

Université Cadi Ayyad
Faculté des Sciences
Marrakech - MAROC

TransTec
**Une translittération arabe/roman
pour un e-document**

Mostafa BANOUNI, Azzeddine LAZREK et Khalid SAMI

rydarab@ucam.ac.ma

www.ucam.ac.ma/fssm/rydarab

CFD'02 - CIDE'5

Plan

1. Introduction

- Présentation
- Caractéristiques

2. Classification

- Translittérations scripturales
- Translittérations électroniques

3. La translittération **TransTec**

- Textuel
- Symbolique

4. Conclusion

1 Introduction

1.1 Présentation

Objet : Une translittération arabe/roman pour un e-document scientifique

Translittération : \mathcal{C} (Alphabet arabe) \longrightarrow \mathcal{C} (Alphabet roman)
 علم \longmapsto Elm

Intérêt : bibliographiques, notices, documents officiels, e-communication, ...

Remarques :

Transcription : \mathcal{C} (Alphabet roman) \longrightarrow \mathcal{C} (API)
 science \longmapsto syans'

Encodage : \mathcal{C} (Alphabet roman) \longrightarrow \mathcal{C} (ASCII)
 science \longmapsto 73 63 69 65 6E 63 65_h
 TROIS (3 ou ʀ) \longmapsto 33_h

Translittération repose sur les correspondances : phonologique et/ou morphologique

Problèmes :

- Langue de référence

خ KHAH \longmapsto j : jota en espagnol castillan
 ج JEEM \longmapsto j : en français

- Correspondance

ض DAD \longmapsto d : non ASCII
 ض DAD \longmapsto .d : symboles, traitement, mémoire
 ع AIN \longmapsto * , ` , e , E , a , A : absence de phonétique

Présence de plusieurs translittérations

1.2 Caractéristiques

Bi-univocité : l'application entre les éléments des deux alphabets est une bijection ;

Complétude : la translittération couvre tous les éléments du système alphabétique arabe ;

Compacité : le nombre de signes utilisés pour la translittération d'un élément est optimal ;

Portabilité : le texte transcrit est transportable sur tout système informatique ;

Lisibilité humaine : le texte peut être facilement lu par un être humain sans trop s'éloigner de sa forme phonétique originale ;

Affichabilité : le texte résultat peut être affiché sur un écran ASCII ou à l'aide d'une imprimante standard ;

Mnémotechnicité : la translittération est facilement mémorisable.

Généralement, ces propriétés présentent quelques antagonismes. Il faut chercher un compromis.

2 Classification

Les translittérations en usage peuvent être classées en deux catégories :

Translittérations scripturales : les translittérations traditionnellement utilisées pour l'écriture manuelle ;

Translittérations électroniques : les translittérations utilisées pour des textes susceptibles d'être traités par des systèmes informatiques, ne comprenant que des caractères ASCII.

2.1 Translittérations scripturales

Orientalist Scholars : la convention internationale de l'Orientalist Scholars adopte une proposition de translittération faite à Rome (1936).

EI (Encyclopedia of Islam) : utilise une translittération basée sur les conventions normatives anglaises. Les diacritiques et les consonnes sont voyellisées (1913-1960).

BS 4280 (British Standard 4280) : propose une translittération consistante mais elle n'a pas été largement utilisée (1969-1983).

Beyrouthamendé : a été adoptée par un grand nombre de pays arabes (1972).

ISO 233 (ou 46-002) : la norme ISO de translittération arabe (1984).

ISO 233-2 : la norme ISO de translittération simplifiée Partie 2 (1993).

ALESCO (Arab League Educational, Scientific and Cultural Organization) : la translittération utilisée par l'ALESCO.

SAWS (Scientific Arabic Writing Systems) : la translittération adoptée lors d'une conférence sur les systèmes de l'écriture arabe scientifique (1995).

Lettre	Nom	EI	ALESCO	SAWS
ا	ALEF	a	a	a
ب	BEH	b	b	b
ت	TEH	t	t	t
ث	THEH	<u>th</u>	<u>t</u>	<u>th</u>
ج	JEEM	<u>dj</u>	j	j
ح	HAH	h	h	h
خ	KHAH	<u>kh</u>	<u>k</u>	<u>kh</u>
د	DAL	d	d	d
ذ	THAL	<u>dh</u>	<u>d</u>	<u>dh</u>
ر	REH	r	r	r
ز	ZAIN	z	z	z
س	SEEN	s	s	s
ش	SHEEN	<u>sh</u>	<u>s</u>	<u>sh</u>
ص	SAD	ṣ	ṣ	ṣ
ض	DAD	ḍ	ḍ	ḍ
ط	TAH	ṭ	ṭ	ṭ
ظ	ZAH	ẓ	Ḍ	ẓ
ع	AIN	‘	‘	‘
غ	GHAIN	<u>gh</u>	<u>g</u>	<u>gh</u>
ف	FEH	f	f	f
ق	QAF	q	q	q
ك	KAF	k	k	k
ل	LAM	l	l	l
م	MEEM	m	m	m
ن	NOON	n	n	n
ه	HEH	h	h	h
ة	TEH MARBUTA		h ou t	t

Suite

Lettre	Nom	EI	ALESCO	SAWS
و	WAW	w	w	w
ي	YEH	y	y	y
ى	ALEF MAKSURA			
-	FATHA	a	a	a
ـ	DAMMA	u	u	u
-	KASRA	i	i	i
=	FATHATAN	an	an	(an)
ــ	DAMMATAN	un	un	(un)
=	KASRATAN	in	in	(in)
ــ	SHADDA	a	a	a
◌	SUKUN	b	b	b
ء	HAMZA ON LINE	.	.	.
ء	HAMZA ON ALEF	a	a	a ou 'a
ء	HAMZA UNDER ALEF	i	i	i ou 'i
ء	HAMZA ON WAW	u	u	u ou 'u
ء	HAMZA ON YEH			'i
ء	MADDA ON ALEF	ā	ā	ā
ء	WASLA ON ALEF			
-	KASHIDA			

Table 1: Quelques translittérations scripturales

^aLa voyelle SHADDA d'une lettre est indiquée par le dédoublement de la lettre en question.

^bLa voyelle SUKUN est omise.

2.2 Translittérations électroniques

Qalam : la translittération morphologique développée par A. Heddaya en contribution avec W. Hamdy et M. H. Sherif, (1985-1992).

ditroff/ffortid : la translittération utilisée dans le ditroff/ffortid développé par J. Srouji et D. Berry.

Arab $\text{T}_{\text{E}}\text{X}$: la translittération utilisée dans l'extension de $\text{T}_{\text{E}}\text{X}$, Arab $\text{T}_{\text{E}}\text{X}$ développée par K. Lagally. Cette translittération utilise certains symboles tels que $\hat{\ } , _$ et \cdot ayant une signification particulière en mode mathématique en $\text{T}_{\text{E}}\text{X}$.

Ω : la translittération utilisée dans le système multilingue Ω développé par Y. Haralambous et J. Plaice. Ω offre à l'utilisateur la possibilité de définir sa propre translittération.

Buckwalter : la translittération de l'orthographe de l'arabe standard moderne développée par le lexicographe T. Buckwalter. Cette translittération est une véritable transcription orthographique.

ARABVISL (ARABic Visual Interactive Syntax Learning) : la translittération utilisée dans le projet de recherche VISL de l'Institut de Langage et de Communication de l'Université du Danemark (1996).

Lettre	Nom	QALAMditroff/ffortid	ArabTeX	Ω	Buckwalter	ARABVISE	
ا	ALEF	aa	a	a ou A ou _a	A	A	aa
ب	BEH	b	b	b	b	b	b
ت	TEH	t	t	t	t	t	t
ث	THEH	th	c	_t	th	v	th
ج	JEEM	j	j	^g	j	j	j
ح	HAH	H	h	.h	H	H	H
خ	KHAH	kh	x	_h	kh	x	kh
د	DAL	d	d	d	d	d	d
ذ	THAL	dh	Z	_d	dh	*	dh
ر	REH	r	r	r	r	r	r
ز	ZAIN	z	z	z	z	z	z
س	SEEN	s	s	s	s	s	s
ش	SHEEN	sh	C	^s	sh	\$	sh
ص	SAD	S	S	.s	S	S	S
ض	DAD	D	D	.d	D	D	D
ط	TAH	T	T	.t	T	T	T
ظ	ZAH	Z	Z'	.z	Z	Z	Z
ع	AIN	'	e	'	'	E	*
غ	GHAIN	gh	R'	.g	gh	g	gh
ف	FEH	f	f	f	f	f	f
ق	QAF	q	q	q	q	q	q
ك	KAF	k	k	k	k	k	k
ل	LAM	l	l	l	l	l	l
م	MEEM	m	m	m	m	m	m
ن	NOON	n	n	n	n	n	n
ه	HEH	h	H	h	-h	h	h
ة	TEH MARBUTA	t ou h	t'	T	"h ou "t	p	-a

Suite

Lettre	Nom	QALAM	ditroff/ffortid	ArabTeX	Ω	Buckwalter	ARABVISL
و	WAW	w	w	w ou U ou _u	w ou U	w	w ou uu
ي	YEH	y	y	y ou I ou _i	y	y	y ou ii
ى	ALEF MAKSURA	ae	A'	Y ou _A	I	Y	aa
-	FATHA	a	'	a	a	a	a
ـ	DAMMA	u	u	u	u	u	u
-	KASRA	i	E	i	i	i	i
ـ	FATHATAN	aN		aN	aN	F	an
ـ	DAMMATAN	uN		uN	uN	N	un
ـ	KASRATAN	iN		iN	iN	K	in
ـ	SHADDA	a	~~	a		~	a
ـ	SUKUN	-	O	b	<>	o	b
ء	HAMZA ON LINE	'	'	'		'	'
أ	HAMZA ON ALEF		A	'a	'a	>	a
إ	HAMZA UNDER ALEF		i	'i	'i	<	i
ؤ	HAMZA ON WAW		w'	'w	'u	&	
ي	HAMZA ON YEH		Y'	'y	'y	}	
آ	MADDA ON ALEF	~aa	a~	'A	'A		~aa
أ	WASLA ON ALEF	e	U		"A	{	

Table 2: Quelques translittérations électroniques

^aLa voyelle SHADDA d'une lettre est indiquée par le dédoublement de la lettre en question.

^bLa voyelle SUKUN est omise.

3 La translittération TransTec

Besoin : une même translittération en mode mathématique et en mode texte pour un système de composition du **document mathématique arabe** où les expressions symboliques arabes sont composées à l'aide de *symboles arabes spécifiques* et dont l'écriture se déroule de la *droite vers la gauche*.

$$9 + \frac{\sqrt{2} \sqrt{s} + \left(\sum_{1=s}^{\tau} \text{جتا} \right)}{4 - \tau 3}$$

Règles :

- faire une analyse phonétique et morphologique ;
- conserver l'utilisation des lettres minuscules ASCII qui sont habituellement utilisées dans les translittérations ;
- utiliser les lettres majuscules pour translittérer toutes les autres lettres alphabétiques. Le trait majuscule n'est plus utilisé pour respecter les conventions typographiques (les noms propres, le début d'une phrase, etc.) ;
- limiter, le plus que possible, l'utilisation aux lettres romanes du code ASCII ;
- minimiser le nombre de lettres utilisées.

3.1 Textuel

• Lettres

– *Phonétique communément d'usage*

Lettre	Nom	Translittération
ا	ALEF	a
ب	BEH	b
ت	TEH	t
ج	JEEM	j
د	DAL	d
ر	REH	r
ز	ZAIN	z
س	SEEN	s
ف	FEH	f
ك	KAF	k
ل	LAM	l
م	MEEM	m
ن	NOON	n
ه	HEH	h

– *Phonétique d'usage restreint*

Lettre	Nom	Translittération	Remarques
خ	KHAH	X	Xavier (espagnol), lettre χ (grec)
ش	SHEEN	C	Cinco (italien)
ص	SAD	S	Saut (\neq six) (français)
ض	DAD	D	Dalle (\neq dix) (français)
ط	TAH	T	Taux (\neq titre) (français)

– *Phonétique approchée*

Lettre	Nom	Translittération	Remarques
ث	THEH	F	tooth (anglais)
ذ	THAL	Z	mother (anglais)
ظ	ZAH	V	

– *Phonétique rare*

Lettre	Nom	Translittération
ح	HAH	H
ع	AIN	E
غ	GHAIN	G
ق	QAF	q

– *Correspondance morphologique*

Lettre	Nom	Translittération
ة	TEH MARBUTA	Q
ع	AIN	E

• Voyelles

Lettre	Nom	Translittération
و	WAW	w
ي	YEH	y
ى	ALEF MAKSURA	Y
-	FATHA	e
و	DAMMA	u
-	KASRA	i
=	FATHATAN	eN ou ee
»	DAMMATAN	uN ou uu
=	KASRATAN	iN ou ii
ّ	SHADDA	W
°	SUKUN	o

• HAMZA, MADDA et WASLA

Lettre	Nom	Translittération
ء	HAMZA ON LINE	A-
أ	HAMZA ON ALEF	Aa
إ	HAMZA UNDER ALEF	AY
ؤ	HAMZA ON WAW	Aw
ي	HAMZA ON YEH	Ay
أ	MADDA ON ALEF	Ma
أ	WASLA ON ALEF	La

• KASHIDA

La KASHIDA est l'allongement de certaines lettres. Ce n'est pas une lettre de l'alphabet.

KASHIDA \longmapsto - ou K ou rien

3.2 Symbolique

• **Ponctuation** ، . : ! ? \longrightarrow , . ; : ! ?

• **Chiffres** 0 1 2 3 4 5 6 7 8 9

ou \longrightarrow 0 1 2 3 4 5 6 7 8 9

٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩

• **Délimitants** } ||] |) > \longrightarrow { || [| (<

{ || [| (< \longrightarrow } ||] |) >

• **Opérateurs** + - * \longrightarrow + - *

> \longrightarrow <

< \longrightarrow >

• Symboles littéraux de désignation

En mathématique arabe :

- il y a moins de lettres arabes. En effet, les lettres ne portent pas de point diacritique et elles ne sont pas voyellisées. Par exemple, les trois lettres BEH, TEH et THEH (ب ت ث) donnent le seul caractère (ب) prononcé BEH.
- l'écriture symbolique n'est pas cursive. Par exemple, l'angle ALEF, BEH et JEEM se note (ا ا ب ou ا ب ا).
- les lettres ne suivent pas le même ordre que celui de l'alphabet, par exemple, (ح) pour JEEM et non HAH mais (ز) pour REH et non ZAIN.
- les traits de caractères : minuscule, MAJUSCULE, **gras** et *italic* n'existent pas en arabe. Par contre, il y a la forme initiale et isolée des lettres (ح et ح), la forme avec queue (ح) et la forme avec contour (ح, ح et ح). Ces formes permettent d'élargir la liste des symboles de désignation mais elles doivent être **translittérées différemment**.

La translittération proposée pour les symboles littéraux arabes dans une expression mathématique repose sur :

- l'utilisation des lettres minuscules pour la forme finale, qui est la forme la plus utilisée en mathématique arabe

$$\text{ج} \longmapsto j$$

- l'utilisation des lettres majuscules pour la forme initiale, si le glyphe de ce symbole diffère de celui de la forme finale

$$\text{ج} \longmapsto J$$

- l'utilisation des commandes pour les formes avec les attributs queue et/ou contour

$$\text{ج} \longmapsto \backslash\text{JEEM}$$

$$\text{ج} \longmapsto \backslash\text{jjeem}$$

$$\text{ج} \longmapsto \backslash\text{JJeem}$$

$$\text{ج} \longmapsto \backslash\text{JJEEM}$$

Exemple :

adrs alda1Q \$d\$ altalyQ \$:\$

```

$$d(c) = {\cases{
-4{\sqrt{c}}&{{\hbox{AYZa kan}}c<0}\cr
\quad 4\cos c&{{\hbox{AYZa kan}}c>0}\cr
-2&{{\hbox{Gyr Zlk}}}\cr}}

```

ادرس الدالة د التالية :

$$\left. \begin{array}{l} \text{إذا كان } c > 0 \\ \text{إذا كان } c < 0 \\ \text{غير ذلك} \end{array} \right\} = d(c) \begin{array}{l} -4\sqrt{c} \\ 4 \\ -2 \end{array}$$

$$\left. \begin{array}{l} \text{إذا كان } c > 0 \\ \text{إذا كان } c < 0 \\ \text{غير ذلك} \end{array} \right\} = d(c) \begin{array}{l} -4\sqrt{c} \\ 4 \\ -2 \end{array}$$

$$d(c) = \begin{cases} -4\sqrt{c} & c < 0 \\ 4 \cos c & c > 0 \\ -2 & \text{غير ذلك} \end{cases}$$

$$d(c) = \begin{cases} -4\sqrt{c} & \text{si } c < 0 \\ 4 \cos c & \text{si } c > 0 \\ -2 & \text{sinon} \end{cases} \quad d(c) = \begin{cases} -4\sqrt{c} & \text{if } c < 0 \\ 4 \cos c & \text{if } c > 0 \\ -2 & \text{otherwise} \end{cases}$$

Lettre	Nom	Texte	Expression mathématique					
ا	ALEF	a	ا	a				
ب	BEH	b	ب	b	ر	B	ب	\BEH
ت	TEH	t						
ث	THEH	F						
ج	JEEM	j	ح	j	ج	J	ج	\JEEM
ح	HAH	H						
خ	KHAH	X						
د	DAL	d	د	d				
ذ	THAL	Z						
ر	REH	r	ر	r				
ز	ZAIN	z						
س	SEEN	s						
ش	SHEEN	C	س	c	س	C	س	\SHEEN
ص	SAD	S	ص	s	ص	S	ص	\SAD
ض	DAD	D						
ط	TAH	T	ط	T			ط	\TAH
ظ	ZAH	V						
ع	AIN	E	ع	e	ء	E	ع	\AIN
غ	GHAIN	G						
ف	FEH	f	ف	f	و	F	ف	\FEH
ق	QAF	q	ق	q				
ك	KAF	k	ك	k			ك	\KAF
ل	LAM	l	ل	l	ل	L	ل	\LAM
م	MEEM	m	م	m	م	M	م	\MEEM
ن	NOON	n	ن	n	ن	N		
ه	HEH	h			ه	H	ه	\HEH
ة	TEH MARBUTA	Q						

Suite

Lettre	Nom	Texte	Expression mathématique				
و	WAW	w	و	w			
ي	YEH	y	ی	y	ے	Y	ی \YEH
ى	ALEF MAKSURA	Y					
-	FATHA	e					
-	DAMMA	u					
-	KASRA	i					
=	FATHATAN	eN ou ee					
»	DAMMATAN	uN ou uu					
=	KASRATAN	iN ou ii					
»	SHADDA	W					
o	SUKUN	o					
ء	HAMZA ON LINE	A-	ء	A			
أ	HAMZA ON ALEF	Aa					
إ	HAMZA UNDER ALEF	AY					
ؤ	HAMZA ON WAW	Aw					
ي	HAMZA ON YEH	Ay					
آ	MADDA ON ALEF	Ma					
أ	WASLA ON ALEF	La					
-	KASHIDA	K ou -					

Table 3: La translittération **TransTec**

Code	Commande	Glyphe	Code	Commande	Glyphe	Code	Commande	Glyphe
32	\alef	ا	64	\aalef	ا	96		
33	\beh	ب	65	\bbeh	ب	97	\BEH	ب
34	\jeem	ج	66	\jjeem	ج	98	\JEEM	ج
35	\dal	د	67	\ddal	د	99		
36	\waw	و	68	\wwaw	و	100		
37	\reh	ر	69	\rreh	ر	101		
38	\tah	ط	70	\ttah	ط	102	\TAH	ط
39	\yeh	ي	71	\yyeh	ي	103		
40	\lam	ل	72	\llam	ل	104	\LAM	ل
41	\meem	م	73	\mmeem	م	105	\MEEM	م
42	\noon	ن	74	\nnoon	ن	106		
43	\sheen	س	75	\ssheen	س	107	\SHEEN	س
44	\ain	ع	76	\aain	ع	108	\AIN	ع
45	\feh	ف	77	\ffeh	ف	109	\FEH	ف
46	\sad	ص	78	\ssad	ص	110	\SAD	ص
47	\qaf	ق	79	\qqaf	ق	111		
48	\hamza	ء	80	\hhamza	ء	112	\HEH	ء
49	\lamalef	لا	81	\llamalef	لا	113	\KAF	ك
50	\Yeh	ي	82	\YYeh	ي	114	\TTAH	ط
51	\Noon	ن	83	\NNoon	ن	115		
52	\Reh	ر	84	\RReh	ر	116		
53	\MEem	م	85	\MMEem	م	117		
54	\Beh	ب	86	\BBeh	ب	118	\BBEH	ب
55	\Jeem	ج	87	\JJeem	ج	119	\JJEEM	ج
56	\Heh	ه	88	\HHeh	ه	120	\HHEH	ه
57	\Kaf	ك	89	\KKaf	ك	121	\KKAF	ك
58	\Lam	ل	90	\LLam	ل	122	\LLAM	ل
59	\Meem	م	91	\MMeem	م	123	\MMEEM	م
60	\Sheen	س	92	\SSheen	س	124	\SSHEEN	س
61	\Ain	ع	93	\AAin	ع	125	\AAIN	ع
62	\Feh	ف	94	\FFeh	ف	126	\FFEH	ف
63	\Sad	ص	95	\SSad	ص	127	\SSAD	ص

4 Conclusion

La translittération **TransTec** :

- se base sur les deux approches : phonétique et morphologique ;
- utilise quasi exclusivement les lettres romanes sous leurs deux traits d'écriture majuscule et minuscule ;
- préserve au maximum la correspondance en usage dans la plupart des translittérations existantes ;
- est non ambiguë, complète, compacte (une correspondance lettre à lettre sauf pour les lettres ou les signes composés), portable (ASCII standard) mnémotechnique et relativement lisible ;
- est utilisable en mode texte et en mode mathématique.

mais

- aucune translittération n'est simple d'utilisation ;
- besoin d'interfaçage pour une translittération non transparente.

TransTec est un outil pour la composition et la traduction du document scientifique arabe.