INFORMATIQUE ET ÉCRITURE

Alain VUILLEMIN

La défense d'une langue passe aussi par celle de son écriture. Or, jusqu'à une date récente, les progrès de l'informatique paraissaient beaucoup plus menacer la pureté de la langue, surtout écrite, qu'ils ne semblaient pouvoir jamais devenir capables de la préserver. Et il est vrai que les ordinateurs n'ont été pendant longtemps que des calculateurs qui ne savaient guère opérer que sur des nombres. Puis, les processus de traitement se sont étendus à de premières catégories de caractères littéraux, à des chaînes de caractères et enfin aux textes et aux mots. Les ordinateurs ont même commencé depuis le début des années 1980, grâce aux développements de l'intelligence artificielle, à comprendre les langues naturelles, à les déchiffrer et à les traduire, à reconnaître la parole et à traiter du langage, à reproduire des voix humaines et à identifier des écritures. En 1986 sont apparues les premières tablettes graphiques capables de convertir instantanément une écriture manuscrite en caractères dactylographiques ou typographiques. L'écriture numérisée de demain existe déjà. Quelles en sont les caractéristiques principales ?

Il s'agit d'abord d'une écriture « ? numérisée » ?, ce qui signifie que tous les signes écrits qui la composent (et, éventuellement, tous les sons entendus qui y correspondent) se trouvent transposés et représentés à l'intérieur des systèmes modernes de traitement de l'information en de longues suites de nombres binaires, exclusivement constitués par les chiffres 0 et 1. Ces nombres sont définis en fonction de codes variés. Les codes les plus anciens ne comportaient jusqu'en 1950 que les dix chiffres du système décimal et quelques signes mathématiques. En 1951, l'UNIVAC I fut le premier ordinateur doté d'un code rudimentaire de 3 9 caractères alphanumériques, dont les 26 lettres majuscules de l'alphabet anglais. Dès 1952, la firme IBM créait un code de 68 caractères, qui permettait de travailler en typographie pauvre, puis en 1965 le code EBCDIC composé de 137 caractères. Dans le même temps s'imposait comme norme internationale le code ASCII, dont la version multilingue « Data Interchange » comportait en 1986 un jeu de 128 caractères

standard et un autre jeu de 128 caractères spécifiques. Ce n'est qu'à partir de 1980 toutefois que sont apparus en France les premiers claviers en typographie riche. Les progrès concevables sont néanmoins vertigineux. Dès 1984, la firme américaine XEROX a réussi à mettre au point un système de codage souple sur plusieurs octets d'information, qui permettaient de gérer jusqu'à 16 millions de caractères différents, soit pratiquement la quasi-totalité des systèmes d'écriture existants, y compris idéographiques.

En ce domaine, les travaux des pionniers sont plus anciens qu'on ne le soupçonne. Dès le début des années 1960, des chercheurs français du Centre national de la recherche scientifique avaient réussi à transcrire ainsi la plupart des écritures de l'Orient ancien, du lihyanite au méroïtique ou au sanscrit en passant par l'akkadien, l'araméen ou l'hébreu. En 1987, le CATAB (Centre d'analyse ou de traitement automatique des bases de données) de l'université Jean Moulin à Lyon travaillait à la conception d'un terminal, le Calligraphe 44, susceptible d'être connecté à n'importe quel modèle d'ordinateur et capable en principe de gérer tous les systèmes d'écriture existants, anciens ou modernes, quels qu'en soient les modes de présentation typographique et les natures alphabétiques, syllabiques, phonétiques, idéographiques, pictographiques, hiéroglyphiques, voire cunéiformes ou punctiformes. Des générateurs d'écriture en latin, en hébreu, en arabe et en chinois étaient déjà disponibles. Et, d'ores et déjà, un nombre croissant de systèmes de traitement de texte unilingues ou multilingues comportent en complément des polices de caractères dactylographiques ou typographiques, les unes prédéfinies, les autres programmables.

La dernière voie de recherche qui est explorée porte sur l'analyse et la synthèse de l'écriture. Les premiers systèmes de reconnaissance de caractères par des procédés optiques, tels le système Omni-Reader, remontent aux années 1950-1960. Les progrès ont été assez lents ensuite pendant une vingtaine d'années. Ces techniques sont désormais bien maîtrisées, y compris la reconnaissance de l'écriture manuscrite. Un système japonais, le CLL-2000 diffusé en 1985, était capable de reconnaître 2 400 caractères japonais dans les trois syllabaires japonais Kanji, Hiragana et Katakana. En 1986, la firme japonaise Toshiba proposait des systèmes de reconnaissance d'écriture capables d'identifier aussi bien les écritures cursives occidentales que japonaises. En 1986 également, était diffusée sur des 'Matériels Mac-Intosh une tablette graphique, le « Personal Writer 15 », qui-pouvait transformer immédiatement une écriture manuscrite quelconque en caractères typographiques. Il n'est

pas jusqu'à l'analyse et la synthèse des écritures manuscrites qui ne soient désormais concevables. L'expérience a déjà été menée en France, dès 1980, sur des manuscrits de P. Claudel.

De la typographie, dès le milieu des années 1950 par l'intermédiaire des procédés de photocomposition, à la calligraphie, depuis le début des années 1980 avec les progrès de l'imagerie numérique, ce sont toutes les techniques d'écriture manuscrite ou imprimée qui commencent ainsi à être bouleversées. On le méconnaît, mais, parce qu'elle est une science du traitement de l'information qui opère par le détour de processus de représentation symbolique, l'informatique est un autre système d'écriture. Et peut-être en est-elle un système privilégié. La numérisation d'une écriture n'en change pas la nature. L'on continuera d'écrire et de lire dans chacun des systèmes d'écriture qui existent exactement comme on le faisait dans le passé. En revanche, l'informatisation d'une quelconque écriture en facilite les tâches ultérieures de transcription, de translittération et même de traduction. L'informatique en devient un nouvel invariant. Les contraintes du plurilinguisme et la nécessité de préserver l'originalité des langues et parlers français pourraient enfin se trouver conciliées. Et qui sait si l'intégration de ces systèmes d'écriture numérisés à des systèmes de traduction écrite interactifs, dont les progrès récents sont aussi spectaculaires, et à des systèmes de traduction orale instantanée, dont les premiers prototypes fonctionnent depuis 1986, ne Permettra pas aux hommes de réaliser le vieux rêve d'une « pasigraphie », d'une écriture universelle et absolue qui permettrait à chacun de comprendre d'une façon immédiate des textes écrits en n'importe quelle langue ?

Alain VUILLEMIN.

Cet article est paru dans le numéro 140 de la revue *Défense de la Langue Française*, 8 rue Roquépine 75008 PARIS, Tél : (1) 42 65 08 87