

L'INFORMATISATION DES SYSTÈMES D'ÉCRITURE

Alain VUILLEMIN

Il est des révolutions technologiques dont on ne s'aperçoit pas. En quatre décennies et en moins de cinq générations de matériel, l'informatique est parvenue à conquérir presque tous les systèmes d'écriture jusqu'à devenir à son tour une autre écriture, la première écriture universelle peut-être. En 1989, nul n'en avait encore écrit l'histoire depuis le début des années 1950. Entre 1946 et 1951 en effet, les premiers ordinateurs n'étaient que des calculateurs, qui n'étaient capables d'opérer que sur des nombres et les dix chiffres du système décimal. En 1984, la firme américaine Xerox avait inventé un système particulier, le système Star, capable de traiter plus de 16 millions de caractères typographiques et idéographiques différents, soit pratiquement la quasi totalité des systèmes d'écriture qui existent. En 1990, de nouvelles normes internationales de description de caractères, limitées à une soixantaine de milliers de ces caractères typographiques et idéographiques, étaient en cours d'élaboration. Les progrès accomplis sont impressionnants. Aujourd'hui, tous les systèmes d'écriture, y compris sous leur forme manuscrite, sont susceptibles d'être informatisés, numérisés, générés, retraités ou reconnus.

I. LA NUMÉRISATION DE L'ÉCRITURE

C'est évidemment par métaphore qu'il faut comprendre le fait qu'un ordinateur sache désormais "écrire" et même "lire". Dans le secret du fonctionnement de son unité centrale, un ordinateur n'est toujours en effet qu'un calculateur. Il ne sait travailler que sur des nombres et seulement dans un système de numérisation binaire. Mais, quelles qu'aient été les langues concernées et les démarches de raisonnement ou de traitement adoptées, le principe général de la "numérisation" ou encore de la "digitalisation" de l'écriture a historiquement consisté à ramener de proche en proche tous les traitements exigés à des opérations effectuées sur des nombres exprimés à l'aide des deux seuls chiffres "zéro" et "un". Pour y parvenir, on a recouru dès 1950 à des systèmes de

codification arbitraire, longtemps différents d'un constructeur à un autre, qui se sont cependant lentement perfectionnés et normalisés. C'est ainsi qu'ont été progressivement numérisées les écritures alphabétiques, d'abord en typographie pauvre dès 1951 puis en typographie riche à partir de 1965 avec l'adoption des codes Ascii et Ebdic. Les écritures consonantiques, comme celle de l'arabe par exemple, ont commencé de l'être dès 1960, et ont abouti depuis à l'élaboration de premières normes internationales unifiées. Les premières recherches sur des modes de transcription informatisée des caractères idéographiques japonais et chinois remontent au début des années 1950 mais ce n'est qu'entre 1978 et 1982 que sont enfin adoptées de véritables normes, distinctes toutefois entre le Japon, la Chine continentale et la Chine nationaliste. L'informatisation des systèmes d'écriture phonétiques, tels que ceux qui sont utilisés dans la recherche sur l'histoire des langues écrites ou sur les caractéristiques des langues sans écriture, celles des civilisations orales, a suivi enfin des cheminements parallèles et, dès 1981, de premiers alphabets phonologiques universels informatisés avaient été conçus aux États-Unis. Mais le rythme de tous ces progrès a été inégal. Il faut le rappeler, aucun des matériels informatiques utilisés pour numériser ces systèmes d'écriture n'était adapté au départ. Il fallut attendre 1975 environ pour que les perfectionnements apportés aux équipements par les constructeurs permettent aux recherches de progresser. Le désordre des normes nationales et internationales de transcription qui ont été adoptées depuis en sont encore un frein. Il n'empêche que par le détour de ces processus de numérisation de l'écriture, l'informatique était devenue en 1989, en puissance du moins, un mode d'écriture nouveau et le dénominateur commun de tous les autres systèmes d'écriture.

II. LA GÉNÉRATION DE L'ÉCRITURE

Une fois décomposée à l'intérieur des dispositifs physiques qui constituent les circuits internes de traitement d'un ordinateur, une écriture numérisée se retrouve transcrite selon les techniques utilisées en des successions d'impulsions électriques, de phénomènes électro-niques, d'états magnétiques ou d'incrustations opto-magnétiques qu'il faut être capable de retraduire à l'inverse en des perceptions visuelles ou sonores qu'un être humain puisse déchiffrer. La "génération" de l'écriture désigne en termes informatiques cette reconstruction, effectuée par l'intermédiaire d'autres dispositifs spécialisés, les "éditeurs" ou les "générateurs" de caractères. Tel est le principe général de ce processus de

reconstitution, qui vaut quelle que soit la nature des écritures concernées. En ce qui concerne les caractères typographiques, la démarche repose sur une correspondance stricte entre les types de caractères qui sont reconstitués sur les écrans de visualisation ou sur les dispositifs d'impression et les nomenclatures standard ou normalisées qui définissent les jeux de caractères des codes de référence alphanumériques que choisissent les constructeurs. Ces codes varient d'un constructeur à un autre. Pendant longtemps, les ordinateurs se sont contentés d'afficher 48 puis 64 et 96 caractères. En 1965, les codes Ascii et Ebcdic ont permis de représenter 128 et 256 caractères respectivement. Depuis ces codes sont devenus multilingues et, à partir de 1979, sont apparus les premiers logiciels de description de caractères et de création de fontes ou de polices typographiques qui permettaient de modifier la forme même de ces caractères. La génération des caractères idéographiques recourt en revanche à des techniques et aussi à des équipements informatiques plus complexes. Les idéogrammes sont assimilés à des représentations figurées, à des dessins stylisés dont le nombre, le sens et l'ordonnance sont définis par les règles de la calligraphie extrême-orientale. Sur les 60 000 caractères idéographiques chinois ou japonais qui ont pu être recensés, seuls 10 000 sont usuels et 4 000 très fréquents. Ces caractères sont ramenés à 200 tracés fondamentaux et recomposés par les ordinateurs, aussi souvent qu'il est nécessaire, à partir de ces tracés élémentaires grâce à des éditeurs de caractères spécialisés. Des normes ont été progressivement adoptées, pour chacune des langues extrême-orientales, aussi bien pour les caractères pi-ying chinois, kanji japonais ou hangful coréens. L'on peut même modifier, recréer ou inventer des polices d'idéogrammes. La génération des caractères vidéographiques par les systèmes dits "d'acquisition numérique et de télévision d'images organisées en pages d'écriture", les systèmes Antiope ou Vidéotex en France, font appel par contre à d'autres techniques. Les pages-écrans employées sont décomposées en une succession de matrices emboîtées, de 24 lignes sur 40 colonnes, de manière à déterminer une première grille de 960 cases, décomposables à leur tour en damier de 6 à 8 carrés. Il en résulte une grille de composition, à interligne et à chasse fixes, dont il suffit de noircir ou de colorer les surfaces élémentaires pour reconstituer des caractères. Les codes de représentation correspondants étaient loin d'avoir été normalisés sur un plan international en 1989, même si de premières normes européennes avaient été adoptées dès 1987. Il n'est pas enfin jusqu'à la génération de calligrammes qui ne recoure à d'autres instruments informatiques, des tablettes ou des palettes graphiques,

dérivés de l'infographie. Ces instruments sont devenus très variés. Ils présentent la particularité de retrouver, de conserver, de prolonger ou d'améliorer le geste même du calligraphe par l'intermédiaire de toutes sortes d'effets pré-définis de transformation de taille, de forme, de dimensions, d'échelle, de position, de coloriage ou de texturage des calligrammes. Les logotypes animés ne sont qu'une préfiguration de ce que pourraient être, demain, des systèmes d'écriture calligraphiques animés, dynamiques et interactifs.

III. LA RECONNAISSANCE DE L'ÉCRITURE

Avec la reconnaissance de l'écriture se produit une autre révolution, intellectuelle plus que technologique. L'expression de "reconnaissance de l'écriture" désigne le fait que les ordinateurs sont devenus progressivement capables d'identifier des caractères écrits, manuscrits et imprimés, de les "reconnaître" et de les transformer immédiatement en telle ou telle représentation numérisée selon les codes utilisés. Les tout premiers systèmes de reconnaissance de caractère par des procédés de lecture magnétique ou optique ont été mis au point entre 1950 et 1960. C'étaient des systèmes de saisie indirecte, qui exigeaient que les caractères à identifier eussent été soit imprimés ou dactylographiés auparavant à l'aide d'encre magnétisable soit convertis au préalable en codes à chiffres ou à barres. Ces systèmes de lecture optique sont devenus directs entre 1960 et 1970, mais nécessitaient toujours que l'on eût recours à des normes de transcriptions dactylographiques ou typographiques particulières, spécialement conçues à l'époque pour faciliter ces modes de saisie nouveaux. À partir de 1980, l'apparition de langages informatiques inédits de description de caractères permit d'assouplir l'utilisation de ces systèmes d'identification. Dès 1985, la plupart des polices de caractères alphabétiques étaient susceptibles d'être ainsi décrites et identifiées et, dès 1987, les polices de caractères idéographiques japonaises ou chinoises l'étaient à leur tour. La reconnaissance de l'écriture manuscrite a accompli de même des progrès spectaculaires, en particulier depuis 1980, jusqu'à permettre sa saisie dans l'instant même de sa transcription par le biais de tablettes de télé-écriture qui se confondent parfois avec l'écran même de visualisation. Ces systèmes ont commencé à se répandre depuis 1988. Ils permettent désormais d'effectuer à volonté la synthèse de l'écriture manuscrite de tout un chacun. L'apparition également, depuis 1988, des premiers systèmes de traitement de texte (et donc d'écriture) à entrée vocale a

enfin achevé de bouleverser les frontières admises entre l'écrit et l'oral. Ces systèmes permettent de produire en effet directement des signes écrits, des mots et des phrases, à partir de leur énoncé oral. En ce domaine, les progrès ont été lents pourtant. Les premiers systèmes de reconnaissance des mots qui aient été réellement opérants remontent au début des années 1970. En 1982, ces systèmes étaient déjà devenus capables de reconnaître entre une centaine et un millier de mots prononcés en différentes langues, de l'anglais au japonais en passant par le russe, l'espagnol, l'allemand ou l'italien. Ils réussissaient à identifier entre 3 000 et 5 000 mots en 1985, plus de 20 000 mots en 1988. Ces chiffres et ces dates donnent une idée de la rapidité des progrès récents de ces systèmes. L'ambition proclamée de ces recherches est immense puisqu'il s'agit de rendre les ordinateurs capables de "comprendre" le langage humain, de le parler et de l'écrire du moins, et cela dès la décennie 1990-2000. En cette fin du XX^e siècle, ce sont ainsi la place, le statut et le rôle que tous les systèmes d'écriture ont pu acquérir lentement au cours de ces cinq ou six derniers millénaires qui se trouvent sur le point d'être radicalement bouleversés.

Cette conquête de l'écriture par l'informatique aura été extraordinairement rapide à l'échelle de l'histoire de l'humanité. De 1950 à 1985, les ordinateurs se seront en effet substitués, en moins de trente cinq ans, à pratiquement tous les procédés antérieurs connus de transcription, de transmission, de communication, de conservation ou de transformation d'un message écrit ou imprimé. De surcroît, en ramenant par des biais successifs tous les systèmes d'écriture concevables à un système unique de représentation à l'aide de deux chiffres, de deux caractères numériques, le "zéro" et le "un" seulement, la numérisation de l'écriture a brusquement révélé l'existence d'une espèce d'unicité des modes de transcription des langages humains qu'on ne faisait que soupçonner auparavant. En multipliant également les techniques de restitution, de régénération et de télécommunication ou de télédiffusion, l'informatique en aura aussi accentué la multiplicité. En ce sens, l'informatisation de ces systèmes d'écriture recèle autant de promesses d'éclatement que de perspectives de fusion vers un système de transcription unique, universel. Dès l'instant où ils ont commencé à être informatisés et numérisés, ces systèmes d'écriture ont cessé d'être des réalités stables, permanentes et figées comme on le concevait jadis. Ils sont devenus des entités instables, éphémères et mouvantes, sujettes à d'innombrables métamorphoses par le fait même de leur médiatisation et de leur numérisation. Les frontières avec ce qui était purement oral

auparavant se déplacent également. C'est le statut même de ce qui était écrit, imprimé, lu, vu, entendu ou dit jusqu'à présent qui se déplace.

Alain VUILLEMIN
Institut de Littérature Française
Université de Paris-Sorbonne
(Paris IV)
Membre du Bureau national de l'EPI