

## **Essai sur la Standardisation Ecrite de la Langue Tamazight : Adaptations Computationnelles**

*Par Iliass AHALLI*

*Professeur d'Enseignement Supérieur  
Université Abdelmalek Essaadi, Tanger, Maroc*

*Et Mounir MAOUENE*

### **Introduction**

Nous proposons ici un essai de standardisation de l'écriture de la langue Tamazight afin qu'elle soit adaptée au traitement informatique. Notre objectif derrière ce travail est de développer un système informatique d'aide à l'apprentissage de la langue tamazight surtout pour les berbères qui vivent dans des régions montagneuses dépourvues d'écoles.

Ce travail fournit une approche d'écriture qui se base d'un côté sur les caractères latins dans le but de conserver la cohérence et l'homogénéité de la langue, d'un autre côté, faciliter le système d'écriture pour une utilisation informatique aisée.

Par exemple, nous observons que les linguistes amazighs rifains et kabyles utilisent un mélange de caractères latins, grecs et spéciaux comme  $\gamma$ ,  $\varepsilon$ ,  $\check{c}$ ,  $\check{z}$ ,  $z$ ,  $h$ ,  $\check{g}$ ,  $d$ , etc. Un autre exemple est donné par les ambiguïtés et les anomalies induites par le système linguistique utilisé actuellement, qui rend très complexe tout traitement informatique du Tamazight.

La langue berbère, ou Tamazight, est l'une des langues de la famille linguistique chamito-sémitique, ou afro-asiatique selon la terminologie de Joseph H. Greenberg [1]. La langue amazighe peut être considérée comme la langue « autochtone » de l'Afrique du nord [2]. Elle est présentée actuellement dans une dizaine de pays de l'ensemble Maghreb-Sahara-Sahel : Maroc, Algérie, Niger, Mali etc. Le Tamazight se présente sous forme d'un ensemble de dialectes, fonctionnant comme langue vernaculaire dans une aire régie essentiellement par l'oralité. Un passage à l'écrit a été déclenché dès le 19<sup>ème</sup> siècle par les précurseurs et surtout dans le cadre de l'anthropologie coloniale. Et c'est vers la composante de la langue qui est l'orthographe que notre intérêt est dirigé.

En traitement automatique du langage naturel (TAL), la langue amazighe, comme la plupart des langues non-européennes, souffre de la rareté, voire de l'absence, des outils de traitement automatique de la langue [3].

Nous nous sommes basés, pour formuler nos propositions sur les travaux qui ont été réalisés sur la transcription de Tamazight, dans le but d'en profiter de leurs aboutissements, en leur faisant une adaptation computationnelle pour qu'ils soient en conformité avec les exigences du traitement automatique des langues naturelles.

En premier lieu, nous avons opté pour l'utilisation des caractères latins ; en se débarrassant de tout caractère à signe diacritique ; ensuite, nous avons entamé la question de l'article en Tamazight et

nous avons opté pour son séparation du substantif. Puis après, nous avons abordé les emphatiques, les pharyngales, les consonnes tendus, la voyelle neutre et les règles de son apparition vue son statut d'instabilité, et en terminant par la question des emprunts en Tamazight.

Vers un Système d'écriture Cohérent : Propositions d'écriture.

Le débat sur la graphie pour Tamazight ne cesse d'enfler. Entre partisans de la graphie arabe, latine et tfinagh, le débat enfle, mais sans pour autant se débarrasser des convictions idéologiques et surtout sans prendre en compte les décennies de travaux de production en Tamazight transcrites en latin.

Trois types de graphies peuvent être envisagés pour la transcription de la langue amazighe, à savoir : la graphie arabe, la graphie tfinagh et la graphie latine.

Le choix de l'écriture en graphie arabe, a été exclu durant notre démarche du choix de la graphie, pour les raisons suivantes : (

Impossibilité de combinaison de caractères (comme cela se fait pratiquement dans toutes les langues utilisant le caractère latin) pour représenter les phonèmes étrangers à l'alphabet arabe ;

Obligation de créer de nouvelles lettres par ajout de points ou de signes diacritiques aux lettres déjà existantes.

Absence de voyelles courtes sous forme de lettres d'où l'obligation d'utiliser tout le temps la vocalisation.

En cas d'utilisation des caractères représentant les voyelles longues arabes pour représenter les voyelles amazighes, on aura des difficultés à représenter les semi-voyelles « w » et « y »). (Ces arguments sont avancés par H. SAHKI [4])

A tout cela, nous rajoutons que la graphie arabe génère un alourdissement au niveau du traitement informatique, vu que l'ensemble des compilateurs sont à base du caractère latin ;

De même, le choix de la graphie tfinagh a été exclu, vu qu'il présente un alourdissement énorme quant au traitement informatique. Nous avons constaté que, quelque soit la valeur symbolique et historique de cette graphie, ses usages restent réduites et emblématiques (titres d'ouvrages, slogans, enseignes...), aucune œuvre littéraire ou publication scientifique n'a été publiée en tfinagh au cours des dernières années.

Alors, la graphie latine s'impose comme graphie d'écriture, découlant d'une analyse technico-scientifique, que nous avons accompli, et qui tient compte, dans un premier temps, de la dimension traitement informatique et écartant autant que possible, toute préférence d'ordre idéologique qui ne servent en rien la langue amazighe et sa diffusion via les technologies de l'information.

Les défenseurs de la graphie latine voyaient dans les caractères latins, un outil pour diffuser la langue amazighe dans la présence d'écrits scientifiques en caractères latins [5]. De notre part nous nous

rejoignons à cette réalité annoncée qui indique que la graphie latine permet une large diffusion.

On considère qu'un grand nombre de problème d'écriture sont déjà réglés, dans la notation usuelle [6], et dans les études d'écriture menées par H. SAHKI [7]. Nous nous contentons de traiter ceux que nous jugeons, ont besoin d'adaptation computationnelle, et qui génèrent des problèmes délicats en traitement informatique.

L'alphabet Amazigh Latin

L'alphabet qu'on se propose d'utiliser est l'alphabet latin :

a b c d e f g h i j k l m n o p q r s t u v w x y z

Il n'y aura ni caractères supplémentaires, ni lettres appartenant à un autre alphabet, ni signe diacritique.

Certains phonèmes seront représentés par deux caractères. Cette façon de faire existe pratiquement dans toutes les langues utilisant les caractères de l'alphabet latin.

<b>Caractère</b>	<b>Valeur</b>	<b>Exemple</b>
A	« a » anglais dans « man »	a man (eau/l'eau)
B	« b » français dans « barbe »	rbu(Porter sur le dos)
C	« ch » français dans « chacal »	amcic (un chat/ le chat)
D	« d » français dans « danse »	adrar (une montagne/ la montagne)
E	voyelle vide ou neutre	i rem(peau/la peau)
F	« f » français dans « fatale »	afus (main/la main)
G	« g » français dans « gaz »	argaz

		(homme/l'homme)
H	« h » anglais dans « home » « h » arabe dans « Mohammed »	tahandith (cactus/ les cactus) rehmu (Il fait chaud)
I	« i » anglais dans « hill »	thiri (ombre/l'ombre)
J	« j » français dans « jolie »	ajenna (haut/le haut)
K	« k » français dans « kilomètre »	akannay (une pierre/ la pierre)
L	« l » français dans « élite »	malita (une valise/ la valise)
M	« m » français dans « maison »	ameggiz(joue/la joue)
N	« n » français dans « navire »	anu (puits)
O	« o » français dans « olive »	tomatic (tomates/les tomates)
P	« p » français dans « proie »	rpakith(boite/ la boite)
Q	« q » arabe dans « Iraq »	aqemmm (bouche/ la bouche)
R	« r » anglais dans « rate »	iri (un cou/ le cou)
S	« s » français dans « sable »	asemmum (amande vert)
T	« t » français dans « table »	tomatic (tomates/les tomates)
U	« ou » français	rum(foin)
V	« v » français dans « vélo »	ivawen (fèves/ les fèves) azaevur (cartable/ le cartable)
W	« w » anglais dans « way »	awridji (une araignée/ l'araignée)
X	« x » arabe dans « khartoum »	axxam (une chambre/ la chambre)
Y	« y » français dans « yogourt »	ayaw (un neveu / le neveu)

Z	« z » français dans « zèle »	amazigh
DJ	« dj » français dans « adjudant »	yedja (il existe)
GH	« gh » dans « Maghreb »	aghrum (le pain)
DZ	-----	adzayri (un algérien)
TC	-----	ta tcint (une orange)

**Table1**  
**Les valeurs des caractères et digrammes du système proposé**

### **L'article en Amazigh**

A l'instar des langues européennes, Tamazight garde un lien fort entre le substantif et son article. Cependant son écriture pose problème, entre ceux qui préconisent son séparation du radical, et ceux qui opte pour une écriture attachée. Les linguistes amazighes sont scindés sur ce point. Nous proposons de l'écrire séparé. Son séparation de l'article est dictée par une nécessité computationnelle dans le but de réduire le nombre des règles dans la phase de reconnaissance des classes grammaticales des termes d'une phrase donnée. La séparation permet aussi de différencier certains homonymes homophones et homographes (Exemple : azal (valeur, prix), a zal (grand-jour) ; adtili (elle sera) t'ili (l'ombre))[8].

Il est à signalé que l'article en Tamazight n'exprime pas la notion de définition, c'est le sens contextuel qui le détermine. On dira « a mcic » (un chat / le chat).

Pour le cas de l'article au masculin singulier et pluriel à l'état lié, nous aurions pu l'orthographier respectivement "u" et "i". Mais, pour éviter toute ambiguïté ou confusion, on leur a adjoint la voyelle neutre

‘e’ pour les différencier respectivement des caractères alphabétique ‘u’ et ‘i’.

Genre Nombre	Masculin		Féminin		
	Etat	Libre	Lié	Libre	Lié
Singulier		a amcic (un chat / le chat)	ue yenduuemcic	ta ta mcict (une chatte / la chatte)	te tensute mcict
Pluriel		i i mcicen (des chats / les chats)	ie ndwienmcicen	ti ti mcicin (des chattes / les chattes)	te ndwentte mcicin

**Table 2**  
**L'article amazigh en fonction du genre, nombre et de l'état (libre ou lié),  
aux cas réguliers**

Il reste à savoir, comment faire pour différencier l'article de la voyelle initiale ? H. SAHKI [9], répond en proposant une règle simple à appliquer qui est : (si la voyelle initiale est conservée dans toutes les



déterminations du nom, alors elle appartient au radical. Voici quelques exemples à voyelles initial appartenant au radical)(H. SAHKI) :

a: as,awal, akal, aman;

i: iri, imi, izi, irem

u: ul, ussan, ubrin...

Devant ces noms, dont la voyelle initiale fait partie du radical, nous aurons des articles particuliers :

“w” pour ceux qui commencent par “a” et “u”. On écrira donc : itaqsaswul ; litro en w aman.

“y” pour ceux qui commencent par “i”. On écrira : ghasudemy icena.

Il existe des noms verbaux qui ne prennent pas d'article, par exemple : fad (soif), rhazh(faim) etc.

### Les Emphatiques

Cette partie porte sur l'écriture des caractères emphatiques en Tamazight, qui sont : /d/, /r/, /s/, /t/ et /z/.

L'obtention de l'emphase sera marquer par l'ajout d'un “h” au caractère non emphatique, on obtiendra donc : ‘dh’, ‘rh’, ‘sh’, ‘th’ et ‘zh’.

Cette façon de faire pose problème pour certains digrammes qui ont déjà des valeurs quasi-universelles. Nous avons acquis des habitudes scripturales, par l'intermédiaire de la langue française et anglaise. Ces

habitudes enracinées peuvent perturber l'apprenant et lui causer des confusions aux valeurs de certains digrammes. Par exemple:

Le 'sh' en anglais: Shell, shower, shave, etc.

Le 'th' en anglais: The players, think, throw, etc. En français: thermodynamique, mathématiques, athée, etc.

Alors, il est à bannir de représenter le 's' et le 't' emphatique par l'ajout d'un 'h' [10]. Et pour pallier à ce problème, nous proposons de représenter le 's' emphatique par l'ajout d'un 'c', comme cela se fait pratiquement en français : sciences, sciastique, scintigraphie, etc. Il était possible de représenter le 's' emphatique par le caractère 'ç'(en Français, on trouve : garçon, agaçantetc.), mais cela contredit notre objectif qui est l'adaptation computationnelle de Tamazight aux caractères latins, son contradiction s'affiche clairement par la présence du signe diacritique qui lui est associé (la cédille), et qui génère un alourdissement en traitement informatique. Et pour le 't', nous laissons le soin au lecteur de faire la différence de réalisation, pour découvrir s'il s'agit d'un 't' emphatique ou non emphatique (cela se fait pratiquement en anglais : table, dans ce cas le 't'est non emphatique ; tag : dans ce cas le 't'est emphatique).

Nous écrirons, donc :

dh : adhaghar (aveugle), a dhawwar (beau-père), a zhambidh (connerie)

rh : yerhwa (il est rassasié)

zh : izhi (vésicule biliaire), a zhekka (tombe)

t : atanyu (grosse tête) le 't' est emphatique; ta ntala (dialecte) le 't' est non emphatique

Il est à signaler que le choix des consonnes composant chaque digramme est dicté par le fait que pratiquement, ils ne se rencontrent jamais (à quelques exceptions près) dans l'ordre ou ils le sont dans le digramme, en langue amazighe, donc pas de risque de se tromper.

Les Pharyngales /ħ / et /ε / et la Vélaire /γ /

Les pharyngales /ħ / et /ε / sont d'origine arabe, elles ne sont attestées que dans les emprunts à cette langue. Le sens du mot est parfois différent de l'original mais l'origine arabe est chaque fois attestée, il suffit pour cela, de consulter un dictionnaire de langue arabe [11].

Pour la pharyngale /ħ /, elle sera représentée tout simplement par la lettre « h », et nous laissons le soin au lecteur et apprenant en Tamazight, de déterminer s'il s'agit d'une simple consonne comme dans le mot « tahandith » (cactus), ou bien d'une pharyngale comme dans le mot « rehmu » (il fait chaud).

La pharyngale /ε/ sera représenté par un doublet « aa », comme en transcription française (kaaba, daawa, ...).

Pour la vélaire « γ », qui déroge à la règle déjà lancée, qui dit que ce système doit se débarrasser des caractères grecs et de tout signe diacritique. Elle sera représentée par le digramme « gh ». Cette façon de représenté la vélaire est beaucoup utilisé dans les systèmes

existants de transcription de Tamazight. Nous écrivons, donc :  
aghrum(le pain);aghrar (coquillage) ; gher (lire, étudier) etc.

### Les Consonnes et Digrammes Tendus

Les consonnes tendues seront rendues par un doublet de la lettre correspondante. La tension sur les semi-voyelles sera aussi rendue par le redoublement du caractère. Nous écrivons, donc :

akannay (caillou), a bbic(sein), a yyaw (neveu), ta wwurt (porte), a gaddim (falaise), a maggiz (joue).

Et pour les emphatiques représentés en digramme, nous doublons les caractères significatifs, nous obtiendrons donc :izzhar (il est reposé).

### La Voyelle Neutre

La voyelle neutre ou le 'e' muet, qui n'a pas de grande consistance phonétique, sera utilisé pour séparer la succession de plus de deux consonnes, dans le but de faciliter le décodage et par la suite la lecture de cette suite qui risquera d'être imprononçable, si on ne procède pas à l'insertion de ce 'e' muet. On évitera de la multiplier inutilement, surtout à l'initiale: on notera zger "traverser" et non ezger. Dans cette position, on ne la mettra systématiquement que devant les séquences mono- ou bilitères sans voyelle pleine /i, a, u/ :

eg, "faire", ecc "manger", ens "passer la nuit"... (mais : ccigh, nsigh...)[12].

La voyelle neutre se caractérise par son mobilité dans le mot et même dans la chaîne écrite. La règle d'apparition de la voyelle neutre sera

donnée tout de suite .H. SAHKI, a lancé cette règle dans son ouvrage, T'utlayt ta mazight:

«La voyelle neutre apparaît avant la dernière consonne du mot (sauf lorsque c'est l'indice du féminin « t ») puis toutes les deux consonnes à partir de la droite, en comptant à partir de n'importe quelle voyelle.

imsbridn>ims-e-brid-e-n VCCeCCVCeC (les passants)

imzwura>im-e-zwura VCeCCVCV (les premiers)

tasrdunt>tas-e-rdunt CVCeCCVCt(la mule)

asgrarab>as-e-grarab VCeCCVCVC(le cerceau)

De plus, les tendues (non précédées d'une voyelle pleine) sont toujours précédées de la voyelle neutre sauf en début de mot.

yssn>yessen (il connaît)

ssnn>ssnen(ils connaissent)

nttwassn>nettwassen (nous sommes connus) »

Remarque : Les tendues sont comptabilisée comme simples consonnes dans les suites de plus de deux consonnes étant donné qu'elles forment un phonème distinct.

Si on applique cette règle aux deux exemples vus plus haut, on obtient :

ygrarb> y-e-grar-e-b CeCCVCeC

ykrz> y-e-kr-e-z CeCCeC

ykrz as >yk-e-rz as CCeCCVC

## **Les Emprunts en Tamazight**

L'étude des emprunts est une phase essentielle dans le processus d'aménagement de toute langue, et particulièrement pour la langue amazighe qui est connue par son taux d'emprunt élevé. L'étude permet d'en dénicher les phonèmes étrangers à la langue amazighe, et par la suite proposer des représentations de ces phonèmes par des digrammes ou des simples consonnes.

Les études berbérisantes, prône pour l'idée qui indique, que les phonèmes /o/, /p/ et /v/ ne font pas partie du champ phonétique de l'amazighe, et pourtant ces phonèmes sont présents dans le tableau présentant l'alphabet que nous proposons (voir ci-haut tableau1). Ces deux faits ne se contredisent en rien, vu que ces phonèmes ne s'affichent qu'en mots empruntés, qu'ils soient assimilés ou pas.

H. SAHKI dans son ouvrage « T'IRA EN TE MAZIGHT, Méthode d'écriture de la langue amazighe », classe les emprunts en emprunts assimilés et emprunts non assimilés.

Les emprunts assimilés sont ceux qui sont en conformité avec la structure de la langue d'accueil avec les règles de sa grammaire, Exemple : ta macint> pl. ti macinin. H. SAHKI, recommande que ces emprunts assimilés soient considérés comme des mots amazighs a part entière, mais avant de faire, l'emprunt doit être adapté graphiquement : un graphème non attesté dans le système d'écriture de la langue d'accueil doit être remplacé par son équivalent dans ce système,

Exemple : vidéo, emprunté au français devient en Tamazight : avidyu ; de même moteur devient amutur.

Toutes les langues (même celles dénommées vivantes/dominantes) absorbent les emprunts. L'anglais qui a emprunté des mots à au moins 130 langues [13], est un meilleur exemple à cet égard, et c'est tout à fait normal dans le cadre d'inter-influence des civilisations, par l'intermédiaire de la langue.

Les emprunts non assimilés, gardent leur forme d'origine et parfois aboutissent à des formes non amazighes, Exemple : lbiru>lbiruwat, et qui posent des problèmes, qui sont dus au non-respect de la structure de la langue amazighe. Les emprunts non assimilés aussi sont censés être adapté graphiquement au système d'écriture de la langue d'accueil.

## **Conclusion**

Ces nouvelles propositions d'écriture décrites dans ce papier, seront exploitées pour réaliser des applications informatiques de pointe ; à savoir : la construction d'une ontologie fondamentale pour concevoir une application machine-Learning, destinée à l'apprentissage des nouvelles technologies par les berbères.

Le présent système est destiné à tout type d'utilisateurs. Ainsi, il vise la simplicité de l'écriture dans le but de se servir des technologies existantes dans le domaine du traitement automatique de la langue, au profit de la langue amazighe.

Pour vérifier l'efficacité de ce travail de standardisation nous avons développé une application informatique écrite en langage de programmation Python qui calcule la complexité de traitement d'un texte amazigh classique et du même texte standardisé, nous avons trouvé que le texte "standardisé" est moins complexe que le texte "classique"



## **Références Bibliographiques :**

1. Berment Vincent. 2004. Méthodes pour Informatiser Des Langues et des Groupes de Langues « Peu Dotées ».
  2. El Barkani Bouchra. 2012. Le Choix de la Graphie Tifinaghe pour Enseigner Apprendre L'Amazighe au Maroc, Conditions, Représentations et Pratiques : thèse de Doctorat Université de Jean Monnet SAINT ETIENNE,
  3. Greenberg, J. H... 1963. The Languages of Africa: Bloomington Indiana University.
  4. Hacène, Sahki. 1997. T'iraenTeMazight, Méthode d'écriture de la Langue Amazighe (Parler Kabyle).
  5. Hacène .Sahki. 2004. Etudes Eparses sur L'écriture Amazighe.
  6. Hacène, Sahki. 1998. T'utlayt Ta Mazight Analyse et écriture.
  7. Hacène, Sahki. 1998. T'utlayt Ta Mazight Analyse et écriture.
  8. Hacène. Sahki. 1980. Le Problème du Classement dans A Mawal, Lexique du berbère moderne : Editions Imedyazen, Paris.
  9. Salem .Chaker. Juillet 1996. Propositions pour la Notation Usuelle à Base Latine du Berbère:
  10. Salem, Chaker. 2003. La Langue Berbère : les langues de France (sous la direction de Bernard Cerquiglini, Paris, p.215-227).
  11. Salem, Chaker. Juillet 1996. Propositions pour la Notation Usuelle à Base Latine du Berbère.
- Tournier, J. 1988. Précis de lexicologie anglaise Paris : NATH