autres cultures ou savoir-faire sectoriel d'écriture devraient nous aider. Ainsi, l'écriture musicale occidentale et son ensemble de règles grammaticales (ou solfège) qui préfigure sur bien des points la forme et la syntaxe de langages comme XML. On peut trouver de nombreux autres exemples.

La promesse que se développe une *aptitude culturelle superlative* de l'utilisation des codes graphiques et d'écriture sera de mon point de vue une conséquence quasi certaine du développement des technologies multiscripturales. C'est avec le multimédia et les réseaux la troisième facette fondamentale du *nouveau système technique* qui se met en place et qui aussi certainement que l'a fait en son temps la *galaxie Gutenberg* provoquera des évolutions très importantes de notre technoculture, de nos capacités cognitives et communicationnelles. C'est pour cela que nous devons veiller à favoriser l'appropriation de ces techniques aussi souvent et avec autant de détermination que nous le pouvons.

68. En effet, une simple articulation sur 2 unités de code à 4 octets ouvre le potentiel de codage sur des grandeurs insondables.

17. Conclusion

L'appropriation technique et sociale de l'Unicode est certes un enjeu primordial doublé d'une interrogation pour que puisse se développer une informatique planétaire communicante, mais les questions ouvertes vont bien au-delà. Il apparaît d'année en année plus évident que nous sommes condamnés, pour survivre à quelques milliards d'humains sur une si étroite planète, à réussir un humanisme laïc en parvenant à faire communiquer toutes les cultures, toutes les langues, tous les particularismes et toutes les individualités sur un espace de dialogue ouvert dans lequel la différence de l'autre (ou des autres) ne doit pas nous appauvrir mais nous enrichir. Le respect réciproque de l'étranger, de celui qui n'écrit pas pareil, qui ne parle pas la même langue et n'a ni les mêmes croyances ni les mêmes coutumes constituent les fondements mêmes des principes d'universalité humaniste. Le web, mais aussi l'Unicode et les technologies qui permettent de coder toutes les écritures dans un universalisme lié à un encyclopédisme opératoire constituent des éléments fondamentaux pour favoriser les échanges interculturels et interlinguistiques entre les peuples du monde.

Le taoïsme et la philosophie socratique, ne sont pas nés par hasard en Chine et en Grèce quelques cinq siècles avant le Christ. Ils sont nés en des temps et en des lieux où se faisait jour l'idée et où grandissait l'espoir que l'écriture pouvait précisément participer de la communication interculturelle et interlinguistique. En Chine c'était l'idéogramme qui fédérait des populations parlant des langues différentes mais comprenant la même écriture. En Grèce s'affirmait l'idée que l'assemblée démocratique de la cité, sa fédération olympique pouvait s'élargir à une beaucoup plus large échelle si une seule écriture permettait de phonétiser toutes les langues du monde, donc d'ouvrir le dialogue interlinguistique en permettant qu'il s'établisse sur une même base écrite. Cette analyse et cette foi en l'Universel est notamment développée dans le *Cratyle* de Platon.

Certes, il fallait ensuite traduire des énoncés transcrits en diverses langues sur le même alphabet, mais les premières pistes d'une utopie universaliste étaient ouvertes⁶⁹.

À la Renaissance, l'imprimerie a permis que se diffuse la totalité des écrits de L'Humanité. En cela on peut dire que cette technique a placé sur un pied d'égalité à la fois tous les écrits des religions révélées et les multiples essais de réflexion philosophique métaphysique et scientifique. De là naquit dans ce nouvel ensemble technique de l'écriture induit par l'imprimerie une deuxième et plus ample proposition humaniste.

Nous sommes à l'orée de la mise en place cohérente et interopérante de diverses techniques translinguistiques qui constitue sans doute les composants principaux de la transformation radicale du système technique de l'écriture imprimée qui devient

69. Cette facilité était notamment très utile pour le commerce maritime.

mode universel d'encodage numérique. Nous ne savons pas de quoi et comment sera fait l'avenir mais nous savons que nous ne devons pas ménager nos efforts pour perfectionner chacun des composants de cette technoculture multilinguistique dont l'Unicode et ISO 10 646 sont des éléments essentiels. André Malraux affirmait que le XXI^e siècle serait religieux ou ne serait pas ; souhaitons qu'il le soit en effet au sens étymologique du terme : un siècle religieux non pour enfermer chaque communauté dans le particularisme de ses dogmes, de ses diversités linguistiques et culturelles, mais au contraire pour relier les hommes entre eux et les enrichir de leurs différences.

18. Bibliographie

- Claudel (Paul), *Cent phrases pour un éventail*, Paris, Gallimard, 1942. Première édition : Tokyo, 1927.
- Derrida (Jacques), *De la grammatologie*, Paris, éd. de Minuit, 1967.
- El Bokari, *L'authentique tradition musulmane : choix de h'adiths*, Paris, éd. Fasquelle, 1964, (25-17), p. 54.
- Eisenstein (Elizabeth), *La révolution de l'imprimé dans l'Europe des premiers temps modernes*, Paris, éd. la Découverte, 1991.
- Gelb (I. J.), *Pour une théorie de l'écriture*, Paris, Flammarion, 1973.
- Gille (Bertrand), *Histoire des techniques, techniques et civilisations, techniques et sciences*, Paris, éd. Gallimard, Encyclopédie de la Pléiade, 1978.
- Griollet (Pascal), *La modernisation du Japon et la réforme de son écriture*, Paris, Publications orientalistes de France, 1985.
- Hudrisier (Henri), *La norme de codage des caractères sur plusieurs octets (ISO/IEC 10 646) un enjeu pour le Sommet de Hanoi*, in MEDIA'COM inforoutes 97, Aupelf, Beyrouth, avril 1997.
- Hudrisier (Henri), *Champollion 2000, une encyclopédie et une base de ressources typographiques normalisées sur les écritures du monde*, Cahiers de la Francophonie, oct. 1997.
- Hudrisier (Henri), *Des abacistes aux algoristes : les risques potentiels d'un hypermédia mal compris*, in Passerelles n° 24 (spécial NTIC), revue de l'Université de Paris 8, fév. 1999.
- Hudrisier (Henri), *L'ère des machines grammatologiques, la normalisation des technologies de l'information comme attracteur de leur convergence*, mémoire d'habilitation à diriger des recherches, Université de Paris 8, 2000.
- Ifrah (Georges), *Histoire universelle des chiffres, l'intelligence des hommes racontée par les nombres et le calcul*, Paris, éd. Robert Laffont, 1994.
- Lexique des règles typographiques en usage à l'Imprimerie nationale*, sans auteur, Paris, Imprimerie nationale, 1990.

Mac Luhan (Marshall), *La galaxie Gutenberg*, Paris, éd. universitaires Jean-Pierre Delarge-Mame, 1967.

Manguel (Alberto), *Une histoire de la lecture*, Arles, éd. Actes Sud, 1998, p. 60.

Marti (Bernard) et coauteurs, *Télématique - techniques, normes, services*, Paris, Dunod, 1990.

Perriault (Jacques), *La logique de l'usage essai sur les machines à communiquer*, Paris, éditions Flammarion, 1989.

Rifkin (Jeremy), *L'âge de l'accès*, la révolution de la nouvelle économie, Paris, éd. la Découverte, 2000.

Saussure (Ferdinand), *Cours de linguistique générale*, éd. Critique par Tullio de Mauro, Paris, Payot, 1972 (1^e édition : Genève 1916), p. 107 puis p. 48.

Simondon (Gilbert), *Du mode d'existence des objets techniques*, Paris, éd. Aubier, 1958, p. 12- 13.

Solaris, revue en ligne de l'URFIST de Paris, numéro spécial sur la normalisation, oct. 1999.