

La translittération arabe-latin à l'ère du numérique

Mahmoud Fawzi MAMMERI

Ecole supérieure de commerce ESC, Koléa, Algérie

mf.mammeri@esc-alger.dz

Reçu: 09/11/2022,

Accepté: 18/12/2022,

Publié: 30/12/2022

Arabic-Latin Transliteration in the Digital Era

ABSTRACT: *This article focuses on the Arabic-Latin transliteration alphabet. Interest in transliteration here is digital literacy: the multitude of transliteration solutions and dedicated Unicode characters make knowledge and manipulation in terms of transliteration difficult to any person not accustomed to advanced techniques of digital. This article answers three essential questions. What is transliteration and what are its character needs? What does digital offer as tools for transliteration? And how can a writer access transliteration characters to meet writing needs?*

KEYWORDS: Transliteration, Arabic-Latin Transliteration, Unicode, Character, Script

RÉSUMÉ: *Cet article s'intéresse à l'alphabet de translittération arabe-latin. L'intérêt porté à la translittération ici est de l'ordre de la littératie numérique : la multitude des solutions pour la translittération et des caractères Unicode qui lui sont consacrés rendent la connaissance et la manipulation en termes de translittération difficile à toute personne inhabituée des techniques avancées du numérique. Cet article répond à trois interrogations essentielles. Qu'est ce que la translittération et quels sont ses besoins ? Que propose le numérique pour la translittération ? Et comment un scripteur peut-il accéder aux caractères de translittération pour satisfaire ses besoins ?*

MOTS-CLÉS: Translittération, Translittération arabe-latin, Unicode, Caractère, Écriture

1. Introduction

Pour écrire du texte, tout scripteur a besoin de symboles. Ces symboles peuvent être des lettres de l'alphabet, des ponctuations, des caractères de transcription, des caractères de translittération, des émojis, etc. L'utilisation de tous ces caractères n'est pas toujours facile, soit parce qu'ils ne sont pas toujours à portée de main du scripteur soit parce que celui-ci n'est même pas au courant de leur existence. Certains scripteurs, faute de parvenir à trouver le bon caractère, recourent à un substitut. On utilise souvent un trait d'union à la place d'un vrai tiret, trois points successifs au lieu d'un point de suspension, des guillemets dactylographiques au lieu des bons guillemets (français, anglais, allemands, etc.)... Dans certaines situations, on change carrément de clavier pour introduire un caractère qu'on ne retrouve pas dans son clavier habituel, comme c'est le cas du point d'exclamation absent du clavier arabe 102 touches azerty.

Dans le cas des caractères de translittération, toutes les polices existantes ne présentent pas tous les caractères utilisés par tous les systèmes de translittération, et même si une police réussit à le faire, la création d'un nouveau système de translittération – ce qui n'est pas exclu – peut mettre cette police en difficulté. C'est le cas, par exemple, du nouveau système de romanisation CARS (CJKI Arabic Romanization System), utilisé dans le CJKI Arabic Learner's Dictionary, qui utilise certaines glyphes qu'on ne retrouve dans aucun des systèmes de translittération ou de transcription traditionnels et par conséquent dans aucune des polices de caractères existantes. L'exemple le plus illustratif est le caractère \acute{a} qui est composé avec le caractère LETTRE MINUSCULE LATINE A sur lequel est placé un DIACRIQUE MACRON au-dessus duquel un DIACRIQUE ACCENT AIGU. La composition de ces glyphes a été banalisée depuis l'avènement d'Unicode¹. Les polices pré-Unicode étaient pauvres en caractères de translittération ; ceci est bien sûr lié aux possibilités de codage offertes par les premiers codages informatiques. Avec la généralisation d'Unicode, les polices implémentent de plus en plus de nouveaux caractères, dont les blocs supplémentaires du latin qui contiennent les caractères de translittération communément utilisés et, mieux encore, tous les caractères diacritiques nécessaires pour la composition de caractères de translittération (en lettres sur/sous lesquels sont empilés un ou plusieurs

¹ Unicode est le standard en matière de codage informatique des caractères utilisé par tous les scripts et alphabets du monde. Il fournit un ensemble d'outils dont un répertoire de codes de caractères où toutes les langues du monde sont représentées.

diacritiques). L'avantage de la composition par rapport à la pré-composition présente deux avantages intéressants :

- une offre élargie de caractères diacritiques pour la composition de nouveaux caractères ;
- la possibilité d'afficher les caractères composés avec toutes les polices supportant les caractères servant à leur composition.

Dans ce travail, nous allons nous intéresser à l'alphabet de translittération arabe-latin, qui est un sous ensemble de l'alphabet latin étendu, en tant qu'outil complémentaire à l'alphabet arabe dans plusieurs situations concrètes dont particulièrement l'écriture d'articles scientifiques de linguistiques et l'alignement d'informations multilingues en arabe et en latin dans des documents d'identité et de voyage, des panneaux de signalisation routiers, des plaques d'adresses... L'intérêt porté à la translittération ici est de l'ordre de la littératie numérique. En effet, la multitude des solutions pour la translittération et des caractères Unicode qui lui sont consacrés rendent la connaissance et la manipulation en termes de translittération difficile à toute personne non habituée des techniques avancées du numérique. L'objectif de cet article serait alors de répondre à un certains nombre de questions concernant l'essentiel de ce qu'il faut savoir sur la translittération. Muni de ces réponses, le lecteur serait en mesure de comprendre trois interrogations principales qui lui seront utiles pour circonscrire ses besoins en caractères de translittération : (i) qu'est ce que la translittération et quels sont les besoins en matière de caractères de translittération ?, (ii) que propose le numérique comme caractères pour la translittération ? et (iii) comment un scripteur peut-il accéder à ces caractères pour satisfaire ses besoins en écriture ?

Le reste de l'article est organisé comme suit. Dans la section subséquente, nous présenterons la notion de translittération, ses racines historiques et ses domaines d'application les plus substantiels ; nous donnerons ensuite un aperçu sur le standard de codage Unicode comme l'outil proposant les différents caractères de translittération et les éléments de base pour la composition de nouveaux glyphes utilisés en translittération ainsi que les algorithmes qui leurs sont associés. Dans la section 2, nous passerons en revue les systèmes de romanisation existants pour l'arabe, à travers leurs besoins en caractères. La section 3 est consacrée aux outils Unicode pour la translittération : il sera question de présenter d'une part les caractères proposés dans le standard Unicode et

d'autre part les éléments de base pour la composition de nouveaux caractères.

2. La translittération : Définition, histoire, domaines et outils

Dans l'ère du numérique, la notion de texte monolingue perd peu à peu de l'espace au profit du texte plurilingues, qui devient le prépondérant : il est rare, par exemple, de trouver un texte totalement arabe, ne serait ce qu'à cause de l'utilisation des chiffres qui sont codés en Unicode comme des caractères latins ! Pour écrire des textes « arabes multilingues » – si je me permets de le dire –, renfermant par exemple des noms propres ou communs ou des textes scientifiques renfermant des translittérations, on pourrait avoir besoin de caractères qui transforment des signifiants écrits dans l'alphabet arabe en d'autres signifiants écrits en alphabet latin, facilitant entre autre à des lecteurs ne connaissant pas l'arabe de lire de l'arabe.

Selon le TLFi, *translit(t)ération* désigne l'opération qui permet de faire correspondre à un signe d'un système d'écriture (source) un signe d'un autre système (cible).² Lorsque le système cible est le latin, on parle de romanisation (ou latinisation).³

Lorsque le nombre de caractères existant dans l'alphabet de conversion est moins élevé que celui de l'alphabet converti, il devient nécessaire de recourir à des digraphes ou à des signes diacritiques. (ISO 233, 1984 (fr) : paragraphe 0.2.1.2)

L'opération de translittération est réversible : elle part d'un texte écrit dans une écriture originale, le transforme dans une écriture cible et doit pouvoir aboutir à sa reconstitution exacte (retranslittération ou rétroconversion).

Un système de translittération n'est cependant pas utilisé pour la transcription qui est basée sur la prononciation. La norme ISO 233 définit la *transcription* comme étant l'opération qui :

(...) consiste à représenter les caractères d'une langue, quelle qu'en soit l'écriture originale, par ceux du système phonétique de lettres ou de signes d'une langue de conversion. (ISO 233, 1984)

²TLFi : *Trésor de la langue française informatisé*. Document en ligne, consulté le 2021-07-15 : <http://atilf.atilf.fr/>.

³ On parle de romanisation de l'arabe, chinois, coréen, grec, cyrillique, bulgare, etc.

La translittération trouve son origine dans la diversité des langues qui a fait naître un besoin en outils pour communiquer entre nations :

La barrière de la diversité des langues, l'incommunicabilité qu'elle instaure entre tant de peuples, (...). Cette barrière s'abaisse un peu dans le cas de langues qui s'écrivent en une même écriture. Quand il s'agit d'écritures idéographiques, comme le chinois classique, on aboutit même à une situation où l'incommunicabilité s'abolit totalement sur le plan du graphisme. Le même caractère se lit de façon toute différente à Pékin, à Canton, à Shanghai. Mais le texte lu garde le même sens. Ainsi nos signaux de circulation automobile ou plus simplement encore nos chiffres (...) (Rodinson, 1964)

Ceci amena les premiers chercheurs qui se sont intéressés à la communicabilité entre peuples à tirer avantage de la notion d'idéographie pour développer l'opération de translittération qui sera utilisée par la suite en parallèle à celle de transcription.

La tendance à préférer l'idéographie, c'est-à-dire la reproduction pure et simple de l'orthographe originale, quitte à provoquer des erreurs de prononciation, à la transcription phonétique visant à suggérer une prononciation proche de l'originale n'a fait que progresser depuis le Moyen âge. Elle s'est accélérée avec la diffusion dans des couches étendues des populations européennes de la connaissance de l'écriture et même d'éléments des langues étrangères. Il s'agit plus de reconnaître le nom écrit que le nom prononcé. (Ibid.)

La translittération est utilisée dans plusieurs domaines :

En particulier, cela est d'une très grande utilité en deux domaines (...) : celui des cartes géographiques et celui des catalogues de bibliothèques. (...) Pour les bibliothèques, on peut transcrire les titres, fût-ce sans les comprendre, les classer comme les noms d'auteurs par ordre alphabétique, fût-ce en les prononçant mal. (...) (Ibid.)

(...) On a souvent recours à cette opération dans des textes historiques ou géographiques, dans des documents cartographiques, et particulièrement dans des travaux bibliographiques, chaque fois qu'il est nécessaire d'écrire des mots fournis dans différents alphabets sous une forme qui permette l'intercalation avec d'autres

mots dans un alphabet unique, de façon à rendre possible un classement alphabétique uniforme dans des bibliographies, des catalogues, des index, des listes toponymiques. (...) (ISO 233, 1984)

D'autre part, l'utilisation de la translittération s'est élargie au (i) traitement informatique de données textuelles, y compris l'indexation de ressources en arabe notamment les ressources bibliographiques, le traitement métrique automatique de corpus poétiques en arabe, l'alignement de mots à partir de corpus parallèles (par exemple, français-arabe) ; et au (ii) glosage des exemples dans les textes en linguistique, la translittération des noms propres...

Avec les nouvelles technologies de l'information et de la communication, la translittération est facilitée par l'unification des codes des caractères utilisés en translittération et des méthodes de composition de nouveaux caractères.

La translittération peut exister pour tout couple de scripts ; un script pouvant être utilisé pour écrire plusieurs langues : arabe, latin, etc. Pour l'arabe, nous pouvons considérer une translittération vers l'arabe et une translittération à partir de l'arabe. Dans la section subséquente, nous nous intéressons à la translittération arabe-latin qui est la plus utilisée et la plus soutenue par les organismes nationaux et internationaux.

3. Systèmes de translittération arabe-latin

Pour l'écriture arabe, plusieurs systèmes de translittération ont été proposés. Ces systèmes sont considérés selon deux points de vue : le domaine d'utilisation et la nature des caractères utilisés dans la conversion.

3.1. Domaine d'utilisation

Il y a ceux qui sont utilisés dans les traitements informatiques et les bibliothèques et ceux utilisés pour transcrire les localités géographiques et les noms propres.

3.2. Nature des caractères utilisés dans la conversion

Dans le cas où le nombre de caractères utilisés dans la langue d'origine est supérieur à celui de la langue de conversion, les systèmes de translittération recourent à la composition de nouveaux caractères. Ces systèmes se répartissent, là aussi, en deux grands groupes : (Rodinson, 1964)

— ceux n'utilisant que des signes se trouvant sur les claviers⁴ ordinaires, ainsi pour représenter certains signes arabes on utilise soit des digrammes soit des caractères spéciaux ;

— ceux faisant recours à des signes diacritiques.

Pour recenser l'ensemble des caractères utilisés en translittération arabe-latin, calculer leur nombre et déduire comment ils sont composés, il faut tout d'abord énumérer tous les systèmes proposés. Dans ce cadre, nous distinguons :

— les systèmes censés être lus par des humains : *DIN-31635*, *ISO 233-2*, *International Journal of Middle East Studies (IJMES)*, *Encyclopædia of Islam(3rd ed.)*, la norme *ALA-LC (American Library Association/Library of Congress, 1997)*, *UN (United Nations Group of Experts on Geographical names, 1972)*⁵... et qui sont utilisés dans les textes scientifiques et les dictionnaires et dans les interfaces homme-machine des bibliothèques numériques;

— des systèmes conçus pour être lus essentiellement⁶ par des machines : notamment le système Buckwalter qui est non seulement une translittération de l'orthographe de l'arabe standard moderne développée par le lexicographe T. Buckwalter mais aussi une véritable transcription orthographique.

Dans ce qui suit, nous présentons les premiers systèmes.

DIN 31635 (1982). DIN 31635 est une norme du Deutsches Institut für Normung — organisme allemand de normalisation — pour la translittération de l'alphabet arabe. Elle est, à côté de la norme adoptée par l'Encyclopédie de l'Islam (EI), l'une des deux normes utilisées internationalement par la communauté scientifique. Le standard DIN 31635 et les standards associés (la norme de la revue Arabica, la norme de translittération de Hans Wehr, la norme utilisée par le Bulletin d'études orientales de l'IFPO⁷...) visent à substituer à chaque graphème du système d'écriture arabe un graphème ou un groupe de graphèmes romans. Ce standard n'a pas été conçu initialement pour la retranslittération : en d'autres termes, il est appliqué dans les cas où il faut translittérer le plus

⁴ Historiquement, il s'agissait des machines à écrire ordinaires.

⁵ United Nations Group of Experts on Geographical Names (UNGEGN):
(<http://www.eki.ee/wgrs>).

⁶ Il y a exception ici de certains systèmes tels que ArabTeX — un système de composition pour l'arabe.

⁷ Publications de l'Institut français du Proche-Orient
(<http://books.openedition.org/ifpo/2471>).

fidèlement possible le mot ou le texte arabe en caractères latins, indépendamment de la prononciation, mais ne peut assurer une retranslittération (reconstitution) du texte original sans perte. DIN 31635 est la norme recommandée dans le domaine des études arabes en France. Cette norme utilise :

- des caractères latins minuscules;
- des diacritiques : macron (¯), macron souscrit (̣), point suscrit (̣), point souscrit (̣), caron (ˇ), caron souscrit (̣);
- les deux demi-anneaux (̣) et (̣).

ISO 233-2 (1993). L'ISO⁸ a mis au point une série de normes pour la translittération des caractères arabes en caractères latins. La première norme fut ISO/R 233 (1961), qui a été annulée et remplacée par la norme ISO 233 : 1984, elle-même mise à jour par la norme ISO 233-2 (1993). Cette mise à jour est une simplification des règles strictes établies par la norme ISO 233 : 1984 et est prévue en particulier pour faciliter le traitement de l'information bibliographique (par exemple, catalogues, construction des index, références...) (ISO 233-2, 1993). L'avantage du système ISO 233-2 (1993) sur ses concurrents est qu'il se veut bijectif : à chaque caractère arabe correspond un caractère latin unique (grâce à l'usage de signes diacritiques), ce qui permet de retrouver automatiquement la graphie arabe par un processus inverse de retranslittération lors des traitements informatiques en particulier le traitement de l'information bibliographique. C'est ce qui a permis à cette norme de s'imposer dans le traitement algorithmique de textes. Elle est utilisée notamment par la Bibliothèque nationale de France, le système universitaire de documentation français et la majorité des bibliothèques françaises pour les titres d'ouvrages et le signalement de leurs fonds en langue arabe. De plus, elle se veut légèrement différente de la norme DIN 31635, notamment par un traitement uniforme de l'article *al-* et de la *tā' marbūṭah*.

Cette norme utilise :

- des caractères latins minuscules et majuscules ;
- des diacritiques : accent aigu/circonflexe (´)/(ˆ), macron (¯), macron souscrit (̣), point suscrit/souscrit (̣)/(̣), tréma (¨), caron (ˇ);
- les deux demi-anneaux (̣) et (̣).

⁸ Organisation internationale de normalisation (International Organization for Standardization).

Brill's Encyclopaedia of Islam Three (EI3). L'Encyclopédie de l'Islam (EI, en anglais : *Encyclopædia of Islam*)⁹ est une entreprise internationale de synthèse des connaissances sur l'Islam et le monde musulman, patronnée par l'Union internationale des académies (représentée par l'Académie royale des Pays-Bas) et publié par Brill à Leyde. Elle est à sa troisième édition — celle en ligne — disponible en anglais depuis 2002 (« Encyclopédie de l'Islam | Brill », s.d.) et en français depuis 2010 (Ibid.). Son système de translittération, datant du XIX^e siècle, fut conservé dans la troisième édition mais avec quelques modifications.

Cette norme utilise :

- des caractères latins minuscules et majuscules ;
- cinq digrammes : th (ث), kh (خ), dh (ذ), sh (ش), gh (غ) ;
- deux diacritiques : le point souscrit (.) pour les consonnes et le macron (¯) pour les voyelles ;
- les deux demi-anneaux (') et (') (Ibid.).

Translittérations pour le grand public. Afin de rendre leur contenu plus accessible aux non spécialistes (élèves des niveaux secondaire et collégial, le grand public et même des universitaires non initiés à la translittération) et bénéficier ainsi d'une large audience, certains documents écrits en latin (journalistiques, encyclopédiques¹⁰...) adoptent une translittération implicite (donc non normalisée) où aucun marquage distinctif utilisant des diacritiques n'est utilisé pour les mots d'origine arabe. De plus, les translittérations pour le 'ayn (') et la hamza (') sont elles aussi évitées dans ce type de codage. D'une manière générale, ce type de translittération est recommandé pour les communications écrites des mots et syntagmes arabes dans les rédactions non académiques et est même pratiquée en histoire littéraire comme alternative à la DIN 31635 utilisée dans les études académiques en linguistique.

Ce type de translittération utilise uniquement :

- des caractères latins minuscules et majuscules ;
- des digrammes : th (ث), kh (خ), dh (ذ), sh (ش), gh (غ).

⁹ Attention à la confusion qui pourrait avoir lieu avec l'encyclopédie « Encyclopedia of Islam » de Campo, Juan Eduardo (Campo, 2009) !

¹⁰ Un exemple est celui de « Encyclopedia of Islam » (Campo, 2009 : xxxv).

Tableau 1 : Les caractères Unicode adoptés dans les systèmes de translittération les plus utilisés (DIN-31635, ISO 233-2, EI)¹¹

Caractère	Nom du caractère	Caractère de translittération		
		DIN-31635	ISO 233-2	EI
	LETTERES DE L'ALPHABET	b, t, ḏ, r, z, s, f, q, k, l, m, n, h, w, y, a, u, i	b, t, ḏ, r, z, s, f, q, k, l, m, n, h, w, y, a, u, i B, T, D, R, Z, S, F, Q, K, L, M, N, H, W, Y A, U, I	b, t, j, ḏ, r, z, s, f, q, k, l, m, n, h, w, y, a, u, i B, T, J, D, R, Z, S, F, Q, K, L, M, N, H, W, Y A, U, I
	DIGRAMMES	Non utilisés	Non utilisés	th, kh, dh, sh, gh
U+0301	DIACRITIQUE ACCENT AIGU	Non utilisé	á	Non utilisé
U+0302	DIACRITIQUE ACCENT CIRCONFLEXE	Non utilisé	â	â
U+0304	DIACRITIQUE MACRON	ā, ū, ī	ā, ū, ī	ā, ū, ī
U+0307	DIACRITIQUE POINT EN CHEF	ḡ	ḡ	Non utilisé
U+0308	DIACRITIQUE TRÉMA	Non utilisé	ˆ	Non utilisé
U+030C	DIACRITIQUE CARON	ḥ, š	ḥ, š	Non utilisé
U+0323	DIACRITIQUE POINT SOUSCRIT	ḥ, s, ḏ, t, z	ḥ, s, ḏ, t, z	ḥ, s, ḏ, t, z
U+032C	DIACRITIQUE CARON SOUSCRIT	ḥ	Non utilisé	Non utilisé
U+0331	DIACRITIQUE MACRON SOUSCRIT	ṭ, ḡ	ṭ, ḡ	Non utilisé
U+02BC	LETTRE MODIFICATIVE APOSTROPHE	'	'	'
U+02BD	LETTRE MODIFICATIVE VIRGULE RENVERSEE	'	'	'
U+02BE	LETTRE MODIFICATIVE DEMI-ANNEAU À DROITE	'	'	'
U+02BF	LETTRE MODIFICATIVE DEMI-ANNEAU À GAUCHE	'	'	'

Tableau 2 : Tableau récapitulatif comparatif des différents éléments utilisés dans les normes proposées pour la translittération des caractères arabes en caractères latins¹²

		Norme			
		DIN-31635	ISO 233-2:1993	EI3	Translittération grand publique
Signes utilisés	Caractères alphabétiques	caractères latins minuscules (sauf dans les premières éditions où des majuscules sont employées)	caractères latins minuscules et majuscules	caractères latins minuscules et majuscules	caractères latins minuscules et majuscules
	Digrammes	Aucun	Aucun	Cinq : th (ث), kh (خ), dh (ذ), sh (ش), gh (غ)	Cinq : th (ث), kh (خ), dh (ذ), sh (ش), gh (غ)
	Diacritiques	macron (ˉ) (̄) macron souscrit (˘) (̣) point suscrit (ˆ) (̂) point souscrit (˙) (̣) caron (ˇ) (̂) caron souscrit (˘) (̣)	macron (ˉ) (̄) macron souscrit (˘) (̣) point suscrit (ˆ) (̂) point souscrit (˙) (̣) caron (ˇ) (̂) tréma (¨) (̈) accent aigu (´) (̂) accent circonflexe (ˆ) (̂)	point souscrit (˙) (̣) pour les consonnes macron (ˉ) (̄) pour les voyelles	Aucun
	Autres caractères	demi-anneaux ' (˘) (̣) demi-anneaux ' (˘) (̣)	demi-anneaux ' (˘) (̣) demi-anneaux ' (˘) (̣)	demi-anneaux ' (˘) (̣) demi-anneaux ' (˘) (̣)	Aucun

¹¹ De ces trois normes découlent un certain nombre de systèmes : Arabica, Hans Wehr Transcription, International Journal of Middle East Studies (IJMES), Library of Congress, etc.

En résumé, nous remarquons que l'ensemble des systèmes de translittération arabe-latin présentés ci-dessus utilisent presque les mêmes caractères. Les deux tableaux ci-dessus (Tab. 1 et Tab. 2) donnent respectivement :

- l'ensemble des caractères adoptés dans les systèmes les plus utilisés : DIN-31635, ISO 233-2 et EI ;
- les éléments utilisés pour la composition de ces caractères.

4. Outils Unicode pour la translittération

Par définition, une translittération d'une écriture avec des caractères d'une autre écriture suppose l'utilisation de l'alphabet cible. Ceci garantit trois des critères essentiels à la définition d'un « système de translittération idéal » : (Perrin, 2016 : 4)

- lisibilité : un code lisible par les humains ;
- prédictabilité : une forme phonétique aussi peu éloignée que possible de la forme d'origine pour les non spécialistes ;
- mnémotechnicité : une mémorisation facile.

C'est ce qu'indique explicitement la norme ISO 233-2 :

Toutefois, on doit accepter que le graphisme obtenu ne puisse toujours être correctement prononcé en suivant les habitudes phonétiques de la langue (ou de toutes les langues) qui utilise(nt) habituellement l'alphabet de conversion. Il doit, en revanche, être tel que le lecteur connaissant la langue convertie puisse reconstituer mentalement sans équivoque le graphisme original et, ainsi, le prononcer correctement. (ISO 233, 1984)

Ainsi, un maximum de caractères de l'alphabet latin est utilisé dans la translittération arabe-latin. Ces caractères sont résumés dans le tableau 3 ci-dessous. Ils sont tous localisés dans le bloc « Latin de base ».

Tableau 3 : Caractères de l'alphabet latin utilisés dans la translittération arabe-latin

Bloc	Caractères utilisés uniquement seuls	Caractères utilisés seuls et avec des diacritiques	Caractères utilisés uniquement avec des diacritiques
Latin de base	b, j, r, f, q, k, l, m, n, w, y (consonnes) B, J, R, F, Q, K, L, M, N, W, Y (consonnes)	t, d, z, s, h (consonnes) a, u, i (voyelles) T, D, Z, S, H (consonnes) A, U, I (voyelles)	g (consonnes) G (consonnes)

¹² Dans le tableau, la nature combinatoire des signes diacritiques est illustrée en leur adjoignant le caractère CERCLE EN POINTILLÉ (U+25CC : ◌).

4.1. Caractères précomposés

Dans la translittération arabe-latin, le nombre de caractères latins utilisables est inférieur à celui de l'alphabet de l'arabe. Dans ce cas, les systèmes proposés font recours à la notion de digraphe ou de diacritique :¹³

Lorsque le nombre de caractères existant dans l'alphabet de conversion est moins élevé que celui de l'alphabet converti, il devient nécessaire de recourir à des digraphes ou à des signes diacritiques. (...) (Ibid.)

Toutefois, l'utilisation des digraphes ou des diacritiques est assujettie à ce que pour une même catégorie de phonèmes, les signes diacritiques employés soient les mêmes :

(...) On doit, dans ce cas, éviter le plus possible l'arbitraire et l'emploi de signes purement conventionnels, mais chercher à maintenir une certaine rationalité phonétique afin de rendre le système plus généralement acceptable. (Ibid.)

Ainsi, le point souscrit désigne toujours des lettres emphatiques, le macron les voyelles longues... Ceci renforcera davantage les trois critères précédents : lisibilité, prédictabilité et mnémotechnicité.

Parmi les caractères diacrités proposés dans Unicode, la translittération arabe-latin a bénéficié dès la première version (1.0.0 d'octobre 1991), d'un ensemble de 16 caractères précomposés distribués sur les trois blocs *Supplement Latin-1*, *Latin étendu-A* et *Latin étendu-B* et dans la troisième version (1.1 de juin 1993), de 18 autres ; ce qui fait un total de 34 caractères précomposés. Ces caractères sont présentés dans le tableau 4 ci-après.

Les caractères diacrités les plus fréquents existent sous la forme de caractères précomposés (c.-à-d. codés sous la forme d'un seul caractère Unicode) et ceci pour assurer une compatibilité avec les codages préexistants les plus répandus. Cependant, l'écriture latine est actuellement utilisée pour l'écriture et la translittération d'un grand nombre de langues dans le monde ; et pourra même être utilisée par d'autres langues dans le futur, ce qui pourrait étendre son répertoire en caractères de translittération. La solution de l'adjonction de nouveaux caractères précomposés à chaque nouvelle construction ou révision d'un

¹³ Les translittérations grand public utilisent uniquement des digraphes, DIN-31635 et ISO 233-2 uniquement des diacritiques alors que EI3 utilise les deux en même temps.

alphabet est ainsi non seulement consommatrice d'espace de codage mais imprévisible aussi. Pour gérer cette situation, Unicode ne code plus de caractères précomposés depuis bien longtemps et encourage ses utilisateurs à utiliser les signes diacritiques et le mécanisme de la combinaison des caractères pour exprimer toutes les lettres diacritées existantes ainsi que d'autres encore susceptibles d'apparaître dans l'avenir. C'est ce que nous allons voir dans la sous section subséquente.

Tableau 4 : Caractères diacrités pour la translittération arabe-latin dans Unicode

Bloc	Caractère disponible	Nom du caractère	Point de code	Introduit dans la version
Supplement Latin-1	Á	LATIN CAPITAL LETTER A WITH ACUTE	U+00C1	1.0.0 (October 1991)
	Â	LATIN CAPITAL LETTER A WITH CIRCUMFLEX	U+00C2	« « «
	à	LATIN SMALL LETTER A WITH ACUTE	U+00E1	« « «
	â	LATIN SMALL LETTER A WITH CIRCUMFLEX	U+00E2	« « «
Latin étendu-A	Ā	LATIN CAPITAL LETTER A WITH MACRON	U+0100	« « «
	ā	LATIN SMALL LETTER A WITH MACRON	U+0101	« « «
	Ġ	LATIN CAPITAL LETTER G WITH DOT ABOVE	U+0120	« « «
	ġ	LATIN SMALL LETTER G WITH DOT ABOVE	U+0121	« « «
	Ī	LATIN CAPITAL LETTER I WITH MACRON	U+012A	« « «
	ī	LATIN SMALL LETTER I WITH MACRON	U+012B	« « «
	Š	LATIN CAPITAL LETTER S WITH CARON	U+0160	« « «
	š	LATIN SMALL LETTER S WITH CARON	U+0161	« « «
Latin étendu-B	Ū	LATIN CAPITAL LETTER U WITH MACRON	U+016A	« « «
	ū	LATIN SMALL LETTER U WITH MACRON	U+016B	« « «
Latin étendu-B	Ĝ	LATIN CAPITAL LETTER G WITH CARON	U+01E6	« « «
	ĝ	LATIN SMALL LETTER G WITH CARON	U+01E7	« « «
Latin étendu additionnel	Đ	LATIN CAPITAL LETTER D WITH DOT BELOW	U+1E0C	1.1 (June 1993)
	đ	LATIN SMALL LETTER D WITH DOT BELOW	U+1E0D	« « «
	Ɔ	LATIN CAPITAL LETTER D WITH LINE BELOW	U+1E0E	« « «
	ɔ	LATIN SMALL LETTER D WITH LINE BELOW	U+1E0F	« « «
	Ĥ	LATIN CAPITAL LETTER H WITH DOT BELOW	U+1E24	« « «
	ĥ	LATIN SMALL LETTER H WITH DOT BELOW	U+1E25	« « «
	Ħ	LATIN CAPITAL LETTER H WITH BREVE BELOW	U+1E2A	« « «
	ħ	LATIN SMALL LETTER H WITH BREVE BELOW	U+1E2B	« « «
	Ș	LATIN CAPITAL LETTER S WITH DOT BELOW	U+1E62	« « «
	ș	LATIN SMALL LETTER S WITH DOT BELOW	U+1E63	« « «
	Ț	LATIN CAPITAL LETTER T WITH DOT BELOW	U+1E6C	« « «
	ț	LATIN SMALL LETTER T WITH DOT BELOW	U+1E6D	« « «
	Ț	LATIN CAPITAL LETTER T WITH LINE BELOW	U+1E6E	« « «
	ț	LATIN SMALL LETTER T WITH LINE BELOW	U+1E6F	« « «
	Z	LATIN CAPITAL LETTER Z WITH DOT BELOW	U+1E92	« « «
	z	LATIN SMALL LETTER Z WITH DOT BELOW	U+1E93	« « «
Ƨ	LATIN SMALL LETTER H WITH LINE BELOW	U+1E96	« « «	
ƨ	LATIN SMALL LETTER T WITH DIAERESIS	U+1E97	« « «	

4.2. Caractères combinés

L'alphabet latin étant utilisé — et pourra l'être éventuellement dans le futur — dans beaucoup de langues, soit directement dans leur écriture soit dans la translittération de celle-ci vers le latin, il est alors difficile

d'imaginer le nombre de caractères Unicode précomposés nécessaires pour couvrir les besoins de toutes ces langues ; de nouveaux caractères pouvant toujours venir s'ajouter à ceux existants. Cette problématique fut, dès le départ, prise en charge par Unicode et est garantie par deux des dix¹⁴ principes sur lesquels s'est construit ce standard : (Andries, 2002: 64)

COMPOSITION DYNAMIQUE. On compose les formes à diacritique(s) dynamiquement.

SEQUENCE EQUIVALENTE. À chaque forme statique précomposée correspond au moins une suite équivalente de caractères décomposés que l'on pourra recomposer dynamiquement.

Le premier principe stipule que les formes à diacritiques sont composées dynamiquement ; en d'autres termes, il est toujours possible de combiner un caractère de base avec un caractère combinatoire¹⁵. Ainsi, les caractères de translittération contenant des diacritiques sont obtenus par composition dynamique en adjoignant un (ou plusieurs) caractère(s) diacritique(s), dit(s) caractère(s) combinatoire(s), à une (ou plusieurs) voyelle(s) ou à une (ou plusieurs) consonne(s), dite(s) caractère(s) de base. Le second principe assure qu'il doit exister pour n'importe quel caractère précomposé déjà existant un rendu équivalent sous la forme d'une suite de caractères composables entre eux. Ainsi, tous les caractères de translittération précomposés existants peuvent aussi être obtenus par composition dynamique.

Les caractères combinatoires à chasse nulle sont repris dans Unicode dans le bloc « Combining Diacritical Marks » (« Marques diacritiques

¹⁴ Le standard Unicode énumère dix principes qui ont guidé son élaboration : (i) universalité, (ii) efficacité, (iii) caractères et non glyphes, (iv) sémantique, (v) texte brut, (vi) ordre logique, (vii) unification, (viii) composition dynamique, (ix) séquence équivalente et (x) convertibilité. (Andries, 2002:64-71)

¹⁵ En Unicode, les caractères combinatoires sont des caractères destinés à s'afficher ensemble avec un caractère de base. Il en existe deux types : ceux qui chassent et ceux qui ne chassent pas. (Andries, 2002:61-2) Il convient de distinguer ici les caractères combinatoires à chasse nulle (c.-à-d., ceux qui n'occupent pas de place de leur propre chef mais plutôt un même espace avec les caractères de base auxquels ils s'adjoignent), et qui sont pour la plupart des diacritiques, des lettres modificatives qui, elles, chassent. Une lettre modificative est un caractère qui ressemble à une lettre ou à de la ponctuation et est généralement utilisée pour modifier la prononciation d'une autre lettre ; mais elle peut aussi être utilisée seule.

d'association »).¹⁶ Parmi les 112 caractères de ce bloc (0300..036F), neuf sont utilisés en translittération arabe-latin : (i) accent aigu, (ii) accent circonflexe, (iii) macron, (iv) point suscrit (point en chef), (v) tréma, (vi) caron, (vii) point souscrit, (viii) brève souscrite, (ix) macron souscrit (voir le tableau 5 ci-après).

Tableau 5 : Caractères combinatoires Unicode utilisés en translittération arabe-latin

Bloc	Caractère combinatoire	Nom du caractère	Points de code	Introduit dans la version	Caractère de base	Caractère combinatoire
Marques diacritiques d'association (Signes diacritiques)	◌◌̣	COMBINING ACUTE ACCENT	U+0301	1.0.0 (October 1991)	a	á
	◌◌̂	COMBINING CIRCUMFLEX ACCENT	U+0302	« « «	a	â
	◌◌̄	COMBINING MACRON	U+0304	« « «	a, u, i	ā, ū, ī
	◌◌̇	COMBINING DOT ABOVE	U+0307	« « «	g	ḡ
	◌◌̈	COMBINING DIAERESIS	U+0308	« « «	t	ṫ
	◌◌̆	COMBINING CARON	U+030C	« « «	g, s	ḡ, š
	◌◌̣̇	COMBINING DOT BELOW	U+0323	« « «	h, s, d, t, z	ḥ, ṣ, ḏ, ṭ, ḏ
	◌◌̣̈	COMBINING BREVE BELOW	U+032E	« « «	h	ḥ̣
	◌◌̣̄	COMBINING MACRON BELOW	U+0331	« « «	t, d, h	ṭ̇, ḏ̣, ḥ̣

Les caractères combinatoires suivent toujours le caractère de base auquel ils s'adjoignent. Par exemple, le caractère Unicode U+0064 LATIN SMALL LETTER D « d » suivi du caractère combinatoire U+0331 COMBINING MACRON BELOW « ◌̣̄ »¹⁷ suivi lui-même du caractère Unicode U+0061 LATIN SMALL LETTER A « a » représente sans ambiguïté « ḏa » et non « ḏ̣a ». Si plusieurs caractères combinatoires s'adjoignent à un même caractère de base (l'un à la suite de l'autre), ils sont dessinés en association avec le glyphe de ce caractère selon leur ordre d'apparition. Par exemple, le caractère Unicode U+0064 LATIN SMALL LETTER D « d » suivi du caractère combinatoire U+0331 COMBINING MACRON BELOW « ◌̣̄ » suivis tous les deux du caractère U+0323 COMBINING DOT BELOW « ◌̣̇ » représente sans ambiguïté « ḏ̣̇ » et non « ḏ̣̣̇ ».

4.3. Lettres modificatives

La lettre modificative est un signe qui a un double statut : d'une part, elle est considérée comme une lettre à part entière du fait qu'elle chasse — en

¹⁶ D'autres caractères combinatoires ont été ajoutés par la suite : les blocs « Combining Diacritical Marks Extended » (1AB0-1AFF), « Combining Diacritical Marks Supplement » (1DC0-1DFF) et « Combining Diacritical Marks for Symbols » (20D0-20FF).

¹⁷ Pour bien indiquer qu'il s'agit d'un diacritique, le caractère combinatoire « ◌̣̄ » est représenté « ◌̣̄̇ » : le macron souscrit est placé au-dessous d'un cercle pointillé « ◌̣̄̇ ».

d'autres termes, elle occupe une espace—, et d'autre part, elle possède les propriétés d'un diacritique puisqu'elle a pour rôle essentiel la modification de la lettre à laquelle elle est associée, notamment en phonétique où elle modifie la prononciation de la lettre qu'elle suit. Ces caractères se trouvent dans le bloc Unicode « Spacing Modifier Letters » (« Lettres du modificateur d'espace »). Parmi les 80 caractères de ce bloc (02B0..02FF), quatre sont généralement utilisés en translittération arabe-latin : (i) lettre modificative apostrophe, (ii) lettre modificative virgule réfléchie, (iii) lettre modificative demi-anneau à droite, (iv) lettre modificative demi-anneau à gauche (voir le tableau 6 ci-après). Les deux derniers étant les plus utilisés.

Tableau 6 : Lettres modificatives Unicode utilisés en translittération arabe-latin

Bloc	Caractères	Nom du caractère	Point de code	Lettre arabe représentée
Lettres du modificateur d'espace (lettres modificatives avec chasse)	'	MODIFIER LETTER APOSTROPHE	02BC	Coup de glotte ou <i>hamza</i> : ء (U+0621)
	◌◌	MODIFIER LETTER REVERSED COMMA	02BD	lettre arabe 'ayn : ع (U+0639)
	◌◌◌	MODIFIER LETTER RIGHT HALF RING	02BE	Coup de glotte
	◌◌◌◌	MODIFIER LETTER LEFT HALF RING	02BF	lettre arabe 'ayn

Dans le cadre de la translittération arabe-latin, ces lettres modificatives sont utilisées comme étant des lettres à part entière.

5. Conclusion

Dans cet article, nous avons pu synthétiser tout ce que dont un scripteur a besoin pour appréhender le domaine de la translittération. En particulier, nous avons présenté l'outillage utilisé dans le domaine de la translittération arabe-latin. Le fait que l'intérêt à la translittération n'est pas directement perçu par le grand public ne réduit pas son importance comme une problématique à l'échelle nationale et internationale. Au niveau national, il suffit de voir autour de nous les anomalies attestées dans l'écriture et la prononciation des noms propres et communs et qui s'héritent de génération en génération ; jusqu'à ce jour, aucune solution générale n'a été proposée au débat. Au niveau international, il suffit de voir la multitude de translittérations qui existent pour un même nom propre arabe : l'exemple anecdotique du nom de l'ancien président de la Libye Mouammar Kadhafi qui est transcrit en latin par plus de soixante formes en est un exemple répandu.

Références bibliographiques

- Andries, P. (2002). Introduction à Unicode et à l'ISO 10646. *Document numérique*, 6, 51-88. <https://doi.org/10.3166/dn.6.3-4.51-88>
- « DIN-31635 ». Consulté le 15 juillet 2021. <http://fracademic.com/dic.nsf/frwiki/484318>.
- « Encyclopédie de l'Islam | Brill ». Consulté le 15 juillet 2021.
- Campo, J. E. 2009. *Encyclopedia of Islam*. Infobase Publishing.
- <http://www.brill.com/publications/encyclopedie-de-lislam>.
- ISO 233:1984(fr), Documentation — Translittération des caractères arabes en caractères latins. Consulté le 15 juillet 2021. <https://www.iso.org/obp/ui/#iso:std:iso:233:ed-1:v1:fr>.
- ISO 233-2:1993(en), Information and documentation — Transliteration of Arabic characters into Latin characters — Part 2: Arabic language — Simplified transliteration. Consulté le 15 juillet 2021. <https://www.iso.org/obp/ui/fr/#iso:std:iso:233:-2:ed-1:v1:en>.
- Perrin E. (2016). Signifiants et signifié : la translittération de la langue arabe. In Barthélemy, G. Casajus D. Larzul S. Volait M. (eds), *L'Orientalisme après la Querelle. Sur les pas de François Pouillon*, Karthala. <https://hal.archives-ouvertes.fr/hal-01411439/document>.
- Rodinson, M. 1964. Les principes de la translittération, la translittération de l'arabe et la nouvelle norme de l'ISO. *Bulletin des bibliothèques de France (BBF)*, n° 1, p. 1-24. Disponible en ligne : <<http://bbf.enssib.fr/consulter/bbf-1964-01-0001-001>>.
- The Unicode Consortium. 2020. *The Unicode Standard, 13.0.0 - Core Specification*. <https://unicode.org/versions/Unicode13.0.0/>