

Commission nationale française pour l'UNESCO

# Diversité culturelle à l'ère du numérique

*Glossaire critique*



La  
**documentation**  
Française



# Codes

*Mokhtar Ben Henda et Henri Hudrisier*

Étymologiquement le mot « code » vient de « tablette à écrire, de codex ou *caudex*, assemblage de planches, des planchettes ayant servi à écrire » (Littré). Il apparaît en français en 1236 (Robert). La notion de « Code » a deux acceptions principales:

Le sens 1 est associé à une série de règles, de lois, de textes... qui régulent une conduite ou un comportement : le code Hammourabi de 1750 av. J.-C, le code théodosien de 438, le code Justinien de 529, le code civil, le code de commerce, de la propriété intellectuelle, des assurances, du travail, de la route, etc. Dans le langage moderne, le « code » renvoie à l'ensemble des dispositions légales relatives à une matière spéciale, réunies par le législateur dans des textes d'accords, de conventions et de normes établies par une communauté quelconque pour réguler des domaines d'activité.

Le sens 2 est liée à la codification des nombres et à la cryptologie, mais aussi aux théories linguistiques et sémiotiques. La codification des nombres se confond avec l'histoire de leur écriture mais est aussi liée à une pratique, puis une théorie du système numérique : duodécimal, décimal, binaire... La numération décimale s'est universalisée et le code binaire fonde l'industrie numérique. En Inde (4<sup>e</sup> Sc. Av. JC), en Chine avec les hexagrammes se préfigure déjà une pratique de codification binaire. En 1605, Francis Bacon décrit des séquences de codes binaires, Leibniz élargit la théorie et concrétise le concept. En 1836, Morse inaugure avec son code ce qui deviendra l'électromécanique puis l'électronique numérique. Les étapes et les inventeurs se succèdent, mais il est clair que les progrès du codage informatique sont largement la conséquence de la volonté de casser,

pendant la 2<sup>ème</sup> Guerre mondiale, l'information cryptée dans Enigma. Même si Charles Pierce ou Ferdinand de Saussure utilisent chacun un large vocabulaire conceptuel pour théoriser leur théorie du signe (codage sémiotique et/ou linguistique), il est clair que s'amorce une période d'échanges très féconds entre la linguistique, la recherche-développement des langages formels et aujourd'hui l'e-sémantique. Le code numérique binaire est au cœur de cette dernière période. Il n'est pas obligatoirement certain que des développements ultérieurs de l'informatique (peut-être l'informatique quantique) ne remettent pas en cause cette suprématie binaire actuelle.

### **DOMAINES D'APPLICATION ET APPROPRIATION SOCIALE**

Dans leurs dimensions interprétatives, les deux approches du terme « code » sont aujourd'hui utilisées dans plusieurs domaines d'ordre social, culturel, juridique, économique, linguistique, sémiotique, communicationnel et informatique.

En français, on dit « passer son code », « mettre les 'phares code' » ou même « mettre en 'codes' ». Code d'honneur, code des duels, code moral ou code culturel font aussi sens : « Ces maximes, je l'avoue, doivent être le code du genre humain », Voltaire, Dialogue. XV, 2e entretien. « D'un canton qui l'adore il est souvent l'arbitre. Le bon sens est son code, et l'équité son titre », ST-LAMBERT, Saisons, hiver.

Les données privées d'identification comme le code de sécurité sociale ou le code bancaire, sécurisent nos transactions. Les passeports modernes comportent un encodage crypté (possiblement utilisée dans de nombreux domaines de contrôle) mondialement normalisé (ISO/IEC JTC1SC37) de nos données anthropométriques (empreintes digitales et iris de l'œil) qui deviennent ainsi une clé codée personnelle universelle et (directement dépendante de notre code génétique !). On peut ainsi franchir sans contrôle humain des postes frontières ou y être systématiquement bloqué en cas de problème.

La notion de « code » prend une importance clé dans le secteur juridique. La mondialisation des codes juridiques, administratifs, profes-

sionnels et socio-culturels qui, du fait de leur numérisation et interconnexion notamment sur le Web, induisent des incompréhensions, des contresens, des conflits divers. Par exemple, le droit d'auteur et du *copyright* diffèrent selon les pays et aires culturelles même s'il existe des conventions internationales. Or, le potentiel à pouvoir faire circuler sans frontières et en temps réel des œuvres culturelles dans le monde entier pose clairement la question de l'interconnexion des codes pour garantir les ayants droit. Ceci fait écho à l'utopie d'un droit unique des affaires imposé par l'OMC. Ce code était supposé s'imposer parce que la mosaïque disparate des territoires nationaux freinait « à l'évidence » les échanges capitalistes ne tenant pas compte de toutes les spécificités des codes locaux divers (code du travail, code du droit commercial, conditions économiques locales, écologiques, etc.) y compris les spécificités culturelles et linguistiques qui se sont bien partiellement opposées à cette première fausse évidence. L'interopérabilité numérique sophistiquée de ces codes dans leur diversité est un chantier réaliste et urgent.

Dans le champ linguistique, dès 1909, Saussure s'empare aussi du terme « code », notamment le concept d'arbitraire du code en langue naturelle mais aussi comme code gestuel, code graphique, codes visuel. Dans la linguistique saussurienne, « Le lien unissant le signifiant au signifié est arbitraire, ou encore, puisque nous entendons par signe le total résultant de l'association d'un signifiant à un signifié, nous pouvons dire plus simplement : le signe est arbitraire ». Martinet met en évidence la notion de « double articulation du code » qui élargit de façon quasi-infinie un code, notamment la double articulation du code linguistique dont on a longtemps prétendu à tort qu'elle était le propre des langues humaines.

Dans le domaine des sciences et des techniques, la notion du code a été largement utilisée pour désigner l'ensemble des règles de symbolisation des données et concepts élémentaires. Un code est un système de signes (noms, symboles, signaux...) qui, par convention, sert à représenter et à transmettre l'information entre un émetteur et un récepteur : différents codes sont utilisés, iconiques, scripturaux, sonores (notamment la langue parlée).

---

L'encodage numérique d'un document nécessite un premier niveau signalétique : code langue, code écriture, code pays, etc. Précisons bien que le code langue (fr ou fre, es ou spa, pl ou pol, ja ou jpn, ko ou kor...) n'est ni superposable, ni identique aux codes pays (F ou FR, ES, PL, JP, KR, Corée du sud et KP, Corée du nord..). Il faut aussi savoir distinguer le code d'écriture du code langue. Le code d'écriture latine étendue (avec accents et diacritiques) recouvre de très nombreuses langues du monde ; le code d'écriture arabe recouvre l'arabe, le perse, l'urdu, le kurde... Ces codes normalisés des langues, écritures ou pays sont souvent un acronyme anglais abrégé mais pas obligatoirement.

La communication inter-systèmes et l'encodage numérique systématique (tout numérique) s'architecture en couches superposées de langages allant du langage machine (qui code en binaire les impulsions électriques), puis des langages systèmes sur lesquels se superposent des langages structurels (XML...), puis enfin des langage d'application permettant de coder la captation ou la saisie des documents proprement dit, leur circulation ou leur traitement. À un niveau informationnel plus large, l'encodage des documents n'est plus seulement la numérisation du message (le texte, le son, l'image... et leur possible compression), c'est aussi conjointement le codage de métadonnées (catalographique) et de balisage structurel ou sémantique du document (MPEG 7 ou TEI par exemple).

## **CODE BINAIRE ET DIVERSITÉ CULTURELLE ET LINGUISTIQUE**

Les premiers télégraphes ne sont pas binaires à proprement parler (le morse est faussement qualifié de binaire alors qu'il combine signal long, court et silence, de même le braille qui combine des points sur une matrice de 6 positions). Mais ces codages peuvent se numériser en code binaire, s'articulant sur un 2<sup>ème</sup> niveau de code (Baudot : 5 bits, ASCII, *American Standard Code for information Interchange* : 7 puis 8bits). Seul l'élargissement de la structure du code ASCII pouvait permettre la cohabitation des langues et des systèmes d'écriture du monde. Les premières initiatives venaient des langues asiatiques. Les 2 octets de la première norme japonaise multiécritures JIS (*Japanese Information Standard*), sont vite devenues 3 puis 4 octets. Ceci avait abou-

ti à l'état actuel de l'UNICODE s'articulant hiérarchiquement et fonctionnellement sur :

- 1 octet, soit 256 codes symboliques élémentaires pour l'alphabet latin ou un équivalent d'une trentaine de signes minuscules & majuscules en plus d'une suite des nombres, signes de ponctuation et codes de commandes ;
- 2 octets, soit plus de 65000 codes potentiels suffisant pour coder toutes les écritures du monde y compris les écritures anciennes, différents codes d'écritures musicales (byzantine...), des codes divers (symboles électriques, signalétiques diverses, etc...). Cependant, ces 65000 codes sont insuffisants pour coder les « caractères idéographiques rares », qui, en chinois ou en japonais correspondent notamment à la toponymie et à la patronymie (y compris des noms de firmes).
- 4 octets, soit plus de 4 milliards de codes symboliques potentiels qui posent le problème à un niveau incomparablement plus vaste. Ce codage articulé sur 4 octets (32 bits) est à la fois le fruit d'une nécessité d'unifier dans un même code toutes les écritures du monde, mais c'est aussi le résultat d'un potentiel technique lié à la rapidité du calcul et à la miniaturisation et baisse de coût de la mémoire vive d'ordinateur.

L'articulation double du code binaire de 32 bits est la condition *sine qua non* d'un codage « intelligent » du code informatique. L'argument qui consistait à qualifier de « binaire » le code informatique n'est plus guère d'actualité parce que la culture mondiale a intégré la subtile sophistication du transcodage numérique culturel – appartenances ethniques, pratiques sociales, productions intellectuelles, etc. – par des attributs issus de conventions et de normes internationales connues dans le jargon informatique sous les notions de « localisation » (110n) et « internationalisation » (118n). Il s'agit de valeurs codiques qui induisent l'usage de codes linguistiques et culturels définissant le type de calendrier, le système d'écriture et sa directionnalité, la monnaie d'usage, le code de pays, le code de langue et sa variante nationale, le système numérique, le code de format, etc. L'internationalisation des codes de nommage des domaines (iDNS), des adresses de courrier

électronique et des langages informatiques, rendent les ressources culturelles accessibles à plus de gens dans leurs propres langues.

Le codage numérique culturel touche aussi des activités ou disciplines scientifiques comme les lettres et les arts (i.e. Humanités digitales). La TEI, notamment, propose un niveau avancé de codage dans lequel des collègues de chercheurs partagent des ensembles définis de « codes balises » permettant d'encoder des appareils critiques, des analyses de recherches littéraires (codification de la métrique des vers mais aussi codage de la critique stylistique ou de l'étude narrative...), mais aussi des corpus linguistiques, des manuscrits, etc...

Bref, les deux facettes abordées de la notion du code convergent désormais vers le mode digital. Le gros de la mémoire culturelle, scientifique, technique et juridique accumulée dans les archives, bibliothèques et musées bascule dans les réseaux numériques et les questions liées à la diversité culturelle et linguistique passent au premier plan de la représentation et de l'interprétation des codes. Les développements futurs seront transculturels, transnationaux et transparents dans lesquels d'autres aspects de la communication humaine au-delà de la seule codification des langues entreront sur scène.

**Termes liés : archives, bibliothèques, co-construction, communication, computation, connaissance, documentation, langues, normes**

---

### *Références*

Jacques André et Henri Hudrisier. « Unicode, écriture du monde ? », numéro spécial de *Document numérique* vol 6 –n°3-4, 2002 Edition Hermès-Lavoisier

Patrick Andries. *Unicode 5.0 en pratique : Codage des caractères et internationalisation des logiciels et des documents*, Paris, Dunod, 2008.

Mokhtar Ben Henda. « Langues en danger et multilinguisme numérique », in *Les oubliés de l'internet : Culture et langues sur l'internet, oublié ou déni ?*, A-M. Laulan & A. LeNoble Bart (éd.), Bordeaux, Les Editions Hospitalières, 2014.



Claude Berrou, *Codes et turbocodes*, Paris, Springer Science & Business Media, 2007.

Monica Borda, *Fundamentals in Information Theory and Codin*, Berlin, Springer Science & Business Media, 2011.

Michel Bottin, Yvonne Sallé, « Écritures du monde, un centre de ressources culturelles et techniques pour les applications multi-écritures », *Document numérique*, n°3, vol. 6, 2002, pp. 237-254.

Stuart Hall, Michèle Albaret, Marie-Christine Gamberini, « Codage/décodage », *Réseaux*, vol. 12, n°68, 1994, pp. 27-39.

Henri Hudrisier, « De l'Abécédaire à Unicode », *L'Octogonal Ricochet*, Revue du CIELJ, n°17, avril 2002.

Bruno Martin, *Codage, cryptologie et applications*. Lausanne, PPUR Presses polytechniques, 2004.





En 2005, l'UNESCO adoptait la Convention sur la protection et la promotion de la diversité des expressions culturelles, donnant ainsi une forme juridique à la question de la diversité culturelle. Depuis, la Convention a été ratifiée par 133 États. Pour autant, la diversité culturelle ne saurait être définitivement acquise. Les progrès et les utilisations démultipliées des technologies numériques redessinent un paysage où la Convention de 2005 a plus que jamais sa place.

La Commission nationale française pour l'UNESCO, sensible à « l'importance de poursuivre la réflexion sur les défis et les opportunités posés par les technologies numériques dans le cadre de la mise en œuvre de la Convention, ainsi que leur impact sur la diversité des expressions culturelles », a donc choisi de publier ce nouveau *Glossaire critique*, exploration des relations souvent fécondes et des liaisons parfois dangereuses qui lient la diversité culturelle et le numérique.

Cet ouvrage, réalisé sous la direction de Divina Frau-Meigs et d'Alain Kiyindou et dont la postface a été confiée à Jean Musitelli, aborde soixante termes, d'« Agrégateur » à « Virtuel » et propose des définitions des pratiques et des usages. Véritable outil d'aide à la réflexion et à la décision, il s'adresse à la fois au grand public et à l'ensemble des acteurs et décideurs concernés par les questions de culture et d'éducation, et plus largement de gouvernance du numérique.

© Jürgen Ziewer/Bon images/Corbis

Imprimé en France  
ISBN : 978-2-11-009406-3  
DF : 5HC34180  
Prix : 19 €

Diffusion  
Direction de l'information  
légale et administrative  
La documentation Française  
Téléphone : 01 40 15 70 10  
[www.ladocumentationfrancaise.fr](http://www.ladocumentationfrancaise.fr)

