

Convertisseur pour la langue amazighe : script arabe - latin - tifinaghe

Ataa Allah Fadoua, Boulaknadel Siham

Centre d'Etude Informatique,
Systèmes d'Information et de Communication
Institut Royal de la Culture Amazighe
Avenue Allal El Fassi, Madinat Al Irfane, Rabat – Instituts ;
Adresse postale BP 2055, Hay Riad, Rabat
{ataaallah, boulaknadel}@ircam.ma

Résumé – Abstract

L'amazighe en tant que langue et culture a connu un processus de standardisation et d'intégration dans les nouvelles technologies de l'information et de la communication, qui est passé par plusieurs étapes : un codage spécifié par l'ASCII étendu, un codage propre dans le standard Unicode, l'élaboration des normes appropriées concernant la disposition du clavier amazigh et la création des polices de caractères tifinaghes. Dans ce contexte, le présent article a pour objectif de fournir à la langue un outil de conversion lui permettant de translittérer l'alphabet arabe et latin en caractère tifinaghe Unicode.

The Amazighe language and culture have experienced a process of standardization and integration in the information and communication technologies. This process has gone through several stages: character encoding specified by the extended ASCII, incorporation into the Unicode standard, implementation of a standard keyboard layout for Amazighe, and building new Tifinaghe fonts. In this context, we aim to provide the Amazighe language a conversion tool that could transliterate the Arabic and Latin scripts to Tifinaghe Unicode.

Keywords – Mots Clés

Convertisseur, translittérateur, Langue amazighe, Promotion, Diffusion.
Converter, transliterator, Amazighe language, Promotion, Dissemination.

Introduction

Depuis la création de l'Institut Royal de la Culture Amazighe (IRCAM), la langue amazighe jouie d'un statut institutionnel qui lui a permis d'avoir une graphie officielle et un codage propre dans le standard Unicode. Or dans le contexte de sauvegarder l'héritage littéraire de l'amazighe, et empruntant la politique entreprise par l'IRCAM visant la diffusion de la graphie tifinaghe sur tous les supports médiatiques et pédagogiques, il est intéressant de tirer profit de ce patrimoine culturel par l'élaboration d'outils informatiques permettant la réécriture des documents amazighes.

Cet article s'inscrit dans une démarche progressive qui vise à doter la langue amazighe d'applications capables de traiter de façon automatique les données, particulièrement les outils de conversion. Ainsi, nous avons procédé par le développement d'un outil de translittération qui consiste à convertir l'alphabet arabe en alphabet tifinaghe en partant de l'idée qu'il existe des documents en langue amazighe qui sont écrits en caractères arabes qui peut être exploitable, et en se basant sur l'outil de conversion développé au sein du Linguistic Data Consortium.

Dans la suite de cet article, nous exposons, dans la section 2, les caractéristiques de la langue amazighe à travers la description du système d'écriture et le codage employé. Ensuite dans la section 3, nous présentons notre proposition et une conception de l'outil développé en partant de l'existant. La section 4 est consacrée à la conclusion.

Caractéristique de la langue amazighe

L'amazighe est la langue autochtone de l'Afrique du Nord (Hachid, 200; Charles-André, 1978). Cette langue regroupe une trentaine de variétés dialectales présentes de la Méditerranée jusqu'au sud du Niger, et des Iles Canaries à la frontière ouest de l'Egypte. Particulièrement au Maroc, l'amazighe se répartit en trois grandes variétés régionales qui couvrent l'ensemble des régions montagneuses : le Tarifite au nord-est ; le Tamazighte au centre, le Moyen-Atlas et une partie du Haut-Atlas ; et le Tachelhite au sud et sud-ouest, le Haut-Atlas, l'Anti-Atlas et Sous.

Système d'écriture de la langue amazighe

Avant la standardisation de la langue amazighe, son écriture ainsi que sa modélisation informatique faisait appel à l'alphabet arabe ou l'alphabet latin enrichi par les caractères spéciaux empruntés de l'Alphabet Phonétique International (API). Or, dans l'objectif de fournir à la langue amazighe un système

alphabétique standard plus adéquat et utilisable pour tous les parlers amazighs actuels du Maroc, le Centre de l'Aménagement Linguistique (CAL) a préconisé les caractères tfinaghés, présentés dans le Tableau 1, comme alphabet de la langue amazighe (Ameur et al., 2004). Cet alphabet comporte :

- 27 consonnes dont : les labiales (ⵀ, ⵉ, ⵏ), les dentales (ⵜ, ⵏ, ⵍ, ⵍ, ⵏ, ⵏ, ⵏ, ⵏ), les alvéolaires (ⵔ, ⵙ, ⵓ, ⵙ), les palatales (ⵇ, ⵇ), les vélares (ⵔ, ⵙ), les labiovélares (ⵔ, ⵙ), les uvulaires (ⵔ, ⵙ, ⵙ), les pharyngales (ⵏ, ⵏ) et la laryngale (ⵏ);
- 2 semi-consonnes : ⵙ et ⵏ ;
- 4 voyelles : trois voyelles pleines ⵏ, ⵙ, ⵓ et la voyelle neutre (ou schwa) ⵓ qui a un statut assez particulier en phonologie amazighe.

Pour la ponctuation, les signes conventionnels de l'écriture latine sont employés : « » (espace), « . », « , », « ; », « : », « ? », « ! », « ... », etc. Alors que pour les chiffres, les numéros arabes occidentaux (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) sont retenus.

ⵍⵔⵓⵎⵎⵉⵎⵉⵢⵓⵏ			
	Tifinaghe	Correspondance latine	Correspondance arabe
ya	ⵝ	A	ا
yab	ⵉ	B	ب
yag	ⵔ	G	گ
yag ^w	ⵔⵓ	g ^w	گ
yad	ⵏ	D	د
yaḍ	ⵉ	ḍ	ض
yey	ⵎ	E	
yaf	ⵢ	F	ف
yak	ⵔ	K	ك
yak ^w	ⵔⵓ	k ^w	ك
yah	ⵎ	H	ه
yaḥ	ⵏ	ḥ	ح
yaε	ⵏ	E	ع
yax	ⵔ	X	خ
yaq	ⵎ	Q	ق
yi	ⵔ	I	ي
yaj	ⵢ	J	ج
yal	ⵎ	L	ل
yam	ⵎ	M	م
yan	ⵏ	N	ن
yu	ⵎ	U	و
yar	ⵎ	R	ر
yaṛ	ⵎ	ṛ	ر
yay	ⵢ	Γ	غ
yas	ⵎ	S	س
yaş	ⵎ	ş	ص
yac	ⵎ	C	ش
yat	ⵏ	T	ت
yaṭ	ⵉ	ṭ	ط
yaw	ⵎ	W	و
yay	ⵔ	Y	ي
yaz	ⵔ	Z	ز
yaž	ⵔ	z	ژ

Tableau 1 : L'alphabet Tifinaghe préconisé par le CAL et consacré par l'IRCAM

Le codage du système amazighe

Dans le but d'introduire le système d'écriture amazighe préconisé par l'IRCAM, il a fallu en attendant l'amendement de l'Unicode trouver une solution rapide qui permettra la production des publications, particulièrement les manuels scolaires. Cette solution était le codage ASCII étendu représentée par la table de codage illustré à la Figure 1, où les caractères latins ont été substitués par leur correspondant en tifinaghe.

Figure 1. Codage du Tifinaghe en ASCII étendu

Or après l'amendement, le consortium Unicode a réservé aux caractères tiffinaghes le bloc illustré par la Figure 2, qui précise les codes réservés à chacun des quatre sous-ensembles des caractères tiffinaghes. Le premier sous-ensemble représente les lettres alphabétiques de base préconisées par l'IRCAM dont le nombre est de 33. L'Unicode ne code directement que 31 caractères et le caractère modificatif « ^u », qui permet de former les deux unités labiovélaires « ^uX » et « ^uK ». Le deuxième sous-ensemble contient les 8 caractères de la liste étendue, qui a été définie par l'IRCAM pour l'intérêt historique et scientifique. Le troisième sous-ensemble est formé de 4 lettres néo-tifinaghes utilisés fréquemment dans le reste du Maghreb. Et le quatrième sous-ensemble contient 11 lettres touarègues modernes dont l'usage est attesté (Andries, 2008).

	2D3x	2D4x	2D5x	2D6x	2D7x
0	◌	⊖	+	Δ	
1	⊖	∅	!	⊏	
2	⊕	⋮	∂	∠	
3	×	∠	∞	✱	
4	×	⊏	⊖	⌞	
5	×	×	⊖	✱	
6	⊏	∞	⊏		
7	∧	∠	∞		
8	∨	∞	∞		
9	∞	∞	⊖		
A	∞	∞	⊖		
B	∞	×	⊖		
C	∞	∞	+		
D	∞	∞	×		
E	∞	∞	∞		
F	∞	∞	∞	∞	

Clé

- Tifinaghe Ircam de base
- Tifinaghe Ircam étendu
- Autres lettres néotifinaghes
- Lettres touarègues modernes attestées
- Réservé pour un codage ultérieur

Figure 2. Codage du Tifinaghe en Unicode

Convertisseur pour la langue amazighe

Dès le codage des caractères tifinaghes, il a fallu préparer des outils pour convertir la représentation de ces caractères d'un codage à un autre. Dans ce cadre, le Centre d'Etudes Informatiques Système d'Information et de Communication (CEISIC) ainsi que le Linguistic Data Consortium (LDC) ont réalisé chacun de sa part un convertisseur permettant le passage entre le codage ASCII étendu et le codage Unicode. Alors que dans la perspective de promouvoir la langue amazighe et d'assurer la sauvegarde de son héritage littéraire, il apparaît intéressant de tirer profit des productions amazighes écrites en caractère arabe. Ainsi, nous avons proposé de compléter le convertisseur de LDC afin qu'il puisse traiter les données écrites en caractère arabe.

Convertisseur de LDC

Le convertisseur de LDC est un outil qui permet d'une part la conversion du caractère tifinaghe codé en ASCII étendu (IRCAM Latin) au caractère latin sous le codage Unicode (Unicode Latin), et d'autre part la conversion du caractère Unicode Latin au caractère tifinaghe codé en Unicode (Unicode Tifinaghe). En outre, il permet le passage direct du IRCAM Latin au Unicode Tifinaghe.

Proposition

Dans la perspective de doter la langue amazighe d'un outil de translittération des documents écrits en caractère arabe, nous avons proposé de compléter le convertisseur de LDC d'une manière à répondre à d'autres besoins qui pourraient être lucides dans l'avenir. Ainsi, nous avons opté pour ajouter d'autres fonctionnalités à cet outil afin qu'il assure la conversion entre les différents codes du caractère tifinaghe et la translittération dans les différents sens entre le caractère arabe, latin et tifinaghe.

Conversions existantes	Nouvelles conversions
IRCAM Latin → Unicode Latin	Unicode Tifinaghe → IRCAM Latin
Unicode Latin → Unicode Tifinaghe	Arabe → Unicode Tifinaghe
IRCAM Latin → Unicode Tifinaghe	Unicode Tifinaghe → Arabe
	Arabe → IRCAM Latin
	IRCAM Latin → Arabe

Tableau 2. Types de conversions assurées par le nouvel outil

Conception du convertisseur

L'outil développé assure des fonctionnalités qui consistent à convertir les différents codes du caractère tifinaghe et translittérer les textes amazighs écrits en caractère latin ou arabe au caractère tifinaghe et vice versa, comme il est illustré par la Figure 3.

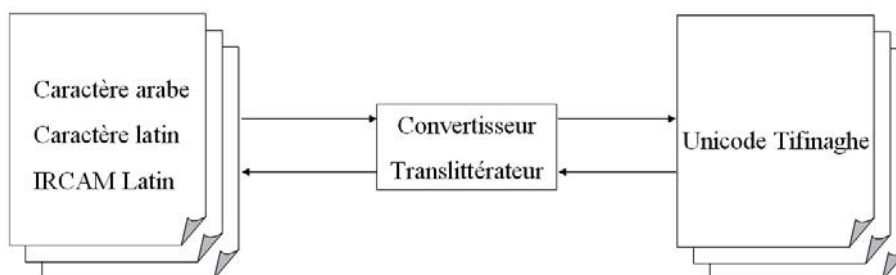


Figure 3. Conception de l'outil de conversion

Pour la translittération des documents écrits en arabe, nous nous sommes basées sur les correspondances utilisées dans le dictionnaire arabe – amazighe (Chafik, 1993), qui diffèrent de celles indiquées dans le Tableau 1 par celles précisées dans le Tableau 3.

	Tifinaghe	Correspondance arabe
Yag	ⵝ	ث
yag ^w	ⵝ ^w	ث
yak ^w	ⵞ ^w	ك
Yi	ⵉ	ي (au début d'un mot (
Yu	ⵓ	و (au début d'un mot (
yaɣ	ⵓ	ر
Yay	ⵣ) ي au début d'un mot (
yaɣ	ⵣ	ر

Tableau 3. Correspondances distinctes établies par Mr Chafik

Au cours de notre implémentation, nous avons remarqué que les caractères { ك } ne figurent pas sur la table Unicode, ce qui nous a amené à remplacer « ك » par le caractère « ك » en s'inspirant de la proposition de Mr Chafik pour le caractère « ك » qui correspond au caractère tifinaghe « ⵞ^w », et à substituer le caractère « ر » par « ر » leur ressemblance. Par ailleurs, en se basant sur les correspondances arabes du caractère tifinaghe pour translittérer les documents écrits en caractère tifinaghe vers celui en arabe, nous avons soulevé quelques anomalies à propos des caractères { ي, و, ر, ك, گ } qui influence sur la cursivité de la langue arabe. Ce qui nous a amené à les remplacer par les propositions de Mr Chafik.

Tandis que pour la translittération du caractère latin au caractère tifinaghe, nous avons adopté les correspondances illustrées dans le Tableau 1.

Conclusion

Le présent article s'inscrit dans une stratégie progressive qui vise à encourager le développement d'outils de traitement automatique de la langue amazighe, particulièrement l'élaboration d'outils informatiques permettant la conversion des documents amazighes. Ainsi, nous avons entrepris de réaliser un outil de translittération qui consiste à convertir l'alphabet arabe et latin en alphabet tifinaghe en exploitant les documents en langue amazighe qui sont écrits en caractères arabes et latins, et en se basant sur l'outil de conversion développé au sein du Linguistic Data Consortium.

Références

- Ameur M., Bouhjar A., Boukhris F., Boukous A., Boumalk A., Elmedlaoui M., Iazzi E., Souifi H. (2004), *Initiation à la langue amazighe*, Rabat, Maroc, IRCAM.
- Andries P. (2008), *Unicode 5.0 en pratique, Codage des caractères et internationalisation des logiciels et des documents*, France, Dunod, Collection InfoPro.
- Chafik M. (1993), *المعجم العربي الأمازيغي*, El Maarif Aljadida, Rabat, Maroc.
- Charles-André J. (1978), *Histoire de l'Afrique du nord des origines à la conquête arabe: Tunisie - Algérie – Maroc*, Editions Payot.
- Hachid M. (2000), *Les premiers berbères: entre méditerranée, Tassili et Nil*, Provence, France, Edisud.