

Vers un Système d'Ecriture Informatique Amazighe : Méthodes et Développements

Ali Rachidi¹ & Driss Mammass²

¹ Ecole Nationale de Commerce et de Gestion, B.P. 37/S Hay Salam Agadir,
Maroc

Laboratoire de Traitement d'Images et Systèmes d'Information (LTISI)

Email : rachidi.ali@caramail.com

² Unité de Recherche en Traitement d'Images et de l'Information (URT2I),
Faculté des Sciences, Université Ibn Zohr, Agadir, Maroc,

E-mail : driss_mammass@yahoo.fr

Mots-Clés – Keywords

Amazighe, Alphabet Tifinaghe, Traitement des langues naturelles.

Amazighe, Tifinaghe Alphabet, Natural Language Processing.

Résumé – Abstract

L'intégration des technologies de l'information et de communication (TIC) à l'apprentissage de la langue Amazighe est absolument nécessaire pour qu'elle ait droit de cité plein et entier sur le Web et dans le monde informatisé.

Nous présentons quelques réflexions sur les stratégies et méthodes d'informatisation de l'amazighe qui est une langue peu dotée informatiquement. Ces réflexions visent surtout l'optimisation de l'effort d'informatisation. En effet, les méthodes proposées tiennent en compte non seulement l'alphabet proposé par l'IRCAM¹ et confirmée par l'ISO (format Unicode) le 21 juin 2004 (IRCAM, 2004 a) mais aussi le contexte francophone des populations berbères.

Learning the Amazighe language would require the introduction of Information and Communication technologies so that the language would have full entire freedom of a mentioning on the Web and in the computerized world.

We present some reflections on strategies and methods of the amazighe computerization which is a language little dowered. These reflections aim particularly the optimisation of the effort of computerization. Indeed, proposed methods hold in account no only the proposed alphabet by the IRCAM¹ and confirmed by the ISO (Unicode format) the 21 June 2004 but therefore the French speaking context of Berber populations.

¹ Institut Royal de la Culture Amazighe, Rabat, Maroc

1. Introduction

Cet article s'inscrit dans un large mouvement international qui vise à ce que chaque peuple puisse disposer de tous les moyens pour communiquer dans sa langue. L'Amazighe fait partie des langues très peu dotées informatiquement. Par conséquent, des recherches scientifiques et linguistiques sont lancées dans ce sens pour améliorer la situation actuelle. La conception et la réalisation d'applications capables de traiter de façon automatique des données linguistiques, exprimées dans la langue naturelle amazighe, sont parmi les objectifs prioritaires de nos travaux de recherche en collaboration avec l'IRCAM. Dans ce contexte, nous proposons quelques réflexions, sur des méthodes et des stratégies à mettre en œuvre pour produire un outil de 1) traitement de texte amazighe sous codage ASCII et un autre sous format Unicode après l'intégration du format Unicode amazighe dans les applications informatiques par les firmes responsables et 2) traduction automatique et de gestion d'une base lexicale Amazighe. Cet article est composé de trois parties. Dans la première partie, nous présentons le contexte linguistique et le système d'écriture de la langue Amazighe. La deuxième partie présente les différentes méthodes d'informatisation. Enfin, la dernière partie est consacrée à la présentation de la mise en œuvre de certaines méthodes et des outils impliqués.

2. Amazighe : langue naturelle

La famille élargie des Berbères qui utilise le Tifinaghe comme écriture traditionnelle et commune est de près de vingt millions. Au Maroc, le berbère (« amazighe ») marocain englobe les trois grandes variantes : le *Tarifite*, le *Tamazighte* et le *Tachelhite*. Plus de 40% de la population marocaine est berbérophone. Nous présentons un petit aperçu sur l'évolution de la langue amazighe tout au long de l'histoire et sur son système d'écriture.

2. 1. Historique

L'alphabet berbère ou Amazighe a subi des modifications et des variations depuis son origine jusqu'à nos jours. Du libyque jusqu'au néotifinaghe en passant par le tifinaghe saharien et le tifinaghe touareg. Nous retraçons ci-dessous les aspects les plus importants de chacune de ces étapes.

2. 1. 1. Le libyque

Il s'agit des variétés du Tifinaghe les plus anciennes. Il existe deux formes du libyque, l'oriental et l'occidental.

2. 1. 2. Le Tifinaghe saharien

Cette variété est également appelée libyco-berbère ou touareg ancien. Elle contient des signes supplémentaires par rapport au libyque, plus particulièrement un trait vertical pour noter la voyelle finale /a/. Cette variété fut utilisée pour transcrire le touareg ancien mais ses inscriptions sont incompréhensibles (IRCAM 2004 b).

2. 1. 3. Le Tifinaghe touareg

Il existe au sein du Tifinaghe touareg quelques divergences dans la valeur attribuée aux signes qui correspondent aux variations dialectales touarègues. Si d'une région à une autre, la forme et le nombre des signes peuvent changer, les textes restent en général mutuellement compréhensibles.

2. 2. 4. Chiffres

L'IRCAM a retenu les chiffres « arabes » occidentaux (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) pour l'écriture Tifinaghe. Cette proposition n'introduit donc aucun nouveau chiffre ou nombre.

2. 2. 5. Directionnalité

L'IRCAM a retenu la direction horizontale de gauche à droite pour l'écriture Tifinaghe.

3. Méthodes pour l'informatisation de l'amazighe

3. 1. Trouver des solutions adaptées : idées forces

3. 1. 1. Bénéficier de développements faits pour des langues liées

L'Amazighe est une langue voisine de l'arabe et du français. Par conséquent, il est très possible de bénéficier des liens qui unissent cette langue au français et à l'arabe pour utiliser leurs ressources et faciliter les développements. Par exemple, on peut intégrer une barre d'outils pour l'écriture Amazighe qui se charge automatiquement avec Ms WinWord version Arabe ou Française. Cette idée sera l'objet d'une section prochaine.

3. 1. 2. S'intégrer à des projets et environnements génériques, open source et à pivot

Projet UNL (www.undl.org): Il est de plus en plus nécessaire de créer des documents multilingues qui intègrent l'Amazighe (Rachidi A. 2004). L'idée actuelle est de faire de la traduction manuelle collaborative sur le Web à l'aide d'une mémoire de polyphrases multilingues (MPM), outils en construction par l'équipe de C. Boitet (GETA, CLIPS, IMAG - Grenoble, France) (Berment V. 2004), puis d'intégrer le résultat (les phrases en Amazighe) dans le document en UNL-XML. Enfin, il faudra construire un déconvertisseur UNL-Amazigh et un enconvertisseur.

Projet Papillon (www.papillon-dictionary.org): L'amazighe intégrera le projet dès que le format Unicode sera bien installé dans les plates formes logicielles. En attendant, il faut préparer une liste exhaustive de catégories (part of speech) qui seront utilisées dans ce dictionnaire.

3. 2. Appliquer une gestion adaptée

1) Déterminer quel produit réaliser ? 2) Déterminer qui réalise les logiciels et ressources et 3) Etablir un plan de développement. En effet, au sein de notre laboratoire et en collaboration avec l'IRCAM, les développements se dérouleront en deux phases distinctes ayant nécessité chacune un plan de développement spécifique : 1) la réalisation d'un traitement de texte bien adapté à l'écriture Amazighe et 2) le développement d'outils d'aide à la traduction et d'une base lexicale en intégrant les projets UNL et le papillon).

4. Mise en œuvre et outils impliqués

4. 1. Phase actuelle (avant le format Unicode)

4. 1. 1. Clavier Amazighe

L'alphabet Tifinaghe est composé de trente trois caractères. Le centre CEISIC de L'IRCAM a proposé un clavier sous format ASCII (police, pilote) comme l'illustre la figure 2 (IRCAM 2003 a) (IRCAM 2003 b). Les 26 premiers caractères sont accessibles directement. Les caractères emphases s'obtiennent en utilisant la case Noir (le « ^ » en clavier latin) de la

même façon qu'on utilise le « ^ » en français (pour taper le « â »). L'IRCAM a proposé un projet qui normalise deux groupes de claviers en précisant deux niveaux de conformité, l'un pour la saisie stricte des trente-deux lettres de l'alphabet Tifinaghe de base tel qu'enseigné dans les écoles marocaines, l'autre pour la saisie de l'alphabet de base, plus les vingt-deux lettres de l'alphabet Tifinaghe étendu et des ligatures (lorsque la technologie sous-jacente permet de traiter deux caractères de commande permettant de former ou d'empêcher la formation de ligatures). Ce projet sera confirmé dans le prochain amendement à la norme internationale ISO/CEI 14651.

				«	•	(-		-)	=	↩
TAB	°	✱	§	Q	+	↵	§	≤	¶	Λ		\$		ENTER
CAPSLOCK	Z	Q	Λ	W	X	Φ	I	R	H	C		*		
SHIFT	^	U	X	C	Y	Θ	I	,	;	:	!			SHIFT
CTRL	⌘	ALT								AltGr	⌘			CTRL

Figure 2 : Clavier Amazighe sous format Ascii

4. 1. 2. Réaliser un traitement de texte pour l'amazighe

On peut réaliser une plate forme logicielle qui permet de traiter un texte Amazighe sous format Ascii avec C / C++ et utilisant uniquement le SDK (Software Development Kit) de Windows, qui visait les premiers niveaux du service de traitement du texte (TALN 2003). Il inclut les fonctionnalités suivantes : 1) saisie de textes amazighe indépendant de la police utilisée et utilisant un clavier intuitif, 2) changement de police (donc transcodage), 3) mise en forme canonique du texte sélectionné (standardisation, saisie non univoque), 4) facilité de sélection à la souris et au clavier des syllabes et des mots Amazighes, 5) formatage de textes Amazighes, pour les rendre utilisables par des traitements de texte commerciaux, 6) export aux formats TeX et RTF, 7) construction d'un lexique à partir de textes (ajouter, modifier, supprimer une entrée dans un lexique local), 8) traduction en français de mots Amazighes et 9) transcription phonétique du texte sélectionné. L'interface proposée de cette plate forme est illustrée dans la figure 3 suivante :

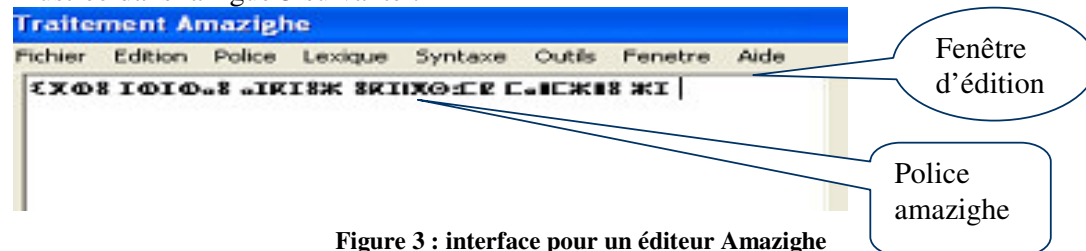


Figure 3 : interface pour un éditeur Amazighe

4. 2. Phase après Unicode : réalisation d'un traitement de texte

4. 2. 1. En utilisant Word Pad

Word Pad peut servir comme base pour une version Amazighe (Amazighe Pad). Ces sources sont disponibles dans C++ de MicroSoft. Word Pad est développé dans C++ et s'appuie sur les classes de la Bibliothèque MFC (Microsoft Foundation Classes) (Berment V. 2004): a) Classe de la fenêtre d'édition CRich Edit Ctrl et b) Format Rich text.

4. 2. 2. En utilisant WinWord

On peut Utiliser le kit du développeur Word (Word's developer's kit). C'est un Ensemble de modules C qui permet de s'interfacer avec Word. Le principe c'est de développer une librairie dynamique qui se charge avec Word et qui appelle les fonctions de Word à travers une API appelée CAPI (Berment V. 2004).



Le format proposé de cette barre est la suivante :

De la gauche vers la droite, les fonctions des boutons sont : a) configuration AmazigheWord ; b) changement de police Amazighe ; c) tri de tableaux en Amazighe ; d) transcriptions Amazighe ; e) dictionnaire électronique ; f) mise en forme du texte ; g) choix Amazighe-latin pour la saisie des textes et h) l'aide en ligne.

5. Conclusion

Née avec l'informatique, l'informatisation des langues a évolué, offrant de plus en plus de services pour de plus en plus de langues. C'est cependant un processus coûteux qui ne bénéficie actuellement qu'à une faible partie des langues du monde (moins de 1 %). Informatiser l'Amazighe demande un effort de plusieurs chercheurs et l'utilisation des leviers permettant d'obtenir rapidement des logiciels de qualité. L'existence du standard Unicode, qui vient d'intégrer l'Amazighe ces derniers mois, a récemment permis la réalisation de systèmes d'exploitation et de logiciels couvrant de nombreux systèmes d'écriture tout en évitant la multiplication des incompatibilités entre plates-formes. L'Amazighe bénéficie ainsi d'outils d'édition performants. Le codage et les principes de base étant communs aux différents systèmes d'écriture. Notre projet, en collaboration avec l'IRCAM, œuvre pour la réalisation d'un traitement de texte bien adapté à l'écriture Amazighe et le développement d'outils d'aide à la traduction et d'une base lexicale.

Références

Berment V. (2004) méthodes pour informatiser des langues et des groupes de langues «peu dotées», thèse de Doctorat de l'université Joseph Fourier, Grenoble 1, UFR d'informatique et mathématiques appliquées, 18 mai 2004.

IRCAM (20004 a), 'Proposition d'ajout de l'écriture Tifinaghe au répertoire de l'ISO/CEI 10646 (format unicode)', 21/06/2004, centre CEISIC, IRCAM, Rabat, Maroc

IRCAM (2004 b), 'Graphie de la langue amazighe' Centre de l'Aménagement Linguistique Coordinateur El Mehdi Iazzi, Publications de l'IRCAM, Rabat, 2004.

IRCAM (2004 c), 'Initiation à la langue amazighe', Centre de l'Aménagement Linguistique, publication de l'IRCAM, manuel N° 1, Rabat, 2004.

IRCAM (2003 a), 'Élaboration d'une première version du clavier amazighe', Centre des études informatiques et des systèmes d'information et de communication, Rabat, plan d'action 2003.

IRCAM (2003 b), 'Conception et mise au point des polices Amazighes', Centre des études informatiques et des systèmes d'information et de communication, Rabat, plan d'action 2003.

Rachidi A. (2004), ' Les Graphes UNL : un concept unificateur pour l'intégration de l'Amazighe dans des Documents Multilingues', séminaire international : La typographie entre les domaines de l'art et de l'informatique, IRCAM, Rabat septembre. 2004.

TALN (2003), 'Traitement automatique des langues minoritaires et des petites langues', l'atelier associé à TALN 2003.