

Chapitre 9

Écritures du Moyen-Orient

9.1 Hébreu

Hébreu : U+0590 – U+05FF

On utilise l'écriture hébraïque pour écrire l'hébreu, le yidiche et le judéo-espagnol (ladino) et un certain nombre d'autres langues. Les voyelles et une série d'autres signes sont indiquées à l'aide de *points* qui s'adjoignent aux lettres-consonnes de base. On omet habituellement ces signes en hébreu, sauf dans les textes liturgiques et quelques autres cas particuliers. Cinq lettres hébraïques s'écrivent différemment quand elles terminent un mot.

Directionnalité. L'écriture hébraïque s'écrit de droite à gauche. Les mises en œuvre conformes de cette écriture doivent utiliser l'algorithme bidirectionnel Unicode. (Voir la *Section 3.12, Comportement bidirectionnel*.)

Cursive. Le standard Unicode utilise le mot *cursive* pour désigner les écritures où les lettres d'un mot sont reliées. On parle parfois de cursive en désignant une forme manuscrite de l'hébreu, mais ses lettres arrondies ne sont pas reliées, le standard Unicode ne les appelle donc pas cursives. Il existe des polices qui s'inspirent de l'écriture manuscrite : on ne les utilise pas uniquement pour illustrer quelques lettres manuscrites, mais également à des fins d'affichage.

Normes. *ISO 8859-8 – Partie 8. Alphabet latin/hébreu.* Le standard Unicode code les caractères alphabétiques hébreux aux mêmes positions relatives que l'ISO 8859-8. La norme ISO ne contient cependant ni points ni caractères de ponctuation hébreux.

Points-voyelles et autres signes de prononciation. Ces signes combinatoires – que l'on désigne habituellement en hébreu sous le nom de *points* – notent les voyelles et les autres modifications apportées aux lettres-consonnes. Le *Chapitre 2.6, Caractères combinatoires* et la *section 3.10, Mise en ordre canonique*, énoncent les règles générales qui régissent l'utilisation des caractères combinatoires. Le comportement de certains caractères particuliers à l'écriture hébraïque est décrit ci-dessous.

On peut diviser les points hébreux en quatre classes : *daguèch*, *points chine et sine*, *points-voyelles* et autres signes de ponctuation.

Le *daguèch*, U+05BC ◌ POINT HÉBREU DAGUÈCH OU MAPIQ¹, prend la forme d'un point qui apparaît à l'intérieur de la lettre qu'il modifie. Il ne s'agit pas d'une voyelle mais d'un diacritique qui altère la prononciation de la consonne. La même consonne de base peut également être pourvue d'une voyelle et d'autres diacritiques. Le *daguèch* est le seul élément qui s'insère au sein d'une lettre.

La consonne pointée *chine* est codée explicitement comme U+05E9 װ LETTRE HÉBRAÏQUE CHINE suivie de U+05C1 ◌ POINT HÉBREU CHINE. On place le *point chine* en haut à droite de la

¹ Encore appelé *chourouk*.

lettre de base sans point. De même, la consonne pointée *sine* se code explicitement comme U+05E9 ץ LETTRE HÉBRAÏQUE CHINE suivie d'U+05C2 ם POINT HÉBREU SINE. Le *point sine* s'écrit en haut à gauche de la lettre de base. Les deux points sont mutuellement exclusifs. La lettre de base chine peut être pourvue d'un daguéçh, d'une voyelle et d'autres diacritiques. L'utilisation de ces deux points avec un autre caractère de base constitue une erreur.

Les *points-voyelles* apparaissent sous le caractère de base qu'ils modifient, sauf pour le *holam*, U+05B9 ן POINT HÉBREU HOLAM, qui s'écrit en haut à gauche. Les points suivants représentent des voyelles : U+05B0–U+05B9, U+05BB.

Le solde des *points* représente trois signes de prononciation : U+05BD ם POINT HÉBREU METEG, U+05BF ם POINT HÉBREU RAPE et U+FB1E ץ POINT HÉBREU JUDÉO-ESPAGNOL VARİKA. Le *meteg*, également appelé *silouq*, se place sous le caractère de base ; le *rafe* et le *varika* au-dessus. Le *varika*, utilisé en judéo-espagnol, est une variante graphique du *rafe*.

Chine et sine. Ce bloc ne comprend pas de caractères séparés pour les lettres *chine* et *sine* pointées. Quand il est nécessaire de distinguer ces deux formes, on codera ces lettres comme un U+05E9 ץ LETTRE HÉBRAÏQUE CHINE suivi du point approprié (U+05C1 ou U+05C2). Cette pratique est conforme à la norme de codage israélienne.

Formes finales (variantes contextuelles). Ce bloc code les variantes de forme de cinq lettres hébraïques comme des caractères séparés, tout comme le font les standards hébreux y compris l'ISO 8859-8. Ces variantes de formes remplacent habituellement les formes génériques en fin de mots. Certains mots, néanmoins, s'écrivent avec les formes génériques plutôt que les formes finales, plus particulièrement les mots et noms étrangers empruntés par l'hébreu, ainsi que quelques mots en yidiche. La forme de lettres finales est régie par des conventions orthographiques, il est donc impossible aux logiciels de remplacer automatiquement les formes génériques en fin de mot par leur forme finale correspondante. En d'autres mots, les variantes positionnelles doivent être codées et affichées comme telles, sans analyse contextuelle.

Digrammes yidiches. En yidiche, on considère les digrammes comme des caractères indépendants. Le standard Unicode les code en tant que caractères séparés afin de pouvoir distinguer certaines combinaisons de lettres dans les textes yidiches. C'est le cas par exemple quand on veut différencier le digramme U+05F0 םם LIGATURE HÉBRAÏQUE YIDICHE DOUBLE WAW de l'occurrence d'une U+05D5 ם LETTRE HÉBRAÏQUE WAW consonne suivie d'un U+05D5 waw voyelle. L'utilisation de digrammes est conforme aux pratiques orthographiques yidiches. On peut composer d'autres lettres de l'alphabet yidiche, comme le *pasekh alef*, à partir d'autres caractères.

Ponctuation. La plupart des signes de ponctuation utilisés dans l'écriture hébraïque ne sont pas codés indépendamment (c'est-à-dire qu'ils sont unifiés avec la ponctuation latine) sauf dans les rares cas où le signe a une forme particulière en hébreu. Il s'agit des U+05BE ׀ PONCTUATION HÉBRAÏQUE MAQAF, U+05C0 ׀ PONCTUATION HÉBRAÏQUE PASEQ (également appelé *pési* ou *légarméh*), U+05C3 ׃ PONCTUATION HÉBRAÏQUE SOF PASOUQ, U+05F3 ׀ PONCTUATION HÉBRAÏQUE GERECH et U+05F4 ׀ PONCTUATION HÉBRAÏQUE GERCHĀĪM. Remarquons que les glyphes correspondant aux signes de ponctuation appariés, comme les parenthèses, dépendront de la directionnalité du texte rendu. Voir la Section 4.7, *Caractères miroirs – normatif*, pour plus de renseignements.

Signes de cantilation. On trouve dans les textes liturgiques, y compris la Bible, des signes de cantilation. Il existe différentes écoles historiques de notation, l'ensemble des signes inclus dans le standard Unicode coïncide à la norme israélienne SI 1311.2.

Positionnement. Les différents signes peuvent s'adjoindre aux points-voyelles et autres points selon des règles typographiques complexes.

Le positionnement vertical (c.-à-d. au-dessus, en dessous ou à l'intérieur) des points et des signes est bien défini. Il en va de même du positionnement horizontal des points (c.-à-d. à gauche, à droite ou au centre). Par contre, le positionnement horizontal des signes n'est pas fixé et il est permis de placer ces signes à différents endroits par rapport au caractère de base.

Quand des points et des signes sont situés en dessous de la même lettre de base, on place toujours d'abord le point (à droite) et ensuite le signe (à gauche), sauf pour les signes *yetiv*, U+059A ◌ ACCENT HÉBREU YETIV et U+05AD ◌ ACCENT HÉBREU DEHI où le signe précède (est à droite) le point (situé à gauche).

On observe les règles suivantes lorsque des points et des signes sont situés au-dessus de la même lettre de base :

- Si le point est un holam (U+05B9 ◌), tous les signes de cantilation le précèdent (sont situés à droite), sauf pour le signe de cantilation U+0599 ◌ ACCENT HÉBREU QADMA².
- Le qadma suit toujours (se situe à gauche des) les points.
- Un holam sur une consonne *sine* (*chine* comme base + *point sine*) suit (est situé à la gauche du) le point sine. Toutefois, on affiche parfois les deux signes combinatoires sous la forme d'un seul point assimilé.
- On représente généralement les points chine et sine plus proches verticalement de la lettre de base que les autres points et signes qui pourraient la chapeauter.

Formes de présentation alphabétiques : U+FB1D – U+FB4F

Les caractères hébreux de ce bloc appartiennent essentiellement à deux types : des variantes de lettres et de marques codées dans le bloc hébreu principal et des combinaisons précomposées d'une lettre ou d'un digramme hébreu et d'un ou plusieurs points. Ce bloc comprend toutes les lettres voyellisées de l'alphabet yidiche. On y trouve également la ligature U+FB4F װ LIGATURE HÉBRAÏQUE ALEF LAMÈD et la variante hébraïque du signe plus, FB29 ⊥ LETTRE HÉBRAÏQUE SIGNE PLUS VARIANT. Celle-ci est plus souvent utilisée dans les textes manuscrits qu'imprimés, mais on les retrouve dans les manuels scolaires. On s'en sert quand on veut éviter les croix et leurs connotations potentielles historiques ou religieuses.

Utilisation des lettres larges. On emploie les lettres larges dans l'écriture manuscrite ou, dans les textes imprimés, pour les textes composés en pavé. Si un texte hébreu doit être composé en pavé, cette justification doit être laissée au logiciel de rendu.

Ces formes de présentation alphabétiques sont incluses pour des raisons de compatibilité. Pour une liste des codes correspondants conseillés, consultez le *Tableau des caractères, Formes de présentation hébraïques*, U+F1BD–U+FB4F.

Pour les symboles inspirés des lettres, consultez U+2135..U+2138.

Pour d'autres signes de ponctuation en usage, consultez la Section 6.1, *Ponctuation générale*. L'U+20AA ׀ SYMBOLE NOUVEAU CHÉKEL est codé dans le bloc des devises.

² Également appelé *azla* ou encore *pachta*.

9.2 Arabe

Arabe : U+0600 – U+06FF

L'écriture arabe, comme son nom l'indique, a en premier lieu servi à écrire l'arabe. Aujourd'hui, grâce à l'adjonction de nouveaux signes, elle sert également à écrire plusieurs autres langues parmi lesquelles : le perse, l'ourdou, le pachtou, le sindhi et le kurde. L'ourdou s'écrit souvent à l'aide du style d'écriture fleuronnée appelé *nastaliq*³. Certaines langues, comme l'indonésien-malais, le turc ou l'ingouche s'écrivirent jadis à l'aide de l'alphabet arabe mais utilisent aujourd'hui l'alphabet latin ou cyrillique.

L'arabe est une écriture cursive, même sous sa forme imprimée (voir Figure 9-1). La même lettre peut donc apparaître sous différentes formes selon la manière dont elle se lie aux lettres voisines. Les voyelles et tout un appareil d'autres signes s'écrivent à l'aide de diacritiques appelés *harakat* qui s'adjoignent à la consonne de base. D'ordinaire, l'écriture omet ces diacritiques.

Directionnalité. L'arabe s'écrit de droite à gauche. Les mises en œuvre conformes de l'écriture arabe doivent employer l'algorithme bidirectionnel Unicode (voir la *Section 3.12, Comportement bidirectionnel*).

Figure 9-1. Directionnalité et liaison cursive

Représentation en mémoire :	o o o _ o
Inversion :	o _ o o o
Liaison :	o _ ا ه و

Normes. *ISO 8859-6 – Partie 6. Alphabet latin/arabe.* Le standard Unicode code les caractères alphabétiques arabes aux mêmes positions relatives que l'ISO 8859-6. La norme ISO, quant à elle, s'inspire de la norme ECMA-14, élaborée à partir de l'ASMO 449.

Principes de codage. L'ensemble des lettres arabes de base est bien défini. Chacune de ces lettres se voit attribuer un seul numéro de caractère Unicode dans le bloc arabe de base, quel que soit le nombre de formes contextuelles qu'elles puissent prendre dans un texte. On peut considérer que chaque lettre arabe du standard Unicode représente l'identité sémantique intrinsèque de cette lettre. Le glyphe représentatif qui apparaît dans le tableau de caractères Unicode correspond habituellement à la forme isolée de la lettre en question. Il ne sert qu'à illustrer le caractère et à le différencier des autres caractères repris dans les tableaux de codes et n'impose pas de restrictions quant aux formes que ce caractère pourrait prendre.

Ponctuation. La plupart des signes de ponctuation utilisés avec l'écriture arabe ne se voient pas attribuer de code particulier (en d'autres mots, ils partagent les mêmes codes que la ponctuation latine), sauf dans les rares où un signe a une apparence sensiblement différente en arabe – à savoir : U+060C ، VIRGULE ARABE, U+061B ؛ POINT-VIRGULE ARABE, U+061F ؟ POINT D'INTERROGATION ARABE et U+066A ٪ SYMBOLE POUR CENT ARABE. Pour la ponctuation appariée,

³ Les scribes ourdous, après avoir longtemps employé le *naskhî* (l'écriture des « copistes ») adoptèrent vers le XV^e siècle le *nasta'liq*, un compromis entre le *naskhî* et le *ta'liq* (écriture « suspendue ») usité dans les chancelleries. Chaque mot commence au-dessus de la ligne et s'achève au-dessous.

comme les parenthèses, le glyphe qui représente, par exemple, U+0028 (PARENTHÈSE GAUCHE et U+0029) PARENTHÈSE DROITE dépendra de la directionnalité du texte rendu.

Liant et antiliant. Le standard Unicode fournit à l'utilisateur deux codes de formatage U+200D LIANT SANS CHASSE et U+200C ANTILIANT SANS CHASSE (voir les *Figures 8-2, 8-3 et 8-4*). L'utilisation d'un antiliant entre deux lettres interdit toute liaison cursive entre celles-ci lors du rendu. Parmi les exemples d'utilisation de ce mécanisme, on trouve le suffixe pluriel perse, certains noms de famille perses et les voyelles en turc ottoman. Pour un examen plus détaillé des liants et antiliants, veuillez vous reporter à la section 14.2, *Commandes de composition*.

Figure 9-2. Utilisation du liant

Représentation en mémoire : $o o o _ \boxed{\begin{smallmatrix} ZW \\ J \end{smallmatrix}} o$

Inversion : $o \boxed{\begin{smallmatrix} ZW \\ J \end{smallmatrix}} _ o o o$

Liaison : $a _ a e \text{د}$

Figure 9-3. Utilisation de l'antiliant

Représentation en mémoire : $o \boxed{\begin{smallmatrix} ZW \\ NJ \end{smallmatrix}} o o _ o$

Inversion : $o _ o o \boxed{\begin{smallmatrix} ZW \\ NJ \end{smallmatrix}} o$

Liaison : $o _ a e o$

Figure 9-4. Combinaisons de liant et d'antiliant

Représentation en mémoire : $o \boxed{\begin{smallmatrix} ZW \\ NJ \end{smallmatrix}} \boxed{\begin{smallmatrix} ZW \\ J \end{smallmatrix}} o o _ o$

Inversion : $o _ o o \boxed{\begin{smallmatrix} ZW \\ J \end{smallmatrix}} \boxed{\begin{smallmatrix} ZW \\ NJ \end{smallmatrix}} o$

Liaison : $o _ a e o$

Signes sans chasse harakat (« voyelles »). Les harakat sont des diacritiques qui servent à indiquer la présence de voyelles ou d'autres modifications apportées aux consonnes de base. Un harakat représenté par rapport à un cercle pointillé signifie qu'un certain mécanisme devra l'adjoindre, à la position relative indiquée, au caractère qui le précède dans le flux textuel, c'est-à-dire au caractère de base. Les règles générales d'adjonction des signes sans chasse sont précisées dans la description du bloc des signes diacritiques (v. la *Section 8.9*). Le standard Unicode considère les rares signes qui s'écrivent à la gauche du caractère de base comme de simples caractères qui chassent. Unicode ne précise pas l'ordre de plusieurs harakats qui s'adjoignent au même caractère de base, car il n'existe pas d'ambiguïté sémantique possible. Pour plus d'informations sur l'ordre canonique des signes à chasse nulle, voir le *Chapitre 2, Structure générale* et le *Chapitre 3, Conformité*.

Chiffres arabo-indiens. Le nom que l'on donne aux formes des chiffres décimaux varie sensiblement d'une culture à l'autre. Le système décimal naquit en Inde (० १ २ ३...dévanâgarî) avant d'être adopté par le monde arabe sous différentes formes (٠ ١ ٢ ٣... arabe). Les Européens empruntèrent à leur tour ces chiffres aux Arabes et, une fois de plus, l'apparence des chiffres changea considérablement (0123 européens). Les formes européennes se propagèrent ensuite de par le monde et elles sont même utilisées aujourd'hui dans de nombreux pays du Maghreb. Chaque fois, on interprète les signes décimaux de la même façon. Toutefois, la forme de ces

signes a subi de tels changements qu'on ne peut plus les considérer comme des caractères identiques. En raison de leur origine, on nomme souvent les chiffres décimaux européens des « chiffres arabes » alors que les chiffres décimaux utilisés dans une grande partie du monde arabe sont appelés « chiffres hindî ».

Le standard Unicode code à la fois les chiffres de l'Inde (les formes utilisées par les différentes écritures de l'Inde), les chiffres arabes (en usage dans la plus grande partie du monde arabe) et les chiffres européens (aujourd'hui utilisés dans le monde entier). Cette décision exclut l'utilisation des noms traditionnels ambigus. En outre, il existe deux variantes principales des chiffres arabes – ceux utilisés en Iran et au Pakistan (nommés dans cet ouvrage, les *chiffres arabo-hindî orientaux*) et ceux utilisés ailleurs dans le monde arabe.

Le standard Unicode sépare les chiffres perso-ourdous des autres chiffres arabes afin de conserver leurs différences d'apparence et de directionnalité lors du rendu. La forme de présentation des chiffres perso-ourdous dépend de la police utilisée.

Le standard Unicode et la norme ISO 10646 utilisent les noms repris au *Tableau 9-1*. La nomenclature choisie vise à éliminer l'imprécision qu'entraîne l'utilisation simultanée des noms traditionnels des différents chiffres décimaux.

Tableau 9-1. Désignation des chiffres

Nom	Numéros des caractères	Formes
Européens	U+0030..U+0039	0123456789
Arabo-hindî ⁴	U+0660..U+0669	٠١٢٣٤٥٦٧٨٩
Arabo-hindî orientaux	U+06F0..U+06F9	۰۱۲۳۴۵۶۷۸۹
Hindî ⁵ (en l'occurrence dévanâgarî)	U+0966..U+096F	०१२३४५६७८९

Lettres arabes étendues. D'importantes langues, comme le perse ou l'ourdou, s'écrivent à l'aide de l'écriture arabe, mais c'est également le cas de langues relativement obscures comme le baloutche ou le lahnda qui ne possèdent pas de tradition typographique établie. C'est pourquoi l'ensemble des caractères dans ce bloc peut contenir des formes inutiles. Ce manque de tradition établie peut entraîner l'attribution de plusieurs numéros de caractère à un seul caractère arabe étendu ou à des variantes de chiffre identiques. On a cependant choisi cette méthode car elle est la plus pragmatique et elle minimise le risque d'omettre des caractères valables (à savoir réellement distincts).

Signes d'annotation coranique. Ces caractères servent à indiquer dans le Coran la prononciation d'un passage ainsi que d'autres annotations. Les signes U+06DD ◌ FIN DE AYAH ARABE et U+06DE ◌ DÉBUT DE ROUB EL HIZB ARABE entourent les chiffres coraniques. La taille de ces chiffres est alors réduite au rendu.

Langues. Les annotations du tableau de caractères indiquent parfois la langue dans laquelle un caractère s'utilise. Toutefois cette information n'est pas exhaustive. Quand une annotation se termine par des points de suspension (...), cela signifie que l'on ne cite que les langues principales.

⁴ Également appelés arabo-indiens.

⁵ Ou encore chiffres indiens ou de l'Inde.

Exigence minimale de rendu. La nature cursive de l'écriture arabe implique une série de contraintes de rendu et de traitement que l'on n'exige pas habituellement des systèmes qui traitent uniquement des textes en écriture latine. Le processus de rendu doit assurer la conversion entre l'ordre logique dans lequel sont stockés les caractères en mémoire et l'ordre visuel (ou physique) dans lequel ils sont affichés. Se rapporter à la *Section 3.12, Comportement bidirectionnel*, pour une description détaillée de ce processus de conversion.

Un processus de rendu doit au moins également sélectionner le glyphe approprié permettant de représenter chaque caractère arabe selon son contexte de *liaison* immédiat, il doit enfin remplacer certaines suites de caractères arabes par des glyphes représentant des ligatures. Le reste de cette section définit l'ensemble minimal des règles de liaison (ou de contextualisation) et de ligature qui permet d'obtenir un texte arabe lisible.

Liaison cursive

Classes de liaison. Chaque lettre arabe peut apparaître sous la forme de plusieurs glyphes contextuels. La forme appropriée dépend de la classe de liaison de ce caractère ainsi que de celles des caractères attenants. Chaque caractère a une des classes représentées au *Tableau 9-2* (pour une liste complète voir le fichier *Arabicshaping.txt* sur le cédérom). Le tableau 9-3 définit les classes dérivées à partir des classes de liaison arabe de base. Les règles de liaison cursive utilisent ces classes dérivées. *Droite* et *gauche*, dans ces tableaux, font référence à l'ordre visuel.

Tableau 9-2. Classes de liaison⁶ arabe de base

Classe de liaison	Symbole	Membres
Reliée à droite	R	ALIF ا, DAL د, DHAL ذ, RA' ر, ZAÏN ز
Reliée à gauche	L	Aucun
Reliée des deux côtés	D	BA' ب, TA' ط, THÉ' (thâ') ث, DJÏM ج, ...
Liante	C	LIANT SANS CHASSE, _ TATOUIL (<i>kachidé</i> ou encore <i>allonge</i>). Ces deux caractères diffèrent des caractères reliés des deux côtés dans la mesure où ils ne changent pas de forme.
Non liante	U	ANTILIAN SANS CHASSE et tous les caractères avec chasse (autres que ceux mentionnés ci-dessus) y compris HAMZA ء, HAMZA ÉLEVÉ ؤ, les espaces, les chiffres, les signes de ponctuation, les lettres non arabes, et ainsi de suite.
Transparente	T	Tous les signes combinatoires et de formatage, y compris FATHATAN ّ, DAMMATAN ّ, FATAH َ, DAMMA ُ, KASRA ِ, CHADDA (tachdîd) ّ, SOUKOUN (quiescence, djezm) ّ, ALIF EN CHEF ّ, MARQUE DROITE-À-GAUCHE, ...

⁶ Il s'agit dans le texte qui suit d'une liaison graphique (de contextualisation) et non d'une liaison phonétique ou d'une suppression d'attaque vocalique (comme dans le cas du *wasla*).

Tableau 9-3. Classes de liaison arabe dérivées

Classe de liaison	Membres
liante à droite	surensemble des reliées des deux côtés, des reliées à gauche et des liantes.
liante à gauche	surensemble des reliées des deux côtés, des reliées à droite et des liantes.

Règles de liaison. Les règles ci-dessous décrivent le comportement contextuel des lettres arabes en fonction de leur ordre visuel (ou de rendu). En d'autres mots, la position à laquelle les glyphes (ou œils) des exemples ci-dessous apparaissent correspond à celle de l'affichage après l'application de l'algorithme bidirectionnel et le réordonnement des caractères d'une ligne qui l'accompagne.

- Il est permis aux mises en oeuvre de redéfinir ces règles selon l'ordre logique de sorte qu'elles s'appliquent avant la phase de réordonnement de l'algorithme bidirectionnel. Dans ce cas, les mots *droite* et *gauche* dans cette section doivent être remplacés par *précédent* et *suivant*.

Dans les règles ci-dessous, si X correspond à un caractère, le *Tableau 9-4* décrit la notation utilisée pour désigner les différents types de glyphes qui représentent ce caractère.

Tableau 9-4. Types de glyphe arabes

Types de glyphe	Membres
X_n	Forme générique (« isolée » ou nominale) du glyphe tel qu'il apparaît dans le tableau des codes.
X_d	Forme du glyphe relié à droite (« finale »), les caractères reliés à droite et reliés des deux côtés peuvent utiliser cette forme.
X_g	Forme du glyphe relié à gauche (« initiale ⁷ »), les caractères reliés à gauche et reliés des deux côtés peuvent utiliser cette forme.
X_m	Forme du glyphe relié des deux côtés (« médiane »), seuls les caractères reliés des deux côtés emploient cette forme.

R1 Les caractères transparents n'affectent pas le comportement contextuel des caractères de base (à chasse non nulle). Par exemple :

$$\hat{M}\hat{I}M_n + CHADDA_n + LAM_n \rightarrow \hat{M}\hat{I}M_d + CHADDA_n + LAM_g$$

م + ؤ + ل ➡ م + ؤ + ل ➡ لم

R2 Un caractère X qui se lie à droite (« relié à droite ») dont le caractère à sa droite cause une liaison à droite (« liant à droite ») doit adopter la forme X_d . Par exemple :

$$ALIF_n + TATOUIL_n \rightarrow ALIF_d + TATOUIL_n$$

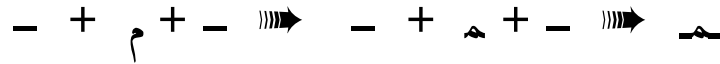
ا + ـ ➡ ا + ـ ➡ لـ

⁷ On préfère le terme « relié à gauche » au plus usuel « initial » car les formes génériques des caractères reliés à droite peuvent également servir de forme initiale, voir *Tableau 9-7*.

R3 Un caractère X qui se lie à gauche (« relié à gauche ») dont le caractère à gauche cause une liaison à gauche (« liante à gauche ») doit adopter la forme X_g .

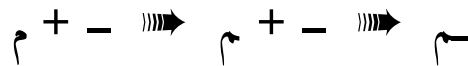
R4 Un caractère X qui se lie des deux côtés dont le caractère à droite cause une liaison à droite et le caractère à gauche cause une liaison à gauche doit adopter la forme X_m . Par exemple :

$$\text{TATOUIL}_n + \text{MÎM}_n + \text{TATOUIL}_n \rightarrow \text{TATOUIL}_n + \text{MÎM}_m + \text{TATOUIL}_n$$



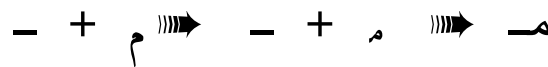
R5 Un caractère qui se lie des deux côtés dont le caractère à droite cause une liaison à droite et le caractère à gauche ne cause pas de liaison à gauche doit adopter la forme X_d . Par exemple :

$$\text{MÎM}_n + \text{TATOUIL}_n \rightarrow \text{MÎM}_d + \text{TATOUIL}_n$$



R6 Un caractère qui se lie des deux côtés dont le caractère à gauche cause une liaison à gauche et le caractère à droite ne cause pas de liaison à droite doit adopter la forme X_g . Par exemple :

$$\text{TATOUIL}_n + \text{MÎM}_n \rightarrow \text{TATOUIL}_n + \text{MÎM}_g$$



R7 Si aucune des règles ci-dessus ne s'applique au caractère X, celui-ci doit alors adopter la forme X_n .

Comme nous l'avons mentionné dans ce chapitre, l'ANTILIAN SANS CHASSE s'utilise en perse (farsi) pour les suffixes pluriels et en turc ottoman pour les voyelles.

Ligatures

Classes de ligature. Certaines ligatures arabes sont obligatoires quelle que soit la police utilisée. De nombreuses ligatures sont optionnelles et dépendent de la police utilisée. Ces ligatures sont facultatives, nous n'en parlerons donc pas ici.

En ce qui a trait aux ligatures arabes obligatoires, certains caractères Unicode se classent dans les deux catégories suivantes (voir la fin de cette description de bloc pour une liste complète) :

Type alif : ALIF MADDA EN CHEF, ALIF HAMZA EN CHEF...

Type lam : LAM, LAM PETIT V, LAM POINT EN CHEF...

On nomme respectivement ces deux classes ALIF et LAM.

Règles de ligature. Les règles suivantes décrivent la formation des ligatures. On les applique après les règles de liaison décrites ci-dessus. Tout comme pour les règles de liaison, ces règles-ci décrivent la formation des ligatures en fonction de l'ordre visuel (ou de rendu).

Dans les règles suivantes, si X et Y correspondent à des caractères, le *Tableau 9-5* décrit la notation utilisée pour désigner les différents types de glyphes qui représentent la combinaison de ces caractères.

Tableau 9-5. Notation des ligatures

Symbole	Membres
$(X.Y)_n$	Forme générique du glyphe qui représente la combinaison des formes X_d et Y_g .
$(X.Y)_d$	Forme reliée à droite du glyphe qui représente la combinaison des formes X_d et Y_m .
$(X.Y)_g$	Forme reliée à gauche du glyphe qui représente la combinaison des formes X_m et Y_g .
$(X.Y)_m$	Forme reliée à gauche et à droite (« médiane) du glyphe qui représente la combinaison des formes X_m et Y_m .

L1 Les caractères transparents n'ont pas d'effet sur la formation de ligature des caractères de base (non transparents). Par exemple :

$$\text{ALIF}_d + \text{FATHA}_n + \text{LAM}_g \rightarrow \text{LAM-ALIF}_n + \text{FATHA}_n$$

L2 Toute suite comprenant un ALIF_d à gauche et un LAM_m à droite doit former la ligature $(\text{ALIF.LAM})_d$. Par exemple :

$$\text{ا} + \text{ل} \Rightarrow \text{ال} \text{ (et non لا)}$$

L3 Toute suite comprenant un ALIF_d à gauche et un LAM_g à droite doit former la ligature $(\text{ALIF.LAM})_n$. Par exemple :

$$\text{ا} + \text{ل} \Rightarrow \text{لا} \text{ (et non لا)}$$

- Considérées en fonction de l'ordre logique (ou de lecture), on nomme les formes de ligature précédentes (ALIF.LAM) les formes LAM-ALIF. Cette différence de nomenclature s'explique par le fait que nous avons utilisé ordre visuel plutôt que l'ordre logique comme point de référence pour désigner ces règles de ligature ci-dessus.

Fonctionnalités optionnelles. De nombreuses autres ligatures et formes contextuelles sont facultatives, elles dépendent de la police et de l'application utilisée. Certaines de ces formes de présentation sont codées dans les intervalles FB50..FDFB et FE70..FEFE. Néanmoins, ces formes ne doivent pas s'utiliser dans l'échange de données ordinaire. En outre, on ne peut supposer que toutes les polices arabes comprendront toutes ces formes, ni que ces formes représentent toutes les formes de présentation qui pourraient apparaître dans une police.

Certains systèmes de rendu plus perfectionnés utiliseront des contextualisations et des placements différents. Ainsi, le placement contextuel des voyelles sans chasse comme le *fatha* produit un résultat plus esthétique. En règle générale, la justification de l'arabe tend à se faire par l'étirement des mots plutôt que par l'ajout de blancs. On peut effectuer un étirement de base en insérant des *tatouils* (*kachidés* ou *allonges*) entre les caractères contextualisés par les règles

R2, R4..R6, L2 et L3. L'endroit le plus approprié pour placer ces allonges dépend de la police et du logiciel de rendu. Les systèmes les plus puissants choisissent des formes différentes pour les caractères comme le *kaf* afin de remplir l'espace nécessaire à la justification.

Types jointifs des caractères arabes. Les tableaux 9-6, 9-7 et 9-8 énumèrent en détail les caractères arabes qui sont reliés à droite ou des deux côtés. Tous les autres caractères arabes (sauf le *kachidé*) ne sont pas liants, y compris la U+06D5 ﺀ LETTRE ARABE E. Pour des raisons de concision, on remplace les mots DEUX, TROIS et QUATRE par leur chiffre correspondant.

La plupart des caractères arabes étendus ne sont que variations sur les formes arabes de base dotées de quelques signes supplémentaires. Pour des raisons de compatibilité, on a inclus de nombreuses formes précomposées.

Certains caractères ne se présentent parfois qu'en fin de mots : on les appelle les *caractères de queue*. Parmi ceux-ci on trouve U+0629 ﺀ LETTRE ARABE TÉ' MARBOUTA, U+0649 ﺀ LETTRE ARABE ALIF MAKSOURA et U+064C ﺀ DAMMATAN ARABE. Quand ces caractères de queue sont liants (comme le té' marbouta), on les classe parmi les reliés à droite, même si certains caractères au dessin similaire sont reliés des deux côtés.

Quand un de ces caractères ne se lie pas et ne cause pas de liaison (comme le dammatan), on le classe parmi les transparents.

- Dans le cas du U+0647 ﺀ LETTRE ARABE HÉ', on trouve souvent dans les tableaux alphabétiques ce glyphe non sous sa forme générique mais sous la forme HÉ'_g (ﻫ) afin de bien le distinguer du U+0665 ٥ CHIFFRE ARABE-HINDI CINQ au dessin très proche. La forme nominale du HÉ' est la forme isolée qui ressemble également à la lettre U+06D5 ﺀ LETTRE ARABE E. Dans le cas du U+06C1 ﺀ LETTRE ARABE HÉ' GÔL, la forme nominale n'est même pas reprise [PA24] dans le tableau des codes, elle aussi ressemble à la LETTRE ARABE E.

Les caractères des tableaux ci-dessous sont groupés selon leur forme et non selon l'ordre alphabétique habituel arabe. Pour une version informatique de ces tableaux, veuillez consulter le fichier ArabicShaping.txt sur le disque optique.

Tableau 9-6. Caractères arabes reliés des deux côtés

Groupe	X _n	X _d	X _m	X _g	Autres caractères au comportement contextuel similaire
BA'	ب	ب	ﺏ	ﺏ	Toutes les lettres qui s'inspirent de la forme BA', y compris TÉ', THÉ' et leurs variantes à diacritique : 0628, 062A, 062B, 0679..0680.
NOÛN	ن	ن	ﻥ	ﻥ	Toutes les lettres qui s'inspirent de la forme NOÛN : 0646, 06B9..06BD.
YA'	ي	ي	ﻱ	ﻱ	Toutes les lettres qui s'inspirent de la forme YA', y compris ALIF MAKSOURA : 0626, 0649, 064A, 0678, 06CC, 06CE, 06D0, 06D1.
HA'	ح	ح	ﺡ	ﺡ	Toutes les lettres qui s'inspirent de la forme HA', y compris KHA', DJÎM et leurs variantes à diacritique : 062C..062E, 0681..0687, 06BF.
SÎN	س	س	ﺲ	ﺲ	Toutes les lettres qui s'inspirent de la forme SÎN, y compris CHÎN et leurs variantes à diacritique : 0633, 0634, 069A..069C, 06FA.
ÇAD	ص	ص	ﺺ	ﺺ	Toutes les lettres qui s'inspirent de la forme ÇAD, y compris DAD et leurs variantes à diacritique : 0635, 0636, 069D, 069E, 06FB.

Groupe	X _n	X _d	X _m	X _g	Autres caractères au comportement contextuel similaire
TA'	ط	ط	ط	ط	Toutes les formes qui s'inspirent de la forme TA', y compris ZZA' et leurs variantes à diacritique : 0637, 0638, 069F.
'AÏN	ع	ع	ع	ع	Toutes les lettres qui s'inspirent de la forme 'AÏN, y compris GHAIÏN et leurs variantes à diacritique : 0639, 063A, 06A0, 06FC.
FA'	ف	ف	ف	ف	Toutes les lettres qui s'inspirent de la forme FA' : 0641, 06A1..06A6.
QAF	ق	ق	ق	ق	Toutes les lettres qui s'inspirent de la forme QAF : 0642, 06A7, 06A8.
MÎM	م	م	م	م	Toutes les lettres qui s'inspirent de la forme MÎM : 0645.
HÉ'	ه	ه	ه	ه	Toutes les lettres qui s'inspirent de la forme HÉ' : 0647.
HÉ' NOUÉ	ه	ه	ه	ه	Toutes les lettres qui s'inspirent de la forme HÉ' NOUÉ : 06BE.
HÉ' GÔL	ه	ه	ه	ه	Toutes les lettres qui s'inspirent de la forme HÉ' GÔL, sauf HÉ' GÔL HAMZA EN CHEF : 06C1.
KAF	ك	ك	ك	ك	Toutes les lettres qui s'inspirent de la forme KAF : 0643, 06AC..06AE.
KAF ÉCRASÉ	ك	ك	ك	ك	Toutes les lettres qui s'inspirent de la forme KAF ÉCRASÉ : 06AA.
GAF	گ	گ	گ	گ	Toutes les lettres qui s'inspirent de la forme GAF : 06A9, 06AB, 06AF..06B4.
LAM	ل	ل	ل	ل	Toutes les lettres qui s'inspirent de la forme LAM : 0644, 06B5..06B8.

Tableau 9-7. Caractères arabes reliés à droite

Groupe	X _n	X _d	Autres caractères au comportement contextuel similaire
ALIF	ا	ا	Toutes les lettres qui s'inspirent de la forme ALIF : 0622, 0623, 0625, 0627, 0671, 0672, 0673, 0675.
WAW	و	و	Toutes les lettres qui s'inspirent de la forme WAW : 0624, 0648, 0676, 0677, 06C4..06CB, 06CF.
DAL	د	د	Toutes les lettres qui s'inspirent de la forme DAL, y compris DHAL et leurs variantes à diacritique : 062F, 0630, 0688..0690.
RA'	ر	ر	Toutes les lettres qui s'inspirent de la forme RA', y compris ZAÏN et leurs variantes à diacritique : 0631, 0632, 0691..0699.

Groupe	X _n	X _d	Autres caractères au comportement contextuel similaire
TÉ' MARBOUTA	ة	آ	Toutes les lettres qui s'inspirent de la forme HÉ' et qui se nouent quand elles sont reliées à droite, y compris le HAMZA SUR HÉ (hé ya en chef) : 0629,06C0.
HÉ' GÔL HAMZA EN CHEF	ه	هٓ	Toutes les lettres qui s'inspirent de la forme HÉ et qui présentent un gôl quand elles sont reliées à droite (en ourdou), y compris le HÉ GÔL HAMZA EN CHEF : 06C2, 06C3.
YA' QUEUE	ي	ي	La forme à queue du YA' : 06CD.
YA' BARRÉ	ي	ي	Toutes les lettres qui s'inspirent de la forme YA' BARRÉ (ourdou) : 06D2, 06D3.

Tableau 9-8. Autres classes de liaison des caractères arabes

Classe	Membres
Liante	LIANT SANS CHASSE (200D) et TATOUIL (0640)
Non liante	ANTILIAN SANS CHASSE (200C) et tous les caractères avec chasse (autres que ceux mentionnés aux tableaux 9-6 et 9-7) sont non liants , y compris le HAMZA (0621) et le HAMZA ÉLEVÉ (0674), les espaces, les chiffres, les signes de ponctuations, les lettres non arabes, etc.
Transparente	Tous les signes combinatoires et de formatage sont sans effet, ceci inclut FATHATAN (064B) et les autres points, MADDA EN CHEF (0653), HAMZA EN CHEF (0654), HAMZA SOUSCRIT (0655), ALIF EN CHEF (0670), MARQUE DROITE-À-GAUCHE (200F) et ainsi de suite.

Formes de présentation arabes A : U+FB50—U+FDFF

Ce bloc contient une liste de formes de présentation (glyphes) codées en tant que caractères à des fins de compatibilité. Nous ne connaissons alors que nous publions ce livre aucune mise en œuvre qui utilise toutes ces formes de présentation. Comme pour tous les codes de compatibilité, ces caractères possèdent des équivalences recommandées qui n'utilisent pas de caractères de compatibilité.

Les formes de présentations de ce bloc représentent des variantes contextuelles des lettres arabes étendues, des variantes contextuelles de ligatures arabes, des formes avec chasse de combinaison de diacritiques, des variantes contextuelles de certaines combinaison de lettres ou de diacritiques et des mots arabes ligaturés. Ces ligatures comprennent une grande gamme de formes de présentation pour des raisons de compatibilité avec des systèmes préexistants. Toutefois, le jeu des ligatures d'une police arabe ne correspondra habituellement pas exactement aux ligatures de ce bloc. Les polices ne comprendront le plus souvent qu'un sous-ensemble de ces glyphes et elles incluent parfois des glyphes qui n'appartiennent pas à ce bloc. En règle générale, ces glyphes ne sont pas accessibles en tant que caractères et ne sont utilisés que par les moteurs de rendu.

La forme festonnée des parenthèses arabes (U+FD3F ﴿ PARENTHÈSE DROITE ORNÉE et U+FD3E ﴾ PARENTHÈSE GAUCHE ORNÉE) ne sont pas des caractères de compatibilité.

Formes de présentation arabes B : U+FE70—U+FEFF

Ce bloc comprend des formes de présentations arabes supplémentaires. Elles consistent en des diacritiques avec chasse ou de type tatouïl (kachidé), des variantes contextuelles de lettres arabes de base et de la ligature obligatoire LAM-ALIF. Elles sont reprises ici pour des raisons de compatibilité avec des normes ou des systèmes préexistants qui utilisent ces formes en tant que caractères. Elles peuvent être remplacées par des lettres du bloc arabes (U+0600..U+06FF). Les mises en œuvre doivent effectuer la contextualisation des glyphes grâce à des règles de rendu lors de l'accès aux glyphes des polices présentes plutôt que de coder les formes contextuelles en tant que caractères.

Diacritiques arabes avec chasse et formes de tatouïl. Pour des raisons de compatibilité avec certaines implantations, on fournit ici un jeu de formes de diacritiques à chasse non nulle. Les formes tatouïl (kachidé) représentent des combinaisons du trait de liaison tatouïl et d'un diacritique.

Espace sans chasse insécable. Ce caractère (U+FEFF) qui ne fait pas partie des formes de présentation arabes est décrit dans la section consacrée au bloc des caractères Spéciaux (U+FFF0..U+FFFF).

9.3 Syriaque

Syriaque : U+0700–U+074F

La langue syriaque appartient au rameau araméen des langues sémitiques. Les premiers documents écrits en syriaque remontent au I^{er} siècle après J.-C. Le syriaque est la langue liturgique de nombreuses communautés du Proche-Orient (syriennes jacobites, nestoriennes, maronites, syriennes catholiques et chaldéennes) et du sud-est de l'Inde (syro-malabares et syro-malankares). C'est également la langue maternelle d'une importante partie de ces communautés.

Le syriaque se divise en deux dialectes. Le syriaque occidental est utilisé par les catholiques syriens, les maronites et les orthodoxes syriens. Le syriaque oriental, pour sa part, est utilisé par les nestoriens (c'est-à-dire l'ancienne Église d'Orient) et les chaldéens. Les deux dialectes diffèrent à peine par leur grammaire et vocabulaire. Ils se distinguent, néanmoins, par leur prononciation et l'utilisation de variantes dialectales de l'écriture syriaque.

Les langues syriaques. Un certain nombre de langues modernes et de dialectes emploient l'écriture syriaque sous une forme ou l'autre. Parmi celles-ci on retrouve :

1. *Le syriaque littéraire.* L'utilisation principale de l'écriture syriaque.
2. *Les dialectes néo-araméens.* Les langues araméennes modernes utilisent largement l'écriture syriaque en concurrence avec les écritures hébraïque, cyrillique et latine. Appartiennent à ces dialectes les araméens orientaux appelés *souadaya* (aussi nommés syriaque vernaculaire, syriaque moderne, assyrien moderne ou sourayé) parlés par les nestoriens et chaldéens d'Iraq, de Turquie et d'Iran ainsi que l'araméen central, le *touroyo* ou *torani*, parlé par les jacobites (Église orthodoxe syrienne) de la région de Toûr Abdîn dans le sud-est turc.
3. *Le garchouni* (l'arabe écrit en syriaque). Utilisé à l'heure actuelle par les chrétiens syriaques dans la rédaction des textes liturgiques arabes. Le garchouni utilise l'ensemble des points-voyelles et diacritiques arabes.
4. *L'araméen palestinien* (également appelé syriaque de Palestine). Dialecte mort.
5. *Autres langues.* L'écriture syriaque s'est également utilisée à différentes époques pour transcrire l'arménien et quelques dialectes perses. Les locuteurs syriaques l'employaient pour écrire l'arabe, le turc ottoman et le malayalam.

L'écriture syriaque. Le syriaque est une écriture cursive et obéit à des règles de contextualisation similaires à celles de l'arabe. Le standard Unicode n'inclut que des caractères abstraits pour le syriaque et donc aucun caractère glyphe (ou forme de représentation).

Directionnalité. Le syriaque s'écrit de droite à gauche. Les mises en œuvre conformes de l'écriture syriaque doivent employer l'algorithme bidirectionnel Unicode (voir la *Section 3.12, Comportement bidirectionnel*).

Styles d'écriture syriaques. Les textes syriaques emploient différents styles de caractère. Cependant tous ces styles utilisent les mêmes caractères syriaques même si leurs formes respectives peuvent différer, c'est pourquoi le standard Unicode ne code qu'une écriture syriaque.

1. *L'estranghelo* (du grec *στρογγύλη*, « ronde », « trapue »), la plus ancienne, dérivé de l'écriture carrée hébraïque par l'intermédiaire du palmyrien. Les plus anciens manuscrits n'utilisent que ce type d'écriture. On l'utilise souvent aujourd'hui dans les textes syriaques orientaux et occidentaux pour les titres et sous-titres. C'est le style privilégié par les orientalistes anglo-saxons.
2. *Le serto* (« ligne », « caractère » en araméen) est le type le plus cursif. Il est utilisé en Syrie occidentale depuis le XVI^e siècle. C'est également le style le plus répandu, celui qu'emploient les Syriens, qu'ils soient unis ou jacobites, les maronites et aussi la plupart des éditeurs orientalistes. Une série de types intermédiaires rattachent le serto à l'estranghelo. On les désigne sous le nom d'écritures de transition ou de cursives jacobites. L'imprimerie ne les a pas adoptés bien qu'ils furent utilisés par les Grecs melkites jusqu'au XVII^e siècle dans leurs textes liturgiques.
3. *Le nestorien ou oriental*, proche de l'estranghelo dont il dérive, est en usage en Syrie orientale depuis le XVI^e siècle. Style utilisé aujourd'hui dans la rédaction des textes syriaques orientaux ainsi qu'en soudaya (néo-araméen oriental). Il côtoie l'estranghelo dans les titres et les sous-titres syriaques occidentaux.
4. *Araméen palestinien chrétien*. Les manuscrits écrits dans ce dialecte emploient ce type proche de l'estranghelo. On le considère comme une sous-catégorie de l'estranghelo.

Figure 9-4b. Les types d'écriture syriaques

	Estranghelo	Serto (occidental, jacobite, maronite)	Nestorien (chaldéen, oriental)
Âlaph	Ⲁ	ⲁ	Ⲃ
Bêth	Ⲅ	ⲅ	Ⲇ
Gâmal	Ⲉ	ⲉ	Ⲋ
Dâlath	Ⲍ	ⲍ	Ⲏ

Le standard Unicode permet d'écrire ces différents types. En outre, Unicode prévoit également des lettres et des diacritiques utilisés dans les langues néo-araméennes, l'araméen palestinien et le garchouni. *Les exemples ci-dessous sont écrits en serto, sauf mention contraire. Par contre, les glyphes représentatifs syriaques choisis au chapitre 15 sont de type estranghelo.*



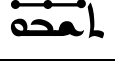
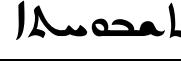


Le symbole d'abréviation syriaque. L'U+070F SYMBOLE D'ABRÉVIATION SYRIAQUE (SAS ou SAM en anglais) est un code de formatage à chasse nulle qui n'a pas d'effet sur la contextualisation des caractères syriaques. Le SAS marque le début d'une abréviation syriaque. On signale l'abréviation à l'aide d'une ligne horizontale tracée au-dessus d'un ou plusieurs caractères. Tout caractère qui n'est ni un caractère ni un diacritique syriaque y met fin (une espace par exemple). Une abréviation syriaque peut donc comprendre des diacritiques syriaques.

En principe, on représente l'abréviation syriaque à l'aide d'une ligne supérieure pommetée à chaque extrémité ainsi qu'au centre. Bien que cela ne soit pas recommandable, on admet aujourd'hui que les logiciels représentent les abréviations dans les textes séculiers à l'aide d'une simple ligne dépourvue de pommettes.

L'abréviation syriaque s'utilise avec les lettres numérales et les contractions. La partie abrégée (et donc remplacée par la marque d'abréviation) s'étend généralement du dernier caractère


élançé d'un mot jusqu'à la fin du mot. Les lettres numériques précédées d'une préposition constituent une exception notable (v. la figure 9-5 ci-dessous).


Figure 9-5. Abréviation syriaque


	=	15 (lettres numériques)
	=	le 15 (nombre précédé d'un préfixe)
	=	
	=	


Le SAS précède le premier caractère de l'abréviation, la fin du mot abrégé termine l'action du SAS. La figure 9-6 illustre l'utilisation du SAS.

Figure 9-6. Utilisation du SAS (SAM)

Représentation en mémoire : 

Inversion : 

Rendu : 

Les textes écrits en syriaque oriental contemporain marquent la fin des mots abrégés d'un signe de ponctuation (U+070A ܐܘܢܐ CONTRACTION SYRIAQUE). Exemple : .

Ligatures et caractères combinatoires. Une seule ligature est incluse dans le bloc syriaque, U+071E ܘܘܢܐ LETTRE SYRIAQUE YOÛDH HÊ. Cette combinaison s'utilise comme un seul caractère, un peu à la manière de la ligature (ou digramme) æ. Le bloc reprend un certain nombre de diacritiques particuliers au syriaque, cependant cette écriture utilise également des caractères combinatoires provenant d'autres blocs, plus particulièrement du bloc arabe.

Signes diacritiques et voyelles. La fonction des signes diacritiques varie : valeur des voyelles (comme en arabe ou en hébreu), attribut grammatical (verbe par opposition à substantif, interjection) ou aide à la lecture ou à la prononciation du texte (quiescence d'une lettre, vocalisation ou non, etc.).

« Le lecteur d'un manuscrit ou d'un livre syriaque ordinaire est confronté à une profusion déroutante de points. Des gros, des moyens, des petits, seuls, doublés ou triplés, suscrits ou souscrits ou encore posés sur la ligne. »

Il existe deux systèmes de voyellement. Le premier, attribué à Jacques d'Édesse (633-708 après J.-C.) utilise des lettres dérivées du grec placées au-dessus (ou en dessous) des consonnes. Le second système est plus ancien, il utilise une série de points de différentes tailles placés en diverses positions. Les textes orientaux n'utilisent que le système à points, alors que les textes occidentaux, plus particulièrement les plus récents, emploient un concours des deux systèmes.

Les signes diacritiques ne chassent pas et sont habituellement centrés au-dessus ou en dessous du caractère affecté. On trouvera ci-dessous les exceptions à cette règle :

1. Les U+0741 ܐ QOUCHCHÂI SYRIAQUE et U+0742 ܐ ROUKKÂKH SYRIAQUE ne s'emploient qu'avec les lettres bêth et gâmal (sous sa forme syriaque ou garchounie), dâlath, kâph, pê et taw.
 - Le *qouchchâi* (« induration ») indique une prononciation dure et non aspirée.
 - Le *roukkâkh* (« amollissement ») indique une prononciation douce et aspirée. Quand on adjoint un roukkâkh au dâlath, on imprime le point légèrement à la droite du point souscrit du dâlath.
2. En syriaque moderne, lorsqu'un mot comprend un rîch et un *seyâmê*⁸, on remplace le point du rîch et du seyâmê par deux points au-dessus du rîch.
3. Le *point féminin* s'adjoint habituellement à la gauche d'un *taw* final.

Ponctuation. La plupart des signes de ponctuations utilisés en syriaque se retrouvent dans les blocs arabe et latin-1. Les autres signes sont codés dans ce bloc.

Chiffres. Le syriaque moderne utilise les chiffres européens, tout comme l'hébreu. L'ordre des chiffres suit le même système que l'hébreu.

Signes héracléens. Les signes héracléens s'utilisent dans les traductions héracléennes du Nouveau Testament. Les U+070B ܘ OBÈLE SYRIAQUE HÉRACLÉEN et U+070D ܘ ASTÉRISQUE SYRIAQUE HÉRACLÉEN indiquent le début d'une proposition, d'un mot ou d'un morphème glosé. L'U+070C ܘ MÉTOBÈLE SYRIAQUE HÉRACLÉEN indique la fin du passage en question.

Dâlath et rîch. À l'époque reculée qui précède l'introduction des points, les textes syriaques ne différenciaient pas le dâlath du rîch qu'un point en dessous ܐ et au-dessus ܐ de ces lettres les distinguera par la suite. Unicode a prévu le caractère ambigu U+0716 ܐ LETTRE SYRIAQUE DÂLATH RÎCH SANS POINT.

Semkath. Contrairement aux autres lettres, le mécanisme de liaison du semkath a changé au cours de l'histoire pour passer de *relié à droite* à *relié des deux côtés*. Il faut insérer un caractère U+200C ANTI-LIANT SANS CHASSE après le semkath si l'on veut être assuré d'obtenir la forme reliée à droite. Il existe deux variantes fréquentes de ce caractère, U+0723 ܘ LETTRE SYRIAQUE SEMKATH et U+0724 ܘ LETTRE SYRIAQUE SEMKATH FINAL. On les retrouve indifféremment dans un même document, à l'instar des sigmas dans les textes grecs.

Signes voyelles. Comme nous le mentionnions plus haut, on peut utiliser les voyelles dites grecques⁹ au-dessus ou en dessous des lettres. Puisque les textes occidentaux (serto) emploient un système mixte de voyellement (grec et pointé), ce bloc code les deux systèmes.

Divers diacritiques. Les textes syriaques utilisent de nombreux diacritiques universels. Leur utilisation est expliquée dans le *Tableau 9-9*.

Tableau 9-10. Utilisation de divers diacritiques

U+0303 ̃, U+0330 ̣	Utilisés en souadaya pour indiquer des lettres inexistantes en syriaque.
U+0304 ̄, U+0320 ̠	Utilisés pour diverses raisons, allant du phonologique au grammatical ou encore comme marques orthographiques.

⁸ Littéralement des « signes placés », ils consistent en deux points placés sous un mot pour indiquer qu'il est au pluriel. Substantifs et adjectifs reçoivent les seyâmês et, pour les verbes, le féminin.

⁹ C'est-à-dire ̂ = A (a), ̃ = E (epsilon = e,ê), ̄ = H (êta = i), ̅ = O (ô), ̆ = OY (ou).

- U+0307 ܐ̇, U+0323 ܐ̇ Points utilisés pour divers usages : grammaticales, phonologiques ou autres. Ils diffèrent typographiquement et sémantiquement du *gouchchâi* et du *roukkâkh* ainsi que des points-voyelles.
- U+0308 ܐ̈ La marque du pluriel. Également utilisé en garchouni comme équivalent du *té' marbouta*.
- U+030A ܐ̈̇, U+0325 ܐ̈̇ Formes concurrentes du *gouchchâi* et du *roukkâkh*. On utilise indifféremment ces formes avec l'U+0741 ܐ̈̇ QOUCHCHÂI SYRIAQUE et le 0742 ܐ̈̇ ROUKKÂKH SYRIAQUE, plus particulièrement dans les grammaires du syriaque occidental.
- U+0324 ܐ̈̈ Ce diacritique se retrouve dans les anciens manuscrits. Il possède des fonctions grammaticales et phonologiques.
- U+032D ܐ̈̈̈ Un des *indicateurs numériques*.
- U+032E ܐ̈̈̈̈ Signe utilisé dans les textes orientaux récents ou modernes, ainsi qu'en souadaya pour indiquer un *pê* fricatif.

Utilisation de caractères provenant du bloc arabe. Le syriaque utilise plusieurs caractères issus du bloc arabe, parmi lesquelles l'*allonge arabe* (U+0640 َ kachidé). Les textes modernes emploient également les U+060C ٫ VIRGULE ARABE, U+061B ٫ POINT-VIRGULE ARABE et U+061F ٫ POINT D'INTERROGATION ARABE. Le *chadda* (U+0651 ّ) fait également partie des signes essentiels du syriaque littéraire ainsi que le *waw* (U+0648 ِ) dans le mot (« Ô »). On emploie l'harakat arabe en garchouni pour indiquer les voyelles et diacritiques arabes correspondants.

Contextualisation syriaque

Exigences minimales de rendu. Elles sont similaires à celles de l'arabe. Le reste de cette section précise un ensemble minimum de règles qui permettent de produire un texte cursif syriaque lisible.

Classes de liaison. Chaque caractère syriaque peut avoir jusqu'à quatre glyphes contextuels. La forme d'une lettre dépend de sa classe de liaison ainsi que de la classe de liaison des deux lettres voisines. Ces classes décrivent des comportements identiques aux classes arabes ; on y ajoute trois classes supplémentaires qui déterminent le comportement des *âlyphs* finaux (v. le *Tableau 9-10*).

Tableau 9-10. Classes de liaisons syriaques supplémentaires

Classes de liaison	Description
A _{fj}	finale jointe (seulement pour âlaph)
A _{fn}	finale non reliée <i>sauf</i> après dâlath et rîch (seulement pour âlaph)
A _{fx}	finale non reliée après dâlath et rîch (seulement pour âlaph)

- R1** Un âlaph qui trouve à sa gauche un caractère séparateur de mots et à sa droite un caractère relié à gauche prend la forme d'un A_{fj}.

} + ܐ̈̈̈̈ |||> | + ܐ̈̈̈̈ |||> ط

- R2 Un âlaph qui trouve à sa gauche un caractère séparateur de mots et à sa droite un caractère non relié à gauche autre que dâlath et rîch prend la forme d'un A_{fn} .



- R3 Un âlaph qui trouve à sa gauche un caractère de séparateur de mots et à sa droite im dâlath ou un rîch prend la forme d'un A_{fx} .



L'exemple ci-dessus est écrit dans un style nestorien (oriental).

Liaison cursive syriaque

Les tableaux 9-11, 9-12 et 9-13 décrivent la forme contextuelle de chaque caractère selon le type de liaison approprié. Les caractères syriaques qui n'apparaissent pas dans ces tableaux ne sont pas reliés. Ces classes contextuelles sont reprises dans le fichier *ArabicShaping.txt* présent sur le disque optique.

Tableau 9-11. Caractères syriaques reliés à droite

Caractère	X_n	X_d
DÂLATH	ܕ	ܕ
DÂLATH RÎCH SANS POINT	ܕ	ܕ
HÊ	ܗ	ܗ
WAW	ܘ	ܘ
ZAÏN	ܙ	ܙ
YOÛDH HÊ	ܝܗ	ܝܗ
SSÂDHÊ	ܣ	ܣ
RÎCH	ܝ	ܝ
TAW	ܬ	ܬ

Tableau 9-12. Caractères syriaques reliés des deux côtés

Caractère	X_n	X_d	X_m	X_g
BÊTH	ܒ	ܒ	ܒ	ܒ
GÂMAL	ܓ	ܓ	ܓ	ܓ
GÂMAL GARCHOUNI	ܓ	ܓ	ܓ	ܓ
HHÊTH	ܚ	ܚ	ܚ	ܚ
TTÊTH	ܛ	ܛ	ܛ	ܛ
TTÊTH GARCHOUNI	ܛ	ܛ	ܛ	ܛ

Caractère	X _n	X _d	X _m	X _g
YOÛDH	ܘ	ܘ	ܘ	ܘ
KÂPH	ܟ	ܟ	ܟ	ܟ
LÂMADH	ܠ	ܠ	ܠ	ܠ
MÎM	ܡ	ܡ	ܡ	ܡ
NOÛN	ܢ	ܢ	ܢ	ܢ
SEMKATH	ܟ	ܟ	ܟ	ܟ
SEMKATH FINAL	ܟ	ܟ	ܟ	ܟ
'Ê	ܐ	ܐ	ܐ	ܐ
PÊ	ܦ	ܦ	ܦ	ܦ
PÊ RÉFLÉCHI	ܦ	ܦ	ܦ	ܦ
QOÛPH	ܩ	ܩ	ܩ	ܩ
CHÎN	ܨ	ܨ	ܨ	ܨ

Tableau 9-13. Formes contextuelles de l'âalph syriaque

Style	A _n	A _d	A _{fj}	A _{fn}	A _{fx}
Estranghelo	ܐ	ܐ	ܐ	ܐ	ܐ
Serto (syriaque occidental)	ܐ	ܐ	ܐ	ܐ	ܐ
Syriaque oriental (nestorien)	ܐ	ܐ	ܐ	ܐ	ܐ

Ligatures

Classes de ligature. Comme pour d'autres écritures, les ligatures en syriaque dépendent du style de la police choisie. Le *Tableau 9-14* identifie les principales ligatures permises pour chaque type d'écriture. Les ligatures facultatives sont marquées d'un astérisque (*).

Tableau 9-14. Ligatures syriaques

Caractères	Estranghelo	Serto (syriaque occidental)	Syriaque oriental (nestorien)	Sources
ÂLAPH LÂMADH	S/O	liée des 2 côtés	S/O	Beth Gazo
GÂMAL LÂMADH	S/O	liée des 2 côtés*	S/O	Armalah
GÂMAL 'Ê	S/O	liée des 2 côtés*	S/O	Armalah
HÊ YOÛDH	S/O	S/O	liée à droite*	Qdom
YOÛDH TAW	S/O	liée à droite*	S/O	Armalah*
KÂPH LÂMADH	S/O	liée des 2 côtés*	S/O	Chhimo

KÂPH TAW	S/O	liée à droite*	S/O	Armalah
LÂMADH ESPACE ÂLAPH	S/O	liée à droite*	S/O	Nomocanon
LÂMADH ÂLAPH	liée à droite*	liée à droite	liée à droite*	BFBS
LÂMADH LÂMADH	S/O	liée des 2 côtés*	S/O	Chhimo
NOÛN ÂLAPH	S/O	liée à droite*	S/O	Chhimo
SEMKATH TTÊTH	S/O	liée des 2 côtés*	S/O	Qourobo
SSÂDHÊ NOÛN	liée à droite*	liée à droite*	liée à droite*	Mouchhotho
RÎCH SEYÂMÊ	liée à droite*	liée à droite*	liée à droite	BFBS
TAW ÂLAPH	liée à droite*	S/O	liée à droite*	Qdom
TAW YOÛDH	S/O	S/O	liée à droite*	

9.4 Thâna

Tanna : U+0780 — U+07BF

L'écriture thâna est utilisée pour transcrire la langue divéhié, langue officielle de la République des Maldives, un essaim d'atolls situé dans l'Océan Indien. Comme l'arabe, le thâna s'écrit de droite à gauche et marque les voyelles à l'aide de signes diacritiques. Contrairement à l'arabe, toutefois, le thâna n'est pas cursif. On a adjoint aux lettres thâna de base quelques lettres pointées supplémentaires pour transcrire l'arabe. L'utilisation de ces lettres modifiées pour écrire l'arabe remonte au milieu du vingtième siècle. Les mots empruntés à l'arabe peuvent s'écrire en écriture arabe bien que cette pratique ne soit plus très courante aujourd'hui (v. la *Section 9-2, arabe*).

Directionnalité. Le thâna s'écrit de droite à gauche. Les mises en œuvre conformes de l'écriture thâna doivent employer l'algorithme bidirectionnel Unicode (voir la *Section 3.12, Comportement bidirectionnel*).

Voyelles. Les consonnes s'écrivent toujours avec un signe-voyelle (U+07A6..U+07AF) ou avec un signe de quiescence¹⁰ (U+07B0 ◌ THÂNA SOUKOUN). La U+0787 ʾ LETTRE THÂNA ALIFOU surmontée du signe d'absence de voyelle représente un coup de glotte.

Chiffres. On utilise à la fois les chiffres européens (U+0030..U+0039) et arabes (U+0660..U+0669). Les chiffres européens sont les plus fréquents, ils possèdent une directionnalité d'affichage gauche-à-droite en thâna. La ponctuation numérale s'utilise avec ces chiffres, qu'ils soient arabes ou européens.

Ponctuation. L'écriture thâna sépare les mots à l'aide d'espaces. Elle utilise un système mixte de ponctuation à la fois arabe et européen, les règles n'en sont pas fixées. La fin de la phrase est aujourd'hui généralement indiquée à l'aide d'un seul point final (U+002E . POINT), mais parfois on le double (U+002E suivi de U+002E). Les propositions sont séparées à l'aide d'une virgule (habituellement une U+060C . VIRGULE ARABE) ou à l'aide d'un seul point (U+002E). On emploie également en thâna le deux-point, le tiret et les guillemets doubles anglais. Enfin, le thâna utilise également le U+061F ؟ POINT D'INTERROGATION ARABE et le U+061B ؛ POINT-VIRGULE ARABE.

Noms et ordre des caractères. Les noms des caractères s'inspirent des noms utilisés dans la République des Maldives. Le caractère de numéro U+0794, ʾ yâ, portent parfois le nom de *yaviyani*, cependant la forme choisie est nettement plus fréquente aujourd'hui. Les caractères sont rangés dans l'ordre alphabétique thâna de *hâ* à *ttâ* pour les lettres thâna, suivie des caractères étendus dans l'ordre alphabétique arabe de *hhâ* à *wâvou*.

¹⁰ Absence de voyelle.