

Chapitre 13

Symboles

L'univers des symboles est riche, voire illimité. Le standard Unicode définit l'ensemble suivant de symboles codés :

- symboles monétaires ;
- symboles de type lettre ;
- formes numérales ;
- flèches et opérateurs mathématiques ;
- symboles techniques ;
- symboles géométriques ;
- divers symboles et casseau ;
- symboles cerclés ou entre parenthèses ;
- combinaisons Braille ;
- symboles mathématiques alphanumériques ;
- symboles musicaux byzantins ;
- symboles musicaux occidentaux.

Le standard Unicode ne comprend pas tous les systèmes de notation. Certains symboles, transitions entre les éléments graphiques et les éléments de texte, ne sont pas codés ici car ils n'occupent pas une place bien définie dans les textes bruts.

Il est permis d'ajouter des diacritiques aux symboles, notamment le jeu compris entre U+20D0..U+20FF (voir *Section 8.9, Diacritiques*).

Les symboles de type lettre, les symboles monétaires ainsi que les formes numérales, suscrites et souscrites, subissent généralement les mêmes changements de police de caractères et de style que le texte environnant. Certains symboles disposés en carré ou entre parenthèses, utilisés dans les textes extrême-orientaux, respectent également, en règle générale, le style dominant.

D'autres symboles possèdent une graphie indépendante, voire complètement différente, du style de caractère du texte régulier avoisinant. Certains symboles tels que les opérateurs mathématiques s'utilisent avec n'importe quelle écriture ou même indépendamment de ces écritures.

Dans un contexte bidirectionnel (voir *Section 3.12, Comportement bidirectionnel*), les caractères de symbole n'ont pas de directionnalité inhérente, celle-ci est établie à l'aide de l'algorithme bidirectionnel Unicode. Lorsque l'image du symbole n'est pas bilatéralement symétrique et que ce symbole fait partie d'un texte suivi composé de droite à gauche, on utilise son image miroir (voir *Section 4.7, Caractères miroirs*).

Les ornements typographiques et les caractères de reconnaissance optique diffèrent de tous les autres caractères Unicode car ils sont codés selon leur apparence précise.

Le Braille forme un cas à part dans la mesure où il peut servir à écrire du texte. Le Braille fait partie du bloc des symboles car Unicode n'en code que les glyphes. L'association entre ces glyphes et les signes des divers alphabets est régie par d'autres normes. Lorsqu'une suite de caractères fournit avant tout de l'information textuelle, elle devrait être codée à

l'aide d'une des écritures. On codera un texte directement en Braille dans le seul cas où l'on souhaite exprimer une relation spécifique entre les signes en Braille et le texte.

Plusieurs symboles codés dans Unicode sont désuets¹. Ils sont inclus pour permettre la prise en charge d'anciennes applications, comme l'émulation de terminal ou d'autres interfaces utilisateur en mode caractères – par exemple, les éléments graphiques destinés à dessiner des cadres et les images de commande.

¹ À distinguer des caractères à éviter.

13.1 Symboles monétaires

Symboles monétaires : U+20A0 – U+20CF

Ce bloc contient les symboles monétaires qui ne sont pas codés dans d'autres blocs (voir *Tableau 13.1*). Quand Unicode respecte la structure d'une norme existante, comme dans les blocs ASCII, Latin 1 et thaï, les symboles monétaires y demeurent codés, on ne les retrouve donc pas dans le bloc des symboles monétaires.

Unification. Unicode utilise un numéro de caractère unique pour chaque symbole même s'il représente plusieurs monnaies. Plusieurs devises sont représentées à l'aide de symboles barrés. Il existe souvent alors de petites variantes d'un même symbole, c'est le cas de U+0024 \$ SYMBOLE DOLLAR qui peut comporter une ou deux barres verticales ou d'autres variantes graphiques. Unicode considère ces variantes comme des variations typographiques et ne prévoit qu'un seul caractère. Des recherches approfondies n'ont pu déterminer s'il est justifié de prétendre que des variantes de glyphes de certains symboles monétaires s'utilisent systématiquement pour représenter une monnaie en particulier². Pour un échantillon des différentes formes de U+00A3 £ SYMBOLE LIVRE, voir l'ISO/CEI PNI 10367, Annexe B (informative).

Polices. Les symboles monétaires ont généralement la même chasse (largeur) qu'un chiffre (en général un demi-cadratin). Semblables à des lettres, ils respectent le style des polices de caractères.

Le *Tableau 13-1* contient les symboles monétaires codés dans d'autres blocs.

Tableau 13-1. Autres symboles monétaires

dollar, milreis, escudo, piastre	\$	U+0024	SYMBOLE DOLLAR
cent ³	¢	U+00A2	SYMBOLE CENTIME
livre	£	U+00A3	SYMBOLE LIVRE
devise générique	¤	U+00A4	SYMBOLE MONÉTAIRE
yen, yuan	¥	U+00A5	SYMBOLE YEN
florin	f	U+0192	LETRE MINUSCULE LATINE F HAMEÇON
roupie bengalie	₳	U+09F2	SIGNE BENGALI ROUPIE
roupie bengalie	₳	U+09F3	SYMBOLE BENGALI ROUPIE
baht	฿	U+0E3F	SYMBOLE MONÉTAIRE THAÏ BAHT
riel	៛	U+17DB	SYMBOLE MONÉTAIRE KHMER RIEL

Pour d'autres formes de symboles monétaires, se rapporter aux formes à pleine chasse (U+FFE0..U+FFE6).

Symbole de l'euro. La nouvelle monnaie unique des pays membres de l'Union monétaire européenne (UME) est l'euro. Unicode attribue le numéro U+20AC (€) à l'euro.

² On trouvera dans <http://www.microsoft.com/typography/developers/fdsspec/monetary.htm> plusieurs recommandations de glyphes pour ces symboles monétaires.

³ Il ne s'agit pas du centième officiel de l'euro.

13.2 Symboles de type lettre

Symboles de type lettre : U+2100 – U+214F

Les symboles de type lettre se forment à partir de lettres normales provenant d'une écriture alphabétique. Ce bloc contient des symboles qui s'inspirent de lettres latines, grecques et hébraïques. Nombre de ces symboles ne sont codés que pour des raisons de compatibilité. En général, il est fortement déconseillé d'utiliser ces symboles de type lettre quand les caractères représentés ne sont qu'une variante de présentation d'autres caractères codés. Dans les équations, les formules ou d'autres contextes, il faut employer les formes alphabétiques normales dans le style approprié pour représenter des lettres-symboles. Ainsi, pour représenter le symbole degré Celsius « °C », utiliser la suite U+00B0 ° SYMBOLE DEGRÉ + U+0043 C LETTRE MAJUSCULE LATINE C plutôt que U+2103 °C DEGRÉ CELSIUS. Pour la recherche, on traite ces deux suites comme identiques.

U+2118 ℘ FONCTION ELLIPTIQUE DE WEIERSTRASS⁴ est dérivé d'un *p* cursif (ou Chancellerie) en bas de casse.

U+2116 № SYMBOLE NUMÉRO s'utilise en cyrillique, où il prend la forme №, ainsi que dans certaines normes asiatiques, où il apparaît sous l'aspect d'un №. En français, on utilise un « N » majuscule ou minuscule, selon le contexte, suivi d'un o supérieur (N^o ou n^o; au pluriel N^{os} ou n^{os}) et non du symbole degré (N^o ou n^o).

En typographie extrême-orientale, les symboles de type lettre s'affichent avec une *pleine chasse* ; ils occupent ainsi une cellule CJC complète. En composition verticale, ces symboles restent « debout » alors que des lettres normales équivalentes pivotent.

Quand des symboles de type lettre ont un équivalent alphabétique normal, ils se trient en ordre alphabétique ; sinon, il faut les traiter comme des symboles neutres. Les symboles de type lettre n'ont pas nécessairement les mêmes propriétés de directionalité que les lettres normales ; ainsi les symboles des quatre cardinaux transfinis (ℵ..ℹ, U+2135..U+2138) utilisés dans les textes mathématiques n'ont-ils pas la directionalité droite-à-gauche forte des lettres hébraïques dont elles dérivent.

Styles. Ce bloc comprend des signes codés qui ne sont en fait que des variantes stylistiques de lettres, phénomène rare dans Unicode. On retrouve ainsi des lettres gothiques, ajourées et de ronde⁵ employées comme symboles mathématiques (voir aussi la *Section 13.10, Symboles mathématiques alphanumériques*). Le codage de ces variantes stylistiques entérine l'usage courant qui les considère comme des symboles autonomes par rapport aux lettres normales. Dans un texte mathématique ou logique, il est permis d'assigner un style particulier à une lettre latine quelconque afin de la distinguer sémantiquement ; les applications requérant ce type de distinctions sémantiques systématiques devraient utiliser des indications de style (grâce à un protocole de niveau supérieur) plutôt que de chercher à grossir le codage Unicode par l'ajout de variantes de caractères⁶.

On parle parfois en se référant aux lettres gothiques de *Fraktur* ou de *Black Letters*⁷. D'un point de vue technique, la *Fraktur* et les *Black Letters* se différencient des lettres gothiques,

⁴ Le nom anglais de ce caractère SCRIPT CAPITAL P est faux : il ne s'agit ni d'une majuscule ni d'une lettre cursive (« script »).

⁵ Il s'agit du style de caractère appelé « script » en anglais.

⁶ Ce paragraphe doit être tempéré à la lumière de l'ajout par Unicode 3.1 de « lettres » mathématiques, voir 13.10, *Symboles mathématiques alphanumériques*.

⁷ Les Américains appellent *Gothics* ce que nous appelons linéales.

ces différences stylistiques ne méritent néanmoins pas un codage distinct. Unicode considère les formes gothiques comme des archétypes.

À des fins de codage, le style ajouré peut être utilisés pour des styles comme l'ombré ou la double frappe.

Normes. Unicode inclut des symboles de type lettre provenant de nombreuses normes nationales et de jeux de caractères privés.

13.3 Formes numériques

Formes numériques : U+2150 – U+218F

Seul un souci de compatibilité avec les normes existantes justifie le codage des formes numériques. Les remarques sur la compatibilité (*Section 13.2, Symboles de type lettre*) s'appliquent également aux formes numériques.

Fractions. Il est permis de représenter les caractères de fractions ordinaires de ce bloc à l'aide de U+2044 / BARRE DE FRACTION.

Chiffres romains. Les chiffres romains peuvent se composer à l'aide de suites de lettres latines appropriées. Afin d'assurer la compatibilité avec les normes extrême-orientales, les variantes des chiffres romains de un à douze et les chiffres L, C, D et M sont codés en bas de casse et en majuscules.

U+2180 𐍀 CHIFFRE ROMAIN MILLE CD⁸ et U+216F 𐍑 CHIFFRE ROMAIN MILLE correspondent à des variantes de glyphes des mêmes chiffres romains, mais ils se distinguent car, en règle générale, l'un ne peut remplacer l'autre et U+2180 𐍀 CHIFFRE ROMAIN MILLE C D n'est pas équivalent en terme de compatibilité à la lettre latine M. Les deux caractères U+2181 𐍁 CHIFFRE ROMAIN CINQ MILLE et U+2182 𐍂 CHIFFRE ROMAIN DIX MILLE s'utilisent comme chiffres romains; ils n'ont pas de décompositions de compatibilité Unicode. U+2183 𐍃 CHIFFRE ROMAIN CENT CULBUTÉ correspond à une forme utilisée de pair avec C et I pour former de grands nombres⁹ dont certains sont les synonymes de chiffres simples tels que D, M, 𐍄, etc.

Pour d'autres formes numériques, se reporter aux nombres à la façon de Hang-tcheou ou Hangzhou (U+3021..U+3029, U+3038..U+303A) et aux fractions présentes dans le bloc Latin-1 (U+00BC..U+00BE).

Exposants et indices : U+2070 – U+209F

Le standard Unicode ne comprend des formes en exposant et en indice qu'afin d'assurer la compatibilité avec des jeux de caractères préexistants. De manière générale, le codage des caractères Unicode ne précise pas la position d'un caractère par rapport à la ligne de base lors de la mise en page ultérieure. Afin de fournir des valeurs de codage compatibles avec l'ISO 8859-1, les exposants des chiffres un, deux et trois en exposant se retrouvent codés dans le bloc du supplément latin-1.

Normes. Les caractères de ce bloc proviennent de jeux inscrits au registre de l'ECMA en vertu de l'ISO 2375 pour utilisation avec ISO/CEI 2022.

⁸ Origine probable du signe infini.

⁹ 𐍀 = D = 500, 𐍑 = M = 1 000
 𐍒 = 𐍄 = 5 000, 𐍓 = 𐍅 = 10 000
 𐍔 = 50 000, 𐍕 = 100 000, etc.

13.4 Flèches et opérateurs mathématiques

Opérateurs mathématiques : U+2200 – U+22FF

Le bloc des opérateurs mathématiques code des caractères représentant des opérateurs, des relations, des symboles géométriques et quelques autres symboles dont l'utilisation se limite aux contextes mathématiques.

En plus des caractères contenus dans ce bloc, on trouve également des opérateurs mathématiques dans les blocs du latin de base (ASCII) et du supplément latin-1. La notation mathématique emploie aussi quelques symboles provenant des blocs *Signes techniques divers* et *Ponctuation générale*. Certaines lettres latines écrites dans un style particulier et utilisées comme opérateurs mathématiques comme U+210B \mathcal{H} MAJUSCULE H DE RONDE, ainsi que la lettre hébraïque U+2135 \aleph SYMBOLE ALEF utilisée comme premier opérateur cardinal transfini, sont codées dans le bloc des symboles de type lettre.

Normes. Les caractères codés dans ce bloc représentent les opérateurs mathématiques de plusieurs normes nationales. Ces normes incluent des ensembles spéciaux comme l'ANSI Y10.20, l'ISO 6862, l'ISO 8879 et des parties du jeu de l'*American Mathematical Society* ainsi que le répertoire de T_EX.

Principes de codage. Les opérateurs mathématiques sont souvent polysémiques. C'est pourquoi on code à dessein les caractères de ce bloc en fonction de leur forme. En conséquence, il se peut, dans certains cas, qu'une valeur Unicode représente plusieurs valeurs sémantiques. Ainsi, U+2218 \circ OPÉRATEUR ROND peut-il être équivalent à *petit cercle blanc*, *fonction composée* ou *beigne apl.* Unicode ne tente pas de distinguer toutes les significations possibles des opérateurs mathématiques et des symboles de relation.

Il arrive parfois que des opérateurs mathématiques et tout particulièrement des symboles de relation apparaissent dans plusieurs normes, manuels et polices de caractères sous un grand nombre de variantes purement graphiques. Quand on a pu déterminer qu'il s'agissait à la source de variantes, ces variantes n'ont pas été codées séparément.

Unifications. On a unifié les opérateurs mathématiques comme *implique* \Rightarrow et *si et seulement si* \Leftrightarrow avec leurs flèches correspondantes du bloc *Flèches* (U+21D2 \Rightarrow DOUBLE FLÈCHE VERS LA DROITE et U+2194 \leftrightarrow FLÈCHE GAUCHE ET DROITE).

L'opérateur U+2208 \in APPARTIENT À peut à l'occasion être rendu sous une forme plus haute que celle illustrée dans les tableaux de codes. Les ouvrages de référence mathématiques et les normes consultées traitent ces caractères comme des variantes du même glyphe. U+220A \in PETIT APPARTIENT À correspond à une version plus petite de *appartient à* provenant des polices de caractères mathématiques Pi.

L'opérateur U+226B \gg BEAUCOUP PLUS GRAND QUE et U+226A \ll BEAUCOUP PLUS PETIT QUE se rendent parfois par le biais d'une forme non imbriquée. Puisqu'il n'existe pas de différence de sens, Unicode ne fournit qu'un seul code pour chaque opérateur.

Une importante partie des unifications concerne les variantes de symboles de relation (égalité, similarité ou négation). Unicode n'assigne pas de codes séparés aux variantes qui comprennent les signes d'égalité à une ou deux barres, les signes de similarité à un ou deux tildes et les barres obliques de négation de différentes longueurs. Ainsi, U+2288 $\not\subseteq$ NI UN SOUS-ENSEMBLE NI ÉGAL À, par exemple, correspond-il à l'archétype d'au moins six variantes de glyphe avérées dans différentes sources.

À deux reprises, on a codé des variantes essentiellement stylistiques: U+2265 \geq PLUS GRAND OU ÉGAL À à distinguer de U+2267 \geq PLUS GRAND QUE PAR-DESSUS ÉGAL À et, de la même manière, U+2264 \leq PLUS PETIT OU ÉGAL À à distinguer de U+2266 \leq PLUS PETIT QUE PAR-DESSUS ÉGAL À. Cette exception se justifie pour des raisons de correspondance bijective avec certaines normes asiatiques qui distinguent ces deux formes.

Symboles dérivés du grec. À des fins de compatibilité avec des normes préexistantes, plusieurs opérateurs mathématiques dérivés du grec se codent séparément. Ces opérateurs apparaissent à l'occasion avec des variables écrites à l'aide de lettres grecques. Parmi ces opérateurs : U+2206 Δ INCRÉMENT, U+220F \prod PRODUIT DE LA FAMILLE et U+2211 \sum SOMMATION DE LA FAMILLE.

Parmi les autres caractères grecs dédoublés, on trouve ceux qui codent le U+00B5 μ SYMBOLE MICRO dans le bloc *Supplément latin-1*, le U+2126 Ω SYMBOLE OHM du bloc *Symboles de type lettre* ainsi que plusieurs symboles de fonction APL¹⁰ du bloc *Signes techniques divers*. On retrouve tous les autres caractères grecs possédant un sens mathématique particulier dans le bloc du grec car leur dédoublement ne s'imposait pas pour des raisons de compatibilité¹¹.

Opérateurs n-aires. Les opérateurs *n*-aires se différencient des opérateurs binaires par leur plus grande taille et par le fait que, en composition mathématique, ils s'accompagnent d'expressions de limites.

Divers symboles. U+2212 – SIGNE MOINS est un opérateur mathématique à distinguer de U+002D - TRAIT D'UNION-SIGNE MOINS hérité de l'ASCII qui peut paraître identique au signe moins ou être un peu plus étroit. Pour une liste complète des tirets et traits Unicode, voir *Tableau 7-2*. L'intervalle U+22EE..U+22F1 correspond à un ensemble de points de suspension utilisés en notation matricielle.

Propriété mathématique. La *Section 4.8, Propriété mathématique*, fournit une liste de caractères ayant la propriété mathématique.

Flèches : U+2190 – U+21FF

Les flèches ont diverses fonctions : indiquer une relation de directionalité, illustrer une dérivation ou une implication logique ou représenter les touches de commande du curseur.

Unicode fournit un jeu assez complet de flèches génériques, surtout quand elles possèdent un usage bien établi et une signification bien définie. Unicode ne tente pas, par contre, de coder isolément toutes les variantes stylistiques possibles, particulièrement si leur usage est au premier chef décoratif. Pour la plupart des variantes de flèches, Unicode assigne des valeurs de caractère aux deux directions horizontales et souvent selon les quatre points cardinaux. Pour les flèches simples et doubles, il code huit directions.

Normes. Unicode regroupe les flèches de plusieurs normes nationales et de divers jeux de caractères privés.

Unifications. Les flèches exprimant des relations mathématiques sont codées dans le bloc des *Flèches*, par exemple U+21D2 \Rightarrow DOUBLE FLÈCHE VERS LA DROITE dont la signification peut être *implique*.

¹⁰ APL : *A Programming Language*, langage de programmation des années 1970 (quelque peu désuet) utilisant un très grand nombre de symboles.

¹¹ L'addition de symboles alphanumériques mathématiques par Unicode 3.1 a quelque peu changé la situation.

Unicode ne représente pas séparément les formes longues et courtes des flèches des répertoires de glyphes ou des systèmes composition comme T_EX.

Principes de codage. Les flèches possédant une telle diversité de fonctions, une même valeur de caractère Unicode peut avoir plusieurs significations. U+21B5 ↙ FLÈCHE VERS LE BAS AVEC COIN VERS LA GAUCHE, par exemple, peut être équivalent à *retour de chariot* ; de même que U+2191 ↑ FLÈCHE VERS LE HAUT peut également signifier *augmente* ou *exposant*.

13.5 Symboles techniques

Pictogrammes de commande : U+2400 – U+243F

Le besoin d'afficher sans équivoque la présence des codes de commandes C0 et de l'ESPACE a donné lieu à l'élaboration de conventions pour représenter ces caractères sans œil (non graphiques).

Par définition, les codes de commande ne se manifestent que par leur action. Toutefois, il faut parfois en connaître la position au sein d'une suite de données. C'est pourquoi on a prévu des pictogrammes qui permettent d'illustrer les codes de contrôle C0 ASCII.

Par définition, l'ESPACE est un caractère sans œil. On a également prévu des pictogrammes permettant de représenter l'espace.

Normes. La norme CNS 11643 définit le codage des pictogrammes de codes de commande. Il existe des représentations normalisées de ces caractères de commande (voir, par exemple, ANSI X3.32 et ISO 2047). Unicode, toutefois, ne code que la sémantique des pictogrammes des codes de commande U+2400..U+241F. Ce choix permet à un logiciel donné d'utiliser la représentation graphique de son choix.

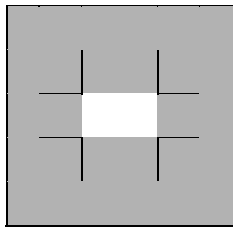
Glyphes représentant l'espace. Deux glyphes peuvent servir à représenter le caractère espace ASCII U+0020 : U+2420 [♠] SYMBOLE POUR ESPACE et U+2422 [♠] SYMBOLE VISUEL POUR L'ESPACE.

Caractères représentant des pictogrammes de commande. On n'associe pas les autres caractères de ce bloc à des glyphes précis ; ils servent plutôt à coder une quelconque représentation picturale du code de commande en question. On suppose que les pictogrammes utilisés pour représenter les codes de contrôle sont souvent propres aux différents systèmes et que ces pictogrammes ne font pas souvent l'objet d'échanges entre ces différents systèmes.

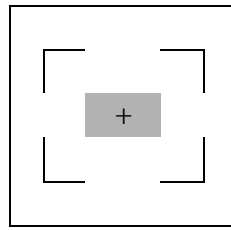
Signes techniques divers : U+2300 – U+23FF

Ce bloc code les signes techniques, y compris les pictogrammes pour les touches de clavier comme U+232B [⌘] EFFACEMENT À GAUCHE. Ne font pas partie de ce bloc les signes qui normalement ne sont pas utilisés dans des textes unidimensionnels, mais plutôt conçus pour des diagrammes bidimensionnels comme les symboles de circuits électroniques. Ce bloc comporte une zone libre exceptionnellement grande, car il est probable qu'un grand nombre de signes techniques viennent éventuellement s'ajouter au standard Unicode.

Repères de coupe. Les hirondelles et les coins dits parfois de Quine s'utilisent dans la mise en page bidimensionnelle : on peut néanmoins y faire référence dans un texte brut. Les ailes des hirondelles restent extérieures à la surface à conserver alors que les coins de Quine (ou parenthèses de Gödel) sont des repères de coupe dont les branches pointent vers la surface à garder tout en restant à l'extérieur de celle-ci. Le diagramme suivant indique comment s'utilisent les hirondelles et les coins de Quine.



Utilisation des hirondelles



Utilisation des coins de Quine

Symboles des fonctions APL. Le langage de programmation APL utilise nombre de symboles de fonction composés à partir de symboles de fonction plus simples. Pour ce faire, les premières mises en œuvre informatiques avaient souvent recours à des mécanismes d'effacement arrière et de surfrappe. En principe, en APL, une composition fonctionnelle est productive ; en pratique, seuls quelques symboles de fonction composés sont devenus des opérateurs APL standards. Ce jeu de symboles relativement restreint est entièrement codé dans ce bloc. On peut coder toutes les autres extensions APL par composition avec d'autres caractères Unicode. C'est ainsi que l'on peut représenter le symbole APL *a souligné* à l'aide de U+0061 **a** LETTRE MINUSCULE LATINE A + 0332 ◌̣ DIACRITIQUE TRAIT SOUSCRIT.

Reconnaissance optique de caractères : U+2440 – U+245F

Ce bloc comprend les caractères issus du jeu de caractères OCR-A qui ne correspondent pas aux caractères ASCII et les symboles de reconnaissance de caractères magnétiques (MICR) utilisés dans le traitement des chèques.

Normes. Ces deux ensembles de symboles sont définis dans la norme ISO 2033.

13.6 Symboles géométriques

Filets : U+2500 – U+257F

Unicode n'a retenu les caractères de ce bloc que dans le seul but de faciliter la prise en charge d'anciennes applications, comme l'émulation de terminal.

Normes. Ce bloc s'inspire des normes industrielles, GB 2312 et KS C 5601.

Pavés : U+2580 – U+259F

Le bloc des pavés constitue une zone de compatibilité graphique dans le standard Unicode. Un certain nombre de normes nationales et de standards privés existants, y compris la page de codes IBM PC-437, contiennent des caractères qui permettent de dessiner simplement des cellules graphiques en remplissant une partie de la cellule ou en remplissant chaque cellule en fonction d'un coefficient d'ombrage particulier. Unicode déconseille ces modèles graphiques à base de caractères, mais il inclut un jeu minimal de ces caractères pour assurer une compatibilité rétroactive avec des normes préexistantes.

Unicode prévoit, pour chaque moitié de cellule d'affichage, des caractères de remplissage ainsi qu'une série de masques horizontaux et verticaux multiples de huitième de cellule. Ces masques fractionnels ne forment pas un ensemble complet et cohérent ; il n'existe que dans le but d'assurer une compatibilité rétroactive.

Formes géométriques : U+25A0 – U+25FF

Ce bloc constitue un ensemble de caractères destinés à représenter plusieurs formes géométriques courantes – plus particulièrement des carrés, des triangles et des cercles. La constitution de ce bloc est quelque peu arbitraire ; il s'agit d'un recueil de formes issues de diverses normes de caractères et de glyphes. Les distinctions les plus systématiques comprennent : noir opposés à blanc, grand à petit, formes de base (carré, triangle ou cercle), l'orientation et enfin haut par rapport à bas ou gauche par rapport à droite.

Les carrés hachurés et quadrillés codés dans l'intervalle U+25A4..U+25A9 proviennent de la norme nationale coréenne (KS C 5601) dans laquelle ils devaient représenter des motifs de remplissage. Cette norme ne précisant pas la signification de ces caractères, Unicode ne retient ces glyphes qu'en tant que formes géométriques afin de permettre une correspondance avec la norme coréenne.

U+25CA ◊ LOSANGE correspond à un symbole typographique provenant de PostScript et du jeu de caractères Macintosh. Il diffère du U+25C7 ◊ LOSANGE BLANC générique, du U+2662 ◊ CARREAU BLANC ainsi que d'un autre caractère parfois aussi appelé losange : U+2311 ▣ PASTILLE CARRÉE.

Les carrés et triangles codés à U+25E7..U+25EE proviennent de la série de polices de caractères Linotype. U+25EF ○ GRAND CERCLE est inclus pour des raisons de compatibilité avec la norme japonaise JIS X 0208-1990.

Normes. Le bloc des formes géométriques provient d'un large éventail de normes nationales et de répertoires de caractères privés.

13.7 Divers symboles et casseau

Divers symboles : U+2600 – U+26FF

Le bloc des divers symboles se compose d'ensemble hétéroclite de symboles qui n'ont pas leur place dans les autres blocs de caractères Unicode et dont la nature est plutôt pictographique. Ces symboles servent habituellement à agrémenter les textes, dans certains cas comme la composition de livres d'échecs, de manuels de jeux de cartes ou d'horoscopes, ils correspondent à des caractères de texte normaux.

Contrairement aux caractères du casseau (« dingbats », ci-dessous) où les caractères correspondent à un glyphe précis, ces symboles peuvent revêtir différentes apparences. Ainsi, les signes U+2641 ♂ TERRE et U+2645 ♃ URANUS se représentent-ils couramment sous les formes suivantes : ♂ ou ⊕ pour la TERRE et ♃ ou ♂ pour URANUS.

L'ordre des divers symboles de ce bloc est complètement arbitraire. Dans la mesure du possible, les symboles similaires sont regroupés en sous-ensembles cohérents. Certains de ces sous-ensembles sont constitués de symboles météorologiques ou astronomiques, de mains pointées, de pièces d'échecs, de symboles religieux et politiques, de trigrammes du Livre des mutations (*Yi-King*), de symboles zodiacaux et planétaires, de cartes à jouer et de symboles musicaux. Pour les autres phases de la Lune, voir les formes circulaires du bloc des symboles géométriques.

Ce bloc ne comprend ni logotypes d'entreprise ni images d'animaux, de véhicules, d'aliment, etc. car leur utilisation est très particulière (logos et symboles de partis politiques) ou leur apparence et signification ne sont pas conventionnelles (images de vache, de chats, de bouteilles de champagne pétillant). Pour ces raisons, ces caractères ne sont pas aptes à être codés dans Unicode. Il est préférable d'incorporer ces éléments dans un texte par le biais de protocoles de niveau supérieur qui permettent le mélange de pictogrammes et de texte, plutôt que par l'adjonction indéfinie des nouveaux symboles dans ce bloc. Remarquons, cependant, que ce bloc réserve une importante zone libre afin de permettre à l'avenir, dans les cas appropriés, l'ajout de jeu de symboles conventionnels.

Normes. Les caractères de ce bloc proviennent d'un large éventail de normes nationales et de jeux de caractères privés.

Casseau : U+2700 – U+27BF

En imprimerie, un casseau est une petite boîte placée sur la casse principale et divisée en compartiments contenant les caractères « spéciaux » pour un travail donné. Il pouvait s'agir de symboles mathématiques, de croix mortuaires ou de décorations pour des travaux de ville. Le contenu de chaque casseau dépendait énormément de l'usage qu'en faisait l'imprimeur : il les choisissait selon ses goûts parmi les caractères spéciaux des spécimens des diverses fonderies. Dans le cas de la norme ISO-10646, il s'agit d'un ensemble de caractères de fantaisie. Il correspond *grosso modo* aux « dingbats », un jeu bien établi de symboles qui comprend la police de caractères « Zapf dingbat » – présente dans la plupart des imprimantes laser. D'autres casseaux existent, mais Unicode ne les code pas, car ils ne sont pas implantés systématiquement en tant que polices de caractères dans le matériel et les logiciels existants. Certains caractères Zapf Dingbats, membres d'autres sources d'Unicode, ont été unifiés avec les caractères de ces sources et se retrouvent codés dans les blocs *Formes géométriques*, *Symboles cerclés* et *Divers symboles*. Cette unification a créé des « trous » dans le répertoire des caractères Zapf Dingbat restants codés dans le bloc

du casseau. Les caractères du casseau sont triés dans leur ordre de leur codage en PostScript.

Unicode traite les caractères du casseau différemment de tous les autres caractères. Il code ces symboles comme des formes particulières de glyphes (ainsi, par exemple, y a-t-il une quarantaine d'étoiles ou astérisques) et non comme des archiglyphes de caractères abstraits qui peuvent être représentés à l'aide de différentes polices et graisses. Ainsi, il serait erroné de remplacer arbitrairement U+279D ➔ FLÈCHE À POINTE EN TRIANGLE VERS LA DROITE par une autre flèche vers la droite du casseau ou par toute autre une flèche générique du bloc des flèches (U+2190..U+21FF). En d'autres termes, les caractères Zapf Dingbat correspondent à un ensemble de glyphes provenant d'une police de caractère spécifique et leur valeur sémantique est leur forme.

Unifications. Un certain nombre de caractères du casseau ont la même forme que les symboles Unicode normaux. Plutôt que de coder à la fois la forme du glyphe Zapf Dingbat et un caractère dont le glyphe ne s'en distingue pas, Unicode les a unifiés. Parmi les caractères concernés, on trouve :

1. les symboles de jeu de cartes, U+2605 ★ ÉTOILE NOIRE, U+260E ☎ TÉLÉPHONE NOIR et U+261B 🖱 INDEX NOIR POINTANT VERS LA DROITE (voir *Divers symboles*) ;
2. U+25CF ● DISQUE NOIR et U+25A0 ■ CARRÉ NOIR (voir *Formes géométriques*) ;
3. les chiffres de 1 à 10 blancs cerclés (voir *Alphanumériques cerclés*) ;
4. et plusieurs flèches génériques (voir *Flèches*).

Les positions de ces caractères unifiés ne sont pas affectées dans le bloc du casseau ; le tableau des caractères (Chapitre 15) renvoie à la position du caractère dans les autres blocs. Les mises en œuvre peuvent utiliser d'autres glyphes, pour représenter ces caractères (comme pour tout caractère Unicode normal), y compris, bien sûr, leur forme exacte dans la police de caractères Zapf Dingbat.

Illustrons cette distinction : une mise en œuvre code un chiffre 1 cerclé à l'aide de U+2460 ① CHIFFRE UN CERCLÉ, elle peut rendre ce chiffre cerclé à l'aide de n'importe quelle police de caractères – avec ou sans empattements, romain ou italique ou au moyen d'un cercle de différentes graisses. Par contre, une application qui code un chiffre un cerclé à l'aide du caractère U+2780 ① VIGNETTE CHIFFRE UN SANS EMPATTEMENT CERCLÉ doit afficher ce caractère à l'aide d'un glyphe sans empattements provenant de la police de caractères Zapf Dingbat.

13.8 Symboles cerclés ou entre parenthèses

Alphanumériques cerclés : U+2460 – U+24FF

Les lettres latines et les chiffres cerclés de ce bloc proviennent de plusieurs sources, essentiellement de normes extrême-orientales ; Unicode les fournit dans le seul but d'assurer une compatibilité avec ces normes.

Normes. Les lettres et chiffres cerclés se retrouvent dans la norme nationale coréenne, KS C 5601, dans la norme nationale chinoise GB 2312, ainsi que dans divers standards industriels extrême-orientaux.

Le jeu de caractères Zapf Dingbat, tel qu'il s'utilise couramment dans l'industrie, contient qua

tre jeux de chiffres cerclés (y compris le zéro cerclé). On code en U+2460..U+2468 et en U+24EA le jeu noir sur blanc des chiffres à empattements. Les trois autres jeux se codent dans l'intervalle U+2776..U+2793 du casseau.

Décomposition. Les lettres ou chiffres entre parenthèses peuvent se décomposer en une suite composée d'une parenthèse ouvrante, d'un ou plusieurs caractères alphanumériques et d'une parenthèse fermante. Les nombres suivis d'un point peuvent se décomposer en une suite de chiffres suivis d'un point. Les lettres et nombres composés d'un seul chiffre cerclé peuvent se décomposer en une lettre ou un chiffre suivi de U+20DD ◉ DIACRI-TIQUE CERCLE ENGLOBANT. Il n'est pas possible de décomposer les chiffres cerclés de 10 à 20 en un texte Unicode brut (voir *Chapitre 2, Structure générale* et *Chapitre 3, Conformité*).

Lettres et mois CJC cerclés : U+3200 – U+32FF

Normes. Ce bloc permet la mise en correspondance d'Unicode et de tous les éléments hangûl cerclés provenant de la norme coréenne KS C 5601, ainsi que des caractères idéo-phonographiques¹² entre parenthèses provenant des normes JIS X 0208-1990, CNS 11643 et de plusieurs répertoires de caractères privés.

Compatibilité CJC : U+3300 – U+33FF

Les mots katakana disposés en carré correspondent à des mots écrits en katakana qui, composés avec des idéophonogrammes CJC, remplissent une cellule de rendu (de la taille du cadratin). De même, mélangés à des idéophonogrammes CJC, les symboles d'abréviation latine disposés en carré sont conçus pour n'occuper qu'une seule cellule CJC.

Seul un souci de compatibilité avec les normes de codage de caractères existantes justifie la présence de ces caractères. Les logiciels modernes peuvent fournir à la volée un répertoire infini de mots épelés en katakana ou d'abréviations disposées en carré.

Normes. Les caractères compatibilité CJC proviennent des normes nationales KS C5601, CNS 11643 et des divers répertoires de caractères privés.

Noms des ères impériales japonaises. Les ères impériales japonaises correspondent aux dates inscrites dans le *Tableau 13-2*.

¹² On parle parfois de sinogrammes ou d'idéogrammes.

Tableau 13-2. Noms des ères impériales japonaises

U+337B	平成	ÈRE HEISEI DISPOSÉ EN CARRÉ	07/01/1989 à nos jours
U+337C	昭和	ÈRE CHÔWA DISPOSÉ EN CARRÉ	24/12/1926 au 06/01/1989
U+337D	大正	ÈRE TAÏCHÔ DISPOSÉ EN CARRÉ	29/07/1912 au 23/12/1926
U+337E	明治	ÈRE MEIJI DISPOSÉ EN CARRÉ	1867 au 28/07/1912

13.9 Combinaisons Braille

Braille : U+2800 – U+28FF

Les aveugles du monde entier utilisent le système d'écriture Braille. Il est fondé sur un système de six ou huit points saillants placés en deux colonnes verticales de trois ou quatre points chacune. Les systèmes à huit points ajoutent deux points supplémentaires au-dessus ou en dessous de la matrice principale du système à six points. Six points Braille permettent 64 combinaisons possibles, alors qu'avec huit points, on peut former 256 combinaisons. Il n'existe pas de correspondance systématique entre une combinaison de points et un caractère ou un symbole d'une écriture donnée. Le contexte et les groupes d'utilisateurs concernés déterminent l'attribution sémantique des combinaisons de points. Une combinaison peut, à elle seule, représenter une abréviation ou un mot court fréquent. Dans un certain nombre de contextes et de communautés d'utilisateurs, une série de normes internationales, à commencer par l'ISO 11548-1, fournissent des tableaux de correspondance standardisés ainsi que des suites de lancement signalant un changement de contexte.

Unicode propose un seul jeu complet de 256 combinaisons de huit points. Ce jeu comprend les 64 combinaisons de points nécessaires aux systèmes Braille à six points.

Les noms de caractère des combinaisons Braille décrivent la position des points, numérotés de 1 à 8. Exemple :

1	● ●	4
2	● ●	5
3	● ●	6
7	● ●	8

Les points 1 à 6 constituent la matrice du système à six points. Les points 7 et 8 s'y ajoutent pour former le système à 8 points. Le nom de caractère d'une combinaison Braille prend la forme suivante : COMBINAISON BRAILLE POINTS-1234567, où seuls apparaissent les chiffres correspondant aux points qui saillent. La combinaison vide se nomme U+2800 $\begin{smallmatrix} \text{⠠} \\ \text{⠠} \\ \text{⠠} \end{smallmatrix}$ COMBINAISON BRAILLE BLANC.

Unicode place les 256 combinaisons Braille dans le même ordre que l'ISO 11548-1, selon la valeur octale formée par la combinaison de points. On calcule cette valeur octale de la façon suivante :

1	● ●	10
2	● ●	20
4	● ●	40
100	● ●	200

On additionne la valeur de tous les points saillants (noirs) de la combinaison. On associe ainsi aux points de la COMBINAISON BRAILLE POINTS-1247 les valeurs octales de 1_8 , 2_8 , 10_8 et 100_8 . La valeur octale de la combinaison équivaut à la somme de ces valeurs, soit 113_8 .

Unicode n'attribue pas de signification aux combinaisons Braille.

Exemple. Selon l'ISO 11548-2, on peut représenter, en Braille à huit points, le caractère LETTRE MAJUSCULE LATINE F par les points 1, 2, 4 et 7 (COMBINAISON BRAILLE POINTS-

1247). Un cercle noirci correspond à un point saillant (« présent »), un cercle évidé indique qu'un point de la matrice ne saille pas (« absent ») :

1	● ●	4
2	● ○	5
3	○ ○	6
7	● ○	8

Modèle d'utilisation. Le standard Unicode permet l'utilisation des deux styles Braille à huit points : les points sept à huit au-dessus ou en dessous. La distinction entre les deux systèmes est purement conventionnelle car, dans un même contexte, on ne mêle jamais ces deux systèmes. On code les 256 combinaisons Braille dans le standard Unicode afin de permettre la mise en œuvre de périphériques d'entrée et de sortie pouvant échanger des données Braille sans avoir recours à des conversions, qui dépendent du contexte, de la valeur sémantique vers la combinaison ou vice versa. On peut, de cette manière, échanger et afficher des documents dans leur forme finale correcte. Par ailleurs, le traitement automatique des données texte nécessitant une prise en charge sémantique devrait se faire à l'aide de caractères normaux Unicode.

Représentation physique. Contrairement aux cercles blancs, les points noirs de la représentation Unicode correspondent aux points tangibles d'un document Braille, alors que les cercles éclairés ne sont pas tangibles. Unicode ne précise pas la dimension physique des caractères Braille.

En l'absence d'un protocole de niveau supérieur, on rend les combinaisons Braille de la gauche vers la droite. Pour le rendu (tangible) de documents, on ne mêle pas, à l'exception des codes de commande, les combinaisons Braille avec d'autres caractères Unicode.

13.10 Symboles mathématiques alphanumériques

Symboles mathématiques alphanumériques : U+1D400 – U+1D7FF

Le bloc des symboles mathématiques alphanumériques comprend un grand nombre de symboles de type lettre utilisés en mathématiques, plus particulièrement pour les variables. Les caractères de ce bloc ne sont destinés qu'à la notation mathématique ou technique ; on ne devrait pas les retrouver dans des textes non techniques. Dans les langages de balisage, comme *MathML (Mathematical Markup Language)*, ces caractères devraient s'utiliser directement, plutôt qu'indirectement par le biais d'entités de référence ou de compositions à partir des lettres de base et de balises de style.

Mots utilisés comme variables. Dans certains domaines, les variables se composent non seulement de lettres, mais aussi de mots. Dans ces cas, il vaut mieux recourir à des balises de style puisque, en mathématiques, la juxtaposition de variables implique généralement la multiplication et non la formation d'un mot, comme dans un texte ordinaire. Baliser permet non seulement de résoudre ce problème, mais rend également possible l'utilisation d'un alphabet plus étendu.

Alphabets mathématiques

Jeu de base des caractères alphanumériques. La notation mathématique emploie un jeu de caractères alphanumériques de base qui comporte :

- l'ensemble des chiffres latins de base (0-9) (U+0031..U+0039) ;
- l'ensemble des lettres latines minuscules et majuscules de base (a-z, A-Z) ;
- les lettres majuscules grecques A-Ω (U+0391..U+0309), ainsi que U+2207 ∇ NABLA et la variante du thêta U+03F4 SYMBOLE GREC THÊTA MAJUSCULE ;
- les lettres minuscules grecques α-ω (U+03B1..U+03C9), ainsi que U+2202 ∂ DÉRIVÉE PARTIELLE et les six variantes glyphiques de ε, θ, κ, φ, ρ et π : U+03F5, U+03D1 ϑ, U+03F0 χ, U+03D5 ϕ, U+03F1 ρ et U+03D6 ω.

Seules les formes non accentuées des lettres s'utilisent en mathématiques parce que les accents généraux, tels que l'accent aigu, font obstacle aux diacritiques mathématiques habituels. Ces diacritiques mathématiques comprennent, entre autres, l'accent circonflexe, le macron et le point (simple ou double) suscrit, ce dernier dénote en physique les dérivés par rapport à une variable de temps. Les symboles mathématiques comportant des diacritiques se représentent toujours par de suites de caractères combinatoires.

Certains caractères du jeu de base des caractères grecs sont codés sous deux variantes. En effet, ces caractères peuvent avoir, en mathématiques, différentes significations alors qu'en grec ils représentent toujours la même lettre

Caractères supplémentaires. En plus du jeu de base, la notation mathématique utilise également quatre caractères dérivés de l'hébreu (U+2135..U+2138). On utilise à l'occasion, d'autres caractères alphabétiques ou numériques, comme U+0428 Ш LETTRE MAJUSCULE CYRILLIQUE CHA, U+306E ノ SYLLABE HIRAGANA NO ou les chiffres arabo-hindî (U+06F0..U+06F9). Toutefois, ces caractères ne s'utilisent que sous leur forme de base.

Distinctions sémantiques. La notation mathématique fait appel à un certain nombre d'alphabets latins et grecs qui peuvent passer au premier coup d'œil pour de simples variantes de polices de caractères. Ainsi, la lettre H peut-elle apparaître en romain (H), en gras (H), en italique (H) ou de ronde (\mathcal{H}). Dans un document donné, chacun de ces caractères possède une signification distincte, en règle générale sans rapport avec celle des autres H. Le H romain représente donc une variable différente du H gras. Si ces attributs devaient disparaître dans un texte normal, le sens du texte s'en trouverait altéré. Sans ces distinctions, une formule pour le hamiltonien comme¹³

$$\mathcal{H} = \int d\tau (\varepsilon E^2 + \mu H^2)$$

se transforme en cette équation intégrale pour l'inconnue H

$$H = \int d\tau (\varepsilon E^2 + \mu H^2).$$

Coder ces deux signes permet de conserver de telles distinctions dans du texte brut (c'est-à-dire sans balisage).

Alphabets mathématiques. Le tableau ci-dessous présente les symboles alphanumériques mathématiques codés dans Unicode :

Style mathématique	Caractères du jeu de base	Emplacement
Romain (droit, avec empattements)	latin, grec et chiffres	PMB
Gras	latin, grec et chiffres	1 ^{er} plan
Italique	latin et grec	1 ^{er} plan [†]
Italique gras	latin et grec	1 ^{er} plan
De ronde	latin	1 ^{er} plan [†]
De ronde gras	latin	1 ^{er} plan
Gothique	latin	1 ^{er} plan [†]
Gothique gras	latin	1 ^{er} plan
Ajouré	latin et chiffres	1 ^{er} plan [†]
Sans empattement	latin et chiffres	1 ^{er} plan
Sans empattement gras	latin, grec et chiffres	1 ^{er} plan
Sans empattement italique	latin	1 ^{er} plan
Sans empattement gras italique	latin et grec	1 ^{er} plan
À chasse fixe	latin et chiffres	1 ^{er} plan

[†] Certains de ces caractères se trouvent dans le PMB, voir le texte ci-dessous.

Les lettres simples sont unifiées avec les caractères présents dans les blocs du latin de base et du grec. On retrouve 25 caractères ajourés, italiques, gothiques et de ronde dans le bloc des *Symboles de type lettre* (U+2100..U+214F). On les unifie explicitement avec les caractères de ce bloc ; les positions correspondantes ne sont pas affectées dans les alphabets mathématiques.

¹³ Il s'agit d'une intégrale de volume et un ν souscrit faciliterait sans doute sa lecture. Les opérateurs en exposant ou indice ne font cependant pas partie d'Unicode à ce stade. Il est impossible de représenter cette intégrale avec cet indice à l'aide d'un texte Unicode brut ; il faut avoir recours à un protocole de niveau supérieur.

Les alphabets de ce bloc ne codent que les distinctions sémantiques sans spécifier les polices de caractères à utiliser pour les glyphes romains, de ronde, gothiques, ajourés, sans empattements ou à chasse fixe. Les styles de ronde et ajouré, notamment, varient considérablement d'une police à l'autre. Les symboles mathématiques alphanumériques ne s'utilisent que dans les textes mathématiques.

Décompositions de compatibilité. Tous les symboles mathématiques alphanumériques possèdent des décompositions de compatibilité vers des lettres latines ou grecques de base. L'élimination des distinctions stylistiques n'est, d'ordinaire, pas souhaitable car elle entraîne la disparition des nuances qui justifient le codage séparé de ces caractères. Voir le *Chapitre 6, Formes de normalisation*, pour plus de renseignements.

Polices utilisées pour les alphabets mathématiques

Les mathématiciens imposent de strictes exigences quant aux polices utilisées dans la représentation des variables mathématiques. Les lecteurs d'un texte mathématique doivent être en mesure de distinguer les variables, même si elles sont éloignées les unes des autres. Ils doivent repérer la lettre, qu'elle fasse partie du texte ou qu'elle soit une variable mathématique, enfin ils doivent savoir de quel alphabet mathématique provient cette lettre.

Les variables mathématiques sont généralement composées en italique, mais toutes les polices italiques ne s'y prêtent pas. Ainsi, dans les polices conçues pour des textes normaux, la lettre italique ν se différencie à peine de la lettre grecque μ . C'est pourquoi la lettre italique ν est de préférence arrondie dans les polices mathématiques. D'autres caractères, parfois de formes semblables, méritent une attention particulière afin d'éviter toute ambiguïté. Le tableau ci-dessous fournit quelques exemples :

a italique	a	α	alpha
v italique (habituel)	ν	ν	nu
v italique (recommandé)	υ	υ	upsilon
X manuscrit	χ	χ	khi
Y simple	Y	Υ	Upsilon

Lettres difficiles à distinguer. Les polices sans empattements ne permettent pas toutes de distinguer la lettre l minuscule de la lettre l majuscule et les polices à chasse fixe (mono-chasse) ne permettent pas toutes de distinguer la lettre l du chiffre 1 . Ces polices sont inutilisables en mathématiques. En gothique, on doit pouvoir en particulier distinguer le I du J. Les lettres gothiques surchargées ne conviennent pas. De la même manière, dans tous les jeux alphanumériques mathématiques, le chiffre *zéro* doit être différent de la lettre O capitale. Certains caractères se ressemblent à un tel point que même les polices mathématiques ne fournissent pas de glyphes différents pour ceux-ci. Typiquement la lettre majuscule A (A) et la lettre majuscule grecque (A). La notation mathématique évite de les utiliser, sauf si, dans un contexte particulier, il est impossible de les confondre.

Répercussions des diacritiques sur les polices. Il est fréquent dans les équations mathématiques de retrouver des caractères auxquels sont adjoints des diacritiques (points, tilde, accent circonflexe, flèches suscrites, etc.) ainsi que des caractères suivis ou précédés de lettres ou de nombres suscrites ou souscrites. Le dessin des lettres italiques est, en conséquence, moins incliné et les caractères de ronde saillent et penchent moins que leurs homologues utilisés dans la composition de certains faire-part de mariage.

Type des caractères de ronde. Dans certains cas, Unicode opta délibérément pour une unification avec un symbole non mathématique : U+2133 \mathcal{M} MAJUSCULE M DE RONDE est unifié au symbole de la monnaie allemande datant d'avant 1949, le Reichsmark ; de même U+2113 ℓ MINUSCULE L DE RONDE est unifié avec le symbole hors SI¹⁴ du litre. Cette unification restreint la gamme de glyphes valables pour ce caractère dans les tableaux. Conformément aux recommandations de l'*American Mathematical Society*, on représente les glyphes de référence dans les tableaux de codes par une anglaise simplifiée. Par souci d'uniformité, les autres caractères de ronde du bloc des symboles de type lettre sont dorénavant représentés à l'aide de cette même police.

Caractères ajourés. Les glyphes ajourés présents dans les versions antérieures d'Unicode tentaient de s'harmoniser avec les autres caractères latins du standard, composés en Times. L'*American Mathematical Society* et certains éditeurs de textes mathématiques, préconisent l'utilisation de la police actuelle aux formes beaucoup plus simples, dérivées de l'écriture manuscrite. On trouve cependant, chez certains éditeurs, des caractères avec ou sans empattements dans des textes mathématiques.

¹⁴ SI = Système d'Unité International.

13.11 Symboles musicaux byzantins

Symboles musicaux byzantins : U+1D000 – U+1D7FF

La notation musicale byzantine fit son apparition au VII^e ou VIII^e siècle apr. J.-C. pour se développer plus complètement à partir du X^e siècle. Les symboles musicaux byzantins s'utilisent essentiellement dans la transcription de musiques et de chants liturgiques orthodoxes, bien qu'on rencontre quelques manuscrits de musiques populaires qui y aient aussi recours. En 1881, le Comité musical du patriarcat orthodoxe redéfinit certains symboles et établit la *Nouvelle notation musicale byzantine analytique* utilisée depuis lors. Si on exclut quelques manuscrits en russe, bulgare, roumain ou arabe, l'immense majorité (environ 95 %) des quelque 7 000 manuscrits utilisant cette notation sont rédigés en grec.

Traitement. La représentation informatique des symboles musicaux byzantins est très récente, même si la publication typographique de livres de musique religieuse remonte à 1820. Il existe deux types de programmes : ceux qui permettent aux musiciens de composer les partitions qu'ils utilisent et ceux qui comparent ou convertissent ce système de notation musicale en notation musicale occidentale (voir *Section 13.12, Symboles musicaux occidentaux, U+1D100..U+1D1FF*).

Les symboles musicaux byzantins se divisent en quinze classes selon leur fonction. Les caractères interagissent horizontalement et verticalement. On définit trois « bandes » horizontales, dans chacune peuvent apparaître diverses classes de notes ; des règles précises régissent l'interaction d'autres caractères au sein de ces bandes. Ces règles sont en cours d'élaboration et, à ce jour, les manipulations de textes bruts contenant des symboles musicaux byzantins, tout comme celles concernant des symboles musicaux occidentaux, dépassent le cadre du standard Unicode¹⁵.

¹⁵ En d'autres termes, Unicode permet de représenter ces symboles de manière isolée dans un texte brut mais n'en précise ni les règles de contextualisation ni la manière de les représenter, dans le cas de la notation occidentale notamment, sur une portée. Un protocole de niveau supérieur peut, bien sûr, utiliser les caractères Unicode et, par exemple, son propre balisage pour représenter de telles partitions.

13.12 Symboles musicaux occidentaux

Symboles musicaux : U+1D100 – U+1D1FF

Les symboles musicaux codés dans le standard Unicode englobent la notation musicale occidentale contemporaine (notation « ovale ») et ses précurseurs : la notation grégorienne (le plain-chant, notation « carrée ») et la notation proportionnelle (notation « losangée »)¹⁶. La notation musicale occidentale (NMO) constitue le langage codé le plus complet couramment utilisé pour la représentation des sons. La notation musicale occidentale forme un ensemble de symboles relativement cohérent et stable, mais qui, comme la musique, continue d'évoluer. Système ouvert, il a su survivre au passage du temps en partie grâce à sa souplesse et à sa ductilité. Les symboles musicaux d'Unicode proviennent surtout de la NMO. S'y ajoutent des compléments courants au répertoire de la NMO : les altérations de quart de ton, les têtes de note « bloc » et les têtes de note « façonnées » à l'américaine (les *shape-notes*)¹⁷.

Le bloc des symboles musicaux n'inclut pas de pictogrammes que l'on retrouve parfois dans les partitions ; il s'agit d'éléments créés pour un répertoire en particulier (parfois pour une seule pièce). Ce bloc ne reprend pas les caractères pouvant également avoir une signification musicale, mais qui sont membres d'autres blocs de caractères. Sont donc absents les chiffres indiquant la mesure, les basses chiffrées, les lettres identifiant les morceaux ainsi que les chiffres romains utilisés en analyse harmonique¹⁸, etc.

Les symboles musicaux s'utilisent à travers le monde de manière à peu près normalisée. On les rencontre souvent dans du texte courant où ils peuvent, sauf rares exceptions, être traités comme de simples caractères à chasse sans propriétés particulières. Ces symboles apparaissent dans divers contextes particuliers comme les ouvrages théoriques, les textes pédagogiques, les dictionnaires terminologiques, les bases de données bibliographiques, les catalogues thématiques ou les bases de données musicales. Les symboles musicaux peuvent également s'utiliser au sein de protocoles de niveau supérieur comme des langages de description musicale ou des formats de fichiers représentant des données musicales ou des partitions de musique.

Les difficultés liées à la représentation graphique des partitions, et notamment celle de la hauteur de ton, sont telles qu'Unicode ne définit pas, à dessein, de codage pour la hauteur de ton musicale. Le bloc des symboles musicaux fournit un jeu d'éléments communs utiles à l'échange et au traitement. Le codage de la hauteur de ton et la mise en page de la structure musicale résultante impliquent, non seulement la définition de la relation verticale entre les diverses notes simultanées, mais également, quand plusieurs portées sont employées, celle entre les diverses parties instrumentales. Ceci est du ressort de protocoles de niveau supérieur qui peuvent puiser parmi les éléments graphiques de ce bloc. L'absence d'indication de hauteur de ton n'est pas une lacune, mais une caractéristique essentielle du codage.

¹⁶ Encore appelée notation mensurale ou mesurée (*musica mensurata*), elle s'oppose au plain chant (*musica plana*) par l'attribution à chaque signe d'une durée relative.

¹⁷ Aide de lecture utilisée principalement dans les livres de cantiques protestants aux États-Unis. Chaque forme de tête de note y correspond à une note de la gamme (il n'est donc pas nécessaire de savoir décoder l'armure de la clé).

¹⁸ L'analyse harmonique schenkerienne, entre autres, en vogue dans le monde anglo-saxon utilise des chiffres romains comme lettres de base. D'autres systèmes d'analyse, comme celui de Riemann, populaire en Europe centrale, utilise plutôt des lettres romaines.

Le bloc des *Symboles divers* (U+2600..U+267F) comprend trois caractères, U+266D ♭ BÉMOL, U+266E ♮ BÉCARRE et U+266F # DIÈSE, qui apparaissent régulièrement dans la notation musicale. Toutefois, quatre symboles codés dans le même bloc doivent être traités comme de simples caractères du casseau (« dingbats ») ou des symboles divers, mais pas comme de véritables notes de musique :

U+2669 ♪ NOTE NOIRE,

U+266A ♪ NOTE CROCHE,

U+266B ♪ DEUX CROCHES RAMÉES,

U+266C ♪ DEUX DOUBLES CROCHES RAMÉES.

Le punctum ou la brève grégorienne, un carré¹⁹, sont unifiés à U+1D147 ■ SYMBOLE MUSICAL TÊTE DE NOTE CARRÉE NOIRE. La semi-brève de la notation grégorienne carrée, un losange, est unifiée à U+1D1BA ◆ SYMBOLE MUSICAL SEMI-BRÈVE NOIRE de la notation proportionnelle²⁰. De ce fait et afin de respecter des nuances de présentation, il se peut que les notations carrée (« grégorienne »), losangée (mesurée médiévale) et ovale (moderne) nécessitent des polices différentes.

Traitement. On peut considérer la plupart des symboles musicaux comme de simples caractères à chasse au sein d'un texte ou d'un exemple, même si, dans une mise en page musicale complète, leur comportement est plus complexe. Certains caractères ne sont conçus que pour s'adjoindre à d'autres et produire ainsi des suites de caractères qui représentent à la fois des notes et leur articulation. On peut saisir, traiter et afficher les symboles musicaux d'une manière analogue à celle utilisée pour les symboles mathématiques. Incorporés à un texte, la plupart des symboles deviennent de simples caractères à chasse sans propriétés particulières. Il existe quelque de caractères dont la fonction est de préciser le formatage, ils sont décrits ci-dessous.

Saisie. La saisie des symboles musicaux peut s'effectuer grâce aux claviers alphanumériques standards, à des claviers de piano ou à tout autre dispositif, y compris une méthode graphique. La saisie des symboles musicaux par clavier peut s'appuyer sur des techniques analogues à celles utilisées pour le chinois, le japonais et le coréen. En outre, les méthodes de saisie à base de dispositifs de pointage ou de claviers de piano peuvent s'inspirer des systèmes de mise en page musicale actuels. Ainsi, en présence d'une interface graphique, l'utilisateur pourra choisir des symboles à partir d'une palette.


Directionnalité. Les symboles musicaux n'ont pas d'incidences bidirectionnelles connues. Incorporés à des textes droite-à-gauche (sinistrogrades), par exemple en hébreu ou en arabe, les symboles musicaux continuent de s'écrire, comme d'habitude, de gauche à droite. On divise les mots en syllabes et on les place en dessous ou au-dessus des notes, comme pour les écritures latines. Toutefois, les mots (monosyllabiques) ou syllabes correspondant à chaque note s'écrivent dans la directionnalité dominante de l'écriture.







Caractères de formatage. De longues rames²¹, structures proche de la ligature, regroupent souvent les notes de courte durée. Cette pratique, répandue et systématique, se prête bien à un traitement informatique. Les caractères de formatage U+1D173 ■ SYMBOLE MUSICAL DÉBUT DE RAME et U+1D174 ■ SYMBOLE MUSICAL FIN DE RAME délimitent la rame. Il arrive parfois qu'une extrémité de la rame demeure ouverte. On indique l'absence

¹⁹ ...ou un losange. On note la confusion possible entre les différentes valeurs liées aux différentes époques. Le *punctum*, un point dans la notation neumatique, indique que le son descend, il sera transcrit par un losange ou carré en notation carrée (dite grégorienne ici).

²⁰ Unification donc purement graphique et non de valeur.

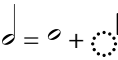
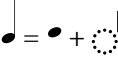
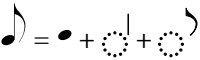
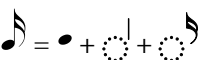
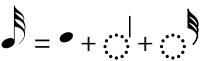
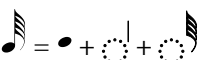
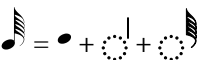
²¹ On les nomme également barres de valeurs, barres transversales ou poutres (cf. l'allemand *Balken*). Exemple de deux croches (♪) ramées : ♪.

de hampe à la fin de la rame à l'aide du caractère U+1D159  SYMBOLE MUSICAL TÊTE DE NOTE NULLE.

Unicode prévoit des caractères de formatage analogues pour d'autres structures de liaison. Les paires de caractères U+1D175  SYMBOLE MUSICAL DÉBUT DE TENUE (liaison de prolongation), U+1D176  SYMBOLE MUSICAL FIN DE TENUE, U+1D177  SYMBOLE MUSICAL DÉBUT DE COULÉ (liaison d'expression), U+1D178  SYMBOLE MUSICAL FIN DE COULÉ, U+1D179  SYMBOLE MUSICAL DÉBUT DE PHRASÉ et U+1D17A  SYMBOLE MUSICAL FIN DE PHRASÉ délimite chacune de ces liaisons. Comme les rames, ces structures se prêtent bien à un traitement algorithmique.

Dans une notation musicale complète, ces paires de caractères modifient la mise en page, le regroupement des notes et le phrasé. Lorsqu'on écrit ou rend des exemples musicaux dans un texte brut sans programme spécial, les caractères de formatage de début ou de fin peuvent se rendre à l'aide de crochets ou rester sans interprétation. Dans la mesure du possible, les logiciels plus perfectionnés qui affichent de la notation musicale au sein de textes peuvent interpréter ces paires de caractères en tant que commandes de formatage et les afficher, suivant le cas, comme des rames, des tenues, des coulés ou des phrasés.

Notes précomposées. Pour une souplesse maximale, le jeu de caractères inclut à la fois des valeurs de note précomposées et des primitives, à partir desquelles on peut composer des notes complètes. On fournit les versions précomposées principalement pour des raisons de commodité. Néanmoins, si on normalise d'une façon quelconque le texte, tous les caractères seront décomposés (voir le *Chapitre 6, Formes de normalisation*, pour plus de renseignements). La *Base de données des caractères Unicode* fournit les équivalents canoniques de ces caractères. Ces équivalents sont également illustrés dans le tableau ci-dessous. Dans ce tableau et dans les exemples suivants, les noms des caractères Unicode symboles musicaux on a omis, pour des raisons de concision, les qualificatifs SYMBOLE MUSICAL, DIACRITIQUE MUSICAL OU SYMBOLE MUSICAL TRAIT D'ORNEMENT.

	U+1D15E BLANCHE	U+1D157 TÊTE DE NOTE ÉVIDÉE + U+1D165 HAMPE
	U+1D15F NOIRE	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE
	U+1D160 CROCHE	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D16E CROCHET-1
	U+1D161 DOUBLE CROCHE	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D16F CROCHET-2
	U+1D162 TRIPLE CROCHE	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D170 CROCHET-3
	U+1D163 QUADRUPLE CROCHE	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D171 CROCHET-4
	U+1D164 QUINTUPLE CROCHE	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D172 CROCHET-5

Têtes de note de remplacement. Des notes plus complexes composées à partir d'autres têtes de note, hampes, crochets et symboles d'articulation sont nécessaires pour l'écriture de toutes les partitions ou une transcription plus précise. On les utilise, par exemple, dans les notes façonnées à l'américaine (*shape-notes*, voir note de bas page ci-devant) et les notations modernes pour percussions. Exemple :

$$\blacksquare = \blacksquare + \circ \quad \text{U+1D147 TÊTE DE NOTE CARRÉE NOIRE + U+1D165 HAMPE}$$

$$\times = \times + \circ \quad \text{U+1D143 TÊTE DE NOTE X + U+1D165 HAMPE}$$

Points d'augmentation et symboles d'articulation. Les points d'augmentation et les symboles d'articulation peuvent s'ajouter à des notes composées ou précomposées. De surcroît, afin de construire un symbole de note complet, il est permis de répéter les points d'augmentation et les symboles d'articulation aussi souvent que nécessaire. Le tableau ci-dessous illustre divers emplois du point d'augmentation :

	U+1D160 CROCHE + U+1D16D POINT D'AUGMENTATION	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D16E CROCHET-1 + U+1D16D POINT D'AUGMENTATION
	U+1D160 CROCHE + U+1D17C DÉTACHÉ	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D16E CROCHET-1 + U+1D17C DÉTACHÉ
	U+1D160 CROCHE + U+1D16D POINT D'AUGMENTATION + U+1D16D POINT D'AUGMENTATION + U+1D17B ACCENT ²²	U+1D158 TÊTE DE NOTE NOIRE + U+1D165 HAMPE + U+1D16E CROCHET-1 + U+1D17B ACCENT + U+1D16D POINT D'AUGMENTATION + U+1D16D POINT D'AUGMENTATION


Ornements ou agréments. Ci-dessous, une liste d'ornements du Grand Siècle (plus précisément des XVII^e et XVIII^e siècles) ainsi que de suites de caractères combinatoires qui permettent de composer ces ornements.

	U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3
	U+1D19C TRAIT NUMÉRO 2 + U+1D1A0 TRAIT NUMÉRO 6 + U+1D19D TRAIT NUMÉRO 3
	U+1D1A0 TRAIT NUMÉRO 6 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3
	U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D1A0 TRAIT NUMÉRO 6 + U+1D19D TRAIT NUMÉRO 3
	U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D1A3 TRAIT NUMÉRO 9

²² Remarquons que la place de U+1717B ACCENT diffère d'une décomposition à l'autre. La seconde colonne représente la décomposition complète qui correspond à l'illustration à droite du signe égal. La décomposition complète requiert également une remise en ordre canonique (forme de normalisation D). Or, étant données les classes combinatoires de U+1D16D (226) et de U+1D17B (220), l'ordre canonique place les points d'augmentation à la suite de l'accent.

En effet, les entrées pour ces deux caractères dans la *Base de données Unicode* sont les suivantes :

1D16D ; DIACRITIQUE MUSICAL POINT D'AUGMENTATION;MC ;226;L;;;;N;;;;;
1D17B ; DIACRITIQUE MUSICAL ACCENT;MN ;220;NSM;;;;N;;;;;

	U+1D1A1 TRAIT NUMÉRO 7 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3
	U+1D1A2 TRAIT NUMÉRO 8 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3
	U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3 + U+1D19F TRAIT NUMÉRO 5
	U+1D1A1 TRAIT NUMÉRO 7 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D1A0 TRAIT NUMÉRO 6 + U+1D19D TRAIT NUMÉRO 3
	U+1D1A1 TRAIT NUMÉRO 7 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3 + U+1D19F TRAIT NUMÉRO 5
	U+1D1A2 TRAIT NUMÉRO 8 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D1A0 TRAIT NUMÉRO 6 + U+1D19D TRAIT NUMÉRO 3
	U+1D19B TRAIT NUMÉRO 1 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3
	U+1D19B TRAIT NUMÉRO 1 + U+1D19C TRAIT NUMÉRO 2 + U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3 + U+1D19E TRAIT NUMÉRO 4
	U+1D19C TRAIT NUMÉRO 2 + U+1D19D TRAIT NUMÉRO 3 + U+1D19E TRAIT NUMÉRO 4