Chapitre 8

Écritures européennes alphabétiques

Toutes les écritures alphabétiques européennes modernes sont issues de l'écriture grecque ou ont subi son influence. Le mot *alphabet* provient du mot grec *alphabetos*, lui-même dérivé du nom des deux premières lettres de l'alphabet grec, alpha et bêta. L'écriture grecque est une adaptation de l'écriture phénicienne. Les Grecs innovèrent en écrivant de gauche à droite, une caractéristique de toutes les écritures dérivées ou s'inspirant du grec.

Les écritures alphabétiques européennes selon le standard Unicode[1] sont :

- latin,
- grec,
- cyrillique,
- arménien,
- géorgien,
- ogam,
- runes,
- italique,
- gotique.

Elles s'écrivent de gauche à droite. Plusieurs font la distinction entre les formes capitales et de bas de casse de leur alphabet. Des espaces séparent les mots. Les accents et autres signes diacritiques indiquent généralement des caractéristiques phonétiques[2] et permettent d'étendre la puissance descriptive des écritures de base et d'écrire ainsi d'autres langues. L'utilisation de ces signes diacritiques est potentiellement ouverte – c'est l'une des raisons pour lesquelles ces signes combinatoires sont repris dans le standard Unicode.

Le latin et le cyrillique servent à écrire ou à translittérer plusieurs langues. L'alphabet latin est issu d'un alphabet étrusque, lui-même inspiré d'une version occidentale de l'alphabet grec classique. À l'origine, il ne contenait que 24 lettres capitales. L'alphabet latin moderne, tel qu'il est codé dans le bloc du latin de base, doit son apparition aux innovations apportées par les scribes médiévaux et les imprimeurs du début de la Renaissance. L'écriture cyrillique, élaborée au IX^{e} siècle, constitue le dernier avatar du grec en Europe.

Les écritures géorgiennes et arméniennes furent inventées au cinquième siècle sous l'influence du grec. Le géorgien moderne ne distingue pas les majuscules des minuscules ; on dit alors que l'écriture est *unicamérale*.

L'alphabet phonétique international est une extension de l'alphabet latin qui permet de transcrire la dimension phonétique de toutes les langues.

---

[1] Dans l'ordre d'apparition de ce chapitre. Cet ordre n'implique aucunement une importance relative des écritures, mais les couches successives d'ajout au standard Unicode. Les dernières écritures dans la liste étant les plus récentes.

[2] Ce n'est pas le cas en français dans des mots comme « où » ou « paraît ».

Les deux écritures historiques de l'Europe du Nord-ouest, les runes et l'ogam, diffèrent foncièrement par leur aspect des autres écritures, conséquence de leur support principal : le bois et la pierre. De manière générale, elles s'écrivent de gauche à droite dans les textes érudits, toutefois à l'origine on les gravait souvent en forme d'arche pour suivre de la sorte la forme de la pierre.

L'alphabet italique synthétise un certain nombre d'alphabets préclassiques provenant tous de la péninsule italienne. Quant au gotique, à ne pas confondre avec le style 𝔻𝕃𝕼𝔼𝔽ℕℝ𝔹, il s'agit de l'écriture des Goths, représentants de la branche orientale des peuples germains installés sur la Mer Noire au IV^{e} siècle. Cette écriture s'inspire du grec pour les sons communs au grec et au gotique, avec quelques aménagements pour les sons propres au gotique.

8.1 Latin

L'écriture latine est dérivée de l'écriture grecque. On utilise à présent pour écrire un grand nombre de langues dans le monde. Au cours de son adaptation, elle a subi différentes extensions. La plus courante est l'ajout de signes diacritiques. La création de digrammes, de formes culbutées ou réfléchies, et de caractères inédits a également enrichi l'écriture latine.

L'écriture latine s'écrit de gauche à droite. Des espaces séparent les mots et permettent le plus souvent de couper les lignes. Pour la coupure de mots en fin de ligne, on emploie des traits d'union. Pour plus d'information, consultez le *Rapport technique d'Unicode n° 14, « Line Breaking Properties »*, présent sur le cédérom ou sur le site Internet du consortium Unicode pour une version tenue à jour. L'écriture latine distingue les majuscules et les minuscules ; on parle donc d'une écriture *bicamérale*.

Signes diacritiques. L'effet d'un signe diacritique sur une lettre de base dépend de la langue qui l'utilise. Certaines langues considèrent cette combinaison comme une lettre à part entière. D'autres, comme l'anglais, permettent que le même mot s'écrive avec ou sans diacritique sans que le sens en soit affecté. Dans la plupart des langues utilisant l'écriture latine, on considère les lettres portant un signe diacritique comme une variante de la lettre de base, sans que cette combinaison ne forme une lettre supplémentaire et indépendante dans son alphabet. Le codage Unicode de l'écriture latine est suffisamment souple pour que les mises en œuvre puissent prendre en charge ces lettres conformément aux attentes des usagers, pour autant que la langue soit connue. Les caractères accentués les plus fréquents existent sous la forme de caractères simples (précomposés) afin de se conformer aux codages préexistants les plus répandus. Toutes ces lettres accentuées, ainsi que d'autres encore, peuvent s'exprimer à l'aide d'une suite de caractères combinatoires.

Unicode précise que les signes diacritiques doivent *suivre le caractère de base auquel il se rapporte*. Pour plus de détails, consultez la sous-section *Diacritiques* dans la *Section 8.9*, *Diacritiques* et la *Section 2.6, Caractères combinatoires.*

Normes. Les lettres latines Unicode jusqu'à U+00FF correspondent dans l'ordre aux lettres de l'ISO/CEI 8859-1. Celle norme reprend elle-même l'ordre de normes antérieures, parmi lesquelles ASCII (ANSI X3.4) identique à ISO/CEI 646:1991-IRV. Comme l'ASCII, l'ISO/CEI 8859-1 inclut des lettres latines, des signes de ponctuation et des symboles mathématiques ; l'utilisation de ces caractères supplémentaires n'est pas restreinte à l'écriture latine. Le *Chapitre 6, Ponctuation,* décrit ces caractères.

Caractères connexes. Pour les autres caractères latins ou dérivés du latin, consultez les *Symboles de type lettre* (U+2100..U+214F), les *Symboles monétaires* (U+20A0..U+20CF), les *Symboles divers* (U+2600..U+26FF), les *Alphanumériques cerclés* (U+2460..U+24FF) et les *Formes à pleine chasse* (U+FF21..U+FF5A).

Lettres latines de base : U+0020 – U+00BF

Rares sont les langues rédigées avec l'écriture latine qui ne s'écrivent qu'avec les 26 minuscules et majuscules latines de ce bloc. Les 26 paires des lettres de base forment l'essentiel des alphabets utilisés par toutes ces langues. Un texte utilisant un de ces alphabets utilisera donc à la fois des caractères du latin de base et des caractères d'autres blocs latins.

Certaines langues omettent quelques paires de lettres de base, comme l'italien qui ne connaît ni le *j* ni le *w.*

Variantes d'œil. La boucle fermée ou ouverte des lettres *a* et *g* en bas de casse constitue une variation typographique courante selon la police utilisée. Des systèmes de transcription phonétique, comme l'API, font la distinction entre ces différentes formes (« a » et « g » par rapport à « ɑ » et « ɢ »).

Lettres du supplément latin-1 : U+00C0 – U+00FF

Le supplément latin-1 ajoute aux 26 paires de lettres de base de l'ASCII les lettres des principales langues d'Europe occidentale (voir la liste au prochain paragraphe). Comme pour l'ASCII, le latin-1 inclut divers autres signes mathématiques et de ponctuation. La ponctuation, les signes et les symboles qui ne sont pas inclus dans le bloc du latin de base ni le supplément latin-1 sont codés dans des blocs de caractères, à commencer par le bloc de ponctuation générale.

Langues. Le supplément latin-1 permet d'écrire l'allemand, le danois, l'espagnol, le finnois, le féroïen, l'irlandais, l'islandais, l'italien, le néerlandais, le norvégien, le portugais et le suédois.

Nombres ordinaux. On peut reproduire 00AA ª INDICATEUR ORDINAL FÉMININ et U+00BA º INDICATEUR ORDINAL MASCULIN accompagnés d'un souligné, toutefois plusieurs polices de caractères modernes les affichent simplement sous la forme d'exposants, sans souligné. Ces caractères devraient être considérés, pour le tri et le repérage, comme des équivalents faibles des caractères latins correspondants.

Clones à chasse des diacritiques. La norme ISO/CEI 8859-1 comprend huit caractères ambigus, car on ne sait précisément s'il s'agit de caractères combinatoires (des diacritiques) ou de caractères à part entière. Les points de code Unicode correspondants (U+005E ^ ACCENT CIRCONFLEXE, U+005F _ TIRET BAS, U+0060 ` ACCENT GRAVE, U+007E ~ TILDE, U+00A8 ¨ TRÉMA, U+00AF ¯ MACRON, U+00B4 ´ ACCENT AIGU et U+00B8 ¸ CÉDILLE) ne peuvent s'utiliser qu'en tant que caractères à chasse. Le standard Unicode prévoit une série de caractères combinatoires univoques dans le bloc des signes diacritiques utilisés pour représenter des lettres latines accentuées par le biais de séquences de caractères composés. Certaines mises en œuvre ISO/CEI 8859-1 utilisent parfois U+00B0 ° SYMBOLE DEGRÉ de façon ambiguë pour représenter un rond en chef à chasse. Pour sa part, Unicode représente de manière univoque ce signe diacritique à chasse par U+02DA ROND EN CHEF. U+007E ~ TILDE est utilisé pour représenter un signe diacritique tilde à chasse, un opérateur ou un signe de ponctuation ; il est alors généralement centré en hauteur par rapport à l'œil de la lettre. On représente sans ambiguïté un tilde à chasse à l'aide d'un U+02DC ˜ PETIT TILDE.

Latin étendu A : U+0100 – U+017F

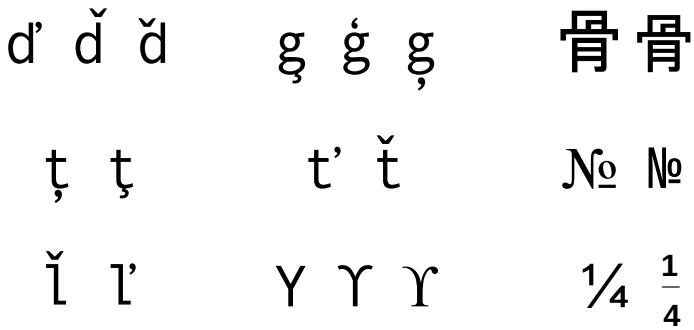
Le bloc latin étendu A contient une collection de lettres qui, jointes aux lettres contenues dans les blocs du latin de base et du supplément latin-1, permettent la représentation de la plupart des langues européennes qui emploient l'écriture latine. Ce bloc permet également d'écrire plusieurs autres langues. La plupart de ces caractères correspondent à des combinaisons précomposées d'un caractère de base et d'un signe diacritique. (Voir *section 2.6, Caractères combinatoires.*)

Normes. Ce bloc reprend les caractères contenus dans la norme internationale ISO/CEI 8859 (2e partie, alphabet latin n° 2 ; 3e partie, alphabet latin n° 3 ; 4e partie, alphabet latin n° 4 ; 9e partie 9, alphabet latin n° 5). Plusieurs autres caractères de ces normes, tels que la ponctuation, les signes, les symboles et les signes diacritiques, sont déjà codés dans le bloc supplément du latin-1. On retrouve d'autres caractères provenant de ces parties de l'ISO/CEI 8859 au sein d'autres blocs, principalement dans le bloc des lettres modificatives (U+02B0..U+02FF) ainsi que dans le bloc de ponctuation générale et dans ceux qui le suivent.

Langues. La plupart des langues supportées par ce bloc font appel à des caractères contenus dans les blocs du latin de base et du supplément latin-1. Lorsque combiné avec ces deux blocs, le bloc du latin étendu A permet d'écrire l'afrikaans, le basque, le breton, le catalan, le croate, l'espéranto, l'estonien, le français, le frison, le gallois, le groenlandais, le hongrois, le latin, le lapon (*sámí*), le letton, le lituanien, le maltais, le polonais, le provençal, le rhéto-roman, le roumain, le slovaque, le slovène, le sorabe, le tchèque, le tsigane (*romani*), le turc et bien d'autres.

Œils de remplacement. Certains caractères peuvent se dessiner de plusieurs façons tout en conservant le même sens. Les tableaux de caractères présentent un glyphe recommandé, même s'il ne s'agit pas de la forme utilisée en toutes circonstances. La *Figure 8-1* présente quelques exemples de ces différents œils.

Figure 8-1. Glyphes de remplacement



En typographie tchèque, on préfère souvent la forme avec apostrophe des lettres U+010F ď LETTRE MINUSCULE LATINE D CARON et U+0165 ť LETTRE MINUSCULE LATINE T CARON à celles qui utilisent un caron (*hacek*, *hatchek*) au-dessus de la lettre de base. En slovaque, cet usage s'applique à U+013E ľ LETTRE MINUSCULE LATINE L CARON. On utilise l'apostrophe pour éviter que la hampe de ces lettres ne se superpose aux signes de la ligne supérieure, pour obtenir une typographie plus lisible. Au contraire, dans des documents écrits à la main ou à la machine, dans du matériel didactique ou pédagogique, on retrouve de manière prépondérante les formes à caron. Il est possible que d'autres langues que le tchèque ou le slovaque utilisent systématiquement les caractères avec caron.

Une situation semblable se présente avec la lettre lettonne U+0123 ģ LETTRE MINUSCULE LATINE G CÉDILLE. La typographie lettonne fine utilise une virgule culbutée au-dessus du *g* et non une cédille sous cette lettre, car il est peu esthétique de placer une cédille sous la boucle inférieure du *g*. Certaines polices de caractères lettonnes incomplètes peuvent substituer un accent aigu à la virgule culbutée. On retrouve cependant la cédille sous le *g* minuscule dans certains manuscrits, voire certains imprimés. La capitale utilise toujours la cédille puisque la forme arrondie de la partie inférieure du *G* est alors propice à l'accrochage de la cédille.

D'autres lettres lettonnes, dont la forme ne se prête pas à l'ajout d'une cédille (U+0137 ķ LETTRE MINUSCULE LATINE K CÉDILLE, U+0146 ņ LETTRE MINUSCULE LATINE N CÉDILLE et U+0157 ŗ LETTRE MINUSCULE LATINE R CÉDILLE), utilisent invariablement une virgule flottante.

En turc et en roumain, la cédille et la virgule souscrites sont interchangeables, selon la police utilisée. Les lettres U+015F ş LETTRE MINUSCULE LATINE S CÉDILLE et U+0163 ţ LETTRE MINUSCULE LATINE T CÉDILLE (ainsi que leurs homologues en capitales) ont été répétées sous la forme de U+0219 ș LETTRE MINUSCULE LATINE S VIRGULE SOUSCRITE et U+021B ț LETTRE

MINUSCULE LATINE T VIRGULE SOUSCRITE. Ces caractères n'existent que pour se conformer à des usages socio-politiques. Les jeux de caractères préexistants (dont l'ISO/CEI 8859-2) ne reprennent qu'une seule de ces deux formes.

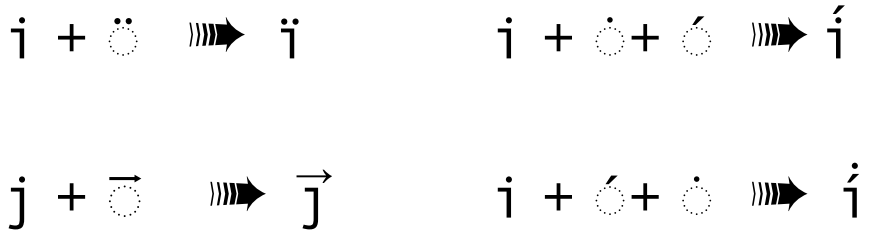
En général, les caractères dont la base est munie d'une cédille ou d'un ogonek sont sujets à des usages typographiques variés, selon l'accessibilité et la qualité des polices utilisées, la technologie et la région. Divers crochets, virgules et autres fioritures peuvent remplacer la forme de référence de ces diacritiques souscrits ; la direction des crochets peut, de surcroît, être inversée. Il faut donc se familiariser avec ces traditions typographiques particulières avant de présumer que des caractères manquent ou sont mal représentés dans les tableaux de caractères Unicode.

Paires de casse remarquables. On considère que les caractères U+0130 İ LETTRE MAJUSCULE LATINE I POINT EN CHEF et U+0131 ı LETTRE MINUSCULE LATINE I SANS POINT (surtout utilisés en turc) ont respectivement pour casse inversée les caractères ASCII « i » et « I ». Ceci signifie que leur correspondance inverse dépend de la langue ; la mise en correspondance bijective (aller-retour) nécessite une attention spéciale de la part du développeur (se rapporter à la *Section 5.17, Tri et repérage*). Voir le fichier *SpecialCasing.txt* sur le cédérom pour plus d'information.

Diacritiques placés sur le i et le j. Un *i* (normal) ou un *j* suivi d'un signe à chasse nulle en chef perd le point. Ainsi, dans le mot *naïf*, le *ï* peut s'écrire à l'aide d'un *i* + *tréma*. De la même façon que le A cyrillique n'est pas équivalent au A latin, un *i* n'est pas équivalent à un *i* turc sans point + un point en chef, pas plus qu'un *i* normal accentué n'est équivalent à un *i* sans point accentué (en d'autres mots, i + ¨ ≠ ı + ¨). La même règle s'applique également au *j*.

Pour exprimer les formes baltes où le point demeure parfois sous l'accent, on utilise *i + point en chef + accent* (voir *Figure 8-2*).

Figure 8-2. Diacritiques sur *i* et *j*



Latin étendu B : U+0180 – U+024F

Le bloc du latin étendu B contient des caractères utilisés pour représenter des langues qui ne peuvent l'être à l'aide du latin de base et du latin étendu A. Il contient également des symboles phonétiques qui ne sont pas inclus dans l'alphabet phonétique international (voir le bloc de l'alphabet phonétique international, U+0250..U+02AF).

Normes. Ce bloc englobe, entre autres, les caractères de la norme ISO 6438 (Documentation, jeu de caractères africains codés pour l'échange d'informations bibliographiques), ceux utilisés par la transcription latine *pinyin* conformément aux normes nationales de la République populaire de Chine GB 2312 et du Japon JIS X 0212, ainsi que les caractères lapons (sámi) de la norme ISO/CEI 8859 : Technologies de l'information — Jeux de caractères graphiques codés sur un seul octet – Partie 10 : Alphabet latin n° 6.

Agencement. Les caractères sont disposés dans l'ordre alphabétique habituel de leur caractère de base, suivi de quelques caractères à forme latine. Les paires de bas de casse et de capitales sont placées côte à côte aussi souvent que possible ; il arrive cependant, dans bien des cas, que l'autre forme de casse soit codée ailleurs ; elle fait alors l'objet d'un renvoi dans le tableau des caractères. Les variantes d'une même lettre se présentent dans l'ordre suivant : culbutée, réfléchie, à crosse ou à hameçon, à trait prolongé ou modifié, au style différent (cursive ou de ronde), petite capitale, à forme de base modifiée, ligaturée et enfin dérivée du grec.

Digrammes croates correspondant aux lettres cyrilliques serbes. Le serbo-croate ne forme qu'une seule langue, mais il s'écrit à l'aide d'alphabets jumelés : une écriture latine (le croate) et une écriture cyrillique (le serbe). Afin de permettre la translittération entre ces deux alphabets, Unicode fournit quelques digrammes. Chaque digramme se présente sous deux formes de capitales possibles : la première pour les capitales initiales (*casse de titre*), la deuxième pour les mots tout en capitales. Unicode code ces deux formes afin que les logiciels puissent passer de l'une à l'autre sans devoir changer de polices de caractères. Un renvoi associé aux minuscules indique les numéros de caractère des casses majuscules correspondantes. Pour plus d'information sur les équivalences canoniques, voir le *Chapitre 3, Conformité*.

Combinaison de voyelles et de diacritiques pinyins. La norme chinoise GB 2312, ainsi que la norme japonaise JIS X 0212, incluent une série de codes pour le pinyin, système de transcription latine du chinois mandarin. On retrouve la plupart des lettres utilisées pour cette romanisation du mandarin (même celles munies de signes diacritiques) dans les blocs latins précédents. Les 16 caractères codés dans ce bloc complètent la série de caractères pinyins définis dans les normes GB 2312 et JIS X 0212.

Paires de casse. Parmi les caractères de ce bloc figurent des majuscules dont la minuscule est codée ailleurs. Plusieurs de ces caractères proviennent de l'alphabet phonétique international ; ils ont acquis une forme majuscule lors de leur intégration à des écritures latines. À l'occasion, cependant, *plusieurs* formes capitales ont ainsi vu le jour. Des recherches ont démontré que ces différentes majuscules ne sont parfois que des variantes d'un même caractère. Ces variantes ne possèdent alors qu'une valeur Unicode, c'est le cas du U+01B7 Ʒ LETTRE MAJUSCULE LATINE EJ. Si ces recherches ont, par contre, établi que les deux formes de capitales possèdent des emplois distincts, chaque forme s'est alors vue attribuer un numéro de caractère différent, c'est le cas pour U+018E Ǝ LETTRE MAJUSCULE LATINE E RÉFLÉCHI et pour U+018F Ə LETTRE MAJUSCULE LATINE SCHWA. La forme bas de casse commune a alors été dédoublée afin de garantir une correspondance de casse univoque : U+01DD ǝ LETTRE MINUSCULE LATINE E CULBUTÉ est donc synonyme à U+0259 ə LETTRE MINUSCULE LATINE SCHWA.

Pour des raisons de fait, les noms de certaines paires de casse diffèrent. Ainsi, U+018E Ǝ LETTRE MAJUSCULE LATINE E RÉFLÉCHI est la capitale de U+01DD ǝ LETTRE MINUSCULE LATINE E CULBUTÉ et non de U+0258 ɘ LETTRE MINUSCULE LATINE E RÉFLÉCHI. (Pour la correspondance de casse par défaut des caractères Unicode, voir le *Chapitre 4, Propriétés des caractères*).

Langues. Les tableaux de caractères fournissent pour la grande majorité des caractères des indications sur les langues qui les utilisent ainsi que d'autres précisions d'utilisation.

Alphabet phonétique international : U+0250 – U+02AF

Le bloc de l'alphabet phonétique international (API) contient principalement les symboles spécifiques de cet alphabet conçu pour représenter graphiquement les sons du langage parlé. Depuis son invention en 1886, son contenu et son utilisation furent modifiés à diverses reprises. Le standard Unicode comprend tous les symboles indépendants et les diacritiques de la dernière version de l'API (publiée en 1989), ainsi que quelques symboles utilisés précédemment par l'API. Quelques symboles employés par les sinologues, les américanistes et d'autres linguistes ont été ajoutés à ce bloc. Certains de ces signes, utilisés dans des contextes étrangers à l'API, peuvent faire appel à des caractères d'autres blocs. Notons que le bloc du latin étendu B reprend quelques symboles phonétiques désuets ou hors normes.

Une des caractéristiques essentielles de l'API est son recours fréquent à des diacritiques. Ces signes diacritiques de l'API sont codés dans le bloc des diacritiques (U+0300..U+036F). L'API permet la libre adjonction de signes diacritiques aux lettres de base afin de représenter les subtiles variations phonétiques nécessaires à une transcription fidèle des langues.

Normes. Les caractères de ce bloc proviennent de l'alphabet phonétique international, publié par l'Association phonétique internationale[3] et révisé en 1989. Cette norme considère l'API comme un alphabet indépendant, ainsi inclut-il l'alphabet latin en bas de casse dans son intégralité (de *a* à *z)*, un certain nombre de lettres latines étendues comme U+0153 œ DIGRAMME SOUDÉ MINUSCULE LATIN OE, quelques lettres grecques et d'autres symboles. Unicode, par contre, n'inclut dans le bloc consacré à l'API ni les lettres latines en bas de casse (de *a* à *z),* ni les autres lettres latines ou grecques. Il est à noter que, contrairement aux autres sources de caractères du standard Unicode, l'API constitue un alphabet étendu et une norme de transcription phonétique et non une norme de codage de caractères.

Unifications. Autant que faire se peut, les signes API ont été unifiés avec d'autres lettres, mais non avec des symboles (qui ne sont pas des lettres) comme U+222B ∫ INTÉGRALE. De nombreuses langues utilisant une écriture latine, dont certaines en Afrique, ont adopté des symboles API. Dans ce cas, il est alors vain d'essayer de distinguer la transcription de l'alphabet lui-même. C'est pourquoi beaucoup de symboles de l'API se retrouvent en dehors du bloc API. Un renvoi à ces symboles figure en début de la liste des caractères du bloc API dans le tableau des caractères.

Formes API équivalentes. Dans certains cas, la pratique de l'API a produit, avec le temps, des formes équivalentes. C'est le cas, par exemple, pour U+0269 ɩ LETTRE MINUSCULE LATINE IOTA qui peut remplacer U+026A ɪ LETTRE LATINE PETITE CAPITALE I. Le standard Unicode propose séparément ces deux formes équivalentes, car les utilisateurs d'API les distinguent habituellement sans que leur valeur phonétique soit différente.

Casse. L'API ne connaît pas de distinctions de casse ; tous ses symboles phonétiques sont en effet en bas de casse. Lorsqu'un signe API est intégré à un alphabet particulier et est utilisé par une langue écrite donnée (ce qui s'est produit, par exemple, en Afrique), il acquiert alors en règle générale une forme majuscule. Ces capitales n'étant pas, à l'origine, des signes API, ils sont généralement codés dans le bloc du latin étendu B (ou dans d'autres blocs du latin étendu). Un renvoi indique la forme API associée.

Variantes typographiques. L'API inclut des variantes typographiques pour certaines lettres latines et grecques qui, d'ordinaire, seraient considérées comme des variations de style de caractères et non comme des caractères ayant leur identité propre, comme c'est le cas des lettres en petites capitales. On peut citer comme exemples une variante typographique de la lettre grecque *phi* ɸ, ainsi que la lettre empruntée au grec iota ɩ, qui possède une forme

[3] <http://www2.arts.gla.ac.uk/IPA/ipa.html>

capitale unique au latin. Ces formes sont codées dans le standard Unicode comme caractères indépendants, car ils possèdent une sémantique distincte.

Ligatures de digrammes affriqués. Officiellement, l'API reconnaît six ligatures de digrammes utilisés dans la transcription des consonnes affriquées (U+02A3..U+02A8). Les ligatures de ces digrammes API sont définies explicitement dans l'API. Elles peuvent de surcroît posséder une valeur sémantique propre ce qui fait d'elles plus que de simples variantes typographiques. U+02A6 ʦ LETTRE MINUSCULE LATINE DIGRAMME TS peut également être transcrite en API sous la forme de « ts » U+0074 U+0073. Le choix de la ligature de digramme peut résulter d'une distinction délibérée effectuée par le transcripteur relativement à la nature phonétique systématique des consonnes affriquées. Ce choix de ligature ne peut dès lors être laissé à un logiciel qui se baserait sur les polices de caractères disponibles. L'œil de cette ligature diffère également de celui de la ligature *ts* présente dans certaines polices de caractères classiques.

Agencement. Les caractères du bloc API sont triés dans l'ordre alphabétique de la lettre latine ressemblant au signe phonétique correspondant. Cet ordre ne dépend donc pas des propriétés phonétiques de ces lettres.

Latin étendu additionnel : U+1E00 – U+1EFF

Ce bloc est constitué d'une série de caractères latins précomposés. Chacun des caractères de ce bloc peut être représenté par une lettre latine de base suivie par un ou plusieurs signes diacritiques. La forme canonique de ces différentes représentations est précisée au *Chapitre* 3, *Conformité*.

Combinaisons d'une voyelle vietnamienne et d'un signe de ton. Une partie de ce bloc reprend les voyelles de l'alphabet moderne vietnamien (*quốc ngữ*) dotées des signes diacritiques représentant le ton phonémique de la syllabe. L'alphabet vietnamien moderne comprend 12 voyelles et cinq signes de tons (voir *Figure 8-3*).

Figure 8-3. Lettres vietnamiennes et signe de ton

a ă â ẹ e ê i o ô ơ u ư y

◌́ ◌̀ ◌̉ ◌̃ ◌̣

Certaines mises en œuvre vietnamiennes préfèrent stocker les combinaisons d'une voyelle et d'un signe de ton sous la forme d'un seul élément codé ; d'autres, au contraire, codent la voyelle et le signe de ton séparément. Le premier type de mise en œuvre utilise les caractères définis dans ce bloc avec les formes combinées définies dans les blocs du supplément latin-1 et du latin étendu A ; le second utilise les voyelles de base des blocs du latin de base, du supplément latin-1 et du latin étendu A, ainsi que les caractères du bloc des signes diacritiques. Cette dernière méthode utilise les caractères U+0300 ◌̀ DIACRITIQUE ACCENT GRAVE, U+0309 ◌̉ DIACRITIQUE CROCHET EN CHEF, U+0303 ◌̃ DIACRITIQUE TILDE, U+0301 ◌́ DIACRITIQUE ACCENT AIGU et U+0323 ◌̣ DIACRITIQUE POINT SOUSCRIT pour représenter les signes de ton vietnamiens. Les caractères U+0340 ◌̀ DIACRITIQUE MARQUE DE TON GRAVE et U+0341 ◌́ DIACRITIQUE MARQUE DE TON AIGU ne doivent plus être utilisés.

Ligatures latines : U+FB00 – U+FB06

Cette section du bloc des formes de présentations alphabétiques (U+FB00..U+FB4F) contient plusieurs ligatures latines courantes, héritées de codages préexistants. De par sa conception, Unicode ne prévoit pas de mécanisme général qui permette d'indiquer l'endroit où une ligature

devrait apparaître. En effet, la formation d'une ligature dépend de règles orthographiques et typographiques particulières à chaque langue. Certaines langues interdisent les ligatures entre les mots. Dans ces cas, il est préférable de stocker en mémoire des caractères non ligaturés et de préciser hors texte à la couche de rendu où les ligatures peuvent avoir lieu.

8.2 Grec

Grec : U+0370 – U+03FF

L'écriture grecque s'emploie pour écrire le grec et (en tant que variante étendue) le copte. L'influence de l'écriture grecque sur le développement des écritures latines et cyrilliques fut décisive.

Le grec s'est écrit de gauche à droite. Il emploie, à l'occasion, des signes à chasse nulle. Les lettres grecques connaissent les deux casses habituelles : minuscules et majuscules, on dit que l'écriture est bicamérale.

Normes. Le codage Unicode du grec se fonde sur la norme ISO/CEI 8859-7, elle-même équivalente à la norme nationale grecque ELOT 928. Unicode met les caractères grecs aux mêmes positions relatives que l'ISO/CEI 8859-7. Un certain nombre de variantes et de caractères proviennent de la norme bibliographique ISO 5428.

Grec polytonique. On peut coder le grec polytonique, utilisé en grec ancien (classique et byzantin), à l'aide de suites de caractères combinatoires ou de caractères de base précomposées auxquels s'adjoignent des diacritiques. Pour plus de renseignements sur cette dernière méthode, consultez la sous-section suivante, *Grec étendu* : U+1F00 – U+1FF.

Signes à chasse nulle. Plusieurs signes à chasse nulle fréquents en grec se trouvent parmi les signes diacritiques (voir *Tableau 8-1*).

Tableau 8-1. Signes à chasse nulle utilisés en grec

Numéro		Nom ISO 10646	Noms optionnels
U+0300	◌̀	DIACRITIQUE ACCENT GRAVE	varia
U+0301	◌́	DIACRITIQUE ACCENT AIGU	tonos, oxia
U+0302	◌̂	DIACRITIQUE ACCENT CIRCONFLEXE	
U+0303	◌̃	DIACRITIQUE TILDE	
U+0304	◌̄	DIACRITIQUE MACRON	long
U+0306	◌̆	DIACRITIQUE BRÈVE	vrakhy, brakhus
U+0308	◌̈	DIACRITIQUE TRÉMA	dialytika, double point en chef
U+0313	◌̓	DIACRITIQUE VIRGULE EN CHEF	esprit doux, psili
U+0314	◌̔	DIACRITIQUE VIRGULE RÉFLÉCHIE EN CHEF	esprit rude, dasia grec
U+0342	◌͂	DIACRITIQUE GREC ACCENT CIRCONFLEXE	tilde, périspoméni
U+0343	◌̓	DIACRITIQUE GREC CORONIS	crase, virgule en chef
U+0345	◌ͅ	DIACRITIQUE GREC IOTA SOUSCRIT	ypogégramméni

Puisque les caractères du bloc des diacritiques n'ont pas de sens particulier mais sont codés selon leur forme ; ils peuvent donc s'utiliser au besoin en grec. Il faut toutefois éviter d'utiliser le caractère U+0344 ◌̈́ DIACRITIQUE GREC DIALYTIKA TONOS. On représente plutôt la combinaison d'un *dialytika* et d'un *tonos* à l'aide de la suite U+0308 ◌̈ DIACRITIQUE TRÉMA + U+0301 ◌́ DIACRITIQUE ACCENT AIGU.

On code les différents signes diacritiques adjoints à un même caractère de base en commençant par les diacritiques les plus proches de ce caractère pour coder ensuite les diacritiques qui en sont plus éloignés (« codage centrifuge »). Voir les règles générales d'adjonction des signes à chasse nulle dans la *Section 2.6, Caractères combinatoires.*

L'accent grec de base, en grec moderne, est le *tonos*. Il est représenté par un accent aigu (U+0301). Cet accent est généralement plus incliné (les formes extrêmes sont presque verticales) que celui utilisé pour les lettres latines. Dans les versions précédentes du standard Unicode, l'accent fut malencontreusement représenté par une ligne verticale au-dessus des voyelles. Le grec polytonique, quant à lui, s'écrit à l'aide de plusieurs accents ; l'accent aigu s'appelle *oxia*, alors que l'accent grave se nomme *varia.*

U+0342 ◌͂ DIACRITIQUE GREC ACCENT CIRCONFLEXE se représente soit par un accent circonflexe (^) soit un tilde (~). Cette variation de forme explique que cet accent est codé indépendamment du U+0303 ◌̃ DIACRITIQUE TILDE.

U+0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF et U+0343 ◌̓ DIACRITIQUE GREC CORONIS prennent tous deux la forme d'une virgule placée au-dessus de la lettre de base. U+0343 ◌̓ DIACRITIQUE GREC CORONIS est inclus pour des raisons de compatibilité ; la forme U+0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF est recommandée dans l'usage courant.

Le signe à chasse nulle *iota souscrit (ypogégramméni)* peut être adjoint aux voyelles *alpha*, *êta* et *oméga* afin de représenter des diphtongues historiques. Adjoint à une voyelle initiale majuscule, l'iota prend alors habituellement la forme d'un iota minuscule collé à droite de la voyelle. On nomme cette forme un *iota adscrit (prosgégramméni)*. Dans des mots écrits entièrement en capitales, l'*iota souscrit* doit être remplacé par un *iota adscrit* majuscule. Voir le fichier *SpecialCasing.txt* sur le disque optique. Les représentations archaïques de mots grecs (qui ne possèdent ni bas de casse ni accents) utilisent un *iota* majuscule à la suite de la voyelle pour ces diphtongues. Ces représentations archaïques exigent une correspondance de casse particulière.

Variantes glyphiques. U+03A5 Υ LETTRE MAJUSCULE GRECQUE UPSILON possède deux formes courantes – une ressemblant à la capitale du Y latin et l'autre à deux branches symétriques rappelant les cornes d'un bélier « ϒ». On a systématiquement choisi la forme en Y dans les tableaux de caractères, à la fois pour le grec monotonique et polytonique. La forme du glyphe en cornes de bélier s'avère utile en mathématique.

L'ISO/CEI 8859-7 et l'ISO 5428 codent d'autres variantes de formes de lettres grecques en tant que caractères indépendants. Unicode hérite de ces formes et les code séparément. Il s'agit de U+03C2 ς LETTRE MINUSCULE GRECQUE SIGMA FINAL, de U+03D0 ϐ SYMBOLE GREC BÊTA ainsi que de formes supplémentaires de la lettre upsilon capitale possédant un crochet asymétrique – par exemple U+03D2 ϒ SYMBOLE GREC UPSILON CROCHET.

Lettres grecques utilisées comme symboles. Pour des raisons de compatibilité, quelques lettres grecques sont codées séparément en tant que symboles dans d'autres blocs de caractères. Par exemple, U+00B5 µ SYMBOLE MICRO se trouve dans le bloc de caractères du supplément latin-1 et U+2126 Ω SYMBOLE OHM est placé dans le bloc des symboles de type lettre. Les lettres grecques sont couramment utilisées comme opérateurs ou variables mathématiques. Les caractères du bloc grec peuvent servir à de tels symboles.

Ponctuation. La distinction entre des caractères de ponctuation exclusivement grecs et ceux qui correspondent à la ponctuation occidentale n'est pas toujours clairement établie. Le point d'interrogation grec U+037E ; POINT D'INTERROGATION GREC est codé dans ce bloc pour des raisons de compatibilité. On recommande plutôt l'emploi de U+003B ; POINT-VIRGULE.

Lettres historiques. Les lettres historiques grecques proviennent de l'ISO 5428.

Lettres propres au copte. On considère habituellement l'écriture copte comme une variante stylistique de l'alphabet grec. Les lettres propres au copte forment un groupe à la fin du bloc grec. Ces caractères s'utilisent de pair avec les caractères grecs de base pour compléter ainsi l'alphabet copte. On peut rendre les textes coptes à l'aide d'une police qui représente les caractères communs au copte et au grec à la manière copte. Les textes écrits à la fois en grec et en copte doivent employer des styles de polices appropriés.

Caractères connexes. Pour les symboles mathématiques, voir les *Opérateurs mathématiques* (U+2200..U+22FF). Pour des signes de ponctuation supplémentaires utilisés avec cette écriture, voir *Commandes C0 et latin de base* (U+0000..U+007F).

Grec étendu : U+1F00 – U+1FFF

Les caractères de ce bloc sont des combinaisons précomposées de lettres grecques munies d'un ou de plusieurs signes diacritiques. On y trouve, en outre, un certain nombre de signes diacritiques à chasse. Ces lettres précomposées facilitent la représentation de textes grecs polytoniques sans avoir recours aux signes combinatoires.

Chacun des caractères de ce bloc peut également être représenté à l'aide d'une lettre de base du bloc grec suivie d'un ou de plusieurs signes diacritiques appartenant au bloc des diacritiques. Le *Chapitre 3, Conformité,* définit une forme canonique pour chaque alternative.

Diacritiques avec chasse. Ce bloc comprend seize autres signes diacritiques à chasse qui permettent de représenter des textes grecs polytoniques. Chacun de ces signes peut prendre une autre forme qu'on utilisera avec des systèmes prenant en charge les signes à chasse nulle. Unicode considère la forme à chasse nulle comme la représentation canonique de l'information codée par la forme à chasse correspondante. Le *Tableau 8-2* illustre ces différentes paires de diacritiques.

Tableau 8-2. Paires grecques avec et sans chasse

Forme avec chasse	Forme sans chasse
037A ͺ CARACTÈRE GREC IOTA SOUSCRIT	0345 ◌ͅ DIACRITIQUE GREC IOTA SOUSCRIT
1FBD ᾽ CORONIS	0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF
1FBF ᾿ ESPRIT DOUX	0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF
1FC0 ῀ CIRCONFLEXE GREC	0342 ◌͂ DIACRITIQUE GREC ACCENT CIRCONFLEXE
1FC1 ῁ DIALYTIKA ET CIRCONFLEXE	0308 ◌̈ DIACRITIQUE TRÉMA + 0342 ◌͂ DIACRITIQUE GREC ACCENT CIRCONFLEXE
1FCD ῍ ESPRIT DOUX ET ACCENT GRAVE	0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF + 0300 ◌̀ DIACRITIQUE ACCENT GRAVE
1FCE ῎ ESPRIT DOUX ET ACCENT AIGU	0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF + 0301 ◌́ DIACRITIQUE ACCENT AIGU
1FCF ῏ ESPRIT DOUX ET CIRCONFLEXE	0313 ◌̓ DIACRITIQUE VIRGULE EN CHEF + 0342 ◌͂ DIACRITIQUE GREC ACCENT CIRCONFLEXE
1FDD ῝ ESPRIT RUDE ET ACCENT GRAVE	0314 ◌̔ DIACRITIQUE VIRGULE RÉFLÉCHIE EN CHEF + 0300 ◌̀ DIACRITIQUE ACCENT GRAVE
1FDE ῞ ESPRIT RUDE ET ACCENT AIGU	0314 ◌̔ DIACRITIQUE VIRGULE RÉFLÉCHIE EN CHEF

Forme avec chasse	Forme sans chasse
	+ 0301 ◌́ DIACRITIQUE ACCENT AIGU
1FDF ῟ ESPRIT RUDE ET CIRCONFLEXE	0314 ◌̔ DIACRITIQUE VIRGULE RÉFLÉCHIE EN CHEF + 0342 ◌͂ DIACRITIQUE GREC ACCENT CIRCONFLEXE
1FED ΅ DIALYTIKA ET ACCENT GRAVE	0308 ◌̈ DIACRITIQUE TRÉMA + 0300 ◌̀ DIACRITIQUE ACCENT GRAVE
1FEE ΅ DIALYTIKA ET ACCENT AIGU	0308 ◌̈ DIACRITIQUE TRÉMA + 0301 ◌́ DIACRITIQUE ACCENT AIGU
1FEF ` ACCENT GRAVE GREC	0300 ◌̀ DIACRITIQUE ACCENT GRAVE
1FFD ´ ACCENT AIGU GREC	0301 ◌́ DIACRITIQUE ACCENT AIGU
1FFE ῾ ESPRIT RUDE	0314 ◌̔ DIACRITIQUE VIRGULE RÉFLÉCHIE EN CHEF

Décomposition des formes avec chasse. Lors de la décomposition des formes à chasse, on doit prendre en compte l'usage qu'on en fera afin d'établir si le résultat chasse ou non. À moins d'indication contraire, ces formes à chasse se décomposent en U+0020 ESPACE suivi de la forme à chasse nulle équivalente apparaissant dans le *Tableau 8-2*.

Dans les textes grecs archaïques, U+0345 ◌ͅ DIACRITIQUE GREC IOTA SOUSCRIT et les formes précomposées qui le contiennent admettent une correspondance de casse particulière.

8.3 Cyrillique

Cyrillique : U+0400 – U+04FF

L'écriture cyrillique fait partie de la famille des écritures fortement influencées par l'écriture grecque. Historiquement, l'alphabet cyrillique a été utilisé pour transcrire diverses langues slaves, dont le russe est le plus important représentant. Aux XIX^{e} et XX^{e} siècles, l'alphabet cyrillique servit également à transcrire des langues minoritaires non slaves de l'ex-Union soviétique. Le cyrillique s'écrit de gauche à droite, il utilise à l'occasion des signes à chasse nulle.

Le cyrillique est une écriture *bicamérale*.

Normes. Le bloc Unicode de l'écriture cyrillique repose sur l'ISO/CEI 8859-5. Unicode place les caractères cyrilliques aux mêmes positions relatives que l'ISO/CEI 8859-5.

Unifications. Les caractères latins comme le *q* et le *w* kurdes, faisant partie d'alphabets où l'on retrouve à la fois des lettres latines et cyrilliques, ne sont pas recodés en cyrillique.

Lettres historiques. On considère les formes historiques de l'alphabet cyrillique comme une variante de style de police par rapport au cyrillique moderne. En effet, ces formes historiques ressemblent non seulement aux formes modernes, mais certaines d'entre elles sont encore utilisées aujourd'hui par des langues autres que le russe (ainsi U+0406 І LETTRE MAJUSCULE CYRILLIQUE I BIÉLORUSSE-UKRAINIEN s'emploie toujours en ukrainien et en biélorusse). Les caractères cyrilliques historiques Unicode (U+0460..U+0486) ne se présentent que rarement dans les textes modernes, c'est pourquoi ils apparaissent sous leur forme archaïque dans les tableaux de caractères. Pour obtenir un jeu cyrillique archaïque complet, il suffit de rendre toute la section de l'alphabet cyrillique (c'est-à-dire U+0400..U+0486) à l'aide d'une police de même style.

Cyrillique étendu. Font partie du cyrillique étendu les caractères propres aux langues minoritaires de l'ex-Union soviétique. Les écritures de certaines de ces langues ont souvent été révisées par le passé. Unicode n'inclut que les alphabets utilisés de nos jours et non les formes de lettres désuètes ou rejetées.

Glagolitique. La genèse et la généalogie des écritures slaves ne sont malheureusement pas documentées. Unicode ne considère pas le glagolitique comme une simple variation stylistique du cyrillique, mais bien comme une écriture à part entière. Ce refus se justifie d'abord par la grande différence d'aspect et de propagation du glagolitique par rapport au cyrillique. À l'heure actuelle, Unicode ne prend pas en charge le glagolitique.

8.4 Arménien

Arménien : U+0530 – U+058F

L'écriture arménienne s'utilise principalement pour écrire l'arménien. Celui-ci s'écrit de gauche à droite et n'utilise généralement pas de diacritiques (sauf pour les lettres modificatives mentionnées ci-dessous). Cette écriture distingue des paires de bas de casse et de capitales, on parle alors d'une écriture *bicamérale*.

Lettres modificatives. En typographie arménienne, les petits signes appartenant au groupe nommé lettres modificatives arméniennes se placent au-dessus et à la droite des autres lettres, occupant ainsi la place de lettres à part entière. Ainsi, le signe d'accentuation, le point d'exclamation et le point d'interrogation se positionnent-ils à la droite de la voyelle de la syllabe mise en relief. L'utilisation de ces lettres modificatives entraîne souvent des modifications d'approche horizontale et verticale; il est donc préférable de recourir aux mécanismes de crénage par paire de caractères décrits à la *Section 5.15, Repérage des frontières d'élément textuel.* Ces lettres modificatives possédant généralement une chasse (largeur) propre, Unicode les traite tout naturellement comme des lettres à chasse plutôt que des signes à chasse nulle.

Il semble que U+0559 ՙ LETTRE MODIFICATIVE ARMÉNIENNE DEMI-ROND GAUCHE ne soit pas utilisée dans les textes arméniens ; sa présence dans ce bloc est donc probablement injustifiée.

Ponctuation. L'arménien utilise des nombreux signes de ponctuation provenant d'autres blocs, comme U+002C , VIRGULE et U+00B7 · POINT MÉDIAN. Dans un texte arménien, ces signes de ponctuation doivent s'afficher dans un style analogue à celui des caractères arméniens du texte. Outre U+055D ՝ VIRGULE ARMÉNIENNE, dont une forme apparaît parmi les lettres modificatives, l'arménien possède deux signes de ponctuation qui lui sont propres : U+058A ֊ TRAIT D'UNION ARMÉNIEN et U+0589 ։ POINT ARMÉNIEN.

Ce dernier caractère agit à la manière d'un U+00AD TRAIT D'UNION VIRTUEL. Il est utilisé pour indiquer une coupure de ligne légitime à l'intérieur d'un mot arménien polysyllabique. Sa forme le distingue d'un trait d'union virtuel.

Ligatures. Le bloc des formes de présentation alphabétiques (U+FB13..U+FB17) contient cinq ligatures arméniennes. En effet, de par sa conception, le standard Unicode n'offre pas de mécanisme pour indiquer où afficher une ligature.

Pour les ponctuations supplémentaires utilisées par cette écriture, voir *Commandes C0 et Ponctuation ASCII* (U+0000..U+007F).

8.5 Géorgien

Géorgien : U+10A0 – U+10FF

L'écriture géorgienne s'utilise principalement pour écrire la langue géorgienne et ses différents dialectes. Elle s'emploie également pour écrire le svane, le mingrélien et, par le passé, l'abkhaze et d'autres langues du Caucase.

Styles d'écriture. L'écriture géorgienne fit son apparition sous la forme d'inscriptions appelées *assomtavrouli ;* elle évolua pour devenir ensuite la forme manuscrite connue sous le nom de *nouskhouri*. Ces deux formes appartiennent à la tradition ecclésiastique *(khoutsouri)*. La forme *nouskhouri* ne s'utilise plus habituellement dans des textes modernes, bien qu'elle se rencontre encore dans des textes liturgiques. Elle fut remplacée, l'histoire est assez vague à cet égard, par un alphabet militaire, le *mkhédrouli,* utilisé dans presque tous les textes géorgiens modernes.

Formes de casse. L'alphabet géorgien est foncièrement unicaméral et c'est ainsi qu'il apparaît dans la plupart des textes. Toutefois, sous l'influence probable d'autres alphabets, le géorgien moderne comporte à l'occasion des lettres capitales. On utilise à cet effet l'*assomtavrouli*, alors que le *mkhédrouli* ou le *nouskhouri* représentent les minuscules. Cette répartition historique coïncide avec celle de l'alphabet latin où le style primitif monumental se transforma petit à petit en nos majuscules alors que certains styles de lettres manuscrites vinrent à représenter nos minuscules. Le codage Unicode du géorgien partage cette évolution avec le latin : la série U+10A0..U+10CF représentent les capitales (*assomtavrouli*), cependant que les lettres de base U+10D0..U+10FF représentent les minuscules (*mkhédrouli* ou *nouskhouri*). Dans des textes géorgiens en bas de casse (c'est-à-dire unicaméraux), le *mkhédrouli* ou le *nouskhouri* diffèrent par leur style de la même manière que le romain et l'italique dans les textes latins en bas de casse.

Style	« majuscules » U+10A0..U+10CF	« minuscules » U+10D0..U10FF
Séculaire	assomtavrouli	mkhédrouli
Ecclésiastique (khoutsouri)	assomtavrouli	nouskhouri

Le géorgien étant foncièrement unicaméral, la *Base de données des caractères Unicode* ne définit pas de correspondance de casse implicite pour cette écriture. Il n'est pas recommandé de convertir les textes *mkhédrouli* en *assomtavrouli* par une simple transformation de casse. Lorsqu'un logiciel considère les formes *assomtavrouli* comme des majuscules, la mise en minuscules devrait s'appuyer sur des transformations de casse particulières, pour former un protocole de niveau supérieur.

Séparateur de paragraphes géorgien. Le séparateur de paragraphes géorgien possède une représentation graphique distinctive ; il se retrouve donc codé à part au U+10FB. Ce caractère, indication visuelle de la fin d'un paragraphe, doit être suivi d'un caractère de passage à la ligne pour mettre fin à ce paragraphe. (Voir *Rapport technique Unicode n° 13, Unicode Newline Guidelines*.)

Autres signes de ponctuation. Le géorgien utilise le U+0589 ։ POINT ARMÉNIEN ou U+002F / BARRE OBLIQUE pour indiquer le point final. Pour des signes de ponctuation supplémentaires utilisés par cette écriture, consulter *Commandes C0 et ponctuation ASCII* (U+0000..U+007F) et *Ponctuation générale* (U+2000..U+206F).

8.6 Runes

Rune : U+16A0 – U+16F0

Les runes sont les caractères dont se sont servies les tribus germaniques avant leur conversion au christianisme. On en retrouve, sous diverses formes, de l'Autriche à la Scandinavie en passant par l'Angleterre et la Frise. Quelque 6000 inscriptions datant du I^er^ au XIX^e^ siècle ont été répertoriées. Elles constituent une source d'information essentielle sur le développement des langues germaniques.

Écriture historique. L'écriture runique est l'une des premières écritures « historiques » ou « mortes » à être incorporée au standard Unicode. Aujourd'hui, la seule utilisation importante des runes réside dans les travaux de recherche, universitaires ou amateurs, relatifs à ces anciennes inscriptions et à leur interprétation. L'écriture runique illustre plusieurs problèmes techniques typiques de ce genre d'écriture. Contrairement à d'autres écritures vivantes du standard Unicode qui desservent principalement les besoins de communautés d'utilisateurs modernes – avec d'occasionnelles incursions vers des formes historiques –, le codage de l'écriture runique tente de répondre à des besoins soulevés par des textes écrits à différentes époques par des sociétés isolées les unes des autres[4].

Directionalité. À l'instar d'autres écritures anciennes, les runes peuvent s'écrire de gauche à droite (*dextrograde*), de droite à gauche (*sinistrograde*), dans une direction puis dans l'autre (*boustrophédon*), ou encore en un ruban qui suit le contour de l'objet portant l'inscription. Quelques fois, les caractères apparaissent culbutés, réfléchis ou les deux à la fois. Dans les textes universitaires modernes, les runes s'écrivent de gauche à droite. Unicode attribue donc implicitement aux textes runiques cette directionalité.

L'alphabet runique. Notre connaissance actuelle des runes demeure incomplète. L'ensemble des graphèmes runiques retrouvés exhibe une variation de forme plus importante que celle des écritures modernes. La forme et le nombre de runes ont été modifiés à plusieurs reprises. Si on peut faire remonter la plupart des runes à une capitale latine de même apparence, ces lettres n'ont pas nécessairement la même prononciation. L'ordre alphabétique des runes diffère considérablement de celui de l'alphabet latin.

L'alphabet runique est connu sous le nom *futhark*, d'après la valeur de ces six premières lettres. Le premier *futhark* contenait 24 runes :

ᚠ ᚢ ᚦ ᚨ ᚱ ᚲ ᚷ ᚹ ᚺ ᚾ ᛁ ᛃ ᛇ ᛈ ᛉ ᛊ ᛏ ᛒ ᛖ ᛗ ᛚ ᛜ ᛞ ᛟ

On les translittère généralement de la façon suivante :

f u þ a r k g w h n i j é p z s t b e m l ŋ d o

En Angleterre et en Frise, on ajouta sept runes supplémentaires du V^e^ au IX^e^ siècle. Dans les pays scandinaves, le *futhark* fut fondamentalement renouvelé au VIII^e^ siècle. Ce nouvel alphabet ne se composait plus que de 16 runes, certaines runes avaient deux formes (à ramille ou à branche). La forme à branche longue est illustrée ci-dessous :

[4] Unicode se base souvent sur la tradition typographique d'une écriture pour établir son répertoire de caractères, sa sémantique et son utilisation

ᚠ ᚢ ᚦ ᚬ ᚱ ᚲ ᚼ ᚾ ᛁ ᛅ ᛋ ᛏ ᛒ ᛘ ᛚ ᛦ

f u þ o r k h n i a s t b m l ʀ

L'usage des runes demeura vivant en Scandinavie tout au long du Moyen Âge. À cette époque, le *futhark,* sous l'influence de l'alphabet latin, s'accrut de nouvelles runes pour permettre une pleine correspondance entre les deux alphabets.

Glyphes représentatifs. Les inscriptions connues comprennent de considérables variantes de formes, à tel point qu'un amateur peut facilement méprendre une rune pour une autre. Certaines runes ne possèdent pas de forme dominante, c'est plus particulièrement le cas pour les runes introduites par le système anglo-frison et le système norrois médiéval. Lors de la transcription Unicode d'une inscription runique, il faut donc éviter de ne considérer que les glyphes de référence idéalisés figurant dans les tableaux des caractères. Il est également impératif d'identifier parmi les quatre systèmes runiques, celui auquel l'inscription appartient, et d'en connaître les variations de formes acceptées pour chaque système. On a choisi les glyphes représentatifs afin de faciliter la distinction des runes appartenant à un même système. En pratique, il est conseillé d'utiliser des polices de caractères différentes pour chaque système runique. En particulier, le glyphe de référence pour U+16C4 ᛄ LETTRE RUNIQUE GER est en fait une forme rare, car la forme la plus commune est déjà utilisée pour U+16E1 ᛇ LETTRE RUNIQUE IOR.

Unifications. Lorsqu'une rune, issue d'un système d'écriture plus ancien, a évolué pour se scinder en plusieurs runes différentes, l'unification de l'ancienne rune avec les runes récentes se fonde sur la ressemblance de graphiques plutôt que la proximité phonétique.Quand un changement radical de la forme graphique s'est produit, l'unification n'a pas été tentée, même si la continuité historique demeure indiscutable. On a attribué des numéros de caractère différents aux runes de même forme issues de systèmes d'écriture différents et de prononciations différentes.

Longue branche et ramille. Neuf des seize runes du système norrois médiéval (l'époque « Viking ») utilisaient deux formes graphiques radicalement différentes : celle à longue branche et celle à ramille. Bien qu'en règle générale une seule forme se présente dans une même inscription, il existe d'importantes exceptions. Les deux formes ont été parfois utilisées par la suite à différentes fins. C'est pourquoi les deux formes ont été séparées dans le standard Unicode.

Runes sans hampe. Les runes sans hampe constituent la troisième forme de runes norroises, une sorte de sténographie runique. On n'en connaît qu'un nombre limité d'occurrences, les formes graphiques de ces runes variant considérablement d'une inscription à l'autre. C'est pourquoi on a unifié les runes sans hampe avec les autres runes norroises correspondantes. On utilisera donc les runes norroises correspondantes pour coder ces caractères, tout spécialement les caractères à ramille lorsque les deux formes (à longue branche et à ramille) existent.

Ponctuation. Unicode n'a conservé que trois caractères de la théorie des signes de ponctuation runique, dont les nuances de sens demeurent largement inconnues, en se fondant sur les caractéristiques graphiques communes à ces signes. Tous les autres signes de ponctuation ont été unifiés avec les signes de ponctuation déjà présents dans le standard Unicode.

Nombres d'or. On a longtemps utilisé les runes dans les computs scandinaves, qu'il s'agisse de symboles représentant les lettres dominicales ou comme nombres d'or sur les bâtons-éphémérides (*rimstock* en danois, *rimur* en norrois). Pour compléter la suite des nombres d'or de 1 à 19, Unicode inclut trois runes supplémentaires. On les trouve après les signes de ponctuation.

Codage. En tout, Unicode comprend 81 caractères de l'écriture runique. Parmi ceux-ci, on compte 75 lettres runiques, 3 signes de ponctuation et trois nombres d'or. L'ordre de ces caractères suit l'ordre traditionnel du *futhark* ; les variantes et runes dérivées suivent directement leur ancêtre.

Le nom des caractères runiques correspond le plus souvent possible aux noms traditionnels, souvent multiples, de chaque rune ; ce nom se termine par la translittération latine de la rune en question.

8.7 Ogam

Ogam : U+1680 – U+169F

L'ogam est une écriture alphabétique destinée à transcrire une forme très ancienne de l'irlandais. Elle consiste en un jeu d'entailles pratiquées à partir d'une arête de la pierre qui lui sert de support. On trouve de ces inscriptions de taille monumentale en Irlande, au pays de Galles, en Écosse et sur l'île de Man. Plusieurs inscriptions écossaises demeurent indéchiffrables, il pourrait s'agir de picte plutôt que de gaélique. Il est probable que les premières inscriptions ogamiques fussent gravées dans le bois. L'ogam « classique », écrit sur des pierres monumentales, connut son apogée aux V^{e} et VI^{e} siècles de notre ère. Ces inscriptions servaient surtout de bornes et de plaques commémoratives ; les exemples les plus anciens sont gravés sur des menhirs.

À l'origine, cette écriture suivait les arêtes de la pierre sur laquelle elle était gravée. Ensuite, lorsqu'elle fut écrite sur du papier, une ligne centrale continue joua le rôle de cette arête. On appelle « scolastiques » les inscriptions gravées sur le plat d'une pierre plutôt que le long de son arête. Elles sont postérieures au VII^{e} siècle. Jusqu'au XVII^{e} siècle, il était courant de retrouver des notes manuscrites rédigées en ogam.

Structure. L'alphabet ogamique se compose de 26 caractères distincts (*feda*), dont les 20 premiers forment les caractères de base, on considère les six derniers comme complémentaires (*forfeda*). Chaque signe porte le nom d'un arbre ou arbuste, nom dont la liste initiale correspond à la valeur phonétique du signe ; c'est ainsi que le signe qui note *b* s'appelle *beithe* (« bouleau ») ou celui qui note *d* se nomme *dour* (« chêne »). L'alphabet ogamique se divise en quatre séries principales nommées *aicmi* (pluriel de *aicme*, signifiant « famille »). Chaque *aicme* porte, à son tour, le nom de son premier caractère (*Aicme Beithe, Aicme Uatha*, signifie « la famille B », « la famille H », et ainsi de suite). Les noms des caractères épousent les noms irlandais modernes, à l'exception de la mutation de *nGéadal,* voir U+168D ᚻ LETTRE D'OGAM NGÉADAL, puisque l'ISO 10646 n'admet que des majuscules dans ses noms.

Rendu. Les textes ogamiques se lisent du coin inférieur gauche de la pierre vers le haut pour redescendre du côté droit (pour de longues inscriptions). L'ogam monumental était surtout taillé de bas en haut, bien qu'il existe quelques exemples d'inscriptions bilingues irlandais-latin écrites de gauche à droite. L'ogam manuscrit adopte la direction dextrograde (gauche à droite) de l'écriture latine, les voyelles sont rendues par des traits verticaux contrairement aux coches des inscriptions gravées dans la pierre. Sur ordinateur, l'ogam doit s'afficher de gauche à droite et de bas en haut (jamais de haut en bas).

Forfeda (caractères complémentaires). Les polices utilisées dans la représentation des textes ogamiques imprimés ou manuscrits sont habituellement conçues avec une arête centrale, cette convention n'est cependant pas essentielle. Le caractère U+1680   ESPACE D'OGAM doit conserver sa chasse habituelle et être laissé vide (ne pas afficher d'œil), à la manière de notre U+0020 ESPACE. On retrouve U+169B ᚛ PLUME D'OGAM et U+169C ᚜ PLUME RÉFLÉCHIE D'OGAM en début et en fin de texte, particulièrement dans l'ogam manuscrit. Parfois, on n'emploie que la *plume d'ogam* afin d'indiquer la direction du texte.

8.8 Lettres modificatives

Lettres modificatives avec chasse : U+02B0 – U+02FF

Les lettres modificatives forment un jeu de petits signes qui, en règle générale, indiquent des modifications apportées à la lettre précédente. Certains de ces signes peuvent modifier la lettre suivante, d'autres peuvent à l'occasion servir de lettres à part entière.

Contrairement aux diacritiques, les lettres modificatives *chassent*. Celles-ci se distinguent des signes de ponctuation ou des symboles d'apparence proche ou identique par leur insécabilité (on ne peut couper un mot avant ou après un de ces caractères). Les lettres modificatives possèdent la propriété « lettre » (voir *Chapitre 4, Propriétés des caractères*). La plupart de ces signes sont des lettres modificatives phonétiques ; ils comprennent ceux nécessaires à l'alphabet phonétique international (API).

Usage phonétique. Les lettres modificatives ont une interprétation phonétique relativement bien définie. Elles modifient généralement la prononciation d'un son représenté par une autre lettre ou apportent une nuance de ton ou d'accentuation (*stress*). En phonétique, on appelle parfois ces signes modificatifs des « diacritiques » puisqu'ils modifient la lettre précédente. Par contre, Unicode et l'ISO 10646 réservent le terme « signe diacritique » aux caractères à chasse nulle, tandis que les codes de ce bloc spécifient des caractères qui chassent. La *Section 15.1, Liste des noms de caractères* indique les signes diacritiques Unicode auxquels correspondent les lettres modificatives de ce bloc.

Principes de codage. Certains des caractères de ce bloc peuvent avoir plusieurs sens selon le contexte. Ce bloc comprend également plusieurs caractères qui représentent une même valeur sémantique. Il n'existe donc pas de bijection entre sens et code. Unicode ne tente pas de clarifier ces variantes d'utilisation ; il fournit néanmoins au développeur un jeu de formes à partir desquelles choisir. Ainsi, le coup de glotte (*hamza*) est-il représenté dans les translittérations latines par les caractères U+02BC ʼ LETTRE MODIFICATIVE APOSTROPHE, U+02BE ʾ LETTRE MODIFICATIVE DEMI-ANNEAU À DROITE ou U+02C0 ʔ LETTRE MODIFICATIVE COUP DE GLOTTE. Réciproquement, une apostrophe peut jouer plusieurs rôles : voir U+02BC ʼ LETTRE MODIFICATIVE APOSTROPHE dans la liste des noms de caractères. La liste des emplois associée à chaque lettre modificative n'est pas exhaustive. Dans certains cas, les lettres modificatives API ont exactement la même signification que les formes de diacritiques à chasse nulle API. Elles ne diffèrent alors que par leur chasse.

Lettres latines suscrites. Certaines lettres phonétiques modificatives sont surélevées ou suscrites, d'autres surbaissées ou souscrites, tandis que les autres sont centrées verticalement. Unicode ne code que les formes utilisées par l'API et d'autres systèmes phonétiques importants.

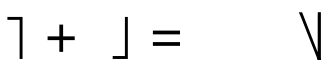
Clone à chasse des diacritiques. Certains standards propriétaires codent les mêmes signes diacritiques sous deux formes : avec ou sans chasse. Au besoin, Unicode alloue des numéros de caractère pour ces deux formes. Un certain nombre de formes à chasse est repris dans les blocs du latin de base et du supplément latin-1. Les six diacritiques européens courants qui n'y sont pas codés figurent ici sous leur forme à chasse. Ces formes peuvent appartenir à des champs sémantiques multiples, comme U+02D9 ˙ POINT EN CHEF, cinquième ton en chinois mandarin.

Crochet de rhotacisme. Le caractère U+02DE ˞ LETTRE MODIFICATIVE CROCHET DE RHOTACISME est défini dans l'API comme une lettre modificative à part entière. Toutefois, on le rencontre habituellement sous la forme d'un crochet rattaché à une lettre de base. La suite

U+0259 ə LETTRE MINUSCULE LATINE SCHWA + U+02DE ˞ LETTRE MODIFICATIVE CROCHET DE RHOTACISME peut donc être considérée comme équivalente à U+025A ɚ LETTRE MINUSCULE LATINE SCHWA CROCHET.

Signes de ton. L'intervalle U+02E5..U+02E9 comprend une série de signes de ton de base définis dans l'API et couramment utilisés dans les transcriptions tonales précises des langues africaines notamment. Chaque signe de ton renvoie à un des cinq niveaux identifiables de ton. Pour représenter un contour tonal, on assemble les signes de ton de base. La *Figure 8-4* illustre un exemple de ces tons de contour, ceux-ci sont soumis à un ensemble de règles de ligature qu'Unicode ne précise pas. Les glyphes de contour, considérés comme des ligatures, ne font pas partie d'Unicode.

Figure 8-4. Signes et contour de ton



8.9 Diacritiques

Diacritiques : U+0300 – U+036F

Les signes diacritiques de ce bloc peuvent s'utiliser avec n'importe quelle écriture. Les signes diacritiques propres à une écriture sont codés dans le bloc correspondant à cette écriture. Les signes diacritiques utilisés habituellement avec des symboles sont définis dans le bloc *Diacritiques destinés aux symboles* (U+20D0..U+20FF).

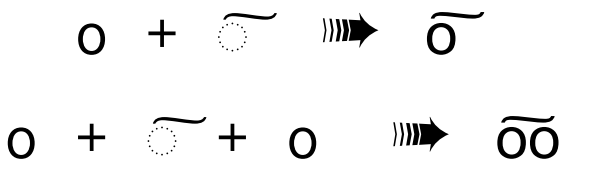
Normes. Les signes diacritiques sont dérivés de multiples sources, parmi lesquelles l'API, l'ISO 5426 et l'ISO 6937.

Ordre des diacritiques par rapport à la lettre de base. Dans le codage de caractères Unicode, tout signe à chasse nulle, y compris les diacritiques, suit le caractère de base. La suite des caractères Unicode U+0061 a LETTRE MINUSCULE LATINE A + U+0308 ◌̈ DIACRITIQUE TRÉMA + U+0075 u LETTRE MINUSCULE LATINE U représente donc sans ambiguïté « äu » et non « aü ».

Cette convention, qui consiste à placer les diacritiques à la suite des caractères de base auxquels ils se rapportent, est conforme à l'ordre logique des autres signes à chasse nulle de la plupart des écritures sémitiques et indiennes. Elle se conforme également à la manière dont les polices de caractères modernes rendent les glyphes à chasse nulle, ce qui simplifie la mise en correspondance de ces caractères. (Pour plus d'information sur l'utilisation des signes diacritiques, voir le *Chapitre 2, Structure générale,* et le *Chapitre 3, Conformité*).

Diacritique chapeautant deux caractères de base. L'API et quelques langues comme le tagalog (ou tagal) utilisent des diacritiques qui s'adjoignent à deux caractères de base. Ces signes s'appliquent au caractère de base qui les précède – comme tout autre signe à chasse nulle –, tout en chapeautant la lettre suivante. La *Figure 8-5* illustre la manière dont doivent s'afficher les deux caractères U+0360 ◌͠ DIACRITIQUE DOUBLE TILDE et U+0361 ◌͡ DIACRITIQUE DOUBLE BRÈVE RENVERSÉE.

Figure 8-5. Diacritiques doubles



Ces diacritiques doubles se lient toujours de façon moins serrée que tout autre signe à chasse nulle, à l'exception de U+0345 ◌ͅ *iota souscrit*, ils se trient près de la fin dans les représentations canoniques. Le diacritique double s'affiche au-dessus des autres diacritiques (à l'exception des diacritiques englobants), comme l'illustre la *Figure 8-6.*

Figure 8-6. Positionnement des doubles diacritiques

Afficher les diacritiques à la façon des lettres modificatives. Par convention, on peut afficher un signe combinatoire de façon isolée en l'adjoignant à U+0020 ESPACE ou U+00A0 ESPACE INSÉCABLE. Cette méthode peut s'utiliser lorsque, par exemple, on souhaite parler du signe diacritique en tant que tel, plutôt que de l'adjoindre comme d'habitude à un caractère de base. L'utilisation de U+0020 ESPACE, par opposition à U+00A0 ESPACE INSÉCABLE, influence le comportement de coupures de lignes.

Dans les tableaux et les illustrations de cette norme, la nature combinatoire de ces signes est illustrée en leur adjoignant U+25CC ◌ CERCLE EN POINTILLÉ.

Unicode code séparément les clones des signes diacritiques européens les plus communs sous la forme de caractères à chasse. Les renvois de la liste des noms de caractères (*Chapitre 15*) lient ces caractères apparentés.

Principes de codage. La grande variété d'emplois des caractères de ce bloc explique leur polysémie potentielle. Ainsi, U+0308 = tréma = *umlaut* = *dérivé double*. À l'inverse, plusieurs caractères Unicode peuvent avoir le même sens : les variantes de la cédille regroupent au moins U+0312 ◌̓ DIACRITIQUE VIRGULE CULBUTÉE EN CHEF, U+0326 ◌̦ DIACRITIQUE VIRGULE SOUSCRITE et U+0327 ◌̧ DIACRITIQUE CÉDILLE. (Pour plus d'information sur les différences entre les signes à chasse nulle et diacritiques, voir *Chapitre 2, Structure générale*.)

Variation glyphique. Rendus dans le contexte d'une langue ou d'une écriture particulière, les diacritiques puvent être, comme toute autre lettre, sujets à des variations stylistiques systématiques. C'est ainsi qu'en polonais, U+0301 ◌́ DIACRITIQUE ACCENT AIGU apparaît plus incliné qu'en français. En grec, l'accent aigu (*oxia*) est parfois presque vertical. U+030C ◌̌ DIACRITIQUE CARON est souvent rendu comme une apostrophe avec certaines formes de lettre. Adjoint à un g minuscule, U+0326 ◌̦ DIACRITIQUE VIRGULE SOUSCRITE s'affiche parfois sous la forme de U+0312 ◌̓ DIACRITIQUE VIRGULE CULBUTÉE EN CHEF, afin d'éviter les conflits avec sa boucle inférieure. Plusieurs polices de caractères ne distinguent pas nettement le U+0326 ◌̦ DIACRITIQUE VIRGULE SOUSCRITE de U+0327 ◌̧ DIACRITIQUE CÉDILLE.

Les accents combinatoires placés au-dessus des glyphes de base sont habituellement ajustés en hauteur selon la casse du glyphe. En l'absence de protocole destiné aux polices, les diacritiques sont souvent conçus pour s'appliquer aux caractères de base typiques de cette police.

Pour plus d'information, voir la *Section 5.14, Rendu des signes à chasse nulle.*

Diacritiques destinés aux symboles : U+20D0 – U+20FF

Les signes diacritiques destinés aux symboles sont généralement jumelés à des symboles mathématiques ou techniques. Ils peuvent être utilisés pour étendre la série des symboles. U+20D3 ◌⃓ DIACRITIQUE LIGNE VERTICALE COURTE COUVRANTE peut, par exemple, exprimer la négation. Dans ce cas-là, son œil peut s'allonger ou s'incliner. Ainsi, U+2261 ≡ IDENTIQUE À suivi de U+20D3 ◌⃓ DIACRITIQUE LIGNE VERTICALE COURTE COUVRANTE est équivalent à U+2262 ≢ NON IDENTIQUE À. Ici, Unicode prévoit déjà une forme précomposée du symbole de négation. Ce n'est pas toujours vrai, car U+20D3 peut servir à exprimer la négation d'autres symboles. Ainsi, U+2258 ≘ CORRESPOND À suivi de U+20D3 ◌⃓ DIACRITIQUE LIGNE VERTICALE COURTE

COUVRANTE peut-il être utilisé pour exprimer *ne correspond pas à*, sans que la forme précomposée ne fasse partie d'Unicode.

Il est à peine besoin de mentionner que d'autres caractères à chasse nulle peuvent s'utiliser dans des expressions mathématiques. U+0304 ◌̄ DIACRITIQUE MACRON est de la sorte couramment utilisé en calcul propositionnel pour désigner la négation logique.

Diacritiques englobants. Ces caractères à chasse nulle sont fournis à des fins de compatibilité avec les normes existantes ; ils permettent de ceindre de diverses façons les caractères de base simples. U+2460 ① CHIFFRE UN CERCLÉ peut, par exemple, être exprimé comme U+0031 1 CHIFFRE UN + U+20DD ◌⃝ DIACRITIQUE CERCLE ENGLOBANT. Comme pour les autres diacritiques, ceux-ci peuvent produire des formes non précomposées (la composition est ouverte); c'est ainsi qu'on obtient la lettre *alef cerclée* à l'aide de la suite U+05D0 א LETTRE HÉBRAÏQUE ALEF + U+20DD ◌⃝ DIACRITIQUE CERCLE ENGLOBANT. Les diacritiques englobants ne peuvent être utilisés pour englober plusieurs caractères de base à la fois dans les *textes bruts*. Ainsi, NOMBRE ONZE n'étant pas un caractère simple, il est impossible de représenter U+246A ⑪ NOMBRE ONZE CERCLÉ à l'aide d'un CERCLE ENGLOBANT sans faire appel à un protocole de niveau supérieur.

Demi-signes diacritiques : U+FE20 – U+FE2F

Ce bloc reprend une série de formes de présentation (glyphes) qui peuvent être utilisées pour représenter des signes diacritiques adjoints à plusieurs lettres de base. Le but de ces caractères est de faciliter la prise en charge de diacritiques hérités de mises en œuvre antérieures.

Contrairement à d'autres caractères de compatibilité, ces caractères ne correspondent ni à un caractère de référence simple ni même à une suite de caractères de référence ; une suite discontinue de ces moitiés de diacritique correspond à un diacritique complet, comme l'illustre la *Figure 8-7*. Unicode recommande toutefois l'utilisation des diacritiques doubles (U+0360 et U+0361).

Figure 8-7. Demi-signes diacritiques

Demi-signe diacritique

Signe diacritique complet

8.10 Italique

Italique : U+10300 – U+1032F

L'écriture italique[5] unifie un certain nombre d'alphabets historiques apparentés et originaires de la péninsule italienne. Quelques-uns furent utilisés pour écrire des langues non indo-européennes (l'étrusque et, probablement, le picénien septentrional) d'autres servirent à transcrire des langues indo-européennes appartenant au rameau italique (le falisque et les langues membres du groupe sabellique parmi lesquelles l'osque, l'ombrien et le picénien méridional). Ces alphabets de l'Italie ancienne remontent tous au grec d'Eubée utilisé à Ischia et à Cumes dans la baie de Naples au VIII^{e} siècle av. J.-C. Malheureusement, aucun abécédaire grec du sud de l'Italie n'a survécu. Le falisque, l'osque, l'ombrien, le picénien septentrional et méridional sont tous dérivés de la forme étrusque de l'alphabet grec.

Il existe des dizaines de milliers d'inscriptions étrusques. Dès le VIII^{e} siècle av. J.-C, époque à laquelle remonte les inscriptions les plus anciennes, des variations locales apparaissent dans l'alphabet. On distingue trois variations styliques importantes : l'étrusque du Nord, l'étrusque du Sud et celui de Caere/Veii. L'évolution de l'étrusque, liée principalement à des changements phonologiques, se divise en deux époques : l'alphabet étrusque archaïque, utilisé du VII^{e} au V^{e} siècle avant J.-C., et l'alphabet néo-étrusque, utilisé du IV^{e} au I^{er} siècle av. J.-C. Les glyphes de huit lettres diffèrent d'une époque à l'autre ; le néo-étrusque avait par surcroît abandonné les lettres KA, KU et EKS.

L'unification de ces alphabets en une seule écriture italique implique l'utilisation de polices de caractères propres aux différentes langues unifiées, car les glyphes varient quelque peu en fonction de la langue.

La plupart de langues ont ajouté quelques caractères au répertoire commun : l'étrusque et le falisque ont adjoint la LETTRE EF ; l'osque, la LETTRE EF, la LETTRE Í et la LETTRE Ú ; l'ombrien, la LETTRE EF, la LETTRE ERSE, et la LETTRE ÇÉ ; le picénien septentrional, la LETTRE Ú et l'adriatique, la LETTRE Í et la LETTRE Ú.

L'écriture latine remonte à une écriture étrusque méridionale, probablement originaire de Caere ou de Veii, vers le milieu du VII^{e} siècle av. J.-C. Cependant les différences de forme, de directionalité et de répertoire présentes entre le latin et le falisque des VII^{e} et VI^{e} siècles av. J.-C. justifient leur codage dans des blocs de caractères différents. Les polices de caractères destinées à représenter le latin archaïque doivent utiliser les points de code correspondant aux majuscules (U+0041..U+005A). L'écriture alpine unifiée, regroupant le vénète, le rhétique, le lépontique et le gaulois cisalpin, ne fait pas encore partie d'Unicode ; on considère qu'elle diffère suffisamment de l'italique pour mériter un bloc séparé. On postule que l'écriture alpine est la source des runes (U+16A0..U+16FF).

Les noms des caractères italiques ne sont pas attestés, ils sont le résultat d'une reconstitution effectuée à partir d'une étude menée à bien par Geoffrey Sampson[6]. Les noms des caractères grecs (*alpha, bêta, gamma*…) furent empruntés directement au phénicien et adaptés à la phonologie grecque. Les Étrusques, par contre, auraient abandonné les noms grecs au profit d'une nomenclature phonétique. Alors que les occlusives se prononçaient avec un *é* prolongé, les liquides ou les sibilantes (qui peuvent se prononcer plus ou moins seules) étaient précédées d'un son *è* (ainsi [k] et [d] vinrent à s'appeler [ke:], [de:], alors que [l:] et [m:] portèrent les noms de [ɛl], [ɛm].) Ces noms auraient été empruntés par les Romains lorsqu'ils adoptèrent l'écriture étrusque.

[5] À ne pas confondre avec les caractères italiques, penchés, d'Alde Manuce.
[6] W. Schulze avait déjà émis une hypothèse similaire en 1904 pour le nom des lettres latines.

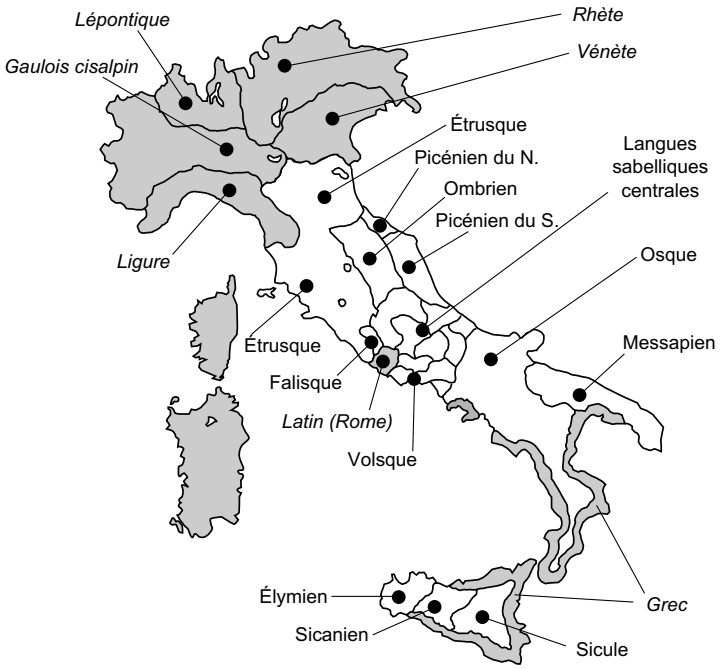
Directionalité. La plupart des textes étrusques se lisent de droite à gauche, ils sont donc sinistrogrades. À partir du III^e^ siècle av. J.-C., des textes écrits de gauche à droite firent leur apparition, démontrant une influence latine. L'osque, l'ombrien et le falisque ont aussi généralement une directionalité droite-à-gauche. Le boustrophédon n'apparaît que rarement et relativement tard (les inscriptions du Forum, par exemple, datent de 550 à 500 av. J.-C.). Malgré ce fait, pour des raisons de simplicité dans la mise en œuvre, la plupart des philologues préfèrent écrire ces textes de gauche à droite, car il s'agit également de la direction utilisée quand ces textes sont transcrits en écriture latine. L'écriture italique possède donc une directionalité implicite dextrograde (de gauche à droite) pour Unicode. Rendu de droite à gauche, on affiche les glyphes sous une forme réfléchie (*spéculaire*) par rapport aux glyphes de référence des tableaux de caractères.

Ponctuation. Les inscriptions les plus anciennes sont écrites en *scriptio continua,* c'est-à-dire sans espaces entre les mots. Il existe plusieurs inscriptions étrusques où des points servent à séparer les mots, et ce depuis la moitié du VII^e^ siècle av. J.-C. Cette ponctuation sert parfois, mais rarement, à séparer les syllabes plutôt que les mots. À partir du VI^e^ siècle av. J.-C., les mots se séparaient souvent par un, deux voire trois points superposés.

Numération. Les chiffres étrusques ne sont pas bien attestés dans les textes à notre disposition, mais ils s'employaient de la même façon que les chiffres romains. On retrouve également d'autres chiffres, mais leur utilisation demeure incertaine. Ils ne font pas encore partie du standard Unicode.

Glyphes. Les glyphes de référence utilisés dans les tableaux de caractères représentent les formes les plus fréquentes de chaque lettre. La plupart d'entre eux ressemblent aux lettres de l'abécédaire de Marsiliana (milieu du VII^e^ siècle av. J.-C.). Remarquez les valeurs phonétiques de U+10317 𐌗 LETTRE ITALIQUE IKS [ks] et U+10319 𐌙 LETTRE ITALIQUE KHÉ [χ] qui trahissent l'influence du grec occidental d'Eubée alors que le grec oriental associe plutôt à des glyphes similaires les lettres U+03A7 Χ LETTRE MAJUSCULE GRECQUE KHI [χ] et U+03A8 Ψ LETTRE MAJUSCULE GRECQUE PSI [ps].

Figure 8-8. Écritures italiques



La *Figure 8-8*, ci-dessus, illustre la distribution géographique de l'écriture italique. Le blanc représente l'aire de distribution approximative des langues anciennes utilisant les alphabets italiques. Le gris symbolise les régions où d'autres écritures dominaient, le nom de ces langues est écrit en *italique*. Il est à noter que les anciennes colonies grecques de l'Italie méridionales et des côtes siciliennes (la « Grande Grèce ») s'écrivaient à l'aide de l'écriture grecque. Les langues septentrionales, comme le ligure ou le vénète, utilisaient des variantes de l'écriture alpine. Rome apparaît en gris, puisque le latin est codé dans un bloc séparé, le bloc latin.

8.11 Gotique

Gotique : U+10330 – U+1034F

L'évêque des Goths Wulfila (311-383 apr. J.-C.) créa l'écriture gotique au IV^e^ siècle dans le but de fournir à son peuple une langue écrite utile, entre autres choses, pour lire sa propre traduction de la Bible. Les écrits en gotique se limitent principalement à des fragments de traduction de la Bible faite par Wulfila ; ces textes ont par ailleurs une grande importance pour l'étude des textes néotestamentaires. Le *Codex Argenteus* ou « livre argenté », conservé à Uppsala, constitue le principal manuscrit ; il est partiellement écrit en feuilles d'or posées sur du parchemin mauve. Le gotique est le seul témoin écrit de la branche germanique orientale. Sa disparition confère aux textes gotiques une importance considérable en linguistique historique et comparative. Il semble que Wulfila s'inspira de l'écriture grecque, comme le démontre l'ordre alphabétique de base. Certaines formes des caractères trahissent une influence runique ou latine, bien que cela puisse n'être que le fruit d'une pure coïncidence.

Diacritiques. Le U+0308 ◌̈ DIACRITIQUE TRÉMA s'adjoint parfois à la dixième lettre U+10339 LETTRE GOTIQUE I au début d'un mot, d'une syllabe ou comme deuxième membre dans des composés verbaux, comme l'illustre l'exemple ci-dessous :

Figure 8-9. Utilisation du tréma en gotique

ï ? ï ? ï ï

swe gameliþ ïst ïn esaïïn praufetau

comme il est écrit dans Isaïe le prophète

U+0305 ◌̅ DIACRITIQUE TIRET HAUT indique la contraction ou l'omission de lettres.

Numération. Les lettres gotiques, comme celles des premiers alphabets occidentaux, peuvent servir de lettres numérales ; deux caractères (90 et 900) ne servent que de chiffres. Pour indiquer la valeur numérale d'une lettre, on place de chaque côté de la celle-ci un U+00B7 · POINT MÉDIAN ou, encore, on fait suivre la lettre de U+0304 ◌̄ DIACRITIQUE MACRON et de U+0331 ◌̲ DIACRITIQUE MACRON SOUSCRIT. (Voir la *Figure 8-10*, ci-dessous.)

Figure 8-10. Lettres numérales gotiques

· · ·e· ou **𐌵̲̅** e̲̅ = 5

Ponctuation. Les manuscrits sont écrits en *scripto continua*, c'est-à-dire sans espaces entre les mots. En règle générale, on sépare les phrases ou les propositions à l'aide de U+0020 ESPACE, U+00B7 · POINT MÉDIAN ou de U+003A : DEUX-POINTS.