

Chapitre 10

Écritures du Sud et du Sud-Est asiatiques

Ce groupe comprend les écritures suivantes :

- la dévanâgarî ;
- le bengâli ;
- le gourmoukhî ;
- le goudjarâti ;
- l'oriyâ
- le tamoul ;
- le télougou ;
- le kannara ;
- le malayâlam ;
- le singhalais ;
- le thaï ;
- le lao ;
- le tibétain ;
- le birman ;
- le khmer.

Les écritures du Sud et du Sud-Est asiatiques partagent un tel nombre de caractéristiques qu'une comparaison directe entre quelques-unes de ces écritures révèle souvent des similitudes structurelles, et ce même dans les formes modernes de leurs lettres. Hormis de rares exceptions historiques, elles s'écrivent de gauche à droite. Plusieurs d'entre elles séparent les phrases à l'aide d'espaces ou de signes particuliers, mais ne séparent pas les mots. Il s'agit d'écritures essentiellement alphabétiques dont la plupart des symboles correspondent à la fois à une consonne et à une voyelle inhérente (souvent le son « a » bref). Dans la plupart de ces écritures, des symboles spéciaux représentent les voyelles initiales, mais, de manière générale, les voyelles internes à un mot s'écrivent en juxtaposant un signe-voyelle à la consonne afférente. L'absence de voyelle inhérente, quand cela se produit, est fréquemment signalée par un glyphe spécial.

De l'Himalaya au nord jusqu'à l'île de Sri Lanka au sud, du Pakistan à l'ouest jusqu'aux plus lointaines îles indonésiennes à l'est, la plupart des écritures du Sud et du Sud-Est asiatiques sont dérivées de l'ancienne écriture brâhmî. Les plus vieilles inscriptions indiennes, et les plus longues, les édits de l'empereur Açoka (Aśoka) remontant au III^e siècle av. J.-C., sont rédigées à l'aide de deux écritures : la kharochthî et la brâhmî. Toutes deux sont d'origine sémitique, probablement dérivées de l'araméen, une importante langue administrative du Moyen-Orient à l'époque. La kharochthî, écrite de droite à gauche, fut supplantée par la brâhmî et ses écritures dérivées. Tout en se modifiant, les écritures descendant de la brâhmî se répandirent dans l'ensemble du sous-continent indien et des îles voisines. Plus de 200 écritures en dériveraient. Au XI^e siècle, l'écriture moderne connue sous le nom de dévanâgarî vint à dominer la littérature sanscrite en Inde. Elle représente l'archétype septentrional de cette famille d'écritures qui regroupe également des écritures modernes comme le birman, le gourmoukhî et le tibétain. La branche méridionale brâhmî inclut, parmi d'autres écritures, le malayalam et le tamoul.

Les principales écritures de l'Inde, y compris la¹ dévanâgarî, épousent une même structure de codage, de sorte que les caractères apparentés des différentes écritures se trouvent dans le même ordre et aux mêmes positions relatives. Cette structure qui, dans une certaine mesure, facilite la translittération entre les langues indiennes s'inspire de la Norme nationale indienne (ISCII) qui code ces écritures. Le singhalais, le birman et le khmer se servent également du modèle brahmî à *virâma* mais sans correspondance avec la norme ISCII. Le thaï et le lao s'appuient sur un autre modèle inspiré de la Norme industrielle thaïe pour le codage du thaï, la structure de ce modèle qui se fonde sur un ordre visuel n'est pas compatible avec celle de l'ISCII. Enfin, reflet de sa structure et de son usage quelque peu différents, le tibétain se démarque de tous ces modèles de codage. De nombreux noms de caractères de ce groupe d'écritures représentent les mêmes sons ; ces écritures utilisent souvent des règles d'affectation de noms similaires.

¹ On remarquera le genre particulier de trois noms d'écriture qui, contrairement à toutes les autres, sont du féminin : kharochthî, brâhmî et dévanâgarî.

10.1 Dévanâgarî

Dévanâgarî : U+0900 – U+097F

Le sanscrit classique ainsi que son descendant contemporain, le hindî, s'écrivent à l'aide de l'écriture dévanâgarî. D'autres langues indiennes apparentées au hindî, comme le marathe (marâthî) et le népalais (népalî), s'écrivent à l'aide de la dévanâgarî et de quelques signes supplémentaires. Les langues suivantes utilisent aussi l'écriture dévanâgarî : awadhi (ou aoudi), bagheli, bhatneri, bhili, bihari, bradj bhasa, djaïpouri, garhwali, gond (betoul, tchhindwara et dialectes mandla), haraouti, ho, katchtchhi, kanaoudji, konkani, kouloui, koumaoni, kourkou, kouroukh, marvari, mouna, néwari, palpa, santâl et tchhattisgarhi.

Toutes les autres écritures de l'Inde, de même que le singhalais de Sri Lanka, l'écriture tibétaine et les écritures du Sud-Est asiatique (thaï, lao, khmer et birman) sont apparentées à l'écriture dévanâgarî, car toutes dérivent de la brâhmî. Les écritures de cette famille partagent un grand nombre de ressemblances structurelles.

Cette introduction à la dévanâgarî expose en détail les principes des écritures indiennes. L'introduction à chacune des autres écritures indiennes est abrégée, elle signale toutefois, le cas échéant, les différences par rapport à la dévanâgarî.

Normes. Le bloc dévanâgarî d'Unicode s'inspire de la norme ISCII:1988 (*Indian Standard Code for Information Interchange*). La norme ISCII de 1988 représente une évolution de ses versions précédentes publiées en 1983 et en 1986. Unicode dispose les caractères dévanâgarî aux mêmes positions relatives que les caractères A0₁₆.F4₁₆ de la norme ISCII. Les autres écritures du sous-continent indien agencent leurs caractères de la même manière. Cette similarité dans l'agencement des codes permet de faire ressortir les ressemblances structurelles des écritures brâhmî et respectent les principes, énoncés par le bureau indien des normes, qui consistent à permettre l'aller-retour entre des positions analogues de code parmi les écritures indiennes. La structure du singhalais, du thaï, du lao, du khmer et du birman diffère davantage de celle de la dévanâgarî ; c'est pourquoi Unicode ne tente pas de fournir de correspondances directes entre ces écritures et la dévanâgarî.

En novembre 1991, au moment de la publication de la version 1.0 du standard Unicode, le bureau indien des normes publiait une nouvelle version de la norme ISCII (IS 13194:1991). Cette version modifiait en partie l'agencement et le répertoire de la norme ISCII:1988. Depuis lors, le standard Unicode ne respecte plus exactement l'agencement de la version actuelle de l'ISCII. Toutefois, Unicode demeure un sur-ensemble du répertoire de l'ISCII:1991, hormis quelques caractères védiques² ajoutés par l'IS 13194:1991, annexe G, *Extended Character Set for Vedic*. Les textes modernes, autres que védiques, codés à l'aide de l'ISCII:1991 peuvent donc être convertis automatiquement en valeurs de code Unicode et reconvertis en leurs valeurs de code d'origine, sans perte d'information.

Principes de codage. Les systèmes d'écritures indiens forment des systèmes hybrides, croisement entre systèmes syllabiques et alphabétiques. L'unité réelle de ces systèmes d'écriture est la tranche syllabique (également appelée syllabe graphique) formée à la base d'une consonne et d'une voyelle (CV). Ce noyau peut être précédé d'une ou plusieurs consonnes pour former la structure canonique [[C]C]CV. La tranche ne doit pas nécessairement correspondre de façon exacte à la syllabe phonémique, particulièrement dans le cas de groupes consonantiques, mais ce système d'écriture est conçu à partir de principes phonologiques et il tend à respecter d'assez près la prononciation. Chaque tranche est

² Une forme archaïque du sanscrit.

construite à partir d'éléments alphabétiques, les caractères de la dévanâgarî. Il existe trois types distincts de caractères : les lettres-consonnes, les voyelles indépendantes et les signes-voyelles dépendants. L'ordre visuel de ces caractères respecte, en règle générale, l'ordre logique (phonétique).

Principes d'écriture

Apparence. La caractéristique la plus visible de la dévanâgarî est la grande barre horizontale, la « potence », qui unit non seulement les syllabes, mais souvent les mots, au-dessous de laquelle viennent se pendre les caractères, à l'exception de quelques signes-voyelles qui la surmontent.

Rendu des caractères dévanâgarî. Les caractères dévanâgarî, comme les caractères de nombreuses écritures, peuvent se combiner ou changer de forme en fonction du contexte. L'apparence d'un caractère change en fonction des caractères voisins, de la police utilisée pour le rendre³ ainsi que du logiciel ou du système d'exploitation employé. Pour ces différentes raisons, l'aspect d'un caractère dévanâgarî peut différer de l'aspect du glyphe de référence utilisé dans les tableaux de caractères.

En outre, quelques caractères dévanâgarî modifient l'ordre des caractères affichés. Ceci se produit rarement dans les écritures non indiennes. Cette modification est indépendante d'un réarrangement bidirectionnel qui pourrait s'avérer nécessaire.

Lettres-consonnes. Chaque lettre-consonne représente une consonne simple, mais on lui associe implicitement une voyelle inhérente, le plus souvent la voyelle brève /a/. Ainsi, U+0915 क LETTRE DÉVANÂGARÎ KA ne représente pas uniquement /k/, mais bien /ka/. Cependant, en présence d'une voyelle dépendante, la voyelle inhérente est remplacée par cette voyelle dépendante.

Une lettre-consonne peut également afficher sous sa demi-forme. Elle prend cette apparence quand elle forme le début de groupe consonantique. Les consonnes mi-formées ne possèdent pas de voyelles inhérentes. En dévanâgarî, elles ressemblent souvent aux lettres-consonnes pleinement formées si ce n'est que manque la hampe qui, habituellement, marque le noyau syllabique. Le trait de la hampe est apparenté, en ce qui concerne le dessin et son origine, au signe représentant la voyelle inhérente /a/.

Certaines lettres-consonnes dévanâgarî adoptent plusieurs formes de présentation (œils) différentes selon les consonnes adjacentes. Ceci est particulièrement vrai pour U+0930 र LETTRE DÉVANÂGARÎ RA dont la forme varie grandement tant comme consonne initiale que finale. Seule la forme de référence de chacune des lettres-consonnes, et non ses variantes contextuelles, est illustrée dans le tableau des caractères.

L'ordre alphabétique traditionnel des consonnes sanscrites (dévanâgarî) respecte les principes de la phonétique articulatoire. L'ISCI ainsi que la norme Unicode suivent cet ordre traditionnel⁴.

³ Terme typographique utilisé pour désigner le fait d'afficher, d'imprimer, etc.

⁴ Voir, par exemple, le tableau de la dévanâgarî (U+0901..U+097F). Dans l'ordre alphabétique traditionnel, on trouve d'abord les voyelles, les palatales (TCHA) succèdent aux gutturales (KA), puis viennent les rétroflexes (TTA), les dentales (TA) et les labiales (PA). Pour les consonnes, cet ordre répond aux situations respectives des organes en commençant par le plus intérieur, pour finir par le plus extérieur. Chaque série articulatoire comprend 4 consonnes outre la nasale. Les deux premières sont sourdes, c'est-à-dire sans résonance des cordes vocales (PA), les suivantes sont sonores, c'est-à-dire accompagnées d'une résonance des cordes vocales (BA). Pour chaque série couplée (sonore/sourde) il existe une aspirée (PHA) et une non aspirée (PA). Viennent ensuite les semi-voyelles (YA, RA, LA, VA), les spirantes (ÇA, CHA, SA) et enfin l'aspiration (HA).

Lettres-voyelles indépendantes. En dévanâgarî, les voyelles indépendantes se suffisent à elles-mêmes. Cette écriture considère les voyelles indépendantes comme des syllabes CV où la consonne est nulle. Ces voyelles servent à noter les syllabes commençant par une voyelle.

Signes des voyelles dépendantes (matras). Dans la pratique courante, les voyelles dépendantes servent à écrire les voyelles explicites ; on les nomme souvent *signes-voyelles* ou *matras* en sanscrit. Les voyelles dépendantes sont toujours adjointes à une lettre de base. La voyelle dépendante s'adjoint à une consonne ou à un groupe consonantique pour indiquer la qualité vocalique de la tranche syllabique quand celle-ci ne correspond pas à la voyelle inhérente. L'utilisation explicite d'une voyelle dépendante entraîne la disparition de la voyelle inhérente d'une seule lettre-consonne.

La plus grande différence entre les écritures indiennes réside dans la manière dont les voyelles dépendantes s'adjoignent aux lettres de base. La dévanâgarî emploie un grand nombre de signes de voyelles dépendantes. Certaines sont à chasse nulle, en chef ou souscrites, les autres chassent à droite ou à gauche de la lettre-consonne ou du groupe consonantique. En général, les autres écritures indiennes possèdent une ou plusieurs de ces formes ; un même signe pourra chasser dans une écriture alors qu'il s'agira d'un diacritique dans une autre. Il existe, de surcroît, des écritures indiennes dont les voyelles dépendantes s'écrivent en plusieurs glyphes séparés. Ces glyphes peuvent entourer une lettre-consonne verticalement ou horizontalement.

La dévanâgarî ne possède qu'un signe-voyelle (matra) s'écrivant à gauche : U+093F ि VOYELLE DIACRITIQUE DÉVANÂGARÎ I. Pour les autres écritures indiennes, ou de tels signes voyelles n'existent pas (télougou et kannara), ou il en existe jusqu'à quatre (khmer).⁵

Il existe une correspondance bijective (aller-retour) entre les voyelles indépendantes et les signes-voyelles dépendants. À l'occasion, on représente une voyelle indépendante à l'aide une tranche composée de la forme indépendante de la voyelle /a/ suivie de la matra correspondant à la voyelle indépendante. La *Figure 10-1* illustre cette relation (voir la notation formelle décrite dans les « Règles de rendu », ci-après dans cette section).

Figure 10-1. Voyelles dépendantes et indépendantes

/a/ + voyelle dépendante					voyelle indépendante	
A_n	+	I_{vd}	→	I_{vd} + A_n	≈	I_n
अ	+	ि	→	अि	≈	इ
A_n	+	O_U_{vd}	→	A_n + O_U_{vd}	≈	O_U_n
अ	+	ु	→	अु	≈	उ

La conjonction de la forme indépendante de la voyelle sous-entendue /a/ (pour la dévanâgarî, U+0905 अ LETTRE DÉVANÂGARÎ A) et d'une matra peut être considérée comme une autre manière de représenter l'information phonétique habituellement notée par la forme isolée d'une voyelle indépendante. Pour ce qui est du rendu, par contre, il ne faut pas considérer ces deux représentations comme équivalentes. Il se peut que certains processus de textes de niveau supérieur, considérant ces graphies au niveau du contenu

⁵ Voir la classe antéposée du *Tableau 4.-3, Classes combinatoires*. Le bengali, le tamoul, le malayâlam et le cinghalais, par exemple, ont chacun trois voyelles antéposées.

informationnel, les traitent comme équivalentes. Unicode, cependant, ne stipule pas cette équivalence.

Virâma. L'écriture dévanâgarî, ainsi que d'autres écritures indiennes, emploie un signe nommé virâma⁶. Adjoint à une consonne, le virâma (par exemple, U+094D ् SIGNE DÉVANÂGARÎ VIRÂMA) élimine la voyelle inhérente. Il se comporte comme un diacritique dont la forme varie selon le système d'écriture. Lorsqu'une consonne perd sa voyelle inhérente à la suite de l'adjonction d'un virâma, elle prend le nom de *consonne dévoyellée* (morte ou encore nue) ; inversement, une *consonne voyellée* (ou vive) conserve sa voyelle inhérente ou s'écrit à l'aide d'une matra. Unicode définit une consonne dévoyellée comme une lettre-consonne suivie d'un virâma. Par défaut, on représente une consonne dévoyellée sous la forme de la consonne vive correspondante à laquelle on adjoint, à la manière d'un diacritique, un virâma.

Soit C_n la forme générique (ou nominale) de la consonne C et soit C_d sa forme dévoyellée (nue), on code alors la consonne nue sous la forme illustrée à la Figure 10-2.

Figure 10-2. Consonnes dévoyellées (ou « mortes »)

$$\begin{array}{ccccccc} T A_n & + & V I R \hat{A} M A_n & \rightarrow & T A_d \\ त & + & ् & \rightarrow & त् \end{array}$$

Conjointes consonantiques. Les écritures indiennes sont célèbres pour le grand nombre de conjointes consonantiques qu'elles emploient ; il s'agit de ligatures représentant plusieurs lettres adjacentes au sein d'une même tranche. Une tranche (ou groupe consonantique) correspond à une suite de caractères représentant une lettre-voyelle indépendante ou une ou plusieurs consonnes mortes (C_d) suivies d'une consonne vive C_v .

D'ordinaire, on représente cette tranche à l'aide d'un glyphe conjoint, si ce glyphe existe dans les polices de caractères actives. Si le glyphe conjoint est absent, la ou les consonnes nues faisant partie de la tranche sont alors affichées sous leur demi-forme. En cas d'absence de la demi-forme, les consonnes dévoyellées sont représentées sous leur forme référence à laquelle on adjoint un virâma visible (voir Figure 10-3).

Figure 10-3. Formation de conjointes

$$\begin{array}{ll} (1) \quad G A_d + D H A_v \rightarrow G A_m + D H A_n & (3) \quad K A_d + C H A_v \rightarrow K.C H A_n \\ ग् + ध \rightarrow गध & क् + ष \rightarrow कष \\ (2) \quad K A_d + K A_v \rightarrow K.K A_n & (4) \quad R A_d + R I_n \rightarrow R I_n + R A_{sus} \\ क् + क \rightarrow क्क & र् + ऋ \rightarrow र्ऋ \end{array}$$

Les exemples suivants illustrent différents genres de conjointes :

- (1) un GA mi-formé soudé à la forme complète du DHA ;
- (2) la conjointe verticale K.KA ;
- (3) la conjointe pleinement liée K.CHA dans laquelle on ne distingue plus les composantes et
- (4) une conjointe rare formée en partie d'une lettre-voyelle indépendante, dans ce cas, la lettre-voyelle RI (également appelée *r vocalique*).

⁶ On l'appelle également halant, signe de quiescence ou de dévoyellement

Remarquons que, dans l'exemple (4), la consonne dévoyellée RA_d est représentée à l'aide d'un diacritique suscrit RA_{sus} (*répha*).

Une police de caractères indiens bien conçue contient des centaines de conjointes; Unicode ne code toutefois pas ces glyphes puisqu'ils ne sont que le résultat d'une ligature de lettres distinctes. En fonction du contexte, le logiciel de rendu des écritures indiennes doit être en mesure d'associer les combinaisons de caractères aux glyphes conjoints appropriés présents dans les polices de caractères.

Lorsqu'une voyelle indépendante apparaît en fin de groupe consonantique, comme c'est le cas dans l'exemple (4) de la *Figure 10-3*, la voyelle indépendante doit prendre la forme d'une voyelle indépendante et non d'une matra.

Virâma explicite. Les virâma servent d'ordinaire à dévoyeller des consonnes qui, à leur tour, se joignent à la consonne suivante pour former une conjointe. Une fois conjointe, la consonne ne fait plus apparaître de virâma. À l'occasion, il se peut, par contre, qu'il faille que la consonne nue reste exclue de la formation d'une conjointe, auquel cas le signe virâma doit rester visible. Pour ce faire, Unicode place le caractère U+200C ANTI LIANT SANS CHASSE immédiatement après la consonne dévoyellée à exclure de la conjointe. Le signe virâma apparaît alors correctement sous la consonne à laquelle il est adjoint.

Par exemple, dans la *Figure 10-4*, le signe antiliant sans chasse inhibe la formation implicite de la forme conjointe $क्ष(K.CHAn)$.

Figure 10-4. Éviter la conjonction

$$KA_d + ALSC + CHA_v \rightarrow KA_d + CHA_n$$

$$क् + \text{[ZW]} + ष \rightarrow क्ष$$

Demi-consonnes explicites. Lorsqu'une consonne morte participe à la formation d'une conjointe, la forme de la consonne dévoyellée est souvent absorbée dans la forme conjointe de telle sorte qu'elle n'est plus distinctement visible. Ailleurs, toutefois, la consonne dévoyellée peut rester visible sous la forme d'une demi-consonne. En général, une forme de demi-consonne se distingue de sa forme de référence par la perte de la hampe de la voyelle inhérente; une hampe apparaît, en effet, à la droite des consonnes de référence en dévanâgarî. Dans d'autres cas, la hampe demeure mais une partie de son côté droit est alors amputée.

Il est parfois souhaitable que la consonne dévoyellée ne prenne ni une forme complètement conjointe ni que le virâma apparaisse. On utilise alors la demi-forme de la consonne. Pour coder explicitement cette demi-forme, le standard Unicode adopte la convention suivante: on place le caractère U+200D LIANT SANS CHASSE immédiatement après la consonne dévoyellée codée. Le LIANT SANS CHASSE représente une lettre sans œil qui se lie (sous forme cursive ou ligaturée) de part et d'autre (c'est-à-dire à la lettre précédente ou suivante). Ainsi, dans ce contexte, le signe LIANT SANS CHASSE permet à la consonne morte de se lier pour prendre son aspect mi-formé.

Ainsi, si C_m dénote un glyphe mi-formé de la consonne C, la *Figure 10-5* illustre comment on code la consonne mi-formée.

Figure 10-5. Demi-consonnes

$$KA_d + LSC + CHA_v \rightarrow KA_m + CHA_n$$

$$क + \boxed{ZW} + ष \rightarrow कष$$

- En l'absence de signe LIANT SANS CHASSE, cette suite produit normalement la forme conjointe pleine क्ष (K.CHA_n).

Ce codage des consonnes mi-formées s'applique de la même façon en l'absence d'une forme de lettre de base. En d'autres mots, cette technique peut s'utiliser pour coder des demi-formes indépendantes (voir la Figure 10-6).

Formes consonantiques. En résumé, chaque consonne peut être codée de sorte qu'elle apparaît sous la forme d'une consonne vive, d'une consonne dévoyellée fusionnée au sein d'une conjointe ou d'une consonne dévoyellée mi-formée (voir Figure 10-7).

Figure 10-6. Demi-formes indépendantes

$$GA_d + LSC \rightarrow GA_m$$

$$ग + \boxed{ZW} \rightarrow ङ$$

Figure 10-7. Formes consonantiques

क	→	क	KA _v
क + क्	→	क	KA _d
क + क् + \boxed{ZW}	→	क	KA _m

Rendu

Règles de rendu. Cette section fournit des règles plus formelles et plus détaillées qui permettent d'afficher correctement de la dévanâgarî dans une suite de texte brut. Elles décrivent la correspondance qui existe entre les caractères Unicode et les glyphs d'une police de caractères dévanâgarî typique ainsi que la combinaison et l'ordre de ces glyphs.

Ces règles représentent l'ensemble minimal de règles pour rendre correctement un texte dévanâgarî échangé. Comme pour toute autre écriture, un processus plus complexe peut y ajouter des fonctionnalités de rendu en fonction de la police de caractères utilisée et de l'application.

Il est important de remarquer qu'une police capable d'afficher la dévanâgarî comporte plus de glyphs qu'Unicode ne propose de caractères dévanâgarî.

Notation. Les règles de cette section utilisent la notation suivante.

C _n	Œil de référence (« glyphe nominal ») de la consonne C, c'est donc la forme que cette consonne adopte dans les tableaux de caractères.
C _v	Consonne voyellée ou vive, identique au rendu à C _n .
C _d	Glyphe représentant la forme dévoyellée ou morte de la consonne C.

C_m	Glyphe représentant la demi-forme de la consonne C.
L_n	Forme de référence de la ligature composée de deux ou plusieurs consonnes. Une ligature conjointe formée des deux consonnes X et Y s'écrit également X.Y _n .
RA_{sus}	Forme du diacritique U+0930 र̣ LETTRE DÉVANÂGARÎ RA suscrite ou adjointe à la partie supérieure du glyphe de base. On nomme également cette forme <i>répha</i> .
RA_{sous}	Forme du glyphe du diacritique à chasse nulle souscrit ou attaché à la partie inférieure de la base de la forme du glyphe de U+0930 र̣ LETTRE DÉVANÂGARÎ RA.
V_{vd}	Glyphe illustrant la forme dépendante de la voyelle V.
VIRÂMA	Forme de référence du glyphe du diacritique illustrant U+094D ॠ SYMBOLE DÉVANÂGARÎ VIRÂMA.

Le caractère virâma n'est pas toujours visible ; lorsqu'il l'est, il prend la forme d'un signe à chasse nulle.

Règle de dévoyement. En toute logique, cette règle précède toutes les autres dans la formation des consonnes dévoyées. Une fois formée, la consonne dévoyée peut faire l'objet des autres règles décrites ci-dessous.

R1 *Toute consonne C_n qui précède un VIRÂMA est considérée comme une consonne morte (dévoyée). Une consonne C_n qui n'est pas suivie d'un VIRÂMA est considérée comme une consonne vive (ou voyellée) C_v .*

$$TA_n + VIRÂMA \rightarrow TA_d$$

$$\bar{त} + ॠ \rightarrow \bar{त}$$

Règles relatives à la consonne RA. Le caractère U+0930 र̣ LETTRE DÉVANÂGARÎ RA peut prendre une parmi plusieurs formes visuelles possibles, selon le contexte de la tranche où il apparaît. Par défaut, cette lettre est représentée à l'aide de son glyphe de référence (voir les tableaux de caractères). Dans certains cas, on le représente sous la forme d'un glyphe à chasse nulle qui se combine à une forme de lettre de base.

R2 *Si une consonne morte RA_d précède une consonne ou une voyelle indépendante, on la remplace par un signe RA_{sus} à chasse nulle, on le suscrit de telle sorte qu'il s'applique à l'élément logique suivant en mémoire.*

$$RA_d + KA_v \rightarrow KA_v + RA_{sus} \quad \text{Résultat affiché}$$

$$\bar{र} + क \rightarrow क + ॠ \rightarrow \bar{क}$$

$$RA^1_d + RA^2_d \rightarrow RA^2_d + RA^1_{sus}$$

$$\bar{र} + \bar{र} \rightarrow \bar{र} + ॠ \rightarrow \bar{र}$$

R3 *Quand on adjoint le symbole suscrit RA_{sus} à une consonne morte qui forme une conjointe ligaturée avec une autre consonne, ce symbole se place de manière à s'appliquer à toute la conjointe.*

$$\begin{array}{l} \text{RA}_d + \text{DJA}_d + \tilde{\text{NA}}_v \rightarrow \text{DJ.}\tilde{\text{NA}}_n + \text{RA}_{\text{sus}} \quad \text{Résultat} \\ \text{र्} + \text{ज्} + \text{ञ} \rightarrow \text{ज्ञ} + \text{ँ} \rightarrow \text{ज्ञँ} \\ \text{affiché} \end{array}$$

R4 Quand on adjoint le symbole suscrit RA_{sus} à une consonne morte qui sera remplacée par une demi-consonne, ce symbole se place de sorte qu'il s'applique à la forme servant de base à la tranche consonantique.

$$\begin{array}{l} \text{RA}_d + \text{GA}_d + \text{GHA}_v \rightarrow \text{GA}_m + \text{GHA}_v + \text{RA}_{\text{sus}} \quad \text{Résultat} \\ \text{र्} + \text{ग्} + \text{घ} \rightarrow \text{र्ग} + \text{घ} + \text{ँ} \rightarrow \text{र्घ} \\ \text{affiché} \end{array}$$

R5 Conformément à la norme ISCII, on représente la consonne mi-formée RRA_m à l'aide d'un RA-cil. On emploie souvent cette forme de RA pour écrire le marathe.

$$\begin{array}{l} \text{RRA}_n + \text{VIRĀMA} \rightarrow \text{RRA}_m \\ \text{र्} + \text{ँ} \rightarrow \text{ँ} \end{array}$$

R5a À des fins de compatibilité avec la version 2.0 du standard Unicode, si une consonne RA_d précède un LIANT SANS CHASSE, la consonne mi-formée RA_m prend alors les traits d'un RA-cil plutôt que RA_{sus} .

$$\begin{array}{l} \text{RA}_d + \text{LSC} \rightarrow \text{RA}_m \\ \text{र्} + \text{ꣳ} \rightarrow \text{ँ} \end{array}$$

R6 Sauf pour la consonne morte RA_d , lorsqu'une consonne morte C_d précède la consonne vive RA_v , on remplace C_d par sa forme de référence C_n et le signe à chasse nulle souscrit RA_{sous} remplace le RA, ce signe se place de sorte qu'il s'applique à C_n .

$$\begin{array}{l} \text{TTHA}_d + \text{RA}_v \rightarrow \text{TTHA}_d + \text{RA}_{\text{sous}} \quad \text{Résultat} \\ \text{ठ्} + \text{र} \rightarrow \text{ठ्} + \text{ँ} \rightarrow \text{ठ्र} \\ \text{affiché} \end{array}$$

R7 Pour certaines consonnes, le signe RA_{sous} peut se confondre graphiquement à la consonne pour former une ligature conjointe. Ces conjointes, comme celles illustrées ici, font également l'objet de règles de ligature décrites ci-après.

$$\begin{array}{l} \text{PHA}_d + \text{RA}_v \rightarrow \text{PHA}_d + \text{RA}_{\text{sous}} \quad \text{Résultat} \\ \text{क्} + \text{र} \rightarrow \text{क} + \text{ँ} \rightarrow \text{क्र} \\ \text{affiché} \end{array}$$

R8 Quand une consonne dévoyellée (autre que RA_d) précède RA_d alors RA_{sous} remplace bien RA comme décrit ci-dessus; toutefois, le VIRĀMA responsable de la formation du RA_d demeure pour former une conjointe morte.

$$\begin{array}{l} \text{TA}_d + \text{RA}_d \rightarrow \text{TA}_n + \text{RA}_{\text{sous}} + \text{VIRĀMA} \quad \text{Résultat} \\ \text{त्} + \text{र्} \rightarrow \text{त} + \text{ँ} + \text{्} \rightarrow \text{त्र्} \\ \text{affiché} \end{array}$$

Une forme conjointe dévoyellée contenant un RA_d absorbé peut, par la suite, se fusionner à une ou plusieurs autres lettres pour former une tranche conjointe.

$$\begin{array}{l} \text{T.RA}_{\mathbf{d}} + \text{YA}_{\mathbf{v}} \rightarrow \text{T.R.YA}_{\mathbf{n}} \\ \text{त्र्} + \text{य} \rightarrow \text{त्र्य} \end{array}$$

Règles des signes modificateurs. En plus des signes voyelles, on peut adjoindre trois autres types de diacritiques (le noukta, le bindou et le svara) à un élément d'une tranche ou à la tranche entière.

R9 Le signe noukta modifie les consonnes, il se place immédiatement après la consonne dans la représentation en mémoire et s'y attache au rendu. Si la consonne représente une consonne dévoyellée, alors le NOUKTA devrait précéder le VIRÂMA dans la représentation en mémoire.

$$\begin{array}{l} \text{KA}_{\mathbf{n}} + \text{NOUKTA} + \text{VIRÂMA} \rightarrow \text{QA}_{\mathbf{d}} \\ \text{क} + \text{◌̣} + \text{◌̣̣} \rightarrow \text{क्} \end{array}$$

R10 Les autres signes modificateurs, comme les bindous et les svaras, s'adjoignent à la tranche complète et devraient suivre (dans la représentation en mémoire) tous les caractères constituant la syllabe. Les bindous devraient notamment suivre les signes voyelles et les svaras venir en dernier. La position relative de ces signes est plutôt horizontale que verticale; des considérations typographiques peuvent influencer le rendu horizontal.

$$\begin{array}{l} \text{KA}_{\mathbf{n}} + \hat{\text{A}}_{\mathbf{vd}} + \text{TCHANDRABINDOU}_{\mathbf{n}} \\ \text{क} + \text{◌ा} + \text{◌̣̣̣} \rightarrow \text{काँ} \end{array}$$

Règles de ligature. Une fois les règles précédentes appliquées, un autre ensemble de règles régit la formation de ligatures. Une application précise de ces règles dépend de la disponibilité de glyphes dans les polices de caractères utilisées pour afficher le texte.

R11 Si une consonne dévoyellée précède directement une autre consonne morte ou vive, la première consonne dévoyellée peut se joindre à l'élément suivant pour former une ligature conjointe à deux parties.

$$\begin{array}{l} \text{DJA}_{\mathbf{d}} + \text{ÑA}_{\mathbf{v}} \rightarrow \text{DJ.ÑA}_{\mathbf{n}} \quad \text{TTA}_{\mathbf{d}} + \text{TTHA}_{\mathbf{v}} \rightarrow \text{TT.TTHA}_{\mathbf{n}} \\ \text{ज्} + \text{ञ} \rightarrow \text{ज्ञ} \quad \text{ट्} + \text{ठ} \rightarrow \text{ट्ठ} \end{array}$$

R12 Une ligature conjointe peut se comporter comme une consonne dévoyellée et faire partie de ligatures plus complexes.

$$\begin{array}{l} \text{SA}_{\mathbf{d}} + \text{TA}_{\mathbf{d}} + \text{RA}_{\mathbf{n}} \rightarrow \text{SA}_{\mathbf{d}} + \text{T.RA}_{\mathbf{n}} \rightarrow \text{S.T.RA}_{\mathbf{n}} \\ \text{स्} + \text{त्} + \text{र} \rightarrow \text{स्} + \text{त्र} \rightarrow \text{स्त्र} \end{array}$$

Une ligature conjointe peut également produire une demi-forme.

$$\begin{array}{l} \text{K.CHAd} + \text{YAv} \rightarrow \text{K.CHm} + \text{YAn} \\ \text{क्ष} + \text{य} \rightarrow \text{क्षय} \end{array}$$

R13 Si la forme de référence d'une consonne ou une ligature conjointe précèdent RA_{sous} , à la suite de l'application de la règle R2, alors la consonne ou la ligature se joignent à RA_{sous} pour former une ligature conjointe complexe (pour plus d'informations, voir la règle R2).

$$\begin{array}{l} \text{KA}_n + \text{RA}_{\text{sous}} \rightarrow \text{K.RA}_n \quad \text{PHA}_n + \text{RA}_{\text{sous}} \rightarrow \text{PH.RA}_n \\ \text{क} + \text{◌} \rightarrow \text{क्र} \quad \text{फ} + \text{◌} \rightarrow \text{फ्र} \end{array}$$

R14 Dans certains cas, d'autres diacritiques se lient à une consonne de base pour s'attacher à un endroit inhabituel ou changer de forme. Si l'on considère des rendus élémentaires, il n'existe que deux cas : RA_v avec OU_{vd} ou OÛ_{vd} .

$$\begin{array}{l} \text{RA}_v + \text{OU}_{vd} \rightarrow \text{ROU}_n \quad \text{RA}_v + \text{OÛ}_{vd} \rightarrow \text{ROÛ}_n \\ \text{र} + \text{◌} \rightarrow \text{रु} \quad \text{र} + \text{◌} \rightarrow \text{रू} \end{array}$$

Représentation en mémoire et ordre de rendu. On stocke, en règle générale, les textes bruts écrits en dévanâgarî ou toute autre écriture indienne dans l'ordre phonétique. En d'autres mots, une syllabe CV composée d'une voyelle dépendante se stocke toujours en mémoire de la manière suivante : consonne C suivie du signe-voyelle V. Cette convention, préconisée par la norme ISCII, correspond à la fois à l'ordre phonétique et à l'ordre de saisie de données textuelles (voir la Figure 10-8).

Figure 10-8. Ordre de rendu

Ordre des caractères (ordre logique)	Ordre des glyphes (ordre visuel)
$\text{KA}_n + \text{I}_{vd}$	$\text{I}_{vd} + \text{KA}_n$
क + ि	डि

Certaines voyelles dépendantes, en dévanâgarî, comme dans d'autres écritures indiennes, s'affichant à la gauche de leur lettre-consonne, le logiciel de rendu de ces écritures doit être en mesure de réorganiser ces éléments et d'assurer la correspondance entre l'ordre logique (caractère) et l'ordre visuel (glyphe). Ainsi, si C_n représente la forme de référence d'une consonne C et V_{vd} une forme dépendante de la voyelle V placée à gauche, alors la remise en ordre des glyphes, par rapport aux caractères codés, s'effectue comme l'illustre la Figure 10-8.

R15 Quand la voyelle dépendante I_{vd} sert à remplacer la voyelle inhérente d'une tranche, elle se code toujours à l'extrême gauche de la tranche syllabique. Si la tranche contient un groupe consonantique, la voyelle inhérente s'affiche toujours à gauche de ce groupe. Exemple :

$$\begin{array}{l} \text{TA}_d + \text{RA}_v + \text{I}_{vd} \rightarrow \text{T.RA}_n + \text{I}_{vd} \rightarrow \text{I}_{vd} + \text{T.RA}_d \\ \text{त्} + \text{र} + \text{ि} \rightarrow \text{त्र} + \text{ि} \rightarrow \text{त्रि} \end{array}$$

Exemples de demi-formes. Le Tableau 10-1 illustre un certain nombre de formes de demi-consonnes d'un emploi fréquent en dévanâgarî. Ces formes sont des glyphes et non des caractères. Comme il est illustré ci-dessous, elles peuvent être codées explicitement à l'aide d'un LIANT SANS CHASSE ; lors de la formation normale de conjointes, on peut les utiliser librement pour représenter une consonne dévoyellée jointe à des consonnes suivantes.

Tableau 10-1. Exemples de demi-formes

क	◌	ZW J	व	न	◌	ZW J	न
ख	◌	ZW J	ख	प	◌	ZW J	प
ग	◌	ZW J	ग	फ	◌	ZW J	फ
घ	◌	ZW J	घ	ब	◌	ZW J	ब
च	◌	ZW J	च	थ	◌	ZW J	थ
ज	◌	ZW J	ज	म	◌	ZW J	म
झ	◌	ZW J	झ	य	◌	ZW J	य
ञ	◌	ZW J	ञ	ल	◌	ZW J	ल
ण	◌	ZW J	ण	व	◌	ZW J	व
त	◌	ZW J	त	श	◌	ZW J	श
थ	◌	ZW J	थ	ष	◌	ZW J	ष
द	◌	ZW J	द	स	◌	ZW J	स

Exemples de ligatures. Le Tableau 10-2 illustre un certain nombre de formes conjointes de ligatures utilisées couramment en dévanâgarî. Ces formes sont des glyphes et non des caractères. Tout système employant cette écriture n'emploie pas nécessairement toutes ces formes ; en particulier, plusieurs de ces formes ne s'utilisent qu'en sanscrit. En outre, certaines polices de caractères offrent plus de ligatures que celles reprises ci-dessous, d'autres moins⁷.

Tableau 10-2. Exemples de ligatures

क	◌	क	क्क	ट	◌	ठ	ट्ठ
क	◌	त	क्त	ठ	◌	ठ	ठ्ठ
क	◌	र	क्र	ड	◌	ग	ड्ग
क	◌	ष	क्ष	ड	◌	ड	ड्ड
ड	◌	क	ड्क	ड	◌	ढ	ड्ढ
ड	◌	ख	ड्ख	त	◌	त	त्त
ड	◌	ग	ड्ग	त	◌	र	त्र
ज	◌	ज	ज्ज	न	◌	न	न्न
ज	◌	ञ	ज्ञ	फ	◌	र	फ्र

⁷ Voir la note de bas de page n° 27 de ce chapitre.

द	◌̣	घ	ङ
द	◌̣	द	ढ
द	◌̣	ध	ञ
द	◌̣	म	ञ
द	◌̣	य	य
द	◌̣	व	व
ट	◌̣	ट	ट

श	◌̣	र	श्र
ह	◌̣	म	ह्र
ह	◌̣	य	ह्य
ह		ृ	हृ
र		ु	रु
र		ू	रू
स	◌̣	त्र	स्त्र

Exemples de demi-ligatures. En plus des demi-formes de consonnes individuelles, on utilise également des demi-formes de ligatures conjointes. Le *Tableau 10-3* en présente un certain nombre. Ces formes sont des glyphes et non des caractères. Comme le tableau l'illustre, elles se codent de manière explicite à l'aide d'un LIANT SANS CHASSE ; lors de formation normale de conjointes, elles s'utilisent librement pour représenter une ligature jointe aux consonnes suivantes.

Tableau 10-3. Exemples de demi-ligatures

क	◌̣	ष	◌̣	◌̣ZWJ	क्ष
ज	◌̣	ञ	◌̣	◌̣ZWJ	ज्ञ
त	◌̣	त	◌̣	◌̣ZWJ	त्त
त	◌̣	र	◌̣	◌̣ZWJ	त्र
श	◌̣	र	◌̣	◌̣ZWJ	श्र

Signes combinatoires. La dévanâgarî ainsi que d'autres écritures indiennes se servent d'un certain nombre de signes combinatoires que l'on peut considérer comme des diacritiques. Les bindous, une parmi les classes de signes combinatoires, se représentent à l'aide de U+0901 ◌̣ SYMBOLE DÉVANÂGARÎ TCHANDRABINDOU ou de U+0902 ◌̣ SYMBOLE DÉVANÂGARÎ ANOUSVÂRA. Ils indiquent une nasalisation ou la terminaison nasale d'une syllabe. U+093C ◌̣ symbole dévanâgarî noukta est un véritable diacritique. Il modifie les consonnes (à l'aide d'un point dévanâgarî suscrit) et sert, ainsi, à créer de nouvelles lettres à partir du jeu de lettres-consonnes de base. Les caractères U+0951...U+0954 forment un ensemble de signes combinatoires utilisés dans la transcription du sanscrit.

Chiffres. Chaque écriture indienne possède ses propres formes de chiffres. Il se peut que ces chiffres ne soient pas utilisés dans des textes ordinaires. De nos jours, les chiffres européens ont, pour beaucoup de ces écritures, remplacé les formes indiennes. Certaines écritures indiennes, notamment le tamoul, ne connaissent pas le chiffre zéro.

Ponctuation et symboles. U+0964 । DANNDDA DÉVANÂGARÎ est, par sa fonction, semblable à notre point ; en poésie, il marque la fin de la demi-stance. De nombreuses autres écritures indiennes possèdent une forme correspondante à ce signe. Dans les textes traditionnels, U+0965 ॥ DOUBLE DANNDDA DÉVANÂGARÎ marque la fin d'une stance.

Plusieurs langues modernes écrites en dévanâgarî émaillent leurs textes de signes de ponctuation latins. Ainsi, U+002C , VIRGULE et U+002E . POINT s'utilisent-ils couramment pour écrire le hindî, le dannda étant généralement réservé aux textes plus traditionnels.

Structure de codage. Chacune des neuf principales écritures indiennes correspond, pour Unicode, à un bloc de 128 points de codage. Les six premières colonnes de chaque écriture sont isomorphes à l'ISCII:1988 sauf pour les 11 dernières positions (c'est-à-dire U+0955..U+095F dans le cas de la dévanâgarî) qui ne sont pas affectées ou définies dans l'ISCII:1988, mais qui le sont dans Unicode.

La septième colonne de chaque écriture et les 11 dernières positions de la sixième colonne représentent des positions de caractères supplémentaires, harmonisées pour les neuf écritures. Les positions U+xx66..U+xx6F et U+xxE6..U+xxEF, par exemple, correspondent aux chiffres de chaque écriture indienne.

La huitième colonne de chaque écriture est réservée aux signes particuliers à chaque écriture qui ne correspondent pas d'une écriture à l'autre.

10.2 Bengali

Bengali : U+0980 – U+09FF

L'écriture bengali, originaire du nord de l'Inde, s'apparente à la dévanâgarî. On l'utilise pour écrire le bengali, essentiellement dans l'État du Bengale-Occidental et au Bangladesh. Elle s'utilise également pour écrire l'assamais dans l'État d'Assam (Inde) et d'autres langues minoritaires du nord-est de l'Inde (daphla, garo, khasi, manipouri, mizo, mouna, nâga, rian et santâl).

Signes-voyelles bipartis. Comme d'autres écritures indiennes, le bengali fait usage de signes-voyelles bipartis (par exemple, U+09CB ঔ VOYELLE DIACRITIQUE BENGALI ô et U+09CC ঔ VOYELLE DIACRITIQUE BENGALI AOU). Les deux moitiés de la voyelle se placent de part et d'autre de la consonne ou de tranche consonantique qu'elles modifient. On code les voyelles biparties à la même position que la voyelle dévanâgarî correspondante. U+09CC ঔ VOYELLE DIACRITIQUE BENGALI AOU et U+094C ঔ VOYELLE DIACRITIQUE DÉVANÂGARÎ AOU sont ainsi isomorphes. Afin d'assurer une compatibilité entre les mises en œuvre existantes des écritures à voyelles biparties, Unicode prévoit également un numéro de caractère pour la moitié droite de ces signes de voyelles : U+09D7 ঔ SIGNE DE LONGUEUR BENGALI AOU représente la moitié droite du glyphe U+09CC ঔ VOYELLE DIACRITIQUE BENGALI AOU.

Caractères spéciaux. Les caractères U+90F2..U+09F9 représentent des devises et des fractions.

Rendu. Pour rendre l'écriture bengali, consulter les règles de rendu de la *Section 10.1, Dévanâgarî*.

10.3 Gourmoukhî

Gourmoukhî : U+0A00 – U+0A7F

L'écriture gourmoukhî⁸, originaire du nord de l'Inde, dérive d'une ancienne écriture appelée lahnda (/laŋɖa/). Elle s'apparente structurellement à la dévanâgarî. Le gourmoukhî⁹ s'utilise au Pendjab (Inde) pour écrire la langue pendjabi¹⁰.

Remarque. Pour rendre l'écriture gourmoukhî, consulter les règles de rendu de la *Section 10.1, Dévanâgarî*.

Tons. Le trait le plus caractéristique du pendjabi est son système à trois tons. Le gourmoukhî ne prévoit pas de symboles particuliers pour représenter ces tons, bien qu'ils correspondent en règle générale à certains signes-consonnes. C'est la fonction principale des signes destinés aux aspirées sonores, de l'*h* et des conjointes consonantiques contenant un *h*.

⁸ Également appelée gurumukhî.

⁹ Le premier livre sacré des Sikhs, l'Âdi Granth est noté en gourmoukhî. Il consiste principalement en hymnes écrits par les cinq premiers gourous et par le neuvième.

¹⁰ Le pendjâbî (pañdjâbî) standard est assez proche de l'hindoustâni (l'hindî et l'ourdou en tant que langue). Il s'en distingue toutefois tant au niveau lexicographique que grammatical. Plus archaïque et plus irrégulier, il sonne à la fois plus dur, plus emphatique (conservation des consonnes doubles du prâkrit), plus nasal et plus chantant (présence de tons).

10.4 Goudjarâtî

Goudjarâtî : U+0A80 – U+0AFF

L'écriture goudjarâtî¹¹, originaire du nord de l'Inde, est proche parente de la dévanâgarî. Elle s'en distingue principalement par le fait que ses lettres sont dépourvues de barre horizontale (de « tête »¹²), une caractéristique de l'antique écriture *kaïthî* à laquelle le goudjarati s'apparente. Le goudjaratî sert à dénoter la langue goudjaratî parlée dans l'État Goudjerate (Inde).

Rendu. Pour rendre l'écriture goudjarati, consulter les règles de rendu de la *Section 10.1, Dévanâgarî*.

¹¹ On dit également goudjerate, goujrati ou gouzrati.

¹² On parle parfois de lettres acéphales.

10.5 Oriyâ

Oriyâ : U+0B00 – U+0B7F

L'écriture oriyâ, originaire du nord de l'Inde, s'apparente structurellement à la dévanâgarî, mais la plupart des lettres possèdent une ligne supérieure semi-circulaire et non la barre horizontale (la « potence ») typique à la dévanâgarî. La forme des lettres, en particulier des signes-voyelles, ressemble aux lettres tamoules¹³. L'écriture oriyâ sert à dénoter la langue oriyâ, parlée dans l'État indien d'Orissa, ainsi que des langues minoritaires telles que le khond et le santâl.

Caractères spéciaux. U+0B57 ଌ SIGNE DE LONGUEUR ORIYA AOU représente la moitié droite de la voyelle bipartie U+0B4C ଌ VOYELLE DIACRITIQUE ORIYA AOU

Rendu. Pour rendre l'écriture oriya, consulter les règles de rendu de la *Section 10.1, Dévanâgarî*.

¹³ Ce caractère hybride de l'écriture s'explique sans doute par l'ethnographie et l'histoire de l'Orissa, État situé aux confins de deux mondes : l'aryen et le dravidien. Cet ancien royaume au nord-est du Deccan est le dernier royaume conquis par l'empereur Açoka, lors d'une conquête particulièrement sanglante qui aurait amené Açoka à embrasser ensuite la foi bouddhique non violente. Au 11^e siècle av. J.-C., la région retombe sous la domination politique du Nord et se divise sur le plan linguistique : seul le Sud reste dravidien (de parler télougou ou andhra), tandis qu'au nord l'aryanisation l'emporte avec la constitution de l'oriyâ, langue du pays d'Orissa.

10.6 Tamoul

Tamoul : U+0B80 – U+0BFF

Le tamoul est une des écritures indiennes méridionales. Si ces écritures s'apparentent structurellement aux écritures du nord de l'Inde, elles servent à écrire les langues dravidiennes parlées au sud du sous-continent et au Sri Lanka, langues qui ne sont pas apparentées aux langues indo-aryennes du nord de l'Inde (par exemple, le hindî, le bengalî ou le goudjérate). L'aspect des lettres méridionales se différencie fortement des lettres de la dévanâgarî. Alors que la dévanâgarî apparaît carrée, les écritures méridionales ont un tracé commun : le *ductus* des caractères tend à s'arrondir, à former des courbes, des volutes. On attribue traditionnellement cette différence au fait que, à l'origine, les écritures méridionales ayant probablement été gravées sur des feuilles de palmiers à l'aide d'un poinçon de métal, les lettres auraient alors adopté un aspect sinueux pour éviter de suivre les veines des feuilles.

L'écriture tamoule sert à écrire la langue tamoule parlée dans l'état du Tamoul Nâdou en Inde ainsi que des langues minoritaires comme le Badaga. Le tamoul s'écrit également au Sri Lanka, en île Maurice, à la Réunion, à Pondichéry et dans certaines régions de la Malaisie. Cette écriture contient moins de consonnes que les autres écritures indiennes. Sauf de rares exceptions, elle ignore les consonnes conjointes. En tamoul, U+0BCD ◌̣ SYMBOLE TAMOUL VIRÂMA demeure apparent et remplace ces consonnes conjointes.

Conventions d'appellation des voyelles moyennes. Les voyelles dravidiennes *é* et *o* peuvent être longues ou brèves, le sanscrit ne distingue pas de longueur pour ces voyelles. C'est pourquoi Unicode nomme différemment les voyelles moyennes dravidiennes ([e] et [o]). Ces différences dans l'appellation apparaissent nettement quand on considère les noms de U+0B8E ஃ LETTRE TAMOULE É et U+0B8F ஄ LETTRE TAMOULE Ê¹⁴ et qu'on les compare aux positions isomorphes de la dévanâgarî : U+090E ए̄ LETTRE DÉVANÂGARÎ É BREF et U+090F ए̆ LETTRE DÉVANÂGARÎ É. Les langues dravidiennes distinguent systématiquement la longueur des voyelles moyennes, distinction inexistante pour la dévanâgarî. Le caractère U+090E ए̄ LETTRE DÉVANÂGARÎ É BREF a été ajouté à la dévanâgarî afin de pouvoir transcrire les voyelles dravidiennes brèves. Ces conventions d'appellation permettent d'identifier au mieux la véritable nature des voyelles concernées dans les écritures dravidiennes, ainsi que dans la dévanâgarî ou les autres écritures septentrionales de l'Inde.

Caractères spéciaux. Unicode prévoit un U+0BD7 ஶ SIGNÉ DE LONGUEUR TAMOUL AOU afin de coder le côté droit de la voyelle bipartie U+0BCC ஶஶ VOYELLE DIACRITIQUE TAMOULE AOU.

Rendu. Les écritures indiennes méridionales fonctionnent *grosso modo* de la même façon que la dévanâgarî, sauf pour ce qui a trait aux voyelles biparties, typiques des écritures dravidiennes. Comme dans le cas de la dévanâgarî, les termes « lettre tamoule » et « voyelle diacritique tamoule » sont omis lorsque cela ne cause pas d'ambiguïté.

Il est important de souligner que, dans une police de caractères tamouls, le jeu de glyphes est supérieur en nombre aux caractères tamouls¹⁵.

Le *Tableau 10-4* présente un aperçu des lettres tamoules.

¹⁴ Rappelons que, dans la transcription adoptée par l'ISO 10646, l'accent circonflexe indique la longueur et non l'ouverture d'une voyelle. Ê est donc un é long [e:].

¹⁵ Voir la note de bas de page n° 27 de ce chapitre.

Figure 10-4. Aperçu des lettres tamoules

க KA	ங NGA	ச TCHA	ஐ DJA	ஞ ÑA	ட TTA	ண NNA	த TA	ந NA	ன NNNA	ப PA	
ம MA	ய YA	ர RA	ற RRA	ல LA	ள LLA	ழ LLLA	வ VA	ஷ SSA	ஸ SA	ஹ HA	
அ A	ஆ Â	இ I	ஈ î	உ OU	ஊ oû	எ É	ஏ Ê	ஐ AÏ	ஓ O	ஔ ô	ஔள AOU
அ A	ஶ Â	ஶி I	ஶீ î	ஶு OU	ஶூ oû	ஶே É	ஶே Ê	ஶை AÏ	ஶொ O	ஶோ ô	ஶைள AOU
ஶ VIRÂ MA	ஶள SIGNE AOU										

Voyelles dépendantes et indépendantes. Comme pour la dévanâgarî, les signes-voyelles dépendants ne correspondent pas à une suite *virâma + voyelle indépendante*. Exemple :

$$\text{ன} + \text{ி} \neq \text{ன} + \text{ஶ} + \text{இ}$$

À l’instar de la dévanâgarî, un groupe (ou tranche) consonantique est constitué d’une suite d’une ou de plusieurs consonnes séparées de virâmas, éventuellement terminée par un virâma.

Voyelles biparties. Certaines voyelles indiennes sont constituées de deux éléments séparés qui flanquent une consonne ou une tranche consonantique. Comme pour tous les éléments séparés, il est possible d’exprimer une même graphie à l’aide de deux suites de valeurs Unicode, situation analogue à celle de la lettre « â » qu’on peut écrire à l’aide d’un « a » suivi du signe à chasse nulle « ^ » ou grâce au caractère Unicode « â ».

$$\text{ொ} \text{ (OBCA)} \approx \text{ெ} + \text{ா} \text{ (OBC6 + OBBE)}$$

$$\text{ோ} \text{ (OBCB)} \approx \text{ே} + \text{ா} \text{ (OBC7 + OBBE)}$$

$$\text{ௌ} \text{ (OBCC)} \approx \text{ெ} + \text{ள} \text{ (OBC6 + OBD7)}$$

Remarquons que le ஶள du troisième exemple n’est pas U+0BB3 ஶள LETTRE TAMOULE LLA, mais bien U+0BD7 ஶள SIGNE DE LONGUEUR TAMOUL AOU.

Si on utilise les formes précomposées en mémoire plutôt que les caractères séparés, il faut alors effectuer une transformation semblable lors du rendu. Comme l’indiquent les conversions ci-après, la forme précomposée de gauche se transforme en deux formes séparées, équivalentes à celles illustrées à droite, lesquelles sont alors sujettes à un réarrangement de voyelles lors de l’affichage.

$$\text{ொ} \rightarrow \text{ெ} + \text{ா}$$

$$\text{ோ} \rightarrow \text{ே} + \text{ா}$$

$$\text{ௌ} \rightarrow \text{ெ} + \text{ள}$$

Rendu des voyelles. Comme le *Tableau 10-5* l'illustre, les voyelles suivantes sont placées devant le groupe consonantique précédent, comportement analogue à celui du *SIGNE VOYELLE DEVANAGARI I*.

ெ (OBC6) ே (OBC7) ை (OBC8)

Tableau 10-5. Réordonnement des voyelles

Ordre des caractères (ordre logique)				Ordre des glyphes (ordre visuel)
க	+	ெ	→	கெ
க	+	ே	→	கே
க	+	ை	→	கை

Un fractionnement des voyelles produit le même effet (voir *Tableau 10-6*).

Tableau 10-6. Fractionnement et réordonnement des voyelles

Ordre des caractères (ordre logique)				Ordre des glyphes (ordre visuel)
க	+	ொ	→	கொ
க	+	ெ	+ ா	→ கொ
க	+	ோ	→	கோ
க	+	ே	+ ா	→ கோ
க	+	ெள	→	கௌ
க	+	ெ	+ ாள	→ கௌ

Dans les deux cas, l'ordre des éléments est sans ambiguïté : la consonne (la tranche) apparaît en premier dans la représentation en mémoire. La voyelle ூள possède également deux éléments séparés et peut se composer à l'aide du *SIGNE DE LONGUEUR AOU*.

Ligatures. Les exemples suivants illustrent la gamme de ligatures disponibles en tamoul. Ces changements ont lieu à la suite du réordonnement et du fractionnement des voyelles. Contrairement à la dévanâgarî, le tamoul ne comporte que de rares consonnes conjointes ; la plupart des ligatures se forment entre une voyelle et une consonne voisine.

1. Consonnes conjointes.

க + ஃ + ஷ → கஷ

Comme pour la dévanâgarî, le réordonnement des voyelles a également lieu autour des consonnes conjointes. Exemple :

க + ஃ + ஷ + ெ + ா → கௌஷா

2. La voyelle ஈ se lie facultativement aux consonnes ண, ன ou ற situées à sa gauche :

$$\begin{aligned} \text{ண} + \text{ஊ} &\rightarrow \text{ணூ} \\ \text{ன} + \text{ஊ} &\rightarrow \text{னூ} \\ \text{ற} + \text{ஊ} &\rightarrow \text{றூ} \end{aligned}$$

Ce processus se produisant après le réordonnancement et le fractionnement, les ligatures suivantes peuvent également se former.

Voyelles séparées

$$\begin{aligned} \text{ண} + \text{ஊ} + \text{ஊ} &\rightarrow \text{ணூஊ} \\ \text{ண} + \text{ஊ} + \text{ஊ} &\rightarrow \text{ணூஊ} \\ \text{ன} + \text{ஊ} + \text{ஊ} &\rightarrow \text{னூஊ} \\ \text{ன} + \text{ஊ} + \text{ஊ} &\rightarrow \text{னூஊ} \\ \text{ற} + \text{ஊ} + \text{ஊ} &\rightarrow \text{றூஊ} \\ \text{ற} + \text{ஊ} + \text{ஊ} &\rightarrow \text{றூஊ} \end{aligned}$$

Voyelles précomposées

$$\begin{aligned} \text{ண} + \text{ஊ} &\rightarrow \text{ணூ} \\ \text{ண} + \text{ஊ} &\rightarrow \text{ணூ} \\ \text{ன} + \text{ஊ} &\rightarrow \text{னூ} \\ \text{ன} + \text{ஊ} &\rightarrow \text{னூ} \\ \text{ற} + \text{ஊ} &\rightarrow \text{றூ} \\ \text{ற} + \text{ஊ} &\rightarrow \text{றூ} \end{aligned}$$

3. Les signes-voyelles U+0BBF ி VOYELLE DIACRITIQUE TAMOULE I et U+0BC0 ீ VOYELLE DIACRITIQUE TAMOULE î se lient à un U+0B9F ட LETTRE TAMOULE TTA situé à leur gauche.

$$\begin{aligned} \text{ட} + \text{ி} &\rightarrow \text{டி} \\ \text{ட} + \text{ீ} &\rightarrow \text{டீ} \end{aligned}$$

Souvent, afin de se greffer au mieux à la consonne située à leur gauche, ces voyelles changent légèrement de forme ou de position :

$$\begin{aligned} \text{ல} + \text{ி} &\rightarrow \text{லி} \\ \text{ல} + \text{ீ} &\rightarrow \text{லீ} \end{aligned}$$

4. Les signes-voyelles U+0B89 ூ LETTRE TAMOULE OU et U+0B8A ௃ LETTRE TAMOULE oû changent habituellement de forme ou se lient (voir *Tableau 10-7*).

Tableau 10-7. Voyelles liantes

x	x + ூ	x + ௃
க	கூ	கூ
ங	ஙூ	ஙூ
ச	சூ	சூ
ஞ	ஞூ	ஞூ
ட	டூ	டூ
ண	ணூ	ணூ
த	தூ	தூ
ந	நூ	நூ
ன	னூ	னூ

x	x + ூ	x + ௃
ப	பூ	பூ
ம	மூ	மூ
ய	யூ	யூ
ர	ரூ	ரூ
ற	றூ	றூ
ல	லூ	லூ
ள	ளூ	ளூ
ழ	ழூ	ழூ
வ	வூ	வூ

À la droite de ஐ, ஐ, ஸ ou ஶ, ces voyelles chassent (voir Figure 10-9).

Figure 10-9. Voyelles avec chasse

$$\begin{aligned} \text{ஐ} + \text{஁} &\rightarrow \text{ஐ஁} \\ \text{ஐ} + \text{஁ள} &\rightarrow \text{ஐ஁ள} \end{aligned}$$

5. Le signe-voyelle ஁ se change en ஁ à la gauche de ண, ன, ல ou ள.

$$\begin{aligned} \text{஁} + \text{ண} &\rightarrow \text{஁ண} \\ \text{஁} + \text{ன} &\rightarrow \text{஁ன} \\ \text{஁} + \text{ல} &\rightarrow \text{஁ல} \\ \text{஁} + \text{ள} &\rightarrow \text{஁ள} \end{aligned}$$

Remarquons que ce changement se produit après le réordonnancement de la voyelle. Dans le premier exemple, la voyelle ஁ suit ண en mémoire. Après le réordonnancement des voyelles, elle se trouve à la gauche de ண et change alors d'œil. Le processus complet est donc :

$$\text{ண} + \text{஁} \rightarrow \text{஁} + \text{ண} \rightarrow \text{஁ண}$$

6. La consonne ர prend la forme ற.

Ce changement se produit lorsque la forme ற de U+0BB0 ற LETTRE TAMOULE RA ne risque pas d'être méprise pour la forme de référence U+0BBE ஁ற VOYELLE DIACRITIQUE TAMOULE Â (par exemple, lorsque ற est combiné à ஁, ஁ி ou ஁ீ).

$$\begin{aligned} \text{ர} + \text{஁} &\rightarrow \text{ற} + \text{஁} \\ \text{ர} + \text{஁ி} &\rightarrow \text{ற} + \text{஁ி} \\ \text{ர} + \text{஁ீ} &\rightarrow \text{ற} + \text{஁ீ} \end{aligned}$$

10.7 Télougou

Télougou : U+0C00 – U+0C7F

Le télougou¹⁶, une écriture méridionale de l'Inde, sert à noter le télougou parlé dans l'État indien Andhra Pradesh et à Yanaon ainsi que des langues minoritaires comme le gond¹⁷ (dialectes adilabad et koī) et le lambadi.

Rendu. Pour rendre l'écriture télougou, consulter les règles de rendu présentées dans la Section 10.6, *Tamoul*. Il faut remarquer que, dans l'écriture télougou, contrairement au tamoul, les consonnes conjointes s'écrivent à l'aide de lettres souscrites. De nombreuses lettres-consonnes, utilisées au sein d'une conjointe, prennent des formes contextuelles. Certains signes-voyelles changent également de forme lorsqu'ils font partie de certaines combinaisons précises.

Caractères spéciaux. U+0C55 ీ SIGNE DE LONGUEUR TÉLOUGOU permet de coder le deuxième élément de la voyelle bipartite U+0C47 ీ VOYELLE DIACRITIQUE TÉLOUGOU ê. U+0C56 ు SIGNE DE LONGUEUR TÉLOUGOU aĩ permet de coder le deuxième élément de la voyelle bipartite U+0C48 ు VOYELLE DIACRITIQUE TÉLOUGOU aĩ. Ces deux signes de longueur ont une chasse nulle.

Matras. En télougou, la même matra peut se greffer différemment selon la consonne à laquelle il s'adjoint. Le logiciel de rendu doit choisir l'aspect correct de la matra et la greffer correctement à la consonne. La même matra U+0C3F ృ VOYELLE DIACRITIQUE TÉLOUGOU I sert dans les mots suivants :

క రా క రి → క్కి
 స ర్ వ రా త రి → స్వతి
 య రా మ రి న రి → యామిని

¹⁶ Langue dravidiennne également appelée télंगा.

¹⁷ On baptisera la partie méridionale de la Pangée du nom de la région du Deccan habitée par les Gonds : le Gondwana. Le Gondwana se serait par la suite divisé en de nombreux continents qui effectueront une lente migration les amenant à leur position actuelle. Ces continents constitueraient de nos jours la moitié des terres émergées : l'Amérique du Sud, l'Afrique, l'Arabie, Madagascar, l'Inde, l'Australie, la Nouvelle-Zélande et l'Antarctique.

10.8 Kannara

Kannara : U+0C80 – U+0CFF

Écriture méridionale de l'Inde, le kannara¹⁸ s'utilise pour écrire la langue kannara parlée dans l'état de Karnataka, ainsi que des langues minoritaires comme le toulou. Elle s'apparente largement à l'écriture télougou pour ce qui a trait à la forme des lettres et au comportement des consonnes conjointes.

Caractères spéciaux. U+0CD5 ೆ SIGNÉ DE LONGUEUR KANNARA permet de coder la partie droite de la voyelle séparée U+0CC7 ೆ VOYELLE DIACRITIQUE KANNARA ê. U+0CD6 ೆ SIGNÉ DE LONGUEUR KANNARA aĩ représente la partie droite de la voyelle bipartie U+0CC8 ೆ VOYELLE DIACRITIQUE KANNARA aĩ. Les voyelles séparées kannara sont constituées d'un élément à chasse nulle placé au-dessus de la lettre-consonne et d'une ou de plusieurs lettres à chasse placées à droite de la lettre-consonne.

Lettre kannara LLLA. U+0CDE ೆ LETTRE KANNARA LLLA est une lettre kannara désuète translittérée, dans les textes universitaires occidentaux, sous la forme d'un *ṛ* (un r marqué du diacritique U+0324 ೆ DIACRITIQUE TRÉMA SOUSCRIT). Cette lettre ne s'utilise plus en kannara depuis la fin du XIX^e siècle. À des fins de classement, U+0CDE doit être traité comme s'il suivait U+0CB3 ೆ LETTRE KANNARA LLA .

¹⁸ On écrit également *canara* ou *kanaḍa* ; cette langue dravidienne est parlée dans l'ancien Maïssour (Mysore) qui prit le nom de Karnātaka en 1973.

10.9 Malayâlam

Malayâlam : U+0D00 – U+0D7F

L'écriture malayâlam¹⁹, originaire du sud de l'Inde, s'utilise pour écrire la langue malayâlam parlée dans l'état du Kérala, à Mahé sur la côte de Malabar et aux îles Laquedives.

La forme des lettres malayâlam ressemble à celles du tamoul. Toutefois, le malayâlam comprend un jeu riche et complexe de formes de consonnes conjointes. Quelques-unes sont représentées ci-dessous. Comme pour les autres langues indiennes, le U+0D4D ് SYMBOLE MALAYALAM VIRÂMA permet la formation de ces conjointes.

ക	്	ക	→	ക	ക	്	ല	→	ക	
ക	്	ഷ	→	ക	ഷ	ഗ	്	ഗ	→	ഗ
ഗ	്	ല	→	ഗ	ല	ങ	്	ക	→	ക
ങ	്	ങ	→	ങ	ച	്	ച	→	ച	
ഞ	്	ച	→	ഞ	ഞ	്	ഞ	→	ഞ	
ട	്	ട	→	ട	ണ	്	ട	→	ട	
ണ	്	ണ	→	ണ	ത	്	ത	→	ത	
ത	്	ഥ	→	ത	ദ	്	ദ	→	ദ	
ദ	്	ധ	→	ദ	ന	്	ത	→	ത	
ന	്	ദ	→	ന	ന	്	ന	→	ന	
ന	്	മ	→	ന	പ	്	പ	→	പ	
പ	്	ല	→	പ	ബ	്	ബ	→	ബ	
ബ	്	ല	→	ബ	ന	്	പ	→	പ	
മ	്	മ	→	മ	മ	്	ല	→	മ	
യ	്	യ	→	യ						

Caractères spéciaux. U+0D57 ് SIGNÉ DE LONGUEUR MALAYALAM AOU permet de coder la partie droite de la voyelle bipartite U+0D4C ൌ VOYELLE DIACRITIQUE MALAYALAM AOU.

¹⁹ On écrit parfois maleiyâlam ou malayâlam.

10.10 Singhalais

Singhalais : U+0D80 – U+0DFF

Le singhalais permet d'écrire la langue singhalaise, principale langue parlée du Sri Lanka, autrefois Ceylan. Elle s'utilise également pour écrire le pâli et le sanscrit. Cette écriture émane de la brâhmî et ressemble aux écritures méridionales de l'Inde quant à la forme et à la structure. Le singhalais n'est cependant pas une langue dravidienne, mais indo-aryenne.

Le singhalais se distingue des autres langues de la région par une série d'occlusives pré-nalisées qu'il faut distinguer de la combinaison d'une nasale suivie d'une occlusive. Ces deux formes existent et s'écrivent de différentes manières – par exemple : අඬ <U+0D85, U+0DAC> /a.Nda/ « bruit » comparé à අනඬ <U+0D85, U+0DAB, U+0DCA, U+0DA9> /a.n.da/ « œuf ». En outre, le singhalais distingue deux signes correspondant aux voyelles antérieures ouvertes, brèves et longues, proches de la voyelle initiale du mot anglais « apple », représentée en API par U+00E6 /æ/ (digramme soudé ae). L'anglais ne différencie cependant pas les variantes longues et brèves. Les formes indépendantes de ces voyelles portent les numéros U+0D87 et U+0D88 ; les matras correspondantes U+0DD0 et U+0DD1.

Ces lettres supplémentaires signifient que le singhalais ne respecte pas dans le détail la structure de codage des autres écritures indiennes (notamment la dévanâgarî), mais il en suit les grandes lignes : ordre phonétique, réordonnement des matras et utilisation de virâmas (U+0DCA ච් SYMBOLE SINGHALAIS AL-LAKOUNA) pour identifier les tranches consonantiques conjointes. Contrairement à la dévanâgarî, le singhalais n'utilise pas de demi-formes. On y retrouve cependant un grand nombre de ligatures.

Autres lettres pour le tamoul. L'écriture singhalaise permet également de noter le tamoul. Dans ce cas, quelques lettres supplémentaires s'imposent. Certaines lettres, telles que U+0DBB ඌ LETTRE SINGHALAISE RAYANNA et U+0DB1 ඍ LETTRE SINGHALAISE DANTADJA NAYANNA, peuvent se voir adjoindre l'équivalent d'un noukta. Néanmoins, il n'y a pas à ce jour de noukta dans le bloc singhalais.

Symboles historiques. Ni les chiffres singhalais ni le signe de ponctuation U+0DF4 කුඹුණ PONCTUATION SINGHALAISE KOUNDDALIYA ne s'utilisent couramment de nos jours, on les remplace par des chiffres et des signes de ponctuation occidentaux. Le kounddaliya s'utilisait autrefois pour indiquer un point final ; on l'a inclus dans Unicode pour son utilité dans les textes d'érudition. Les numéros singhalais ne sont pas repris dans cette version d'Unicode.

10.11 Thaï

Thaï : U+0E00 – U+0E7F

L'écriture thaïe s'utilise pour écrire la langue thaïe ainsi que d'autres langues du Sud-Est asiatique, comme le kuy (kuoy) lavüa (lawa), le pâli²⁰. Elle fait partie de la famille des écritures brâhmî. L'écriture thaïe a modifié la forme des caractères brâhmî, tout en ajoutant de nouveaux pour se conformer aux spécificités de la langue thaïe, notamment les signes de ton dérivés de numéros suscrits. Par contre, l'écriture thaïe ignore les consonnes conjointes et les lettres voyelles indépendantes présentes dans la plupart des écritures brâhmî. Comme toutes les autres écritures de cette famille, le thaï s'écrit de gauche à droite.

L'écriture lao s'apparente au thaï et les principes de codage décrits dans cette section s'appliquent également au codage du lao.

Normes. Le codage Unicode pour le thaï repose sur la Norme industrielle thaïe 620-2529 et sa mise à jour 620-2533.

Principes de codage. Comme pour beaucoup d'autres écritures brâhmî, chaque consonne thaïe possède un son de voyelle inhérente. De surcroît, chaque lettre thaïe possède un ton inhérent. La voyelle inhérente et le ton inhérent peuvent être modifiés par l'adjonction à la lettre-consonne de signes-voyelles et de signes de ton. Certains signes-voyelles et tous les signes de ton s'affichent sous la forme de diacritiques suscrits ou souscrits. En mémoire, ces signes suivent la consonne modifiée.

On rend la plupart des signes-voyelles thaïs à l'aide de glyphes pleinehauteur placés avant (à gauche), après (à droite), ou de part et d'autre du glyphe de la lettre-consonne. Pour le codage du thaï, les glyphes de la taille d'une lettre placés avant (à gauche de) la lettre-consonne et qui représentent complètement ou partiellement le signe-voyelle se codent sous la forme de caractères séparés. On les saisit et les stocke avant le caractère-consonne. Cette manière de coder les signes-voyelles thaïs (ou laos) placés géométriquement à gauche diffère fondamentalement de celui des autres écritures brâhmî, où toutes les voyelles sans exception se codent après leur(s) consonne(s). Cette différence s'explique par les pratiques de codage du thaï en cours, pratique recommandée par la Norme industrielle thaïe.

Ponctuation thaïe. Le thaï emploie divers signes de ponctuation qui lui sont propres. U+0E4F ◉ LETTRE THAÏE FONGMAN correspond à notre puce ; on l'utilise pour marquer les éléments d'une liste ou le début d'un vers, d'une phrase, d'un paragraphe ou de toute autre portion de texte. U+0E46 ¶ LETTRE THAÏE MAÏYAMOK dénote la répétition des lettres précédentes. U+0E2F ๑ LETTRE THAÏE PAÏYANNOÏ indique l'élision ou l'abréviation de lettres ; on la considère comme une lettre et on l'utilise très fréquemment, car elle apparaît dans des mots comme le nom thaï de Bangkok. Le *païyannoï* s'utilise associé à d'autres caractères (U+0E2F U+0E25 U+0E2F) pour former ce qu'on nomme *païyanyaï*, c'est-à-dire « *et cetera*, et ainsi de suite. » On peut comparer le *païyanyaï* thaï à son homologue khmer : U+17D8 ្ក្ក្ក SIGNE KHMER BEYYAL.

²⁰ Si toutes les écritures de ce chapitre peuvent être qualifiées de type brâhmî car elles en sont, en dernier ressort, toutes dérivées, les dernières écritures de ce chapitre (à l'exception du tibétain) portent parfois le nom collectif d'écritures pâli. Ces écritures, situées à l'est de l'Inde, ont servi ou servent encore à noter des langues non aryennes. Il s'agit principalement des écritures birmane, khmère, thaïe, môn, kavi, javanaise et bougi. Elles portent ce nom car elles ont souvent servi ou servent encore à écrire le pâli, langue indo-aryenne traditionnelle du bouddhisme méridional.

U+0E5A ๗ CARACTÈRE THAÏ ANGKHANKHOU indique la fin d'une longue portion de texte. Il peut se joindre à U+0E30 ๘ LETTRE THAÏE SARA A pour signaler la fin d'un passage encore plus long ; on le retrouve généralement en poésie, à la fin d'un vers. U+0E5B ๘๗ CARACTÈRE THAÏ KHOMOUT (« pisse de vache ») marque la fin d'un chapitre ou d'un document, il est alors suivi de la combinaison *angkhankhou + sara a*. L'*angkhankhou* joint au *sara a* indique une coupure dans un texte : on rencontre des caractères analogues dans nombre d'autres écritures brâhmî. Ils s'apparentent ainsi au U+17D4 𑄛 SIGNE KHMER KHAN et au U+17D5 𑄜 SIGNE KHMER BARIYÔSAN, eux-mêmes apparentés au *danndda* et *double danna* dévanâgarî.

On ne sépare pas les mots thaïs par des espaces. Dans un texte, les espaces se trouvent entre des segments de texte où, dans la typographie occidentale, on retrouverait normalement des virgules ou des points. Cependant on utilise également des signes de ponctuation latins (virgule, point, deux-points) plus particulièrement mêlés à d'autres lettres latines ou dans la composition des nombres, des adresses, etc. Le caractère U+200B ESPACE SANS CHASSE peut servir à délimiter les mots – par exemple, lors de l'utilisation d'algorithmes de coupure de mots automatique – en intercalant d'invisibles signes qui facilitent alors de telles coupures. Dans un texte justifié, U+200B ESPACE SANS CHASSE peut chasser.

Transcription thaïe du pâli et du sanscrit. L'écriture thaïe sert souvent à noter le pâli et le sanscrit. Dans ce cas, on représente les tranches consonantiques grâce à U+0E3A ๘๒ LETTRE THAÏE PHINTHOU (*virâma*), cette lettre indique la disparition de la voyelle inhérente. Cependant contrairement aux autres écritures indiennes, les conjointes n'existent pas en thaï, le *virâma* reste donc visible. U+0E4D ๘๕ LETTRE THAÏE NIKHAHIT correspond au *nigghahita* pâli et à l'*asnouvâra* sanscrit. U+0E30 ๘ LETTRE THAÏE SARA A équivaut au *visarga* sanscrit. U+0E24 ๘๓ LETTRE THAÏE ROU et U+0E26 ๘๓ LETTRE THAÏE LOU coïncident aux liquides /r/ et /l/, elles s'utilisent de pair avec U+0E45 ๘๕ LETTRE THAÏE LAKKHANGYAO et indiquent leur longueur phonétique.

10.12 Lao

Lao : U+0E80 – U+0EFF

La langue et l'écriture laotiennes sont étroitement liées au thaï. Le standard Unicode code les caractères laos dans le même ordre relatif que le thaï.

Quelques lettres laos n'ont pas d'équivalent en thaï :

U+0EBB ◌̂ VOYELLE DIACRITIQUE LAOTIENNE MAÏ KON

U+0EBC ◌̆ DEMI-VOYELLE DIACRITIQUE LAOTIENNE LO

U+0EBD ◌̇ DEMI-VOYELLE DIACRITIQUE LAOTIENNE ÑO

Les deux demi-voyelles ci-dessus sont les derniers vestiges d'un système de médianes souscrites qui, en birman, conserve d'autres formes. En effet, le birman et le khmer comprennent un jeu complet de consonnes conjointes généralement souscrites. Le thaï n'utilise plus aucune de ces formes ; le lao n'en possède que deux.

Dans Unicode, il n'existe que deux ligatures laos : U+0EDC ວ໊ SYMBOLE LAOTIEN HO NO et U+0EDD ວ໋ SYMBOLE LAOTIEN HO MO. Ces signes correspondent aux suites [h] plus [n] ou [h] plus [m] sans ligature. Ils ont pour fonction de fournir des versions des consonnes [n] et [m] avec différents tons inhérents.

10.13 Tibétain

Tibétain : U+0F00 – U+0FBF

L'écriture tibétaine s'utilise pour écrire le tibétain dans plusieurs pays et régions de l'Himalaya. En plus du Tibet, cette écriture s'utilise au Ladakh, au Népal et au nord de l'Inde, près de la frontière du Tibet où habitent d'importantes populations d'expression tibétaine. Le tibétain s'utilise également au Bhoutan pour dénoter le dzongkè, la langue officielle du pays. Le tibétain constitue aussi la langue philosophique et liturgique de la tradition bouddhiste qui s'étend du Tibet jusqu'à la zone culturelle mongole, laquelle englobe la Mongolie, la Bouriatie, la Kalmoukie et Touva.

C'est au VI^e siècle, par décret du roi tibétain Songtsèn Gampo, que seize hommes partirent en Inde étudier les langues et la grammaire indiennes. On attribue la création du système d'écriture tibétain à un de ces hommes : Teunmi Sambota. L'intention initiale du roi était de répandre la foi bouddhique au Tibet. Le nouveau système d'écriture devait donc permettre à la fois d'écrire le tibétain et de translittérer les langues indiennes (principalement le sanscrit) afin de représenter aussi précisément que possible les textes bouddhistes. De par cette conception initiale, depuis 1 500 ans, l'écriture tibétaine a été largement utilisée pour représenter des mots indiens dont certains se sont intégrés à la langue tibétaine tout en conservant leur graphie translittérée originale.

Remarque sur la translittération latine : la graphie tibétaine est en grande partie figée et ne reflète pas la prononciation moderne des mots. Dans cette section, les mots tibétains en italique représentent une *transcription* phonétique française. Ils sont accompagnés, lors de leur première apparition dans le texte d'une *translittération* en petites capitales. Dans ces translittérations, la présence du point intersyllabique *ts'ék* (TSHEG²¹) s'exprime à l'aide d'un trait d'union.

Dans son traité de grammaire tibétaine, Teunmi Sambota définit deux styles d'écritures. Le premier, nommé *ou-tchèn* (DBU-CAN, « munie de tête, céphale »), correspond au style officiel fondé sur une forme ancienne de la dévanâgarî. C'est l'écriture que l'on retrouve dans les livres tibétains xylographiques et dans les tableaux de caractères de cet ouvrage. Le deuxième style, nommé *ou-mé* (DBU-MED, « sans tête, acéphale »), plus cursif, s'inspire apparemment du tracé des lettres Ourtou. Depuis lors, plusieurs styles d'*ou-mé* se sont développés, aussi bien des styles calligraphiques solennels utilisés dans divers manuscrits que des styles d'écritures manuscrites plus libres. Toutes les écritures tibétaines respectent les mêmes règles de formation de lettres, même s'il existe de légères différences dans la manière dont on forme certaines piles de lettres dans l'*ou-tchèn* et l'*ou-mé*.

Principes généraux de l'écriture tibétaine. La grammaire tibétaine divise les lettres en consonnes et en voyelles. Il existe 30 consonnes, chacune d'elles se représente à l'aide d'un caractère séparé. Il existe cinq voyelles, seules quatre s'écrivent. Ces quatre voyelles explicites apparaissent comme un signe adjoint au-dessus ou en dessous de la consonne modifiée. En l'absence d'une de ces quatre voyelles, le son de la première voyelle équivaut à un « a » bref en français. Trois des quatre signes s'écrivent au-dessus de la consonne ; le quatrième s'écrit en dessous.

²¹ Les noms ISO 10646, en petites capitales, (WA, TSA 'PHRU) utilisent une translittération dite « internationale », à savoir d'inspiration anglaise, les transcriptions en italique (*oua*, *tsa tr'ou*) correspondent à leur prononciation approximative classique en français (ces transcriptions sont fréquemment utilisées dans les ouvrages de vulgarisation). La différence entre la translittération et la transcription est souvent frappante.

Les voyelles tibétaines

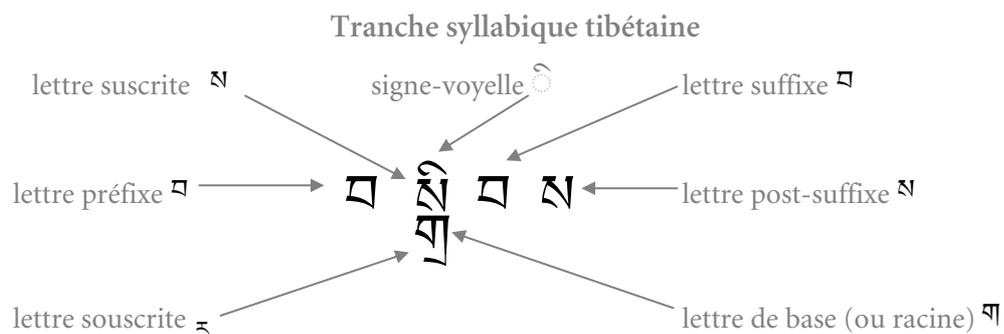
Position relative à la lettre racine	Œil	Numéro Unicode	Nom traditionnel	Transcription
au-dessus	◌̈́	U+0F72	guigou	i
en dessous	◌̈́	U+0F74	chapkyou	ou [†]
au-dessus	◌̈́	U+0F7A	dréngbou	é
au-dessus	◌̈́	U+0F7C	naro	o [†]

[†] Par un phénomène d'inflexion vocalique proche de l'*umlaut* (ö = eu), ces voyelles peuvent se prononcer respectivement *u* [y] et *eu* [ø] selon les affixes (voir plus bas) présents dans la tranche syllabique.

Chaque mot tibétain possède une consonne de base ou consonne-racine. La consonne de base peut s'écrire seule ou d'autres consonnes peuvent s'y ajouter au-dessus ou en dessous pour former une « pile » de lettres. La grammaire tibétaine précise un jeu complet de règles sur le genre²² des lettres et ces règles déterminent quelles lettres peuvent s'écrire côte à côte. Ces règles codifient donc toutes les piles de consonnes possibles. Certaines combinaisons sont interdites par les règles de genres des lettres et, par conséquent, elles n'existent pas dans les mots d'origine tibétaine. Par contre, lors de la transcription d'autres langues (par exemple, le sanscrit ou le chinois) en tibétain, ces règles ne sont pas d'application. Dans certains cas, n'importe quelle consonne peut donc se combiner avec n'importe quelle autre consonne souscrite. Les mises en œuvre sont donc tenues de permettre toutes ces combinaisons.

Le modèle de codage du tibétain comprend les groupes d'éléments suivants: les consonnes, les voyelles, les chiffres, la ponctuation, les signes ornementaux tibétains ainsi que les consonnes et voyelles sanscrites translittérées en tibétain. Chacun de ces groupes sera décrit par la suite.

Tranche syllabique tibétaine. Prenons un exemple pour nous faire une première idée de la structure d'une syllabe tibétaine et introduire quelques notions sur lesquelles nous reviendrons plus tard.



Ce mot se prononce « drib » et signifie « voilé, obscurci ».

On nomme *affixes* les consonnes jointes à la lettre de base, on parle aussi de consonnes adscrites. Chacune des 30 consonnes de l'alphasyllabaire tibétain peut servir de lettre-racine, mais seules 11 consonnes peuvent être adscrites. Tous les affixes ne peuvent être

²² On classe traditionnellement les lettres tibétaines en lettres dites masculines (prononcées tendues, d'un ton haut), les neutres (prononcées non tendues et d'un ton normal avec une aspiration), les lettres féminines (relâchées, d'un ton bas), les très féminines (très relâchées, d'un ton très bas) et, enfin, les stériles et la lettre sans signe (très relâchées d'un ton naturel). Il faut souligner que les termes « masculin » et « féminin » n'ont ici aucun rapport avec la classification française des mots, notion qu'on ne retrouve d'ailleurs pas en tibétain.

combinés avec toutes les lettres de base. Certaines lettres de base peuvent être associées avec tel ou tel affixe, mais pas avec d'autres.

Les tranches syllabiques s'écrivent et s'épellent dans l'ordre suivant : lettre préfixe, suscrite, racine, souscrite, voyelle, suffixe et postsuffixe. Au niveau de la prononciation l'ordre est identique même si de nombreuses lettres adsrites peuvent être muettes tout en modifiant parfois le timbre ou le ton de la lettre de base (et donc son genre).

On peut formellement définir une tranche syllabique tibétaine de la façon suivante :

TRANCHE ::= RACINE AFFIXES* SIGNE-VOYELLE?

où AFFIXES comprend au maximum un seul affixe de chaque catégorie (préfixe, suscrit, souscrit, suffixe ou postsuffixe). Tous les différents affixes peuvent se présenter simultanément.

Les lettres appartenant à chaque catégorie adsrite sont généralement bien définies. Pour les préfixes, on a : GA, DA (*ta*), BA, MA, Â (*a-tchoung*) ; pour les suscrites : RA LA et SA ; pour les souscrites : YA, RA, LA et WA (*oua*) ; pour les suffixes : GA, NGA, DA (*ta*), NA, BA MA, Â (*a-tchoung*), RA LA et SA et, enfin, pour les postsuffixes : SA et DA (*ta*).

La lettre suscrite se place géométriquement au-dessus de la lettre-racine. La lettre suscrite se pend à la ligne de base imaginaire. L'alignement horizontal des autres lettres s'effectue à partir de la tête de la lettre suscrite et, de ce fait, la lettre de base est décalée vers le bas. Les signes voyelles འ ཡ ར coiffent toujours la pile consonantique.

Subjointe. En tibétain, les termes *ba* (-BTAGS, « souscrit ») et *go* (-MGO, « tête, suscrit ») ont plusieurs sens. D'un point de vue structurel, ils indiquent une position précise dans la graphie tibétaine d'origine (« souscrit » et « suscrit »). D'un point de vue géométrique, par contre, ils font référence à la position des glyphes une fois empilés ; toute consonne au-dessus de la pile s'appelle *go*, celles en dessous sont qualifiées de *ba*. Afin d'éviter cette polysémie, nous utiliserons les termes « subjointes » et « tête » dans leur sens géométrique et conserverons les termes « suscrit » et « souscrit » pour désigner les positions structurelles. Ainsi, dans une conjointe telle que « rka », la lettre KA est la racine mais, puisqu'elle n'est pas sur le dessus de la pile, elle s'exprime à l'aide d'un code de caractère subjoint²³, alors que RA, qui est suscrite, s'exprime à l'aide d'un code de caractère de référence. Par contre, dans la conjointe « kra » où KA est une racine, KA se code avec un code de caractère de référence parce que cette lettre se trouve au-dessus de la pile²⁴.

Composition. L'écriture tibétaine possède ses propres conventions de mise en page ; on trouvera ci-après les détails de ce système relatifs à Unicode. Toutefois, un nombre croissant de publications en tibétain ne respecte plus l'ensemble de ces conventions. Ce changement s'explique par le passage progressif de l'impression sur de longs feuillets horizontaux mobiles à celle sous forme de livres reliés écrits verticalement. Les signes de ponctuation tibétains répondent à des besoins fort différents de ceux de la ponctuation développée pour les écritures occidentales. Avec l'apparition, au cours des vingt ou trente dernières années, de journaux, de magazines, de livres scolaires et d'ouvrages de références à l'occidentale, les Tibétains utilisent de nos jours colonnes, blocs de textes avec alinéas, titres et sous-titres à l'européenne et notes de bas de page. On retrouve également dans ce type d'ouvrages certains signes de ponctuation européens comme les crochets, les parenthèses et les guillemets. Depuis l'invention de systèmes d'impression électroniques

²³ Unicode code deux formes des lettres-consonnes tibétaines : une forme de référence (pour les contextes de « tête » ou simples) et une forme subjointe (placée sous une forme de tête). Un caractère-racine peut donc être subjoint.

²⁴ En effet comme on l'a vu plus haut, le KA ne peut être suscrit. D'ailleurs, KA n'est jamais adsrite, il s'agit toujours d'une lettre de base.

plus perfectionnés, on assiste à une renaissance de la publication de volumineux ouvrages religieux et philosophiques dans leur format traditionnel horizontal, en feuillets mobiles – souvent composés à l’aide de caractères numériques qui respectent les proportions de la calligraphie traditionnelle.

Consonnes. Cette sous-section décrit le système Unicode, mentionné ci-dessus, de codage des consonnes tibétaines simples ou empilées.

Dans Unicode, toutes les consonnes tibétaines sont codées une première fois de U+0F40 à U+0F69. Il s’agit des consonnes tibétaines de base et de six consonnes composées utilisées pour représenter les consonnes indiennes GHA (*gaha*), DZHA (*dzaha*), DDHA (*draha*), DHA (*daha*), BHA (*baha*) et KÇA (*k’ya*). Ces codes représentent des consonnes autonomes ou des consonnes de tête dans une pile verticale. Les glyphes produits par ces codes sont toujours en position normale, ils pendent à la ligne de base des caractères. Le standard code une seconde fois les consonnes; ces formes (U+0F90..U+0FB9) représentent les consonnes subjointes d’une pile.

On représente les consonnes simples dans un texte à l’aide de leur valeur de référence. Pour représenter une pile de consonnes, on fait suivre le code de consonne de référence par un ou plusieurs codes de consonnes subjointes. La pile ainsi formée se poursuit aussi longtemps que des codes de consonnes subjointes suivent.

Cette méthode a été préférée à un codage qui aurait utilisé des virâmas, comme c’est le cas en dévanâgarî. Deux raisons justifient ce choix.

1. Tout d’abord, le tibétain n’utilise pas habituellement de virâma pour créer des combinaisons de lettres. L’écriture tibétaine comporte un virâma dans le seul but de représenter la dévanâgarî; il se nomme SROG-MED (*rok-mé*, U+0F84). On n’utilise jamais ce virâma pour écrire des mots tibétains et il ne peut constituer, ce qui ne se produit pratiquement jamais, qu’une autre manière d’empiler les mantras sanscrits écrits à l’aide du système tibétain.
2. Ensuite, le modèle choisi réduit les besoins en mémoire, car les mots d’origine tibétaine privilégient l’empilement des consonnes.

De surcroît, d’autres langues que le tibétain utilisent des piles consonantiques; il s’avère donc nécessaire d’avoir à sa disposition un système qui permette d’empiler n’importe quelle consonne avec n’importe quelle combinaison de consonnes. C’est pourquoi on a choisi un modèle d’empilement qui respecte la méthode traditionnelle tibétaine de formation de ces combinaisons, sans se limiter à un ensemble précis de combinaisons permises.

Voyelles. On code chacun des quatre signes-voyelles tibétains de base comme une entité indépendante. Il s’agit de U+0F72, U+0F74, U+0F7A et U+0F7C. Pour des raisons de compatibilité, certains points de codage entre U+0F71 et U+0F7D représentent des voyelles composées (par exemple, U+0F73 ཨྱ VOYELLE DIACRITIQUE TIBÉTAINE î) utiles à la transcription du sanscrit. La plupart des utilisateurs tibétains ne considèrent pas ces caractères composés comme des caractères simples et ne les utilisent qu’au sein de mots sanscrits. Il est permis de saisir ces caractères composés comme une série de caractères élémentaires et de laisser un logiciel les rendre correctement. Unicode prévoit des équivalences canoniques pour tous ces signes, à l’exception de U+0F77 et U+0F79. Tout signe-voyelle est un signe à chasse nulle placé au-dessus, en dessous ou, à l’occasion, de chaque côté de la pile de consonnes.

Il est permis d’adjoindre un signe-voyelle à une consonne ou à une pile de consonnes. Conformément aux règles de l’écriture tibétaine, on place toujours le code du signe-voyelle après la consonne nue ou après la pile de consonnes auxquelles il s’adjoit.

Tous les symboles et signes de ponctuation se codent sans difficulté. On trouvera plus d'information au sujet de la plupart d'entre eux ci-après.

Ordre de codage. En général, l'ordre de codage correct d'un texte respecte l'épellation tibétaine et l'ordre dans lequel les caractères seraient écrits. Ainsi, l'ordre de codage adéquat pour la plus complexe des piles tibétaines serait :

consonne de tête
première consonne subjointe
...consonne subjointe intermédiaire, le cas échéant
dernière consonne subjointe
voyelle subjointe a-tchoung (â, U+0F71)
virâma ou voyelle ordinaire ou composée

Le signe de lénition optionnel U+0F39 འ SIGNÉ TIBÉTAÏN TSA 'PHRU (tsa trou) apparaît après la consonne qu'il modifie.

Considérations allographiques. Lorsqu'on réunit plusieurs consonnes pour former une pile, l'une d'entre elles conserve le statut de consonne de base (ou racine) de la pile. La consonne-racine conserve toujours sa forme de référence. Les consonnes de tête ou subjointes peuvent, par contre, conserver leur forme de référence ou prendre une nouvelle forme. Sont concernées les consonnes U+0F5D ར WA (*oua*), U+0F61 ལ YA et U+0F62 ལ RA.

« *Ra* » **suscrit**. Suscrite (RA-MGO)²⁵, la consonne *ra* peut changer de forme. Elle peut être affichée sous sa forme complète (ལ) ou amputée de son trait inférieur (elle ressemble alors à un « T » trapu) selon la consonne qu'elle coiffe. Le *ra* suscrit se code à l'aide d'un *ra* isolé (U+0F62) suivi d'une ou plusieurs consonnes subjointes. Dans les combinaisons *ra go* normales, le *ra* suscrit prend généralement la forme d'un demi-*ra* (le T trapu) mais, dans le cas de *ra + ña subjoint*, le *ra* conserve sa forme pleine. Les formes tibétaines habituelles de *ra go* se codent donc convenablement avec la consonne *ra* normale (U+0F62), puisque sa forme est contextuelle. Il revient aux fondeurs²⁶ de fournir les glyphes appropriés afin de permettre la représentation des caractères qui changent de forme en présence d'un *ra* suscrit²⁷.

« *Ra* » **suscrit pleine forme**. Il arrive parfois qu'il faille absolument représenter un *ra* suscrit sous une pleine forme invariable. Ceci *n'est pas* coutumier pour les mots d'origine tibétaine, mais se produit lors de translittérations ou de transcriptions. Dans ces cas, on ne peut utiliser que le code U+0F6A ལ RA À FORME FIXE. Ce *ra* doit toujours être affiché comme un *ra* pleine forme et jamais sous sa demi-forme amputée du trait inférieur. La combinaison de lettres *ra + ya*, par exemple, lorsqu'elle apparaît dans des ouvrages sanscrits translittérés, doit s'écrire à l'aide d'un *ra* pleine forme suivi d'un *ya* subjoint modifié ou d'un *ya* subjoint pleine forme. Il faut remarquer que la forme fixe du *ra* ne devrait s'utiliser que lorsque le *ra* prendrait normalement une forme courte et l'utilisateur désire empêcher ce changement. C'est ainsi que la suite *ra + ña subjoint* ne nécessite jamais l'utilisation de la forme fixe du *ra*, car le *ra* conserve habituellement sa forme

²⁵ Pour rappel RA-MGO, norme de translittération, se prononce comme ra-go.

²⁶ C'est-à-dire les concepteurs de polices de caractères (les « fontes ») ou les dessinateurs de caractère numérique.

²⁷ Les créateurs de polices demandent souvent si les caractères à usage privé servent à coder ces glyphes supplémentaires. Il s'agit là d'une méprise. En effet, quand on désire obtenir le glyphe correspondant à un caractère (par le biais d'une CMAP ou de tout autre moyen), l'index qu'on obtient n'est pas un numéro de caractère mais un numéro de glyphe. Peu importe le numéro de glyphe pour Unicode, ce numéro appartient à un autre espace. Un même caractère A peut être associé dans une police à un glyphe n° 0x41, dans une autre le numéro du glyphe correspondant pourra être 0x07 et, dans une troisième, 0x07 et 0x08 (pour des variantes typographiques).

pleine au-dessus de *ña*. Les fondeurs sont tenus de fournir les glyphes appropriés pour représenter ces codages.

« *Oua* », « *ya* » et « *ra* » *subjoin*ts. Selon la grammaire tibétaine, ces trois consonnes peuvent être subjointes à la consonne-racine. Dans cette position, elles prennent toutes une nouvelle forme. Subjointe, la consonne *oua* se réduit à son coin inférieur droit greffé à la consonne principale (cf. l'exemple de tranche syllabique ci-devant). Le *oua* subjoint prend souvent le nom parlant de *oua-sour* (WA-ZUR, « coin du oua ») ou, moins fréquemment, de *oua-ta* (WA BTAGS). En position subjointe, les consonnes *ya* et *ra* se nomment respectivement *ya ta* (YA-BTAGS) et *ra ta* (RA-BTAGS). Pour coder ces consonnes subjointes conformément aux règles de la grammaire tibétaine, on emploie les formes modifiées et subjointes suivantes : U+0FAD ཨ WA SUBJOINT, U+0FB1 ཨ YA SUBJOINT, U+0FB2 ཨ RA SUBJOINT.

Dans les textes translittérés ou transcrits, ces trois consonnes subjointes ont également des homologues pleine forme qui ne changent pas d'aspect. À cet effet, et si nécessaire, on utilise les consonnes subjointes à pleine forme invariables suivantes : U+0FBA ཨ WA (*oua*) SUBJOINTE À FORME FIXE, U+0FBB ཨ SUBJOINTE YA À FORME FIXE et U+0FBC ཨ SUBJOINTE RA À FORME FIXE. La combinaison *ra + ya* est un bon exemple, car le *ra* peut prendre une forme courte (YA-BTAGS) ou coiffer un *ya* subjoint à pleine forme.

U+0FB0 ཨ 'A SUBJOINT (*a-tchoung*) ne devrait s'utiliser que dans les rares cas où la lettre *a-tchoung* pleine forme subjointe s'avère nécessaire. À sa place, la petite voyelle U+0F71 ཨ VOYELLE DIACRITIQUE TIBÉTAINE À s'utilise dans les textes tibétains et il est recommandé que les mises en œuvre considèrent ce caractère comme le *a-tchoung* subjoint normal.

Coupure de lignes. Les textes tibétains séparent les unités appelées en tibétain *ts'ék-bar*, une traduction erronée de que l'on pourrait traduire de manière inexacte par « syllabe ». Un *ts'ék-bar* correspond en fait à une unité de texte comprise entre deux *ts'ék* ; elle comprend, de façon générale, une suite consonantique avec tous ses préfixes, suffixes et signes-voyelles. Il ne s'agit pas de l'équivalent exact de la syllabe française, c'est pourquoi on préfère parler, comme dans le cas de la dévanâgarî, d'une tranche.

L'écriture tibétaine ne connaît que deux caractères de coupure. Le premier est le *ts'ék* « intersyllabique », le second l'espace. Toutefois dans un texte tibétain suivi, ces caractères ne sont pas nécessairement des caractères de coupure, il faut se fonder sur le contexte. Le texte qui suit expose les diverses questions relatives à ces deux caractères de coupure potentielle.

Dans les textes tibétains, le *ts'ék* (U+0F0B ཨ SIGNE INTERSYLLABIQUE TSHEG) est le principal délimiteur. En règle générale, un processus automatique de coupure de lignes peut couper une ligne après chaque *ts'ék*, sauf quand ce *ts'ék* suit un U+0F44 ཨ NGA (avec ou sans signe-voyelle) et précède un *ché* (U+0F0D ཨ SHAD), ou encore lorsque les règles de grammaire tibétaines ne le permettent. Habituellement, le *ts'ék* ne précède pas un *ché*, sauf après *nga*. Ce type de *ts'ék*-après-*ché* s'appelle « *nga-p'yé-ts'ék* », il peut s'exprimer à l'aide d'un U+0F0B ཨ ou du caractère spécial U+0F0C ཨ, une forme insécable de *ts'ék*. Les noms ISO 10646 de ces deux variantes de *ts'ék* sont erronés ; on ne les a conservés qu'à des fins de compatibilité. Le *ts'ék* ordinaire U+0F0B ཨ est sécable, alors que le *nga-p'yé-ts'ék* est toujours insécable. U+0F0C ཨ SIGNE DÉLIMITEUR TIBÉTAIN TSHEG BSTAR n'est jamais un « délimiteur », il ne s'utilise guère.

Il n'existe pas d'autres caractères de coupure tibétains. Contrairement au français, le tibétain ne connaît pas le trait d'union²⁸ qui permettrait la coupure de mots. La tradition tibétaine n'autorise donc pas les coupures de mots.

On retrouve des blancs dans les textes tibétains, ceux-ci doivent plutôt être représentés par U+00A0 ESPACE INSÉCABLE que par U+0020 ESPACE. En effet, les lignes tibétaines se coupent après un *ts'ék* et non après une espace.

Pour un formatage complet des textes tibétains, il est préférable que ce formatage soit effectué grâce à un module de formatage logiciel plutôt que simplement à l'aide des séquences de codes. Si on emploie correctement les *ts'ék* intersyllabiques et les *ts'ék* insécables respectivement en tant que caractères sécables et insécables et si toutes les espaces sont insécables, un logiciel sera alors toujours en mesure de justifier le texte correctement, même si les coupures risquent d'être inélégantes.

Ponctuation tibétaine. La ponctuation tibétaine est relativement restreinte. Les principaux caractères de ponctuation sont le *ts'ék*, déjà mentionné, le *ché* (translittéré SHAD), un trait vertical utilisé pour indiquer la fin d'un passage, l'espace utilisée parcimonieusement comme blanc et deux parmi les diverses formes particulières du *ché*. Plusieurs autres signes existent encore, mais rarement utilisés.

Le *ché* (U+0F0D |) délimite une portion de texte appelée « TSHIG-GRUB » (*ts'ik groub*). La découpe ne correspond pas à nos phrases ou locutions françaises, il ne faut pas la décrire en termes de phrase. Selon la grammaire tibétaine, un *ché* indique la fin d'une expression (BRJOD-PA) et d'une expression complète. Un double *ché* s'utilise à la fin d'un thème complet (DON-TSHAN). Certains auteurs utilisant un espacement différent pour représenter le double *ché* de celui obtenu par deux *ché* consécutifs, Unicode prévoit un *double ché* (U+0F0E ||) dont l'espacement entre les deux *ché* devrait être plus large que celui de deux *chés* écrits côte à côte. Cependant la plupart des écrivains n'utilisent pas cet espacement inhabituel entre le double *ché*. Les logiciels sont donc tenus de permettre aux utilisateurs d'écrire deux codes U+0F0D l'un après l'autre. De surcroît, il revient aux fondeurs de prendre la décision de dessiner ou non le *double ché* avec un grand espacement.

U+0F11 |̄ RIN CHEN SPUNGS SHAD (*rine tch'ène poung ché*) correspond à une variante du *ché* utilisée dans certaines situations de « passage à la ligne ». Les premières grammaires tibétaines ne définissent pas son utilisation, mais la tradition tibétaine lui accorde une fonction bien précise. De même, le *trul ché* (SBRUL-SHAD) n'est pas défini dans les premières grammaires, cependant la tradition tibétaine définit précisément son emploi. Il délimite des unités de sens équivalentes aux thèmes (DON-TSHAN) *deun-tchèn* et sous-thèmes. Bien qu'il n'existe pas de règle explicite, on dispose généralement l'équivalent d'environ trois espaces de chaque côté d'un *trul ché*. Il s'agit alors d'espaces insécables car le *trul ché* ne doit pas apparaître en début de ligne et l'ensemble formé par les espaces et le *ché* ne doit pas, dans la mesure du possible, être fractionné.

Les textes tibétains emploient le *yik go* (« signe de tête », YIG MGO) pour indiquer le début du recto d'un feuillet puisqu'il n'existe pas d'autre moyen, dans les livres traditionnels tibétains à feuilles mobiles, de savoir avec certitude quel côté est le recto. Le signe de tête peut varier d'un texte à l'autre ; on peut l'écrire de nombreuses façons. U+0F04 ≙ YIG MGO MDUN MA OUVRANT (*yik go dun ma*) représente le signe de tête le plus courant ; U+0F05 ≙ YIG MGO SGAB MA FERMANT (*yik go kab ma*) est son pendant. Le signe *yik go* ouvrant peut s'écrire seul ou être suivi d'au plus trois signes fermants. Lorsqu'un signe ouvrant s'écrit accompagné d'un ou de plusieurs signes fermants, les différentes parties de cet ensemble doivent respecter, pour paraître authentiques, un arrangement spécifique les

²⁸ En typographie française, le trait d'union se nomme « division » ce qui est sémantiquement plus proche de son rôle.

unes par rapport aux autres. Pour cette raison, il est fortement conseillé aux fondeurs de créer des glyphes de ligatures précomposées représentant les différentes combinaisons de ces deux caractères. Les signes de tête les moins fréquents se rencontrent surtout dans les ouvrages Ñingmapa et Bonpo. U+0F01 ཨ, U+0F02 ཨྱ et U+0F03 correspondent à trois de ces signes de tête ; toutefois, de nombreux autres signes ne sont pas codés dans Unicode. Compte tenu des besoins de certains utilisateurs, les fondeurs²⁹ doivent gérer les difficultés causées par l'absence dans Unicode de plusieurs types de signes de tête utilisés dans ces ouvrages qui peuvent être représentés par un signe de rechange.

Deux caractères, U+0F3C འ SIGNÉ TIBÉTAÏN ANG KHANG GYON (*ang kang yeun*) et U+0F3D ཡ SIGNÉ TIBÉTAÏN ANG KHANG GYAS (*ang kang yé*), forment une paire de signes de ponctuation. Utilisés d'ordinaire ensemble, ils forment un toit au-dessus d'un ou de plusieurs mots ou chiffres. Pour un rendu correct, dans ce cas, il se peut qu'on doive avoir recours au crénage ou à des ligatures spéciales. On peut également se servir du *ang kang* un peu à la manière dont une parenthèse fermante simple l'est pour former des listes numérotées ; encore une fois, un rendu correct nécessite un crénage approprié. Les signes U+0F3E ཨྱ yar ts'é et U+0F3F ཨྱ mar ts'é forment une paire destinée à se joindre à des chiffres ; leur usage nécessite des glyphes spéciaux ou un crénage.

Les codes de U+0FBE ཨྱ kourouk'a à U+0FCF ཨྱ dé na soum correspondent à un ensemble de signes astrologiques et religieux particuliers au tibétain.

U+0F34 ཨྱ du ta signifie « *et cetera* » ou « ainsi de suite », il s'utilise après les premières barres ts'ék dans des locutions récurrentes. U+0FBE (souvent en groupe de trois) indique un refrain.

U+0F36 ཨྱ dzu ta chi mig tchène et U+0FBF ཨྱ kourouk'a chi mik tchène indiquent l'endroit où un texte doit être inséré ou servent d'appel de notes de bas de page ou de marge.

Autres caractères. La roue de dharma ཨྱ, qui apparaît parfois dans les textes tibétains, est codée dans le bloc des *Symboles divers* au numéro U+2638.

Les svastikas à branches orientées vers la gauche ou la droite se trouvent parmi les idéophonogrammes chinois (U+534D ཨྱ young-droung-tch'i-k'or et U+5350 ཨྱ young-droung-nang-k'or).

U+0F35 ཨྱ SIGNÉ TIBÉTAÏN NGAS BZUNG ÑI ZLA (*ngé zoung ñi da*) et U+0F37 ཨྱ SIGNÉ TIBÉTAÏN NGAS BZUNG SGOR RTAGS (*ngé zoung gor ta*) s'attachent conceptuellement à une barre ts'ék plutôt qu'à un caractère individuel et se comportent plus à la manière d'attributs que de caractères – ce qui est comparable en quelque sorte à souligner un texte pour en faire ressortir l'importance. Dans les annotations, ils peuvent jouer le rôle d'appel de note. Le plus souvent on obtient ce résultat en composant, dans le texte, le ts'ék dans un corps plus gros que la note ou le commentaire. Placer ces glyphes peut être délicat. Si on les considère comme des diacritiques comme les autres, on peut les coder dans le texte à la suite des signes-voyelle de la pile. Les algorithmes de recherche sont tenus de les prendre en compte.

Unités-et-demies tibétaines. Les glyphes destinés aux unités-et-demies sont propres au tibétain, bien que d'autres écritures (par exemple, le bengali) aient adopté des concepts fractionnels similaires. La valeur de chaque unité-et-demie est inférieure de 0,5 à celle du chiffre dans lequel elle apparaît. Ces formes ne se rencontrent que dans certains textes traditionnels et elles apparaissent comme *dernier* chiffre. La suite de chiffres <U+0F24, U+0F2C> « ཨྱ » représente le nombre 42,5 ou quarante-deux et demi.

²⁹ Voir la note de bas de page n° 27 de ce chapitre.

Justification et composition traditionnelles. Les textes d'origine tibétaine (*pétcha*) se composent en drapeau³⁰ et appuyés à gauche tout en essayant, dans la mesure du possible, d'également les justifier à droite. Chaque page possède une marge. Dans les *pétcha*, la marge doit être indiquée à l'aide de filets visibles. Imprimés de nos jours dans des livres composés à l'occidentale, les textes tibétains peuvent substituer aux filets des marges laissées en blanc. Écrit entre des marges, on tente de justifier les lignes de texte jusqu'à la marge de droite. Pour ce faire, les auteurs doivent porter attention à la longueur de la ligne et tenter de justifier le texte jusqu'à la marge de droite manuellement. Malgré tout, il reste souvent un espace près de la marge de droite qu'on ne peut remplir. Si l'espace est étroit, on le laisse tel quel, considérant la ligne suffisamment justifiée même si, en comparaison avec la justification mécanique, la ligne n'est pas parfaitement justifiée à droite. Si l'espace est grand, on le remplit par autant de *ts'ék* nécessaires pour justifier la ligne. Encore une fois, la justification produite de la sorte n'est pas parfaite comparée à celle utilisée dans les textes français; toutefois pour autant que le dernier *ts'ék* soit assez proche de la marge de droite, cela est satisfaisant. Ce processus de justification à droite ne produit pas un bloc et les lignes demeurent légèrement en drapeau à droite.

On utilise pratiquement toujours des *ts'ék* de justification pour compléter les lignes lorsque le caractère précédent est un *ts'ék* – autrement dit, lorsque la fin de la ligne coïncide avec le milieu d'un *ts'ik groub* (voir, plus haut, la section sur la ponctuation tibétaine). Néanmoins, lorsqu'une ligne se termine par un *ts'ik groub*, il est rare qu'on ajoute des *ts'ék* de justification au *ché* en fin du *ts'ik groub*. Ce qui signifie que

xxxx| |

ne se remplit³¹ pas habituellement de la manière suivante (bien que ce soit permis)

xxxx| |.....

Dans ce cas, au lieu de justifier la ligne à l'aide de *ts'ék*, on augmente l'espacement entre les *chés* ou on laisse l'espace suivant le *ché* final tel quel. On ne remplit jamais une ligne lorsque celle-ci se termine par une espace. Si une espace suit le *ché* dans l'exemple ci-dessous, on ne peut compléter cette ligne à l'aide de *ts'ék* de remplissage :

xxxxxxx|

car le *ché* final doit être suivi d'une espace et qu'on ne peut compléter une ligne par des caractères de remplissage après une espace. La même règle s'applique quand la consonne finale d'un *ts'ik groub* en fin de ligne est KA ou GA. On élimine alors le *ché* final, mais, l'espace étant toujours nécessaire après la consonne, on ne peut compléter la ligne par des caractères de remplissage après cette espace. Ce qui suit est donc interdit :

xxxxxga

Deux règles régissent la composition du début de ligne dans les textes tibétains. De rigoureuses contraintes déterminent quels caractères peuvent commencer une nouvelle ligne ; la règle traditionnelle s'exprime comme suit : un *ché*, quel qu'il soit, ne peut pas s'utiliser en début de ligne. Seul un mot peut amorcer une ligne, les seules exceptions étant un *yik-yo* (YIG-MGO) placé au début de la première ligne d'une page recto ou, sur n'importe quelle autre ligne, un *da-tché* (ZLA-TSHE, « croissant de lune », par exemple U+0F05 ⁹) ou une de ses variantes, c'est-à-dire un *yik-yo* (YIG-MGO). Dans de petits textes en prose, un des deux ou trois *chés* ornementaux remplace fréquemment le *da-tché*, plus austère. Cette règle signifie également qu'on ne peut, dans un texte suivi, utiliser une

³⁰ Lignes successives inégales, dont l'alignement n'est réalisé que d'un seul côté (soit à gauche, soit à droite), l'autre côté de la composition restant irrégulier. S'oppose à la composition en bloc (en pavé) et à la composition centrée (en sommaire brisé).

³¹ Les points dans ces exemples représentent les *ts'ék* de remplissage.

espace en début de ligne. Là où une coupure importante existe dans le texte, il est permis de renfoncer la nouvelle ligne.

Si une syllabe (*barre ts'ék*) suit logiquement un ts'ik group et commence une nouvelle ligne, il faut remplacer le *ché* qui suivrait normalement cette syllabe par un U+0F11 *ř rine tch'ène poung ché*. Un *rine tch'ène poung ché* fait mieux ressortir la fin du ts'ik group et rend, par conséquent, le texte plus lisible.

Dans les textes en vers, on remplace parfois le deuxième *ché* suivant le premier *rine tch'ène poung ché* par un second *rine tch'ène poung ché*, bien que cette pratique soit à vrai dire incorrecte. Il s'agit d'une astuce utilisée par les auteurs pour conférer une certaine élégance à leur propre texte. Cette pratique est assez populaire, bien qu'elle enfreigne la règle. Le *rine tch'ène poung ché* n'est pas le seul substitut du *ché*, une gamme complète de *chés* ornementaux joue le même rôle. Tous sont des variantes calligraphiques du *rine tch'ène poung ché* dont la graphie correcte comporte trois points suscrits.

Sténographie tibétaine (*koung yik*) et limites du *codage*. Selon la grammaire tibétaine, un seul signe-voyelle peut être adjoint à la consonne de base d'un mot (*ming-chji*). La sténographie tibétaine (*koung yik*) permet de raccourcir un ou plusieurs mots en une combinaison inhabituelle de consonnes et de voyelles. Cette contraction entraîne souvent l'adjonction de plus d'un signe-voyelle à une seule consonne ou à une pile; et la composition des piles peut donc enfreindre les règles de grammaire tibétaines habituelles. Pour cette raison, les signes-voyelles interagissent parfois typographiquement, ce qui explique certaines classes combinatoires particulières (voir la *Section 4.2, Classes combinatoires – normatif*).

Unicode prend en compte les groupes composés tibétains formés d'au plus une consonne de base suivie d'un nombre quelconque de consonnes subjointes puis un certain nombre de signes-voyelles. Ceci permet de prendre en charge la grande majorité des textes tibétains. Rares sont les piles qui contiennent plus d'une de ces combinaisons consonnes-voyelles disposées verticalement. On considère d'ordinaire que ces piles ne peuvent être prises en charge par des mécanismes de rendu de textes bruts et qu'elles doivent être traitées à l'aide d'un mécanisme de niveau supérieur.

10.14 Birman

Birman : U+1000 – U+109F

L'écriture birmane sert à noter le birman, langue majoritaire de la Birmanie³². Des variantes et des extensions de cette écriture s'emploient également pour écrire d'autres langues de la région, le chan, le môn, le pâli et le sanscrit.

Au xi^e siècle de notre ère, les bonzes et lettrés môn sont déportés par les envahisseurs birmans dans leur ville de Pagan. Ces lettrés dotent alors la langue birmane d'une écriture dérivée de l'écriture môn, elle-même dérivée au viii^e siècle d'une variante méridionale de la brâhmî. Si on exclut l'arrondissement des caractères à l'origine carrés, cette écriture est demeurée presque inchangée. D'aucuns prétendent que les formes plus rondes furent adoptées pour éviter de suivre les veines des feuilles de palmier et de les déchirer.

Grâce à son origine brâhmî, l'écriture birmane partage les mêmes caractéristiques structurelles que les autres écritures indiennes : symboles de consonnes à la voyelle inhérente « a », signes-voyelles diacritiques représentant une voyelle différente, ligatures et conjointes signalant un groupe consonantique, sens dextrograde (de gauche à droite) de l'écriture. Ainsi, malgré de grandes différences superficielles, l'écriture birmane respecte les mêmes principes de base que, par exemple, la dévanâgarî.

Normes. Il n'existe toujours pas de norme nationale officielle régissant le codage informatique du birman. La mise au point du présent codage a nécessité l'apport d'experts émanant du *Myanmar Information Technology Standardisation Committee* (MITSC) à Yangon (anciennement Rangoun). Le MITSC, organisation gouvernementale fondée en 1997, est constitué d'experts informaticiens, linguistes et historiens.

Principes de codage. Comme pour les autres écritures brâhmî, le codage du birman ne représente que les caractères sous-jacents de base ; la forme visuelle finale de chaque syllabe peut nécessiter plusieurs transformations de rendu et de nombreux glyphes. Même des caractères simples, comme la U+102C ◌ VOYELLE DIACRITIQUE BIRMANE Â, peuvent revêtir différentes formes selon les caractères auxquels ils se joignent. Réciproquement, des caractères ou des combinaisons visuellement identiques dans certaines polices de caractères, comme U+101D ◌ LETTRE BIRMANE WA et U+1040 ◌ CHIFFRE BIRMAN ZÉRO, sont distincts au niveau du codage sous-jacent.

Caractères précomposés. À l'instar du latin étendu et de bien d'autres écritures, certains signes ou lettres birmans peuvent être considérés comme la composition de deux ou plusieurs autres caractères ; on ne les a pas codés séparément. Ci-dessous un exemple de lettre birmane représentée par une suite de caractères combinatoires :

voyelle diacritique birmane o = U+1031 ◌ VOYELLE DIACRITIQUE BIRMANE É
+ U+102C ◌ VOYELLE DIACRITIQUE BIRMANE Â

Sous-intervalles de codage. Le bloc birman commence par les consonnes de base, les voyelles indépendantes et les signes-voyelle dépendants essentiels à l'écriture du birman.

³² Appelé officiellement Myanmar depuis 1989. Ce mot écrit မြန်မာ se prononce Myanmâ. Le « r » final dans sa transcription, dite internationale, est destiné à allonger le « a » et à rendre le deuxième ton en imitant la prononciation britannique de « car ». Évidemment cette transcription présente des problèmes pour les lecteurs de toutes les autres langues que l'anglais. Ce nouveau nom fut adopté afin de rendre compte de l'ensemble des ethnies de l'Union, l'ethnie birmane n'en étant que la principale composante. Étymologiquement, cette nouvelle appellation paraît cependant paradoxale puisque Myan semble dériver du mot môn « mren » qui désignait au ix^e les nouveaux arrivants...birmans.

La fin du bloc supplémente ces catégories à l'aide de caractères utilisés pour écrire d'autres langues, comme le pâli et le sanscrit. Les signes de ponctuation et les chiffres forment le milieu du bloc.

Consonnes conjointes et médianes. Comme pour les autres écritures de type brâhmî, la conjonction de deux lettres consonnes s'indique par l'insertion entre elles d'un virâma U+1039 ◌̣ SYMBOLE BIRMAN VIRÂMA ; il en résulte une ligature ou une autre combinaison visible de consonnes, le virâma reste invisible.

On rend la forme conjointe de U+1004 င LETTRE BIRMANE NGA sous la forme d'un signe suscrit nommé *kinzi*. Le kinzi se code en mémoire comme une consonne conjointe précédant la syllabe à laquelle elle s'adjoint, comme le *ra* dévanâgarî (voir *Section 10.1, Dévanâgarî*, règle R2). Ainsi, un kinzi adjoint à U+1000 က LETTRE BIRMANE KA s'écrit-il à l'aide de la suite suivante :

U+1004 င LETTRE BIRMANE NGA
 U+1039 ◌̣ SYMBOLE BIRMAN VIRÂMA
 U+1000 က LETTRE BIRMANE KA

L'écriture birmane distingue un jeu de consonnes « médianes » généralement souscrites ; on considère que ces formes YA, RA, WA et HA modifient la voyelle de la syllabe. Le codage birman considère ces consonnes médianes comme des conjointes ; en d'autres termes, elles se codent à l'aide d'un virâma. Ainsi le mot *tchwe*³³ (« se détacher »), par exemple, s'écrit င̣က to à l'aide de la suite suivante :

U+1000 က LETTRE BIRMANE KA
 U+1039 ◌̣ SYMBOLE BIRMAN VIRÂMA
 U+101B ရ LETTRE BIRMANE RA
 U+1039 ◌̣ SYMBOLE BIRMAN VIRÂMA
 U+101D ဝ LETTRE BIRMANE WA
 U+1031 ◌̣ VOYELLE DIACRITIQUE BIRMANE É

Virâma explicites. Le virâma U+1039 ◌̣ VIRÂMA participe également à quelques constructions courantes, où il apparaît sous la forme d'un signe visible, appelé « signe qui tue » ou « tueur ». Visible, le virâma (U+1039) est suivi de U+200C မ် ANTI-LIANT SANS CHASSE, comme pour la dévanâgarî (voir la *Figure 10-4*). Le mot မ် မ် « Myanmar » s'écrit donc de la façon suivante :

U+1019 မ် LETTRE BIRMANE MA
 U+1039 ◌̣ SYMBOLE BIRMAN VIRÂMA
 U+101B ရ LETTRE BIRMANE RA
 U+1014 န LETTRE BIRMANE NA
 U+1039 ◌̣ SYMBOLE BIRMAN VIRÂMA
 U+200C မ် ANTI-LIANT SANS CHASSE

³³ On remarquera que l'écriture birmane est loin d'être phonétique. Il faut sans doute attribuer cette distance entre les signes et les sons, d'une part, aux différences phonétiques profondes entre les langues indiennes sans ton auxquelles la brâhmî est bien adaptée et une langue à tons comme le birman, d'autre part, à l'évolution de la prononciation du birman : ainsi le RA et le YA, demeurés distincts dans l'écriture, se prononcent tous deux aujourd'hui *ya* ou, précédés d'une occlusives vélaires (c.-à-d. /k/ ou /g/), comme une occlusive palatale (/c/) approximée dans la transcription ci-dessus par *tch*.

U+1019 ◌◌ LETTRE BIRMANE MA

U+102C ◌◌ VOYELLE DIACRITIQUE BIRMANE Â

Les signes suivent les consonnes. On code en mémoire les voyelles dépendantes et autres signes (sauf le kinzi, mentionné ci-dessus) après la consonne à laquelle ils s'adjoignent, quelle que soit, au rendu, la position géométrique de leur glyphe par rapport à la consonne de base. En particulier, U+1031 ◌◌ VOYELLE DIACRITIQUE BIRMANE É suit en mémoire sa consonne (comme dans l'exemple « tchwé » ci-dessus), même si, visuellement, elle apparaît avant (à gauche de) la consonne.

Espaces. Le birman n'utilise pas d'espace entre les mots. Si l'on désire indiquer la frontière des mots – par exemple, pour des algorithmes de mise en page ou de coupure de lignes automatiques – il faut insérer des U+200B ESPACE SANS CHASSE, invisibles, afin d'indiquer ces coupures potentielles. Dans un texte justifié, l'ESPACE SANS CHASSE peut chasser jusqu'à présenter une largeur visible.

10.15 Khmer

Khmer : U+1780 – U+17FF

Le khmer, également appelé cambodgien, est la langue officielle du Cambodge. Des dialectes proches du khmer se parlent dans le nord-est de la Thaïlande et dans le delta du Mékong au Viêt-nam. Le khmer n'est pas une langue indo-européenne, cependant une considérable partie de son vocabulaire provient du sanscrit et du pâli ; les textes religieux de ces deux langues ont été translittérés et traduits en khmer.

Tout comme le thaï, le lao, le birman et l'ancien môn, l'écriture khmère, l'*a'sâ*³⁴ *kmè*, dérive d'une forme méridionale de la brâhmî. Ses racines exactes restent incertaines, mais il existe une grande similitude entre les premières inscriptions cambodgiennes et l'écriture pallava de la côte de Coromandel. Le khmer moderne connaît deux grands styles d'écritures : l'*a'sâ tchrieng* « écriture inclinée » et l'*a'sâ moûl* « écriture ronde ». Les deux ne comportent pas de différences structurelles. Dans le *Chapitre 15, Tableaux de codes*, on a opté pour l'écriture inclinée (variante « debout »).

La structure de l'écriture khmère reste proche de ses racines méridionales brâhmî ; pour prendre des exemples contemporains, on peut dire qu'elle ressemble aux écritures thaïe et birmane. À chaque lettre-consonne correspond une voyelle inhérente ; cette voyelle inhérente peut être remplacée par une lettre-voyelle explicite placée avant, après, en dessous ou au-dessus de la consonne concernée. Les tranches consonantiques sont représentées à l'aide de conjointes où la première consonne du groupe conserve sa pleine forme et les consonnes suivantes sont souscrites. Le khmer s'écrit de gauche à droite. Les caractères sont disposés en dessous d'une ligne imaginaire, à laquelle ils s'accrochent en quelque sorte par leur partie supérieure que les Cambodgiens appellent le « cheveu ». Le « cheveu » est le reliquat de la potence nâgarî. Les signes-voyelles se placent au-dessus de cette ligne.

La langue khmère possède un plus grand éventail de voyelles que les langues dont elle a emprunté l'écriture, elle différencie, par contre, moins de consonnes. Le khmer a tiré parti de ces différences en affectant plusieurs symboles à une même consonne dotée, chaque fois, d'une voyelle inhérente distincte. Cette situation est analogue à celle du thaï où différents signes-consonnes représentent le même son, mais différents tons. Les lettres-consonnes khmères se classent traditionnellement en deux séries ou registres dont les voyelles inhérentes sont essentiellement « a » et « o ». Deux « embrayeurs » font passer une consonne d'une série à l'autre. Les signes-voyelles dépendants possèdent ainsi plusieurs valeurs phonétiques et s'interprètent en fonction de leur consonne de base.

Principes de codage. Comme pour d'autres écritures apparentées, le codage du khmer ne représente que les caractères de base sous-jacents ; le rendu de chaque syllabe nécessite de multiples transformations et glyphes. Même des caractères simples, comme U+1789 𑄀 LETTRE KHMÈRE ÑO, peuvent prendre des œils différents selon les caractères auxquels ils se lient. Inversement, des caractères ou des combinaisons d'apparence identique dans une même police de caractères, tels que U+17A2 𑄁 LETTRE KHMÈRE A et U+17A3 𑄂 VOYELLE PLEINE KHMÈRE A, diffèrent de par leur codage sous-jacent.

Voyelles indépendantes. En khmer, comme pour la plupart des écritures proches (mais à l'inverse du thaï), les voyelles indépendantes possèdent leur propre œil.

³⁴ Littéralement « lettres khmères », cf. akṣara (*akchara*) en sanscrit.

Consonnes conjointes. U+17D2 𑄛 SIGNE KHMER TCHOENG joue le même rôle dans la formation de conjointes que le virâma indien ; il « dévoyelle » la consonne précédente et indique que la consonne suivante doit être souscrite. Il ne faut pas confondre le *signe tchoeng* avec U+17D1 𑄚 SIGNE KHMER VIRIAM dont le nom ressemble à virâma mais dont la fonction est complètement différente : il indique que le caractère de base fait partie du mot précédent.

Remarquons que la subjointe U+179A 𑄑 LETTRE KHMÈRE RO suit en mémoire la consonne précédente et le *signe tchoeng* et ce même si le glyphe de la subjointe *ro* est rendu avant (à gauche de) la consonne pleine-taille. À l'origine, U+17CC 𑄛 SIGNE KHMER ROBAT correspondait au *rêpha* dévanâgarî (c'est-à-dire un /r/ initial) mais, en khmer, il a perdu cette fonction et correspond de nos jours à un simple diacritique et non à une partie de conjointe.

Embrayeur de consonnes. Les signes U+17C9 𑄛 SIGNE KHMER MOÛSIKETOEN et U+17CA 𑄞 SIGNE KHMER TRÎSAP s'utilisent pour faire passer la consonne de base d'un registre à l'autre. En présence d'autres glyphes suscrits, ces deux signes se rendent à l'aide du même glyphe que U+17BB 𑄛 VOYELLE DIACRITIQUE KHMÈRE OU. Le logiciel de rendu doit sélectionner la forme appropriée.

Signes de voyelles dépendantes. La structure des signes-voyelles dépendants khmers ressemble à celle du thaï, mais leur principe de codage diffère. Le khmer s'inspire des écritures dérivées de la brâhmî, où chaque signe-voyelle dépendant se représente à l'aide d'un seul caractère codé à la suite du signe de la consonne de base. Dans cette optique, chaque signe-voyelle possède une identité propre, quels que soient le nombre et l'emplacement de fragments de glyphes utilisés pour rendre ce signe ou son interprétation phonétique contextuelle. Chaque consonne khmère exprime sa voyelle implicite ou est suivie d'un seul signe-voyelle dépendant.

Ordre des composants syllabiques. On peut exprimer en BNF l'ordre régulier des composants d'une syllabe consonantique de la façon suivante :

$$C (ST C^*) \{ E \} \{ V \} \{ A \}$$

où

- C désigne une consonne
- ST désigne un signe tchoeng (= *virâma*)
- E désigne un embrayeur consonantique
- V désigne une voyelle dépendante
- A représente tout autre signe

Ainsi, le mot /khñom/ (« je, moi ») est représenté par la suite suivante :

- U+1781 𑄑 LETTRE KHMÈRE KHA
- U+17D2 𑄛 SIGNE KHMER TCHOENG
- U+1789 𑄑 LETTRE KHMÈRE ÑO
- U+17BB 𑄛 VOYELLE DIACRITIQUE KHMÈRE OU
- U+17C6 𑄛 SIGNE KHMER NIKAHIT

Sauf pour ce qui est de la présence d'embrayeurs de consonnes, cette structure est analogue à celle de la dévanâgarî (voir *Section 10.1, Dévanâgarî*)

Utilisation de caractères spéciaux. Bien que U+17A3 𑄑 VOYELLE PLEINE KHMÈRE A et U+17A2 𑄑 LETTRE KHMÈRE A se représentent de la même manière, il ne faut employer que

U+17A2 dans les textes modernes. U+17A3 et U+17A4 𑄣 VOYELLE PLEINE KHMÈRE Â s'utilisent uniquement pour translittérer le pâli ou le sanscrit.

U+1789 𑄣 LETTRE KHMÈRE ÑO peut prendre deux formes souscrites différentes. En outre, on omet la partie inférieure détachée de la lettre de base U+1789 quand on y adjoint une lettre souscrite. U+17B1 𑄣 VOYELLE PLEINE KHMÈRE Ô TYPE UN est très proche de U+17B2 𑄣 VOYELLE PLEINE KHMÈRE Ô TYPE DEUX. Le mot « donner », utilisé à profusion, doit s'écrire :

U+17B2 𑄣 VOYELLE PLEINE Ô TYPE DEUX

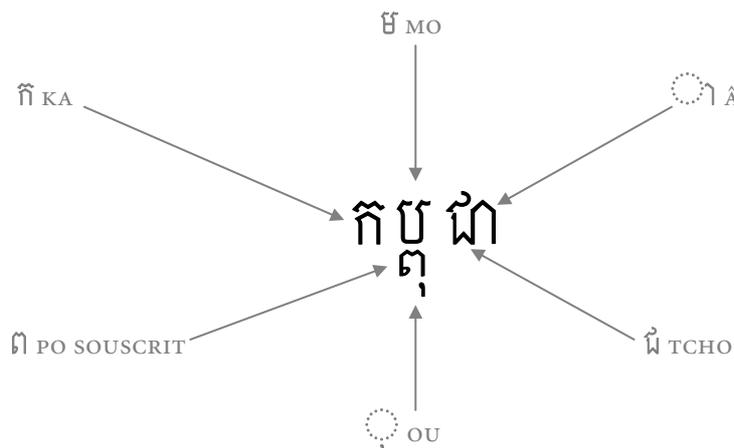
U+17D2 𑄣 SIGNE TCHOENG

U+1799 𑄣 YO

Le logiciel de rendu peut remplacer le glyphe de U+17B2 par le glyphe de U+17B1. Signe extrêmement rare, U+17D3 𑄣 SIGNE KHMER BATHAMASAT sert à exprimer les dates lunaires ; il ne faut pas le confondre avec U+17C6 𑄣 SIGNE KHMER NIKAHIT qui représente un son « m » (cf. *anousvâra*) , signe fréquent dans des textes khmers modernes. Les signes U+17C6 𑄣 NIKAHIT, U+17C7 𑄣 RÉAHMOUK et U+17C8 𑄣 YOÛKALÉAPINTOU correspondent à des parties de voyelle ; toutefois il ne s'agit pas de voyelles complètes.

Espaces. Le khmer ne sépare pas les mots à l'aide d'espaces. Si l'on désire indiquer la frontière des mots – par exemple, pour des algorithmes de mise en page ou de coupure de lignes automatiques – il faut insérer des U+200B ESPACE SANS CHASSE, invisibles, afin d'indiquer ces coupures potentielles. Dans un texte justifié, l'ESPACE SANS CHASSE peut chasser jusqu'à présenter une largeur visible.

Translittération et transcription. Les méthodes de translittération s'inspirent des valeurs historiques (indiennes) des lettres alors que les transcriptions se fondent sur la prononciation moderne. Le nom même du Cambodge constitue un bon exemple de la différence existant entre transcription et translittération khmères. Le mot (ci-dessous) se translittère Kambodjâ (en prenant donc en compte les valeurs théoriques « indiennes » des lettres) mais se transcrit Kampoutchéa (et correspond à notre perception de la prononciation moderne).



Les changements de noms en français (Cambodge, Kampoutchéa démocratique pour revenir à Cambodge) s'opposent à une graphie et prononciation khmères qui demeurèrent constantes pendant toute cette période politiquement agitée. On remarquera que la forme souscrite du PO (le *ba* historique) diffère de sa forme de référence, elle est plus trapue.