Le retour des idéogrammes

Nadine Lucas

Dans Document numérique 2002/3 Vol. 6 , pages 183 à 210
Éditions JLE

ISSN 1279-5127

Article disponible en ligne à l'adresse
https://stm.cairn.info/revue-document-numerique-2002-3-page-183?lang=fr

CAIRN.INFO

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

Distribution électronique Cairn.info pour JLE.
Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.
Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

Le retour des idéogrammes

Unicode CJC vu du Japon

Nadine Lucas

GREYC CNRS UMR 6072
Université de Caen
Boulevard du Maréchal Juin
F-14032 Caen cedex
Nadine.Lucas@info.unicaen.fr

RÉSUMÉ. *Le standard d'usage couramment appelé Unicode, basé sur la norme ISO/CEI 10646 permet de traiter de façon unifiée les écritures codées sur deux octets, notamment les écritures idéographiques, un temps méprisées, Unihan, base d'idéogrammes, et le code de conversion UTF permettent l'échange d'information entre pays du monde sinisé mais aussi avec les autres zones culturelles. Le japonais représente un cas particulier, par la conjonction d'une histoire complexe de l'écriture et d'un savoir-faire technologique de pointe. L'intégration d'Unicode permet l'accès à des ressources documentaires de toute origine, indépendamment du codage, ainsi que leur traitement et stockage informatique. Une aubaine pour le Japon, qui cultive l'ouverture au monde.*

ABSTRACT. *The Unicode standard based on the ISO/IEC 10646 standard allows anyone to collect and read any text in any graphical form including ideographs. Japan obviously benefits from this new standard, both by saving its cultural assets and by opening the way for international information retrieval and processing.*

MOTS-CLÉS : *Unicode, japonais, caractères chinois, idéogrammes, CJC, Unihan, traitement automatique multilingue, fouille de texte.*

KEYWORDS: *Unicode, Japanese, Chinese characters, CJK, Unihan, information monitoring, text mining, multilingual information processing.*

Document numérique. Volume 6 – n° 3-4/2002, pages 183 à 210

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

1. Introduction

CJC[1] est un nouveau sigle du monde JUC[2] pour chinois, japonais, coréen (en anglais CJK pour Chinese, Japanese, Korean), ensemble d'écritures auxquelles on ajoute le vietnamien. Une vaste zone culturelle, qui correspond, dans le standard Unicode[3] à un ensemble comparable à première vue à l'ISO latin. Quels sont les problèmes particuliers rencontrés pour coder ces écritures, qu'est-ce qu'Unihan et quel est l'usage qui est fait de la norme internationale dans les pratiques actuelles ?

Avant d'aborder ces questions, nous rappelons quelques caractéristiques des écritures asiatiques utilisant les caractères chinois. Elles permettent de mieux saisir la façon dont les idéogrammes sont codés et surtout comment est organisée la base de données Unihan.

Nous nous intéresserons ensuite à l'application de ces principes au cas du japonais, écriture mixte utilisant à la fois les caractères chinois et un syllabaire autochtone, pour mieux appréhender une réalité complexe. C'est en effet le Japon qui a pris l'initiative de la normalisation et reste avec Taiwan le champion de l'idéographie à travers son savoir-faire informatique.

Les enjeux sous-jacents ne sont pas seulement culturels, ils ont aussi des retombées technologiques. Souvenons-nous que les imprimantes à laser ont été mises au point pour l'impression fine nécessitée par les caractères chinois et que les premiers systèmes de lecture orale pour aveugles ont été conçus au Japon. La bataille pour la maîtrise de l'information mondiale circulant sur le web passe par l'automatisation des traitements de l'oral et de l'écrit, soit les industries de la langue.

Une illustration récente de l'avantage japonais est celui de l'I-mode proposé par NTT DoCoMo. Alors que le projet occidental de WAP[4] piétine, les Japonais disposent déjà de cet outil de communication portable, qui fait actuellement fureur dans les jeunes générations. L'i-mode compte 10 millions d'abonnés, contre 3 millions d'utilisateurs WAP au Japon. Les capacités réduites d'affichage sur mini-écran pénalisent beaucoup moins les lecteurs d'idéogrammes que les lecteurs de mots alphabétiques. En effet, la densité d'information est l'atout majeur de l'écriture chinoise, et dans une moindre mesure, de l'écriture sino-japonaise.

---

1. Bien qu'il soit encore fréquent d'employer dans des textes français les acronymes anglais comme CJK, UCS, etc., nous préférons employer ici des termes français. On trouvera à la fin de ce numéro spécial un index bilingue de ces sigles et acronymes.

2. Jeu universel de caractères, en anglais UCS Universal Multiple-Octet Coded Character Set.

3. Voir l'article d'Olivier Randier dans ce numéro.

4. *Wireless Application Protocol*, ou protocole d'accès sans fil, pour l'accès à internet à partir de supports mobiles, les assistants personnels ou PDA *Personal Digital Assistant.*

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

2. Les idéogrammes chinois

2.1. *Une domination culturelle historique*

L'invention de l'écriture est la date fondatrice du commencement de l'histoire. L'histoire de la Chine commence très tôt, les plus anciens écrits conservés datent de 1 400 avant Jésus-Christ, mais selon la tradition chinoise, l'écriture aurait été inventée durant le règne de l'empereur Huang Di (2697-2598) par son ministre Cang Jie.

Le mode de représentation est dit « idéographique », c'est-à-dire que les signes d'écriture sont très nombreux et qu'ils correspondent *grosso modo* à des mots ou des idées. Dans l'écriture chinoise, qui est une sorte de langue écrite officielle, on note le signe de l'idée et non le son de la parole dite. Le Chinois du sud écrit le mandarin comme le Chinois du nord même si la prononciation qu'il en donne est différente. De même, le fonctionnaire des pays autrefois vassaux, comme la Corée, a longtemps écrit comme en Chine, même si pour lui les mots étaient prononcés tout à fait différemment. Toute la compréhension du vocabulaire fonctionne sur le même modèle que notre compréhension universelle des signes idéographiques des chiffres écrits 1, *one, un, uno, ichi* ; 2, *two, deux, due, ni* ; 3 *three, trois*…. Ainsi au signe 心 correspond l'idée du cœur, et y sont associées des prononciations telles que *sam, xin, shin, kokoro, sim*…

Notons que cette association simplificatrice, un caractère représente un mot, permet de comprendre le problème essentiel de l'écriture chinoise et de ses dérivées : c'est un système ouvert, en principe, ce qui défie toute tentative de codage se voulant définitive.

La Chine a longtemps été le centre de la civilisation en Asie, et l'écriture l'arme du pouvoir mandarinal, notamment des célèbres dynasties Han (circa 206-24 av. J.-C pour les Han antérieurs, de 25 av. J.-C. à 220 apr. J.-C. pour les Han postérieurs). Les caractères chinois sont appelés caractères Han dans le monde sinisé : la Corée, le Japon, mais aussi le Vietnam les utilisent pendant plusieurs siècles. C'est encore sous ce nom qu'on les trouve dans la terminologie Unicode : Unihan[5] regroupe les caractères chinois qui font partie d'une culture commune, mais qui ont été normalisés précédemment à l'échelle nationale (donc différemment) par la Chine continentale, Taiwan, Hong Kong, le Japon, la Corée du Nord et du Sud et le Vietnam.

Les vicissitudes historiques ont abouti à des situations très hétérogènes à notre époque (Vandermeersch, 1986). Alors que le Vietnam utilise une écriture latine[6]

5. Voir à l'adresse http://www.unicode.org/UnihanDatabase

6. La transcription alphabétique latine du vietnamien a été introduite pour la communauté catholique au 17e siècle par le jésuite Alexandre de Rodhes et la large diffusion de cette notation a été soutenue par les autorités coloniales puis par les communistes.

depuis la colonisation française, et ne conserve plus le *Ham-non*, écriture vietnamienne idéographique que dans une tradition savante, les pays de la zone CJC ont chacun maintenu vivante une part de l'héritage classique. Le Japon a développé une écriture syllabique parallèle, les *kana*, dès le VII^e^ siècle, mais conserve de nos jours l'usage d'une écriture mixte, *kana* et caractères Han, sous une forme parfois différente des caractères chinois classiques (ils ont été simplifiés vers 1900). Après la seconde guerre mondiale, alors que les occupants américains souhaitaient imposer une écriture alphabétique, le gouvernement japonais a résisté et favorisé des travaux conséquents pour informatiser l'écriture mixte (Griolet, 1985 ; Lucas, 1985).

La Corée s'est dotée d'une écriture indigène originale, le *hangul*, en 1446, bien plus tardivement que le Japon. Cette écriture, de type syllabique, qui fut créée par le roi Sejong, combine des éléments graphiques appelés *jamo*[7]. Elle a supplanté presque complètement les caractères chinois dans l'usage actuel. Mais les caractères Han sont conservés dans les noms propres et ils sont toujours enseignés. La possibilité d'accès aux caractères chinois préserve aussi le maniement informatique de textes historiques, comme dans le cas du Vietnam.

La République populaire de Chine a connu une révolution culturelle au XX^e^ siècle, le nombre de caractères courants a diminué et le tracé en a été beaucoup simplifié. Le pinyin, translittération alphabétique du chinois, a été vu comme un premier pas vers l'abandon des idéogrammes, jugés inadaptés au monde moderne (télégraphie, dactylographie). Pendant ce temps, à Taiwan, l'écriture chinoise classique a été conservée sans simplifications, et de gros efforts ont été faits pour mener à bien son traitement informatique.

Actuellement, grâce aux progrès technologiques, et principalement sous l'impulsion de Taiwan et du Japon, relayés par Xerox[8], les caractères chinois ont retrouvé de beaux jours et ont droit de cité sur internet. La communauté CJC existe au sein du consortium Unicode, cependant elle couvre des réalités et des usages bien différents.

2.2. *Idéogrammes et techniques traditionnelles de classement*

Les caractères sont en principe en nombre illimité. Mythologiquement pourrait-on dire car, dans les temps anciens, l'empereur de Chine était chargé d'en inventer de nouveaux périodiquement. Il existait également des caractères frappés d'interdit par l'empereur, mais cette pratique du tabou n'a pas suffi à maintenir un nombre constant de caractères en usage, non plus que les réductions autoritaires qui ont eu lieu au cours de l'histoire. Alors que le plus ancien dictionnaire chinois, le *Cang jie pian,* recense 3 300 caractères vers 210 av. J.-C., le *Xiandai Hanyu da zidian* de

7. Pour de plus amples renseignements, voir par exemple http://www.sigmainstitute.com/koreanonline/hangul_history.shtml

8. Un des fondateurs du consortium Unicode.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

1990 en compte 56 000. En pratique, pourtant, la production de nouveaux caractères s'est ralentie, au profit de la composition de « mots » composés de deux ou trois caractères. Cependant, la nécessité de laisser ce système de signes ouvert est reconnue, c'est sans doute ce qui le différencie le plus de notre système alphabétique. On trouvera une introduction plus nuancée à l'écriture chinoise dans le dossier *Pour la science* intitulé « Du signe à l'écriture » paru en octobre 2001 (Grenié et Belotel-Grenié, 2001) ou dans (Alleton, 1984).

Rappelons que l'écriture chinoise est indissociable du pouvoir politique et que la Chine est un vaste territoire, où coexistent plusieurs langues (qui ne sont pas mutuellement intelligibles) et dialectes (mutuellement intelligibles). Ceci permet de comprendre l'importance attachée à la langue écrite et le souci de séparer, autant que faire se peut, la langue orale quotidienne, et la langue écrite, qui a vocation unificatrice. Ce n'est qu'au XX^{e} siècle, lors de la révolution chinoise, que le fossé entre ces deux formes d'expression a été jugé intolérable, et qu'une réunification a été entreprise. Elle s'est accompagnée d'une purge énergique visant à réduire le nombre de caractères, ainsi que d'une simplification de la graphie.

Les caractères en tant que signes graphiques sont simples ou composés. Pour éviter des confusions avec le terme « composé », utilisé aussi pour les mots composés de plusieurs caractères, nous emploierons ici le terme de glyphe complexe. En fait, les caractères représentent en chinois à la fois une syllabe (forme sonore), un morphème (forme abstraite de la langue), et un glyphe (forme écrite). Les glyphes complexes sont tracés à partir de glyphes existants, combinés entre eux. C'est ce qui permet de créer de nouveaux caractères en variant les compositions. Les caractères Han, en chinois *hanzi*, sont décrits par une « clé »[9], c'est-à-dire une partie du caractère facilement identifiable et qui constitue, isolée, un caractère simple, ou du moins, indécomposable. Par exemple, le caractère du « cœur » est un caractère simple, de 4 traits, et il entre dans la composition de caractères complexes comme « réponse », « sentiment » « pensée », « adorer » ou encore « agréable ». Le tableau 1 propose quelques caractères complexes formés à partir de la clé de « un » et de celle du « cœur ».

Puisque les considérations graphiques priment, la prononciation étant considérée comme trop variable pour être pertinente, le tracé des caractères a une importance capitale. Les caractères sont classés dans un ordre traditionnel, par clé, en commençant par les plus simples. La première des clés est ainsi le chiffre 1 qui s'écrit par un trait horizontal. La dernière clé est celle de la « flûte », caractère jugé « simple » qui compte néanmoins 17 traits : c'est que la flûte a une valeur symbolique importante dans la culture chinoise. Il existe ainsi 214 clés ou caractères jugés indécomposables, puis à l'intérieur de chaque clé, à la suite du caractère simple, les caractères complexes sont ordonnés par nombre de traits (de pinceau) nécessaires à leur écriture. Ces 214 clés proviennent du grand dictionnaire chinois dit *KangXi* publié en 1716 sous le règne de l'empereur du même nom.

9. En anglais *radical.*

Caractère simple	Caractère complexe	clé	sens
一		一	un
	世	一	monde
	七	一	sept
心		心	cœur
	応	心	réponse
	感	心	sentiment
	思	心	pensée
	意	心	désir, volonté
	慕	心	adorer
	快	心	agréable
	性	心	nature

Tableau 1. *Quelques caractères chinois et leur clé*

Pour s'y retrouver parmi des milliers de caractères, il est indispensable de bien mémoriser les clés, qui portent chacune un nom. Les classements de dictionnaire devant être aussi immédiats qu'avec un ordre alphabétique, les ouvrages de référence en usage depuis des générations utilisent tous le même ordre basé sur celui du KangXi. Mais ils proposent chacun leur numérotation pour le caractère, variable suivant la taille du dictionnaire, et la dénomination en usage dans un pays pour les clés. Celle-ci est souvent, mais pas toujours, distincte du nom du caractère simple. Par exemple, le dictionnaire Morohashi classique au Japon (*Daikanwa jiten*) donne 00001 pour le caractère « un » *ichi* et 10295 pour le caractère « cœur » appelé *kokoro* ou *shin* pour le caractère simple et *shinben* pour la clé *;* le dictionnaire Nelson, qui propose des traductions anglaises des caractères sino-japonais et est utilisé par les étudiants européens, donne 0001 pour le caractère « un » *one* et 1645 pour le caractère « cœur » *heart*.

Les clés sont de plus souvent simplifiées, dans leur graphie, quand elles entrent dans un glyphe complexe. On peut observer par exemple dans le tableau 1 des simplifications croissantes de la forme de la clé à partir du glyphe d'origine du « cœur » le caractère simple à 4 traits. Dans le glyphe complexe « réponse », le graphisme du caractère-clé ne change pas. Dans « désir », la forme est déjà un peu écrasée. Dans « adorer », la forme écrasée à l'horizontale de la clé compte toujours 4 traits. Elle reste plus facilement reconnaissable que la forme de clé que l'on trouve dans « agréable », ou encore « nature », l'élément de gauche qui ne compte plus que 3 traits : deux sortes d'apostrophes disposées autour d'un axe vertical. C'est pourquoi une graphie simplifiée peut avoir des statuts différents, comme variante plus ou moins singularisée de la clé d'origine. La forme de clé verticale à 3 traits est

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

appelée *risshinben* (clé du cœur debout) au Japon, mais elle est considérée comme variante de la forme canonique *shinben* par les dictionnaires pour japonophones, tandis qu'elle est traitée comme clé distincte par les manuels qui s'adressent à des débutants étrangers.

La clé a un statut métalinguistique important. C'est à la fois un élément graphique et comme son nom l'indique, une clé de classement. Les glyphes complexes sont formés à partir de plusieurs glyphes existants, mais un seul a statut de clé. Depuis le XVIII^{e} siècle, sur le modèle du monumental dictionnaire *KangXi*, le glyphe choisi comme clé est celui qui permet de constituer des familles de sens. La clé a donc une valeur sémantique, non pas de façon stricte, comme on peut le deviner à travers les exemples, mais suffisamment pour servir de procédé mnémotechnique.

Les caractères chinois ont bien sûr évolué dans leur tracé au cours de l'histoire, et comme on l'a vu, leur nombre n'a cessé de croître. Cependant, l'écriture s'est stabilisée, avec la nécessité de fixer des programmes pour les concours de mandarins. Les techniques d'imprimerie connues depuis des siècles ont également contribué à un figement du nombre et de la forme des caractères d'usage courant. Les machines à écrire n'ont jamais été en usage, en revanche, l'informatique de seconde génération a permis un nouvel essor de ce mode d'écriture, un temps jugé obsolète. C'est cette fois le Japon qui a pris l'initiative et reste avec Taiwan le champion de l'idéographie à travers son savoir-faire informatique. Les autres pays d'Asie ont suivi et ont adopté des normes nationales.

Le codage des caractères se fait sur deux octets, le mode de saisie au clavier est phonétique, ce qui entraîne des protocoles de conversion relativement sophistiqués.

3. Unicode CJC

Le standard dit Unicode, sous-ensemble de JUC (Jeu universel de caractères) regroupe sous la dénomination CJC les glyphes en usage en Chine, au Japon, en Corée et dans une moindre mesure au Vietnam, symboles et idéogrammes codés sur deux octets (JUC-2, plus connu des informaticiens sous le sigle anglais UCS-2, de même que le sigle UTF). La définition du glyphe par Unicode est légèrement différente de celle que nous avons employée jusqu'ici, car les caractères sont tous traités sur un pied d'égalité, en tant que symboles codés, au même titre que les lettres des alphabets. On notera d'ailleurs que dans la nomenclature Unicode les syllabes des syllabaires sont appelées « *letters* ».

Sous le sigle CJC, on trouve une série de tables. Les syllabaires japonais, coréen et yi (langue sino-tibétaine) en constituent une partie, l'autre étant constituée par les idéogrammes. Deux grands ensembles, parmi ceux-ci, l'ensemble des idéogrammes partagés ou *CJK Unified Ideographs* que l'on appelle couramment Unihan, dotée de deux extensions, et l'ensemble *CJK Compatibility Ideographs* qui permet de tenir

compte des formes spécifiques, par exemple celles utilisées dans les noms propres coréens, et de les mettre en correspondance avec un glyphe standardisé. Il existe également une table pour les symboles spécifiques, *CJK Compatibility Symbols*. Enfin, pour tenir compte de la nécessité de gérer des caractères non normalisés, il existe des zones pour l'usage hors norme, dit usage privé (ce qui comprend des usages collectifs ou individuels).

3.1. *Les formats de codage*

Comme les caractères Han sont codés normalement sur deux octets, c'est UTF-16 qui sert de référence, sans changement, pour le code JUC-2. Chaque caractère est représenté par quatre chiffres hexadécimaux formant un ou deux « mots de code » pouvant être représentés en binaire et comprimés par un algorithme d'optimisation. Cependant, certains caractères sont représentés par cinq chiffres hexadécimaux, qui peuvent s'écrire comme un long « mot de code » sur trois octets. La solution défendue par Taiwan dès l'origine pour coder l'ensemble des caractères connus, et laisser une marge pour l'avenir est en effet le passage à un codage en UTF-32, prévue par le consortium Unicode. L'utilisation d'UTF-32 entraîne une taille importante de fichiers, puisque tous les caractères vont occuper systématiquement quatre octets (réductible à trois octets). À l'inverse, partant de JUC-2, en utilisant UTF-8 on enchaîne le nombre strictement nécessaire d'octets, pour représenter un caractère. On a ainsi une série de trois ou quatre octets notés en hexadécimal, que l'on peut traduire en binaire. On en verra plus loin l'illustration (tableau 4).

3.2. *Unihan*

Dans le standard actuel, Unicode 3.1, si l'on consulte la base de données idéographiques Unihan ou *Unihan database*, on se trouvera face à une grande collection de données, qui permettent de gérer le syncrétisme culturel. Unihan recouvre l'ensemble *CJK Unified Ideographs,* les glyphes qui constituent un fonds commun d'usage courant. C'est un gros ensemble qui occupe les cases 4E00 à 9FAF et représente un fichier de 5MB. Il est complété par deux ensembles plus restreints *CJK Unified Ideographs Extension A* et *Extension B* pour les caractères d'usage plus spécialisé.

Une des caractéristiques importantes d'Unihan est que l'ambiguïté entre caractère simple et clé est réduite. Le rôle de clé d'un caractère simple est signalé par un code propre dans un champ différent. Par exemple, le caractère simple qui correspond à « cœur » a pour code JUC-2 (en UTF-16) 5FC3, c'est son identifiant (*code point*). La clé du cœur a le code 61.0, en tant que lien sémantique, dans le champ intitulé *Radical-strokes count*. On aura accès, à partir du caractère simple, à tous les caractères complexes, que la clé soit de même forme générale que le caractère indépendant ou de forme simplifiée. Ils auront en effet, dans le champ

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

Radical-strokes count, un code 61 suivi du nombre de traits supplémentaires nécessaires à leur tracé. Le 0 indique 0 trait supplémentaire, ce qui veut dire que ce caractère est un caractère simple, donc une clé. Le numéro de clé est indispensable à la gestion de la base de données. On voit ici que la tradition de classement par clé et nombre de traits est respectée. La première clé « un » a bien le code 1.0 et la dernière clé « flûte » a bien le code 214.0. Ces numéros de clé proviennent du dictionnaire chinois KangXi, qui comme on l'a noté, sert de référence historique. Ce dictionnaire, utilisé par la norme chinoise continentale, a donc statut d'autorité pour l'établissement des clés de classement dans la base Unihan.

De plus, dans Unihan, les signes ou éléments graphiques entrant dans la composition, décomposition ou recomposition de caractères complexes ont un statut particulier, métalinguistique pourrait-on dire. Il s'agit en particulier des formes simplifiées de caractères. Par exemple, la forme simplifiée *risshinben*, que nous avons vue comme graphie compressée de la clé du cœur, doit pouvoir être imprimée dans un dictionnaire, ou dans un manuel expliquant que cette forme particulière est équivalente à la forme normale de la clé *shinben*. Ces glyphes doivent donc pouvoir être manipulés informatiquement, et visualisés en tant qu'éléments de dessin. Il existe également une procédure spéciale de description de caractères hors norme, utilisée par exemple par les historiens, pour la reconstitution de caractères complexes en désuétude, ou encore pour permettre la visualisation de caractères nouveaux dont on discute la normalisation. Ces éléments graphiques qui entrent dans des processus de décomposition ou de recomposition de glyphes sont dénommés *KangXi radicals,* nous les appellerons « forme de clé ». Ils ont un code particulier, et Unicode recommande de les imprimer de manière à ce qu'ils ne soient pas confondus avec des caractères normaux. C'est ce que nous avons fait en choisissant l'italique dans le tableau 4 pour la forme de clé *shinben*.

Dans la base de donnée Unihan, qui rassemble les caractères normaux d'usage commun, chaque caractère Han est décrit et indexé comme suit :

– le glyphe est donné sous sa forme standard Unicode et sous la forme gérée par le navigateur utilisé (la fonte pouvant être différente) ;

– les formats de codage alignent les formats décimaux[10], UTF-8 (en hexadécimal), UTF-16 et UTF-32. Les sources IRG[11] renvoient aux normes préexistantes, avec leur numéro et le préfixe G pour la Chine, T pour Taiwan, H pour Hong-Kong, J pour le Japon, K pour la Corée du Sud, KP pour la Corée du Nord, V pour le Vietnam et U pour divers dont Singapour ;

– les correspondances avec les codes ou références existants sont ensuite détaillées. On trouvera d'abord les références aux normes successives des pays concernés et le code du caractère dans ces différentes normes. On trouvera ensuite les références aux principaux dictionnaires ayant servi de base aux normes (KangXi

10. Ce n'est pas à proprement parler un format de codage, mais il peut être utilisé notamment comme identifiant dans un document XML.

11. *IRG Ideographic Rapporteur Group* pour ISO/CEI.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

pour la Chine et Morohashi pour le Japon, entre autres). Elles mentionnent le numéro de page/position dans la page ou le numéro du caractère dans ces différents dictionnaires, ainsi que le numéro de clé et le nombre de traits, donnée qui sert comme le verra d'identifiant sémantique. Suivent des données sur la prononciation, la définition et autres, les variantes du caractère (glyphe simplifié ou plus ancien), les contextes, sous la forme d'une liste de mots composés, ou d'expressions dans lesquels apparaît le caractère, et enfin, des données diverses permettant d'établir des liens internes à la base.

Il est à noter que les fichiers CEDICT pour la Chine et EDICT pour le Japon, qui fournissent des contextes d'emploi (mots composés ou expressions) pour les caractères Han, ne font pas partie à proprement parler d'Unihan, mais ils sont accessibles à partir de cette base de données.

Le tableau 2 reproduit les données obtenues en consultant la base de données Unihan pour le caractère « cœur » U+5FC3 avec toutes les informations mentionnées plus haut. La liste des mots composés chinois et japonais a été abrégée.

Unihan 3.1 data for U+5FC3

Glyphs
The Unicode Standard Your Browser

Encoding Forms

Decimal	UTF-8	UTF-16	UTF-32
24515	E5 BF 83	5FC3	00005FC3

IRG Sources

G-source	T-source	H-source	J-source	K-source	KP-source	V-source	U-source
0-5044	1-4540		0-3F34	0-637D	KP0-E5E3	1-5473	

Mappings to Major Standards

Chinese

GB 2312	GB 12345	CNS 11643-1986	CNS 11643-1992	CCCII	Big Five	HK SCS
4836	4836	1-4540	1-4540	213D78	A4DF	

Japanese

JIS X 0208	JIS X 0212	JIS X 0213
3120		

Korean

KS X 1001:1992	KS X 1002:1991	KPS 9566-97	KPS 10721-2000

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

6793		E5E3	

Other

EACC	Xerox	PRC Telegraph	ROC Telegraph
213D78	241:244	1800	1800

Dictionary Information

IRG Indices

KangXi	Morohashi	Dae Jaweon	Hanyu Da Zidian
0375.010	10295	0700.050	42267.010

Other Indices

Cowles	Fenn	Karlgren	Nelson	Mathews	Meyer-Wempe	Lau
			1645	2735		

Radical-stroke Counts

Unicode	KangXi	Japanese	Korean	Morohashi
61.0	61.0			

Phonetic Data

Cantonese	Mandarin	Tang	Japanese On	Japanese Kun	Sino-Korean
SAM1	XIN1		SHIN	KOKORO	SIM

Other Dictionary Data

Definition	Total Strokes	Phonetic	Cangjie
heart; mind, intelligence; soul	4	1118	P

Chinese Compounds (drawn largely, but not exclusively, from the CEDICT dictionary file)

心中		sam1 lei5	inner feeling
心迷		sam1 mai4	to be confused of mind; to be puzzled
円心		yun4 sam1	center of a circle
愛國心	ai4guo2xin		patriotism
安心	an1 xin1		feel at ease; be relieved; set one's mind at rest; keep one's mind on something
筆心	bi3 xin1	bat1 sam1	pencil lead; refill (for a ball-point pen)
人心	ren2xin	yan4 sam1	human heart, will, feeling or emotion; morale
手心	shou3xin	sau2 sam1	center of the palm
心理	xin1 li3	sam1 lei5	mental; psychological
…			

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

Japanese Compounds (drawn from the EDICT dictionary file)

ご心配なく	ごしんぱいなく	(id) don't worry/never mind
愛国心	あいこくしん	patriotism
悪心	あくしん	evil thought/malicious motive
悪心	おしん	nausea/urge to vomit
安心	あんしん	(vs) relief/peace of mind
心強い	こころづよい	sense of security
心魔	しんま	treachery/intrigue
誠心誠意	せいしんせいい	single-mindedly/wholeheartedly/with one's heart and soul
一心	いっしん	one mind/wholeheartedness/the whole heart
心行く	こころゆく	with one mind
円心	えんしん	centre of circle
遠心	えんしん	(an) (vs) centrifuge
遠心分離	えんしんぶんり	(vs) centrifugation/centrifuge
遠心分離機	えんしんぶんりき	centrifuge/centrifugal machine
…		

Tableau 2. *Informations sur le caractère « cœur » U+5FC3*

3.3. *Accès aux informations pour un caractère*

Le site *Unicode Unihan Search Page* propose une consultation en ligne à partir de plusieurs entrées, détaillées ci-dessous. Elles supposent une bonne connaissance des modes traditionnels de classement des idéogrammes et des outils de référence.

Si l'on connaît une prononciation standardisée

Les prononciations répertoriées dans Unihan pour un caractère ne couvrent pas toutes les lectures possibles, mais elles sont tout de même nombreuses : en chinois, japonais, sino-coréen. De plus, pour le chinois, on répertorie trois prononciations standardisées : mandarin (translittérée en pinyin avec un chiffre indiquant le ton), cantonaise (translittération Yale avec un chiffre indiquant le ton) et Tang. Pour le japonais il existe les deux prononciations *on* (lecture sino-japonaise) et *kun* (lecture japonaise), sur lesquelles nous donnons des explications plus bas. Ce sont ces six possibilités qui sont proposées dans un menu à choix multiple. Il ne suffit donc pas de connaître une prononciation répertoriée par Unihan, il faut encore savoir à quel type de lecture elle appartient.

Nous avons testé une entrée par la lecture sino-japonaise *ichi* pour le caractère « un » et japonaise *kokoro* pour le caractère « cœur ». La requête dans la base de

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

données Unihan renvoie le caractère « un » U+4E00 et le caractère « cœur » U+5FC3, avec toutes les informations mentionnées plus haut ainsi que le glyphe en fonte Unicode (non reproduit dans le tableau). En réalité, le système renvoie une liste, et le caractère se trouve parmi des dizaines ou des centaines d'autres. La recherche dans la base se fait uniquement par chaîne de caractères, il faut donc penser à entrer le mot avec des blancs pour obtenir seulement *ichi* et non *nichi, ichigo* etc. De plus, la prononciation n'est pas univoque, comme on l'expliquera plus loin.

Si l'on connaît une définition standardisée

On peut accéder aussi au caractère en fournissant sa « définition » conventionnelle anglaise (en l'occurrence *heart* pour le caractère « cœur »). Là encore, on obtient beaucoup de réponses, car *heart* figure dans plusieurs définitions.

Si l'on connaît le glyphe

Si l'on ne connaît ni la prononciation ni la définition, on peut avoir un accès visuel au caractère recherché en affichant les glyphes rangés dans des tableaux 16*16 avec leur code JUC-2 (*Unihan grid index*). L'affichage normal du tableau se fait par images gif intégrées, ce qui est très lent, mais présente l'avantage de garantir la visualisation du glyphe, même avec des systèmes d'exploitation anciens.

On peut cocher la case UTF-8 pour un accès plus rapide, qui permet la gestion Unicode du caractère par les systèmes d'exploitation récents. Cependant, il faut aussi que le navigateur soit une version récente, par exemple Netscape 5 ou 6, ou Internet Explorer 4, que le navigateur soit bien configuré, et que l'on possède une police japonaise, la police standard étant Osaka. Sinon, le tableau apparaît avec un point d'interrogation dans chaque case.

Pour consulter efficacement ces tableaux, il vaut tout de même mieux connaître la clé (qui apparaît sur fond rose) car il y a des dizaines de pages à consulter sinon. Il vaut mieux aussi connaître l'ordre des clés, pour naviguer entre les pages (*blocks*) en cas d'erreur sur la clé envisagée en première approche[12].

Si l'on connaît la clé KangXi

Si l'on connaît la clé du caractère avec certitude, on peut interroger par la clé et le nombre de traits. Dans le mode de consultation *Unihan Radical-Strokes Index*, on a accès aux clés en sélectionnant d'abord leur nombre de traits. Ensuite, on sélectionne la clé voulue dans une liste et l'on entre un nombre de traits additionnels (traits formant un caractère complexe sans compter ceux de la clé). Ce nombre est nécessairement compris entre un minimum (0) et un maximum (50). On peut ainsi

12. Si l'on connaît le code du caractère ou d'un caractère voisin, l'adresse à consulter pour un accès direct est http://www.unicode.org/cgi-bin/UnihanGrid.pl?codepoint=UUUU où les lettres u sont remplacées par le code hexadécimal JUC du caractère Han.

consulter les listes de caractères et de complexes sous forme de tableaux. On demande par exemple pour la clé du « cœur » 61 les complexes de 0 trait (minimum) à 0 trait (maximum), pour avoir le caractère simple uniquement, ou les complexes de 1 à 50 traits pour avoir la liste exhaustive des caractères complexes formés avec la clé du cœur. Pour retrouver par exemple le caractère complexe « agréable » dont on sait qu'il compte 4 traits additionnels, on demande les complexes dans la fourchette restreinte à 4 traits.

On peut cocher la case UTF-8 pour un accès plus rapide. Rappelons que l'affichage normal du tableau se fait en gif, aussi si l'on interroge avec une grande fourchette de nombres de traits, et qu'il existe un grand nombre de caractères complexes, ceux-ci vont s'afficher sur plusieurs pages, et cela peut prendre beaucoup de temps. A titre indicatif, pour la clé 61, en interrogeant sur une fourchette de 10 traits additionnels (caractères de 2 à 12 traits en plus de la clé) on obtient 6 tableaux de 260 cases chacun. Il est vrai que la clé du cœur est très productive, avec la clé 1 « chiffre 1 », ou la clé 214 « flûte », le nombre de caractères complexes est moindre.

On peut également cocher l'option *Search* lorsqu'on se trouve sur la page d'un caractère jouant le rôle de clé pour sélectionner un sous-ensemble de caractères complexes[13].

L'inconvénient de ce système est qu'il est lourd et lent. L'accès par un nom mnémotechnique serait plus convivial, mais il faudrait alors choisir entre ceux des différents pays de la zone. On n'obtient aucune réponse en entrant par exemple *shinben*, qui est un nom de clé japonais non référencé. Pas davantage actuellement en entrant le numéro de caractère issu du dictionnaire Morohashi. En effet, les champs permettant d'établir les correspondances avec d'autres sources que le *KangXi* sont prévus, mais ils ne sont pas remplis. Cela dit, en passant par un chemin plus long, par exemple en interrogeant par la prononciation japonaise *kokoro*, on arrive aux données sur le caractère et l'on peut rechercher facilement ses complexes.

Si l'on connaît un code

Si l'on connaît déjà le code du caractère en UTF-16, l'accès aux informations est facilité[14]. On peut également avoir accès aux caractères Han si l'on connaît un code d'une norme officielle référencée, par exemple CNS, JIS ou d'un standard comme

13. Lorsque l'on connaît le numéro de clé KangXi, on peut accéder directement à l'information sur les complexes. L'adresse à consulter est http://www.unicode.org/cgi-bin/UnihanRSIndex.pl?radical=X&minstrokes=Y&maxstrokes=Z où la lettre X est remplacée par le code de la clé, Y par le nombre de traits additionnels minimum souhaité et Z le nombre de traits additionnels maximum.

14. Voir à l'adresse
http://www.unicode.org/cgi-bin/GetUnihanData.pl?codepoint=UUUU. Les lettres u sont remplacées par le code hexadécimal JUC du caractère Han. On peut cocher la case UTF-8 pour un accès plus rapide.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

celui de Xerox. De même, on a accès au glyphe si l'on connaît le nom du caractère dans la nomenclature répertoriée dans *Unicode Characters Names List*. C'est cependant un mode d'entrée peu probable, compte tenu de l'arbitraire de ce nom.

Nous allons tenter d'éclairer ces données quelque peu ésotériques, en les replaçant dans leur contexte historique et culturel. Nous prendrons l'exemple du japonais dont l'écriture est héritée de l'écriture chinoise.

4. L'écriture japonaise, un enjeu pour l'informatique

4.1. *Mode de fonctionnement de l'écriture japonaise*

Une graphie mixte

Un texte japonais contemporain se présente comme une suite pratiquement ininterrompue de signes. Le paragraphe est le plus petit ensemble de caractères séparé par des blancs, ce qui, pour nous, est la définition du mot graphique. Point de mots, donc, tels que nous les percevons ordinairement. L'œil occidental non averti, mais néanmoins observateur, reconnaîtra tout d'abord dans cette masse compacte, des ponctuations : le point, qui est un gros rond, la virgule, à l'envers, mais bien identifiable, des crochets éventuellement, qui jouent le même rôle que les guillemets. On peut ensuite distinguer, dans la suite continue des signes d'écriture, quelques zones denses, surchargées de traits et inégalement réparties entre des zones plus aérées. Deux sortes de glyphes en effet : l'écriture japonaise est dite mixte, elle comporte des caractères chinois (*kanji* en japonais) et des caractères autochtones (*kana*) qui correspondent à des syllabes. L'alternance entre ces deux modes d'écriture joue, pour les Japonais, un rôle de différenciation similaire à celui que l'alternance entre espaces et lettres joue pour nous dans un texte écrit : elle facilite la lecture.

En faisant un effort, devant un texte assez riche, on pourra peut-être, dans les parties plus « claires », faire la différence entre des signes cursifs aux contours sinueux (les *hiragana*) et d'autres plus agressifs, anguleux, ces derniers étant cependant plus rares : ce sont les *katakana*. Enfin, on pourra même parfois reconnaître quelques mots en caractères latins.

Telles sont, à première vue, les caractéristiques de l'écriture japonaise, foncièrement syncrétique. Les parties les plus denses du texte, donc plus riches en informations, sont les caractères empruntés à l'écriture chinoise ou *kanji ;* ils permettent une appréhension réputée immédiate du sens (idéogrammes), ce qui est une description certes excessivement simplifiée, passant sous silence les années d'apprentissage, mais que nous ne discuterons pas ici (Tamba, 1986 ; Lucas, 1988 ; Galan, 2001, 2002).

Dans le petit extrait de texte qui suit[15], on reconnaîtra dans la dernière phrase du 1er paragraphe, vers la fin de la deuxième ligne, l'idéogramme du « cœur » donné en exemple plus haut.

「音楽は世界の言葉」だといわれている。たしかにその通りである。音楽の力は偉大だ。国境がないし、民族の違いを超越して人の心を動かす。
しかし、違う意味で、指揮者にとって言葉はとても重要である。
聴衆には、たしかに「音楽は世界の言葉」であるけれど、演奏を聴いてもらうために、指揮者はオーケストラと何回もリハーサルをしなければならない。演奏の内容を高め、自分の解釈をオーケストラに徹底させるために、説明すること、つまり言葉は、絶対に必要である。

Exemple 1. *Texte japonais contemporain*

Voici la translittération et la traduction du premier paragraphe.

« Ongaku wa sekai no kotoba » da to iwareteiru. Tashikani sono tôri dearu. Ongaku no chikara wa idai da. Kokkyô ga naishi minzoku no chigai wo chôetsushite hito no kokoro wo ugokasu.

« La musique est la langue du monde », dit-on. C'est bien vrai. La force de la musique est immense. Elle touche le cœur des hommes, transcendant les frontières et les différences entre les peuples.

Le sens traditionnel d'écriture, en colonnes, de haut en bas et de la droite vers la gauche, est conservé pour la littérature, la presse et les ouvrages à caractère généraliste, tandis que les textes techniques ou scientifiques s'écrivent, en général, en lignes, horizontalement de la gauche vers la droite. Les documents électroniques et sites web ont suivi cette habitude déjà vulgarisée depuis des lustres.

Le japonais écrit s'est fixé à la fin du XIXe siècle sous sa forme mixte actuelle. C'est le produit de l'adaptation de l'écriture chinoise à la langue japonaise. Du point de vue linguistique, le japonais et le chinois sont très éloignés, ils n'appartiennent pas à la même famille. Du point de vue culturel en revanche, l'influence de la Chine est prédominante. Les Japonais, qui avaient encore une culture orale au IVe siècle, ont adopté l'écriture, ainsi qu'un vaste vocabulaire, lors de l'expansion culturelle chinoise. Vers le VIIe siècle les femmes de la cour, qui n'avaient pas accès à la culture chinoise, inventent un syllabaire *(hiragana)* pour noter phonétiquement le japonais, ce qui permet l'éclosion d'une littérature épique et courtoise autochtone. Concurremment, les moines mettent au point un autre syllabaire *(katakana)* volontairement distinct dans la forme, mais qui note les mêmes sons et qui sert à annoter les textes bouddhiques pour faciliter leur diffusion populaire.

15. Il est tiré d'un journal de la Fondation du Japon qui propose des articles faciles à lire pour les apprenants du japonais. Shikisha to bokokugo, Iwaki H. *Nihongo kyôiku tsûshin* 15, 1993.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

Les kanji

Les caractères Han transcrivent le vocabulaire construit sur une base sino-japonaise, les mots en lecture *on* c'est-à-dire suivant la prononciation à la chinoise[16]. C'est le cas de *ongaku*, la musique, que l'on trouve dans la première phrase, dans l'exemple 1, écrit par les deux premiers caractères Han 音楽. Ils sont utilisés aussi pour noter des substantifs, dont la prononciation est japonaise (lecture *kun*, selon le sens) par exemple *kokoro* pour le caractère du « cœur » dans l'expression *le cœur des hommes.* Les *kanji* notent encore la racine des verbes, et on en voit un exemple dans la dernière phrase du 1e paragraphe, juste avant le point 動かす. Ici le caractère sera prononcé aussi à la japonaise (lecture *kun*) *ugokasu,* la terminaison (variable) étant notée en *hiragana*. La traduction serait « *fait bouger* » ou « *émeut* ».

Il est à noter que le système phonétique du japonais n'utilise pas les tons, contrairement au chinois. La prononciation sino-japonaise *on* est appauvrie, car non seulement elle ne reproduit pas les tons du chinois mais elle déforme aussi certaines consonnes. Il s'ensuit que beaucoup de syllabes, différenciées en chinois, ont une même prononciation sino-japonaise, d'où les problèmes d'homophonie. *Kan* par exemple est une lecture *on* qui correspond à plusieurs dizaines de caractères Han différents.

Les caractères Han sont en proportion variable selon la nature du texte (littéraire, scientifique, transcription de la langue parlée, etc.). Jusqu'à récemment, la tendance historique était à la baisse dans l'usage des *kanji*, spécialement dans la presse (60 % de *kanji* en 1878, et seulement 40 % en 1966 pour les quotidiens), et d'aucuns redoutaient leur disparition totale.

Il est difficile de donner le nombre exact des *kanji*, qui n'est pas figé, mais il est beaucoup plus restreint que celui des caractères mandarins. A titre indicatif, le grand dictionnaire Morohashi (*Dai Kanwa jiten*) en propose 48 894. Une police d'imprimerie peut en compter de 10 000 à 15 000. Les dictionnaires courants en comptent 5 à 6 000. Le ministère japonais de l'éducation a défini en 1981 une liste de référence de 1 945 caractères d'usage courant pour l'apprentissage scolaire. Dans la pratique, la presse quotidienne par exemple, on utilise environ 3 000 *kanji*. Cela tient beaucoup aux noms propres, qui introduisent une plus grande variété, ainsi qu'aux mots savants. A ce propos, il arrive que les caractères peu courants soient accompagnés de leur prononciation donnée en syllabaire, écrite en tout petit à côté (des colonnes), ou au-dessous ou au-dessus (des lignes) pour le ou les caractères difficiles. Ces caractères d'annotation sont appelés *furigana*.

Quelques caractères « chinois » ont été en réalité créés au Japon, ce sont les *kokuji*, par exemple le glyphe complexe du croisement des chemins *tsuji* forgé à partir de la clé du chemin (simplifiée) et du caractère « dix » qui forme une croix.

16. Il en existe plusieurs possibles pour un même caractère, les périodes et régions d'emprunt ayant varié au cours de l'histoire.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

Les kana

Les *kana* représentent la cinquantaine de sons du japonais, sous deux variantes (*hiragana* et *katakana*), un peu comme majuscules et minuscules ou encore droit et italique. L'un ou l'autre de ces syllabaires pourraient suffire à noter la langue japonaise. Cependant, les Japonais n'ont pas retenu cette solution dans l'usage courant.

Les *hiragana*, de forme cursive, sont utilisés pour représenter les éléments grammaticaux du japonais (par exemple adverbes, terminaisons de verbe). Ils peuvent remplacer à l'occasion les mots généralement écrits en *kanji*. Ainsi les livres pour enfants utilisent essentiellement les *hiragana*, les caractères chinois étant introduits graduellement au cours de la scolarité (Galan, 2002). Dans un quotidien, de nos jours, on trouve environ 35 % de *kanji* et 65 % de *kana*. Dans l'exemple 1, on trouve au début du 2e paragraphe un adverbe écrit en hiragana しかし *cependant* qui se lit *shikashi* ou encore à la fin du 3e paragraphe, un verbe *dearu* である. On a déjà noté un verbe en graphie mixte *ugokasu*, 動かす dont la terminaison *-kasu* est notée en *hiragana*. On peut voir aussi le signe *ha* は qui est une marque grammaticale indiquant le thème, par exemple après le mot *ongaku* « musique » 音楽は.

Les *katakana*, de forme anguleuse, servent le plus souvent à mettre en relief certains éléments, par exemple à transcrire des onomatopées et les mots d'origine étrangère. Dans l'exemple 1, dernier paragraphe, se trouvent les mots *rehearsal* et *orchestra* transcrits respectivement *rihâsaru* リハーサル et *ôkesutora* オーケストラ.

Quant à l'alphabet latin (*romaji*), il est largement utilisé par la publicité pour son aspect exotique. Il sert également comme notation spéciale, en particulier dans des articles techniques ou scientifiques pour noter des sigles ou encore des mots étrangers dans leur graphie d'origine.

La variété de signes, ainsi que le nombre élevé de *kanji* a longtemps constitué un obstacle à la mécanisation de l'écriture. Une machine à écrire en japonais a été mise au point en 1915, mais elle ne possédait qu'un jeu limité de caractères, de plus son maniement était extrêmement compliqué (Griolet, 1985). Il était plus rentable d'écrire les textes courants à la main et de s'adresser à un imprimeur quand cela était nécessaire. De même, aux tout débuts de l'informatique, le clavier dactylographique constituait à lui seul un barrage pour la saisie du japonais. Les choses ont bien changé depuis, mais c'est dans ce contexte que les Japonais ont développé l'interaction vocale avec les ordinateurs.

4.2. *L'informatisation du japonais*

Dans les années 1950, seul l'anglais était répandu en tant que langue associée à l'informatique, dans le monde entier, et souvent uniquement en majuscules. Au Japon, le problème politique et culturel était d'importance. Les occupants

américains prônaient en effet l'abandon pur et simple de l'écriture idéographique au profit d'une écriture latine. Les intellectuels japonais prennent parti pour ou contre. Commencent alors de difficiles recherches qui débouchent, somme toute rapidement, sur la solution actuelle. Le codage sur deux octets est suivi vers la fin des années 60, par la mise au point d'un procédé de saisie phonétique du japonais, sur un clavier normal (américain), permettant l'affichage immédiat en *kana* et la conversion automatique des *kana* en caractères chinois. Dans les années 70, il n'est plus question de romaniser l'écriture et l'on établit des normes pour le traitement de l'information.

Les normes de codage JIS

Le code ASCII (American Standard Code for Information Interchange) adopté par les constructeurs américains fonctionne sur un octet (un caractère de contrôle et 7 bits utiles) et permet d'imprimer 94 caractères en pratique. On voit immédiatement qu'il est ainsi impossible de traiter les quelques milliers de caractères nécessaires pour écrire en japonais. Le problème allait être résolu par l'utilisation d'un deuxième octet. Après une période de tâtonnements, les Japonais se mirent d'accord sur le nombre de caractères à retenir et sur les codes correspondants. En 1978 l'association de normalisation industrielle du Japon publie la norme JIS (Japan Industrial Standard) C6226.

Cette norme, modifiée plusieurs fois depuis (JIS X0201, JIS X 0208 en 1983, 1990, JIS X 0213 en 1990), établissait un premier canevas, et, sous son apparente complexité, répondait à de réels besoins d'optimisation du traitement des caractères. C'est pourquoi nous revenons à cet ancêtre. La norme JIS C6226 définit un jeu de 6 349 caractères. Ceux-ci sont répartis en 2 niveaux, JIS 1 et JIS 2, que l'on peut représenter dans un tableau de 94 lignes et 94 colonnes.

Le niveau 1 comprend les signes de ponctuation et divers symboles (lignes 1 et 2), les chiffres arabes et l'alphabet majuscule et minuscule (ligne 3) alphabet dit « anglais » car sans signes diacritiques (accents, etc.), les *hiragana* (ligne 4), les *katakana* (ligne 5), les alphabets grec (ligne 6) et cyrillique (ligne 7) et 2 965 *kanji* classés dans l'ordre phonétique (lignes 16 à 47). Ce sont donc les caractères Han les mieux connus, et le niveau 1 suffisait dans la plupart des applications. Le classement phonétique des caractères permet en particulier d'optimiser la conversion *kana-kanji* des traitements de texte, chaque caractère étant appelé par l'intermédiaire de la syllabe correspondante.

Le niveau 2 contient 3 384 *kanji* d'usage moins fréquent ou des versions non simplifiées de certains caractères du premier niveau, classés selon l'ordre des 214 clés traditionnelles et ensuite par nombre de traits (lignes 48 à 83). Lors de la saisie par un logiciel de traitement de texte des caractères du niveau 2, on passe un certain temps à chercher le caractère désiré dans des tables, en remplacement de celui qui est proposé lors de la frappe par le système de conversion *kana-kanji*.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

On constate qu'il reste plusieurs cases vides dans ce tableau : les lignes 8 à 15 (752 cases) et 84 à 94 (1 034 cases). Un certain nombre de lignes est laissé à la disposition de l'usager (expert tout de même) qui peut ainsi ajouter, grâce à un programme annexe, des caractères ne figurant pas parmi les 6 349 *kanji* retenus par la norme ; on peut ainsi introduire des noms de personnes ou de lieux ayant une graphie rare, mais aussi des symboles ou logos particuliers, voire des *kanji* nouveaux, forgés à l'instar de Mishima, par des esprits créatifs. Ce sont les caractères d'usage privé ou *Private Use*.

On voit que cette norme était étudiée pour des japonophones et directement liée à la pratique du traitement de texte. Les caractères du niveau 1 étaient classés en fonction de la fréquence d'usage et accessibles au même titre que les *kana*.

Signalons un problème non résolu, celui de la gestion des *furigana*, les petits caractères d'annotation de caractères chinois. Il s'agit davantage d'un problème de mise en page, qui échappe aux normes JIS et au standard Unicode. Les *furigana* en effet doivent rester collés au caractère annoté. Les imprimeurs professionnels gèrent encore cette mise en forme délicate, qui nécessite un traitement informatique complexe de la chasse et de la taille des *furigana*. Dans l'usage courant, sur les sites web des particuliers et même des quotidiens, on a recours aux parenthèses pour noter les prononciations des caractères Han inhabituels. Toutefois, cette solution de facilité est jugée très inesthétique.

Unicode

Unicode regroupe dans la base Unihan les glyphes en usage en Asie, dans la sphère d'influence culturelle de la Chine, symboles et idéogrammes. Les codes spécifiques pour le japonais reprennent la base établie par la norme JIS X 0208 de 1990, qui recense 6 877 glyphes dont 6 353 caractères Han, amendée par JIS X 0212 et JIS X 0213 de 1990. Les *hiragana* et les *katakana* sont codés dans des tableaux spécifiques accessibles par l'index des codes (*Chart index*). On trouve ainsi les ponctuations, symboles, 91 glyphes *hiragana* (tableau 50) et 94 *katakana* (tableau 51). Le nombre de glyphes est plus élevé que le nombre de *kana*, car il existe des « minuscules » signes plus petits servant pour les allongements, les diphtongues ou les redoublements. La différence entre le nombre de *hiragana* et celui des *katakana* est imputable aux symboles de ponctuation et diacritiques associés aux *katakana*.

Comme on l'a vu, ce sont les tables Unihan qui permettent de retrouver le code unifié pour un caractère Han. Dans l'ensemble, les habitudes JIS sont respectées. Toutefois, dans le souci d'uniformisation allant souvent de pair avec la normalisation, dans le cadre du CJC, les Japonais ont perdu quelques degrés de liberté dans l'optimisation des classements de caractères. La norme commune est évidemment plus lourde à gérer.

On pourra s'étonner de la complexité ou de la lourdeur du système des correspondances. Si l'on prend le cas de « un » (*ichi* en japonais) il est codé en tant

que caractère par son identifiant 4E00 (et tant que clé sémantique 1.0), mais il figure aussi par son identifiant 2F00 en tant que glyphe servant à dessiner des caractères (appelé ici forme de clé), et encore ailleurs avec un autre identifiant 3021 en tant que lettre représentant un chiffre. En passant par les classes de combinaison canoniques (*Canonical combining classes*) on peut établir des équivalences entre la classe Lo pour les « lettres » (en réalité caractère d'écriture) et N1 pour les lettres représentant des chiffres.

Un petit problème est constitué par les caractères « chinois » en réalité forgés au Japon, les *kokuji*, qui sont peu nombreux mais d'usage très courant. Ils sont traités comme de véritables caractères Han, quand ils ont été répertoriés par des dictionnaires chinois ou coréens, sinon, ils ne sont pas accessibles dans Unihan mais figurent dans les tables annexes.

Le tableau 3 montre le code général JUC en UTF-16 pour quelques *kana* et graphies associées.

N°clé. nb de traits	Caractère simple	Caractère complexe	kana	Code JUC (hexa)	commentaire
			あ	3042	A (hiragana)
			ア	30A2	A (katakana)
			は	306F	Ha (hiragana)
			ハ	30CF	Ha (katakana)
			ー	30FC	(allongement, katakana)
1.0	一			U+4E00 3021	*Ichi* un
1.0				U+2F00	**Forme de clé *ichi***
1.4		世		U+4E16	*Se* monde
61.0	心			U+5FC3	*Kokoro* cœur
61.0				**U+2F3C**	**Forme de clé *shinben***
61.0				**U+5FC4**	**Forme de clé *risshinben***
61.3		感		U+611F	*Kan* sentiment
61.4		快		U+5FEB	*Kai* agréable
61.5		思		U+601D	*Omo(u)* penser
61.9		意		U+610F	*I* désir, volonté
61.9		応		U+5FDC	*Ô* réponse
61.11		慕		U+6155	*Shita(u)* adorer
162.2		辻		U+8FBB	*Tsuji* croisement

Tableau 3. *Quelques caractères japonais ou sino-japonais et leur code JUC-2*

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

On voit que le tiret, signe d'allongement pour une syllabe en *katakana,* ressemble beaucoup visuellement au *kanji* de « un ». C'est le signe utilisé pour allonger le son *ha* dans *rihâsaru* リハーサル qui reproduit plus ou moins la prononciation de *rehearsal* de l'exemple 1. Y figurent également les caractères Han déjà mentionnés, ainsi qu'un caractère *kokuji*, celui du croisement des chemins *tsuji,* qui est répertorié comme caractère Han (avec le préfixe U+). On a ajouté dans le tableau le code correspondant au numéro de clé et nombre de traits pour permettre la reconnaissance de la clé de classement sémantique.

De plus, pour montrer la différence entre le glyphe élément de dessin (forme de clé) et le caractère Han, on a donné en gras le code des glyphes visualisables représentant la clé de « un » (*KangXi radical one* dans la nomenclature JUC), la clé du « cœur » normale dite en japonais *shinben* (*KangXi radical heart one* dans la nomenclature JUC) et la clé du cœur debout dite *risshinben* (*KangXi radical heart two* dans la nomenclature JUC). Quoique ces formes de clé soient imprimables, elles ne sont pas insérées dans le tableau, car elles ne sont pas accessibles à partir d'un logiciel de traitement de texte simple commme celui que nous avons utilisé.

4.3. *Les critiques*

Comme toute norme, JIS et Unicode ont suscité des critiques. Au Japon, mais aussi ailleurs, les utilisateurs de l'écriture chinoise relèvent essentiellement l'antagonisme entre la vision normalisatrice qui suppose que le nombre de signes est fini, et la productivité du système Han. En effet, même si notre concept de mot ne recouvre pas exactement le caractère Han, il s'agit bien d'une unité linguistique et le lien entre caractère et lexique est très fort. Puisque le monde change, que le vocabulaire évolue, de nouveaux caractères peuvent apparaître, et on ne saurait limiter cette fonction vitale de renouvellement.

La marge laissée pour les caractères hors norme est un piètre palliatif à la vision figée des systèmes d'écriture proposée par JIS puis Unicode. Même au Japon, où le nombre de caractères répertoriés a été et reste modeste, par rapport à ceux du mandarin, le problème de la limitation est fréquemment évoqué par les défenseurs de la langue. Il est clair que l'invention de caractères nouveaux est devenue exceptionnelle. Au Japon, Mishima est un des rares auteurs a avoir forgé de nouveaux caractères au siècle dernier. N'empêche, répondent certains, il ne faudrait pas que la création soit découragée pour des raisons techniques. Qui sait si les générations futures ne vont pas investir massivement les possibilités de composition offertes par le système idéographique ? On a voulu énumérer exhaustivement les caractères Han, mais ce n'est sans doute pas la bonne unité.

Il est en effet dommage que les caractères Han aient été perçus comme des signes graphiques sinon en nombre fini, du moins énumérables, alors qu'ils constituent un système évolutif. Quitte à définir un ensemble clos, relativement stable dans le temps, il eût été plus judicieux de garder les clés comme base de

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

composition car elles sont effectivement énumérables. Ceci aurait permis de ne pas limiter le nombre des complexes. On pourrait donc imaginer former des glyphes complexes à partir des caractères simples, comme on forme des mots à partir des lettres. Une autre solution, à un niveau de granularité encore plus fin, eût été de traiter les traits comme jeu de formes fini. Les traits sont aussi une base de composition pour former des caractères simples. Ces solutions ont été étudiées au Japon et à Taiwan mais elles n'ont pas été comprises par les partenaires du consortium. Elles sont en effet basées sur une approche à très long terme, et sur l'idée que les périodes de création et les période de réduction des caractères Han se sont succédées et se succéderont dans l'histoire. Si notre époque est plutôt une période de réduction du nombre des caractères Han, cela ne veut pas dire qu'il s'agisse d'un aboutissement.

Les critiques adressées à Unicode sont finalement du même ordre que celles que l'on peut adresser aux visions réductrices et fixistes du traitement informatique des langues. Le *français restreint* ou le *Simplified English* prônés par l'informatique de première génération (et encore en usage dans la documentation technique) supposent un monde clos, où tous les mots sont répertoriés une fois pour toutes. Les limitations du vocabulaire et de la syntaxe imposées aux rédacteurs de notices techniques ou de manuels (ceux-ci devant être traduits automatiquement) aboutissent à un figement de l'expression écrite, sans néologie, sans dérivation, sans glissement ni enrichissement de sens, pour tout dire, à une sorte de code.

On pourra arguer que l'usage évolue partout, et que les normes suivent : les *smileys* ou émoticons des courriers électroniques peuvent être normalisés un jour et passer de mode ensuite, au profit de notations graphiques futures, abrégées ou complexifiées, dont nous ne savons encore rien.

Par ailleurs et sur un autre plan, les utilisateurs favorables à une normalisation regrettent par avance le coût (en espace mémoire) du codage sur trois octets, qui est prévu et constitue une extension logique dans la droite ligne du projet JUC. De même que nous pouvons nous plaindre que la taille des documents codés en UTF-16, même optimisée en UTF-8, oblige à gérer des fichiers beaucoup plus importants que lorsqu'on se contentait de traiter des fichiers ASCII, de même certains Japonais se plaignent que les exigences des Chinois aboutissent à terme à un codage UTF-32 lourd pour tout le monde, alors qu'eux-mêmes pourraient se contenter d'un codage sur deux octets en UTF-16. Même si sur le fond, ils sont favorables à la logique de « qui peut le plus peut le moins », ils ne se privent pas de critiquer ceux qui sont plus gourmands qu'eux en espace de codage.

4.4. *Le traitement de l'information*

L'existence d'Unicode permet de rapatrier des informations, de les restituer à l'écran, mais aussi de traiter informatiquement des contenus sans se soucier du codage d'origine. Voyons ce qui se passe si un Japonais cherche à traiter des

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

documents multilingues, comprenant par exemple du français ou/et du japonais, ce qui nécessite la compatibilité des formats d'échange avec des normes d'origine ISO latin et JIS. Le tableau 4 permet de comparer les codes JUC hexadécimaux et les codes binaires pour différents glyphes alphabétiques, syllabiques ou idéographiques. Les chiffres en gras facilitent la reconnaissance des règles de conversion du code JUC du format UTF-16 au format UTF-8 en binaire (voir l'article de Giguet et Lucas dans le même numéro pour plus de détails techniques).

Indexation

Le Japon fait partie des pays de premier rang pour la collecte et le stockage de l'information mondiale. Cette activité ne saurait que s'intensifier, dans la mesure où les sources deviennent accessibles de façon transparente.

Dans le sens inverse, les banques de données japonaises scientifiques ou économiques sont accessibles depuis l'étranger (Haon, 1995). Les plus importantes et les plus anciennes, comme celle de Nikkei dans le domaine économique, permettent l'accès par des mots anglais. Mais l'interrogation en ligne depuis la France par exemple peut néanmoins poser des problèmes, si la saisie prévue est en caractères japonais ou sino-japonais. De même, les moteurs de recherche locaux sur les sites web et portails s'avèrent très souvent inutilisables, à moins de disposer d'un traitement de texte japonisé, qui permet d'entrer les mots et de les convertir en *kanji* avant de les soumettre *via* UTF-8 ou UTF-16. Ce truchement est souvent efficace, mais pas toujours. La mise à niveau Unicode des systèmes de recherche locaux prendra sans doute un certain temps.

Informatique linguistique

Le Japon s'est trouvé confronté au problème de la conversion automatique *kana-kanji,* ce qui a suscité des travaux importants sur l'analyse automatique du japonais. La saisie s'effectue en transcription phonétique, c'est-à-dire à partir des mots tels qu'ils sont prononcés, soit directement en *kana*, soit le plus souvent en caractères latins, à partir d'un clavier QWERTY. Les syllabes sont automatiquement converties et affichées en *kana*. La conversion en *kanji* était ensuite effectuée à la demande, caractère par caractère, ou mot à mot, mais elle est désormais effectuée en ligne segment de phrase par segment de phrase ou phrase par phrase selon le degré de sophistication du logiciel. La conversion mot à mot des débuts a été abandonnée, en raison du grand nombre d'homophones (mots prononcés de la même manière mais correspondant à des caractères Han différents) car l'utilisateur devait intervenir trop souvent pour choisir parmi plusieurs caractères celui qui convenait, ce qui rendait la saisie très lente. La conversion de segments de phrase ou de phrases est plus efficace, certaines ambiguïtés étant levées par l'analyse syntaxique automatique du contexte. De plus, les caractères fréquemment et récemment utilisés par l'utilisateur sont placés automatiquement en première position dans les fenêtres de choix. Les logiciels en vente actuellement possèdent tous un analyseur syntaxique

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

intégré, qui rétablit la graphie mixte *kanji-kana*, l'utilisateur n'intervenant plus que de temps en temps pour corriger les choix erronés de la machine.

Glyphe	Code JUC (hexa)	Nom du caractère JUC	Code JUC (binaire) UTF-16	Code UTF-8 (binaire)	Code UTF-8 (hexa)
é	00E9	Latin Small Letter E With Acute	**11101001**	11000**11** 10**101001**	
€	20AC	Euro-Currency Sign	**10000010101100**	11100**10** 10**000010** 10**101100**	
あ	3042	Hiragana letter A	**0011 0000 0100 0010**	1110**0011** 10**000001** 10**000010**	
は	306F	Hiragana letter HA	**0011 0000 0110 1111**	1110**0011** 10**000001** 10**101111**	
ア	30A2	Katakana letter A	**0011 0000 1010 0010**	1110**0011** 10**000010** 10**100010**	
ハ	30CF	Katakana letter HA	**0011 0000 1100 1111**	1110**0011** 10**000011** 10**001111**	
一	4E00	Unified ideograph one	**0100 1110 0000 0000**	1110**0100** 10**111000** 10**000000**	E4 B8 80
心	5FC3	Unified ideograph heart	**0101 1111 1100 0011**	1110**0101** 10**111111** 10**000011**	E5 BF 83
⼼	*2F3C*	*KangXi radical heart one*	**0010 1111 1100 0011**	1110**0010** 10**111111** 10**000011**	
思	601D	Unified ideograph thought	**0101 0000 0001 1101**	1110**0101** 10**000000** 10**011101**	E6 80 9D

Tableau 4. *Exemples de codage en UTF-8 ou en UTF-16 d'un caractère JUC*

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

L'anglais est toujours considéré comme la langue dominante, aussi de nombreux travaux en informatique linguistique se focalisent sur cette langue. En témoigne la participation du Japon à de nombreux concours américains comme DARPA (reconnaissance de la parole et interface oral-écrit) ou TREC (recherche d'informations ciblée dans les textes).

Mais le Japon fait aussi figure de leader régional, formant un très grand nombre d'ingénieurs et chercheurs de la région asiatique, ce qui contribue sans doute à contrebalancer son image négative dans la région depuis la guerre. Le centre de recherche international ATR[17] de Kyôto lancé et animé par les Japonais favorise les travaux sur les langues et écritures chinoises, coréennes, thaï, etc. L'intégration d'Unicode a été rapide et une intense activité scientifique est en cours pour traiter de l'interface oral-écrit, et de la traduction. Dans d'autres centres de recherche, on traite des textes de toutes origines à l'aide de modèles éprouvés pour le traitement du japonais, tant pour l'indexation que pour l'analyse syntaxique ou l'extraction thématique. Les langages de programmation ne sont pas encore tous adaptés à Unicode, mais ceux qui le sont comme Java ou Perl bénéficient de cet avantage.

Les congrès internationaux comme CoLing[18] voient donc une participation accrue de jeunes chercheurs des pays asiatiques, aux côtés des Japonais qui y sont activement présents depuis de nombreuses années. Les congrès internationaux spécialisés pour les langues et écritures asiatiques comme PACLIC[19] attirent de leur côté des chercheurs australiens, américains et quelques Européens.

En résolvant les problèmes complexes liés à leur écriture, les Japonais sont allés technologiquement plus loin que les Occidentaux dans les dernières décennies. Les ordinateurs devaient afficher les caractères avec une très grande finesse, il en est résulté les écrans à haute définition. Pour les imprimer lisiblement, il a fallu des imprimantes matricielles très fines ou à laser, qui ont ensuite envahi le marché mondial. Des mémoires rapides étaient indispensables au traitement. Enfin, la constitution de dictionnaires et la modélisation de l'analyse syntaxique qui ont accompagné la mise au point du traitement de texte, ont permis aussi la mise au point d'interfaces sophistiquées, tels les systèmes de lecture pour aveugles, qui prennent en entrée un texte papier, traité par un lecteur OCR, texte ensuite analysé et vocalisé automatiquement.

On assiste ainsi, en moins d'un siècle, à un retournement de situation, qui a mené de la tentation d'abandon des écritures idéographiques, réputées ingérables, à leur éloge appuyé. Les arguments techniques évoqués dans un cas comme dans l'autre ne doivent pas faire oublier les arguments culturels, politiques et sociaux, moins immédiats mais plus décisifs que l'on ne veut souvent l'admettre.

---

17. *Advanced Telecommunications Research Institute International.*

18. Computational Linguistics, congrès international parrainé par l'ACL *Association for Computational Linguistics.*

19. *Pacific Asia Conference on Language, Information and Computation.*

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)

5. Conclusion

Nous avons évoqué l'adoption du codage unifié des idéogrammes par le consortium Unicode, au sein d'un standard basé sur la norme ISO-CEI10646, et en particulier la base de données Unihan. Nous avons cherché à relier les choix entérinés par la norme en éclairant le contexte culturel historique.

Nous avons également évoqué quelques facettes de l'échange, par l'intermédiaire des formats UTF en cours d'intégration.

L'exemple du japonais nous a servi à illustrer les avantages et les inconvénients d'une norme supranationale appréhendée à travers une culture particulière de l'écrit. Au-delà de la commodité pour un utilisateur individuel, nous avons souligné les enjeux du traitement de l'information désormais envisageable à l'échelle planétaire, défi que le Japon entend relever en s'intéressant activement à la diversité linguistique et graphique de la région asiatique. L'adoption du JUC représente une revanche pour les pays dont la culture ancienne, basée sur la connaissance des idéogrammes, était refoulée. Malgré les imperfections du système, cette intégration est certainement aussi une victoire pour la diversité du patrimoine culturel dont tous peuvent désormais tirer parti.

Bibliographie

Alleton, V. *L'écriture chinoise*, Paris: PUF (Que sais-je), 1984.

Février, J. G. *Histoire de l'écriture*. réed. 84. Paris: Payot, 1959. 616 p.

Christin, A.-M. (sous la direction de) *Histoire de l'écriture* Paris: Flammarion 2001.

Du signe à l'écriture *Pour la science* dossier hors-série octobre 2001-janvier 2002.

Galan, C. *L'enseignement de la lecture au Japon*. Toulouse: Presses universitaires du Mirail, 2002. 365 p.

Galan, C. « *Kanji* et *kana* : ce que nous apprennent les travaux récents de psycho et neurolinguistique » N. Lucas et C. Sakai (sous la dir. de), *Japon Pluriel 4*, *Actes du quatrième colloque de la Société française des études japonaises*, Arles: Philippe Picquier, 2001.

Griolet, P. *La modernisation du Japon et la réforme de son écriture*. Paris, Publications Orientalistes de France, 1985. 124 p.

Grenié, M. et Belotel- Grenié, A. « L'écriture chinoise, mythes et réalités ». *In* Du signe à l'écriture *Pour la science* dossier hors-série 33 octobre 2001-janvier 2002.

Haon, H. « L'accès à l'information scientifique et technique japonaise : moyens et problèmes » P. Beillevaire et A. Gossot (sous la dir. de), *Japon Pluriel*, *Actes du premier colloque de la Société française des études japonaises*, Arles: Philippe Picquier, 1995.

Hudrisier, H. et Lucas, N. "Les idéogrammes dopés par l'ordinateur". Allemagne Japon les deux titans *Manière de voir* n° 12, *le Monde diplomatique*, mai 1991, p. 44-46.

Loupy, C. (de). « Multilinguisme et document numérique : la dimension technique à l'épreuve du codage des caractères » *Revue SOLARIS* décembre 1999/janvier 2000.

Lucas, N. « Multivalence de la langue » *in Les bases de la puissance du Japon*, J. Esmein (sous la dir. de) Paris: Collège de France, Fondation pour les études de défense nationale, 1988. p. 91-110.

Lucas, N. « L'informatique dans le sanctuaire de la langue » in *L'évolution des systèmes japonais*, sous la direction de J. Esmein et R. Dubreuil, Paris: CESTA, 1985. p. 211-227.

Origas, J-J. « La langue et son écriture ». *Encyclopédie permanente Japon*, vol. VIII, 1979. Paris: Publications Orientalistes de France.

Tamba, I. « Approche du «signe» et du «sens» linguistiques à travers les systèmes d'écriture japonais ». *Langages* n° 82, juin 1986. p. 83-100.

Vandermeersch, L. Le nouveau monde sinisé. Paris: Presses Universitaires de France, 1986.

JLE | Téléchargé le 03/06/2026 sur https://stm.cairn.info (IP: 216.73.216.232)