

L'amazighe comme la majorité des langues africaines a accumulé un retard considérable dans ce domaine. D'où la nécessité de le combler à travers des actions incitatives de grande ampleur, exploitant les TICs en vue de réajuster et d'accélérer la promotion de cette langue. A cette fin, il convient de développer une méthodologie d'intégration. Cette optique nous amène à réfléchir sur l'apport et les moyens de l'utilisation des TICs sur le plan des différents axes stratégiques de la promotion de la langue amazighe :

- Quelles mesures faut-il prendre pour contribuer à la préservation du patrimoine amazighe par le biais des TICs ?
- Que peut apporter l'intégration des TICs au processus d'aménagement linguistique de l'amazighe afin de sauvegarder la diversité linguistique marocaine ?
- Comment les nouvelles technologies peuvent-elles promouvoir l'apprentissage de la langue amazighe ?
- Comment investir dans l'usage social des médias numériques au profit de la langue et la culture amazighes ?

C'est à ces questions, entre autres, que cet article souhaite apporter des éléments de réponse. Il s'agit de proposer des choix méthodologiques et des principes à retenir dans le processus de la préservation et de la modernisation de la langue amazighe. Nous étayerons notre contribution par des exemples concrets issus de recherches récentes dans le domaine des TICs, et plus spécifiquement celles en traitement automatique des langues qui consiste en la création d'outils et de ressources linguistiques par l'intégration de fonctionnalités du traitement du langage humain.

2. L'utilité des TICs pour la pérennité de la culture et la langue amazighes

Le développement et la mondialisation des technologies de l'information et de la communication offrent à toutes les langues du monde un espace globalisé de communication et d'échange. Dans cette perspective, nous introduisons quelques retombées de l'exploitation des TICs pour l'amazighe.

2.1. Les TICs au service de la sauvegarde du patrimoine amazighe

La langue amazighe comme toute langue africaine dite de tradition orale, dont le patrimoine culturel, scientifique et historique ne se transmettait de génération en génération qu'à travers la mémoire et les traditions, ce qui l'a menacée de la destruction des particularités de sa culture et de son identité nationale. D'où l'intérêt de recenser l'ensemble des éléments patrimoniaux et de pouvoir gérer leur aspect descriptif *via* une banque de données multimédias évolutive répondant à un double objectif de conservation et de valorisation de ce patrimoine vernaculaire. En effet, l'élaboration d'une telle banque de données numérique résoudra le problème de la conservation et contribuera à assurer la pérennité de cet héritage ancestral sous une forme perceptible d'une part et d'autre part reflétera la richesse et la diversité de ce patrimoine tout en garantissant une bonne structuration de ses

composants matériels et immatériels, une meilleure accessibilité universelle et une résurrection de l'intérêt du grand public.

A cette fin, une étape de numérisation est indispensable pour la création de cet espace culturel, suivie d'une phase d'archivage sous un format standard permettant la structure du contenu brut et des métadonnées nécessaires à la gestion. Cette phase repose sur un ensemble d'outils et de méthodes de traitement automatique des langues, à savoir :

- Le transcodage pour en faire un contenu intelligible dont le codage est particulièrement compatible aux évolutions technologiques ;
- L'indexation qui sert à identifier et sélectionner les métadonnées pertinentes décrivant le contexte patrimonial ;
- La classification pour regrouper et classer ce contenu numérique selon le support, le type, le thème et la région.

Par ailleurs, afin de permettre une exploitation effective de ce contenu et de diversifier les supports de stockage, une démarche de transcription automatique ou d'océrisation est essentielle à entreprendre pour accélérer la conversion des archives audio et image en documents sous format texte, suivie par la translittération afin de permettre d'unifier la graphie utilisée pour représenter le contenu.

2.2. Les TICs au service de l'aménagement linguistique de l'amazighe

Après des décennies de négligence, la création de l'Institut Royal de la Culture Amazighe a motivé la revitalisation de la langue amazighe par un processus d'aménagement linguistique qui englobe un éventail d'approches, y compris la codification graphique, l'établissement de règles d'orthographe, de grammaire et de lexiques, la terminologie et la traduction.

La réussite d'un aménagement linguistique dépend de sa visibilité. C'est pourquoi l'utilisation des TICs constitue de nos jours un levier dans ce processus permettant de valoriser les efforts fournis et d'agir sur les présentations linguistiques et ce, à travers :

- L'internet pour contribuer à la diffusion des productions de recherche et faciliter l'échange des connaissances et le suivi de l'évolution du processus d'aménagement linguistique pour d'autres langues dans des situations similaires.
- La codification graphique qui définit une norme standard et une police de caractères adéquate de la graphie afin d'assurer l'exploitation et le partage des contenus.
- La gestion des ressources linguistiques à savoir les corpus, les lexiques et la terminologie, dont le développement est hautement interdisciplinaire, exige la coopération de spécialistes de nombreux domaines. Étant donné la quantité considérable de ressources linguistiques et le nombre de personnes

qui devraient travailler ensemble en unissant leurs efforts ainsi que leurs ressources, les bases de données présentent un outil essentiel de gestion et de structuration dans le processus d'aménagement linguistique.

- L'extraction automatique de terminologie qui permettra l'attestation des termes dans différents genres textuels et l'enregistrement de l'usage des termes en contextes, et ce en exploitant la richesse des ressources langagières telles que les corpus textuels.
- La traduction automatique qui assurera une meilleure compréhension entre des locuteurs de langues différentes et contribuera à la diffusion de la culture à l'échelle internationale en utilisant des outils d'analyse, de génération et de transfert à l'aide des corpus multilingues.

2.3. Les TICs au service de l'enseignement

Malgré l'initiative très significative de l'intégration de l'amazighe au sein du système éducatif marocain en 2003, elle n'est toujours pas suivie de faits concrets permettant à cette langue d'atteindre le niveau de l'arabe ou du français. Pour tenter de remédier à cet état de fait, l'application des TICs à l'enseignement est une solution primordiale pour notre communauté pour accélérer les échanges de savoir et réduire le retard accumulé dans ce domaine.

Dans une situation dominée par l'insuffisance du nombre des enseignants et formateurs et du matériel pédagogique adapté, ainsi que par la restriction de l'expérience aux classes primaires et à un certain nombre limité d'établissements, il est judicieux de pouvoir profiter suffisamment des retombées de l'utilisation des TICs au profit de l'éducation à travers :

- L'apprentissage en ligne qui présentera une piste pertinente pour favoriser la promotion des ressources humaines et la généralisation verticale et horizontale, en résolvant le manque de formateurs par le rassemblement des enseignants dispersés à travers le pays ;
- Les outils multimédias et d'interactivité qui encourageront l'apprentissage autodirigé par la garantie de la correction informatisée, faciliteront la création et amélioreront la qualité des contenus éducatifs tout en assurant la transmission du savoir porteur de la diversité des formes d'expression culturelle ;
- L'enseignement hybride basé sur la conjonction des activités en ligne et en classe qui permettra de favoriser le potentiel motivationnel des apprenants et d'améliorer la qualité de l'enseignement ;
- Un réseau professionnel d'enseignants qui facilitera l'échange du savoir et le débat autour des besoins éducatifs et des utilitaires de production d'outils pédagogiques.

Dans cette perspective, il est important de pouvoir recenser les mécanismes et les techniques nécessaires pour mener à bien la généralisation de cette vocation et la mise en place d'une plate-forme d'apprentissage en ligne assurant la modernisation des moyens pédagogiques, l'élargissement du savoir ainsi que le partage des

connaissances et favorisant l'émergence d'un enseignement de qualité. Toutefois, l'élaboration d'une telle plate-forme dépend principalement du contenu pédagogique, dont les supports du cours sont nécessairement accompagnés par des activités pédagogiques dont la linguistique computationnelle constitue un fondement important, et ce par :

- L'exploitation de la richesse des ressources langagières telles que les corpus textuels à travers des concordanciers permettant l'exploration de la diversité des situations langagières et l'extraction d'exemples authentiques dans le but d'étudier le sens et les règles d'emploi ;
- L'utilisation de correcteurs orthographique et grammatical facilitant l'apprentissage autodirigé en assurant l'automatisation de l'analyse des productions des apprenants ;
- L'exploitation des correcteurs phonétiques pour soutenir l'acquisition de la prononciation à travers des entraînements systématiques.

2.4. Les TICs au service des médias

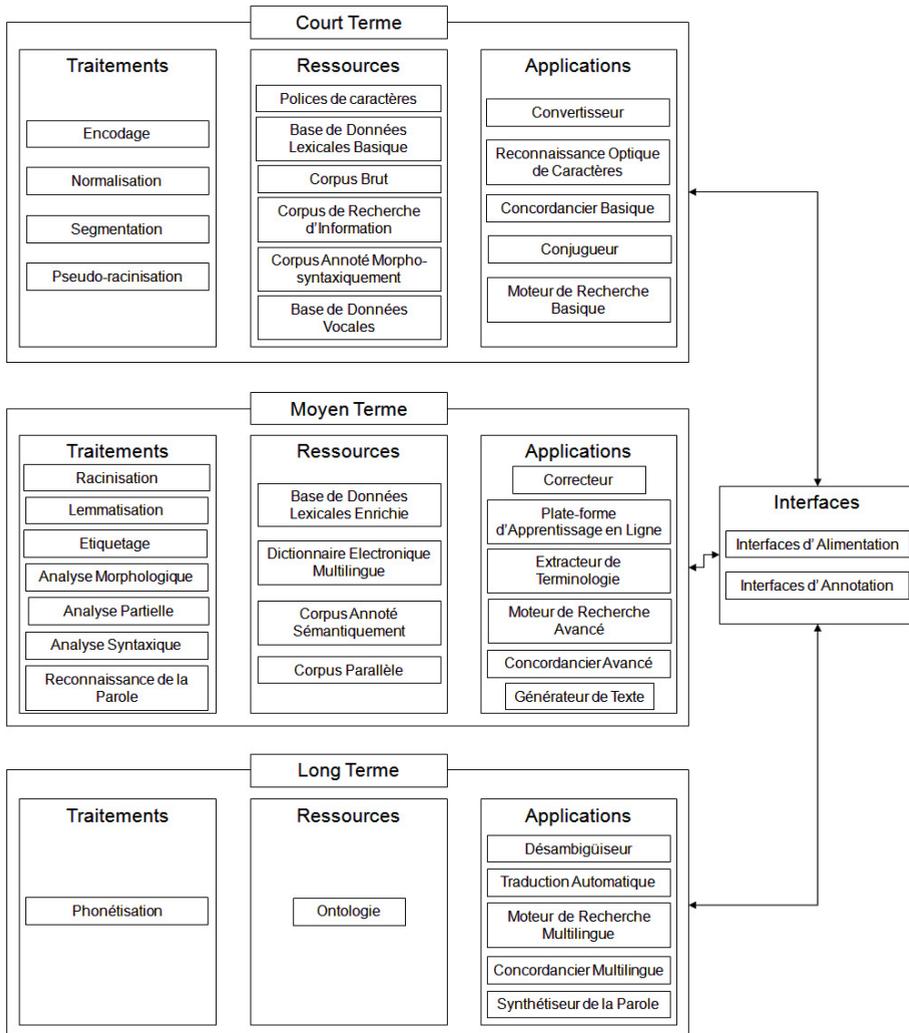
Les technologies de l'information et de la communication, principalement l'internet, ont permis de diffuser plus efficacement les moyens de communication et d'améliorer le traitement, la mise en mémoire, la diffusion et l'échange de l'information. Néanmoins, cette émergence a suscité une mutation technique des médias classiques (télévision, radio, presse, cinéma) aux niveaux de leurs chaînes de valorisation allant de la production jusqu'à la consommation de l'information, et ce à travers le passage de l'analogique au numérique et le développement des techniques du traitement de l'information, qu'elle soit écrite, orale ou multimédia, monolingue ou multilingue pour sélectionner, classer, structurer et rechercher.

Cependant, la place de l'amazighe dans les principaux médias a toujours été et reste très limitée. Dans la perspective de faire face à ce défi et de faire de l'amazighe une langue de communication, il faudrait, d'une part, adopter les outils multimédias pour valoriser et rénover son contenu médiatique et, d'autre part, exploiter les différentes plate-formes de diffusion comme les sites Web, les séminaires et les plate-formes de vidéo en ligne, les médias sociaux et les blogs pour une présentation attrayante de ce contenu.

3. Méthodologie pour l'informatisation de l'amazighe

Dans l'objectif de concrétiser l'apport des TICs aux différents axes stratégiques de l'Institut, nous avons mené une étude pour l'élaboration d'une méthodologie de mise en œuvre assurant la revitalisation et la promotion de la langue amazighe. Cette méthodologie repose sur les travaux élaborés pour l'informatisation des langues peu dotées (Berment, 2004 ; Scannell *et al.*, 2012) ainsi que sur une cartographie des différents traitements et ressources disponibles pour des langues peu, moyennement et bien dotées (Ataa Allah, 2010), afin de définir les priorités permettant de planifier une feuille de route dotant l'amazighe d'outils assurant son fonctionnement comme les langues conformément équipées (Ataa Allah et Boulaknadel, 2012).

La méthodologie proposée structure le développement du projet de l'informatisation de l'amazighe en tâches fondamentales, conçues d'une manière à ce qu'elles se déroulent en trois périodes (courte, moyenne et longue) qui peuvent s'étendre sur 5 à 10 ans selon la disponibilité humaine et les ressources linguistiques. Ces tâches sont représentées par une chaîne qui part des traitements élémentaires allant vers des applications génériques répondant généralement à des besoins et passant par la constitution de briques de ressources linguistiques (Figure ci-dessous). Cette chaîne va de l'adaptation et l'amélioration d'outils en fonction des nouvelles technologies jusqu'au développement d'applications, en se basant sur la recherche fondamentale.



Feuille de route pour l'informatisation de l'amazighe

3.1. La phase à court terme

Cette phase constitue une première étape dans le processus de l'informatisation de l'amazighe. Elle est initiée par un codage de caractères, un clavier et des polices de caractères, assurant la constitution de ressources électroniques telles que les bases de données lexicales et vocales, les corpus de recherche d'information, brut et annoté morpho-syntaxiquement. Ces ressources servent de briques élémentaires pour l'élaboration d'outils et d'applications du traitement du langage humain, en particulier celles planifiées pour cette phase, à savoir la normalisation, la segmentation, la pseudo-racinisation, le concordancier de base, le moteur de recherche, le conjugeur et le système de reconnaissance optique de caractères.

3.2. La phase à moyen terme

Après avoir établi les ressources de base et les outils fondamentaux, cette phase repose sur un traitement linguistique qui consiste en une réalisation d'une racinisation, une lemmatisation, un étiquetage morpho-syntaxique automatique, une analyse morphologique, partielle et syntaxique, et une reconnaissance de la parole. Ces traitements permettront de développer des applications avancées telles qu'un correcteur orthographique, une plate-forme d'apprentissage en ligne, un extracteur terminologique, un moteur de recherche et un concordancier avancés ainsi qu'un générateur de textes. En outre, cette phase représente une étape clef de préparation des ressources nécessaires pour la prochaine étape, y compris les bases de données enrichies, les dictionnaires multilingues, et les corpus alignés et annotés sémantiquement.

3.3. La phase à long terme

La troisième étape de la feuille de route pourrait être considérée comme une synthèse du travail réalisé. A côté d'un traitement de phonétisation et d'une élaboration d'une ontologie, d'un développement d'un désambiguïseur sémantique et d'un synthétiseur de la parole, cette phase se concentre également sur la réalisation d'applications multilingues, principalement la traduction automatique.

4. Stratégies technologiques à respecter

Dans la perspective de mener à bien le développement de ressources et d'outils nécessaires pour l'informatisation d'une langue, des études ont porté sur des stratégies technologiques qui devraient être respectées. Ainsi, nous nous sommes inspirées des travaux de Muhirwe (Muhirwe, 2007) qui suggère de prendre en considération lors du développement des applications la notion d'extensibilité et de documentation ainsi que l'adoption des technologies libres, pour dresser nos stratégies de développement préalablement nécessaires pour réussir et accélérer le processus de l'informatisation de l'amazighe.

4.1. Standardisation des ressources

L'élaboration des ressources numériques est une étape indispensable dans tout processus d'informatisation. Cependant, cette tâche est coûteuse en termes de temps et de compétences humaines. D'où l'utilité de les concevoir sous un format standard permettant à la communauté académique impliquée dans la réalisation ou l'intégration d'outils et ressources, de jouir d'un environnement garantissant la gestion et l'interopérabilité entre de tels composants.

4.2. Adaptation des technologies de langues

Une part sérieuse du temps et d'efforts nécessaires pour informatiser une langue est généralement consacrée à la réalisation de l'environnement. Cependant, cette tâche laborieuse peut être optimisée par l'intégration et l'adaptation des techniques existantes afin de profiter pleinement du potentiel qu'offrent ces technologies pour la réalisation des tâches de manière générique sans avoir recours à « réinventer la roue ».

4.3. Extensibilité

L'extensibilité est un facteur élémentaire dans la conception de tout projet en technologie des langues, assurant l'extension de ses fonctionnalités. Cette propriété permet une plus grande réutilisabilité du projet en facilitant l'ajout de nouveaux composants et l'amélioration ou l'adaptation de l'existant.

4.4. Logiciels libres

En vue de garantir l'adaptation des technologies de langues et leur extensibilité, il est nécessaire d'adopter un mode de conception et d'exploitation qui obéit à des règles de liberté susceptible d'être soumis à modification et redistribution sans restriction. Ces caractéristiques confèrent une certaine fiabilité et réactivité.

4.5. Documentation

La documentation constitue l'une des pierres angulaires dans le développement d'un projet d'informatisation d'une langue. Elle permet d'expliquer le fonctionnement du projet en déterminant ses objectifs, son architecture ou sa conception, les techniques utilisées pour son développement et son manuel d'utilisation. Ce qui peut assurer sa bonne exploitation, son extensibilité et sa réutilisation logicielle.

4.6. Evaluation des systèmes

Le recours à l'évaluation conduit à repenser les objectifs, les pratiques, voire les approches théoriques. Une évaluation peut être apportée avant la mise en œuvre d'un système, pour déterminer à l'aide d'un diagnostic les objectifs attendus et les indicateurs nécessaires à l'évaluation ; au cours de sa réalisation, afin d'ajuster le

système au besoin à travers une série d'évaluations progressives ; ou après l'achèvement du travail, et ce pour déterminer le niveau de satisfaction, la pertinence, la durabilité et l'extensibilité du système.

Par ailleurs, deux types d'évaluation sont distingués :

- L'évaluation objective fournit une mesure des performances du système en dehors de toute considération sur sa perception par les utilisateurs.
- L'évaluation subjective fondée sur le jugement des utilisateurs et qui consiste à mesurer l'adéquation du système à la tâche ainsi que son utilité en termes de convivialité, fiabilité et facilité d'utilisation.

5. Réalisations

Dans le cadre de la promotion de la langue amazighe, de nombreux travaux ont été réalisés afin de fournir à cette langue des ressources et outils permettant son traitement automatique et son intégration dans le domaine des technologies de l'information et de la communication.

Cette section introduit les différentes réalisations au niveau national¹ en les structurant en sept parties selon la nature du traitement visé, passant par le préalable à l'exploitation des TICs ; outils de recherche, d'analyse et d'assistance ; ressources langagières, localisation logicielle et reconnaissance optique des caractères.

5.1. Préalables à l'exploitation des TICs

En amont de l'exploitation des TICs dans le processus de la promotion de l'amazighe, il faut disposer de données numérisées. A cette fin, il est judicieux de définir un codage adapté, un clavier et des polices de caractères.

5.1.1. Codage de Tifinaghe

Le codage du caractère tifinaghe constitue le premier pas vers l'exploitation des TICs, permettant la transcription de l'amazighe à travers une représentation numérique pour chaque caractère. Ce processus s'est déroulé en deux étapes : la première a consisté en une adaptation de la norme ISO 8859-1 pour coder les caractères tifinaghes en codage ANSI, afin de répondre à l'urgence de l'introduction de l'amazighe dans le système éducatif. Cependant, la portée de ce codage privé de l'IRCAM a été limitée, et la gestion des textes comportant plusieurs systèmes d'écriture était difficile. D'ailleurs, les traitements des textes multilingues doivent jongler à la fois avec les différentes normes de codage et avec les polices associées. D'où la nécessité et l'importance d'intégrer le tifinaghe dans le plan multilingue c'est-à-dire un codage plus portable et facile à gérer tel que l'Unicode qui a permis d'affecter un code unique à chaque caractère (Andries, 2008).

¹ Dans cet article, nous nous sommes limitées au niveau national, néanmoins les réalisations au niveau international pourront faire l'objet d'une autre contribution.

5.1.2. Clavier

L'intégration de l'amazighe dans la norme internationale de la prescription des claviers ISO/CEI 9995 a fixé définitivement les claviers de la langue amazighe conçus pour la bureautique. En fait, cette norme a spécifié deux types de claviers : un clavier de base contenant les caractères tfinaghes préconisés par l'IRCAM et un clavier étendu comportant tous les caractères adoptés par l'ISO.

5.1.3. Polices de caractères

Dans l'objectif d'intégrer la langue amazighe dans le système éducatif et de favoriser la publication assistée par ordinateur, huit polices de caractères, associées au codage ANSI, ont été réalisées. Cette réalisation a été suivie par l'élaboration d'une nouvelle génération de polices, qui a constitué un saut qualitatif de l'universalisation de l'écriture amazighe. Cette génération inclut les polices associées à l'Unicode.

5.1.4. Convertisseur

Dès l'encodage Unicode des caractères tfinaghes, il a fallu préparer des outils pour convertir la représentation de ces caractères d'un codage à un autre. Par ailleurs, 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 caractères arabe et latin. A cette fin, un convertisseur a été réalisé, permettant le passage du codage ANSI au codage Unicode et la translittération du caractère arabe et du caractère latin au caractère tfinaghe (Ataa Allah et *al.*, 2013).

5.2. Outils de recherche

Dans la perspective de contribuer à la diffusion de la langue et de la culture amazighes, il est intéressant de développer des applications d'exploration et de recherche.

5.2.1. Moteur de recherche

Les moteurs de recherche sont actuellement la première source d'information sur le Web pour plusieurs domaines et leur utilisation est devenue incontournable. D'ailleurs, les sondages et enquêtes qualitatives effectués en la matière démontrent bien l'importance de ces applications qui répondent aux besoins professionnels et personnels suscités par les internautes².

Conscients de ce fait et afin de contribuer à la promotion et la diffusion de la culture et la langue amazighes à travers le Web et d'assister les internautes à découvrir la diversité et la richesse de la culture amazighe, un moteur de recherche supportant la graphie tfinaghe a été développé (Ataa Allah et Boulaknadel, 2010a).

² http://www.nielsen-online.com/pr/pr_040223_us.pdf

Ce dernier est basé sur un ensemble de traitements linguistiques, à savoir, l'élimination des mots anti-dictionnaires³ et l'exploitation de la pseudo-racination.

5.2.2. Concordancier

Les concordances sont un mode de présentation d'extraits de texte, contenant le même mot ou le même motif linguistique, basées sur une méthodologie d'analyse textuelle laborieuse. Cependant, le recours aux outils de concordance numérique a facilité cette tâche séculaire que nécessitaient les concordances manuelles. Dans cette perspective, un outil de concordance supportant les scripts arabes, latins et Unicode de l'amazighe a été développé. Il permet d'effectuer des recherches dans des textes numériques et répond au besoin d'exploration suscitée par l'utilisateur dans le but d'étudier le sens et les règles d'emploi (Ataa Allah et Boulaknadel, 2010a).

5.3. Outils d'analyse

Afin de doter l'amazighe d'outils d'analyse et de génération, un ensemble de travaux portant sur la morphologie flexionnelle et dérivationnelle a été effectué.

5.3.1. Pseudo-racineur

L'outil de pseudo-racination est basé sur une approche relevant du cas de la morphologie flexionnelle et reposant sur l'élimination d'une liste de suffixes et de préfixes de la langue amazighe dans un ordre prédéterminé. Cette approche légère permettant de regrouper les mots sémantiquement proches à partir de ressemblances pourra être facilement exploitée dans des applications telles que la recherche d'information et la classification (Ataa Allah et Boulaknadel, 2010b).

5.3.2. Conjugueur

Dans un contexte de développement d'outils de génération pour la langue amazighe, un conjugueur a été réalisé pour faciliter l'apprentissage de la langue. Cet outil est conçu à la base de modèles de comportement flexionnel augmenté de règles de régularisation morphologiques (Laabdelaoui et *al.*, 2012). Il permet la conjugaison en ligne des formes verbales simples et dérivées dans différents aspects et modes de l'amazighe.

5.3.3. Analyseur morphologique

Etant donné que toute analyse linguistique passe par une première étape d'analyse morpho-lexicale, qui consiste à tester l'appartenance de chaque mot du texte au lexique de la langue, il a été entrepris de développer des systèmes d'analyse morphologique pour l'amazighe, profitant des apports des modèles à états finis et faisant appel à une formalisation du vocabulaire et à des règles grammaticales à

³ Les mots non significatifs ou non discriminants pour la recherche d'information, tels que les prépositions, les conjonctions et les déterminants.

large couverture. Un premier système a été développé à la base de l'environnement linguistique NooJ. Il s'est focalisé essentiellement sur la formalisation flexionnelle et dérivationnelle de la catégorie nom (Nejme et *al.*, 2013). Cependant, le deuxième système a été réalisé en exploitant l'environnement de développement Xerox en se basant principalement sur la formalisation flexionnelle de la catégorie verbe (Ataa Allah, 2014).

5.4. Ressources langagières

La réalisation de ressources langagières (jeux d'étiquettes, dictionnaires, terminologies, corpus oraux et écrits,...) est une activité de mise en œuvre d'infrastructures qui doit être perçue à un niveau amont étant donné, d'une part, le coût prohibitif d'une telle tâche, et d'autre part le fait de constituer un pré-requis indispensable pour le développement des applications de technologie de langue à haute valeur ajoutée.

5.4.1. Jeux d'étiquettes morpho-syntaxiques

Dans la perspective d'annotation de corpus amazighe, un jeu d'étiquettes morpho-syntaxiques a été élaboré à la base de la *Nouvelle grammaire de l'amazighe* (Boukhris et *al.*, 2008). Ce jeu s'inspire du projet EAGLES⁴ où le choix des étiquettes repose sur une distinction entre étiquettes obligatoires, étiquettes recommandées et extension particulière. Ainsi, il est proposé deux listes d'étiquettes spécifiques aux caractéristiques de l'amazighe. L'une contient les étiquettes obligatoires ; l'autre, des étiquettes recommandées fournissant des indications plus précises (Ataa-Allah, 2011).

5.4.2. Dictionnaire

Dans la perspective d'appuyer l'enseignement et l'apprentissage de l'amazighe, un dictionnaire imagier sonore en ligne a été conçu pour les enfants âgés de 2 à 14 ans (Ataa Allah, 2011). Il se base sur l'interaction entre les connaissances linguistiques et extralinguistiques afin d'offrir à l'enfant un simple moyen pour stimuler son langage et son éveil. Ce dictionnaire comporte actuellement 550 entrées lexicales réparties sur 105 catégories regroupées en 18 thématiques. Chaque entrée est déterminée par 4 langues (anglais, amazighe, arabe, français) et représentée par une image et un son.

5.4.3. Terminologie

Dans l'objectif de doter l'amazighe d'une terminologie couvrant le plus grand nombre de champs lexicaux, une base de données terminologiques a été élaborée. Elle comporte des entrées terminologiques relevant de plusieurs thématiques, y compris les lexiques usuels, des médias et grammaticaux ainsi que leurs équivalents arabes et français (El Azrak et El Hamdaoui, 2011).

⁴ <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>

5.4.4. Corpus

Dans l'objectif de développer des outils de traitement automatique de l'amazighe et pour pouvoir se fonder sur l'usage réel de la langue, un recueil de textes numérisés et codés a été réalisé.

- **Corpus brut**

Dans le cadre d'élaboration de corpus électronique pour la langue amazighe, un corpus a été collecté, composé de 160 textes amazighes représentant différents genres littéraires, à savoir conte, conte pour enfants, poésie et articles de presse contenant les sous-genres journal, magazine et Net. Il comprend 136.093 occurrences dont 23.174 sont des mots distincts (Boulaknadel et Ataa Allah, 2011).

- **Corpus étiqueté**

Les ressources lexicales annotées ne sont que rarement disponibles, particulièrement pour les langues peu dotées telle que l'amazighe. Dans ce contexte, une démarche d'élaboration de corpus étiqueté a été menée. Ainsi, le corpus réalisé comprend environ 20 k mots et se compose de textes extraits d'une variété de sources telles que la version amazighe du site Web de l'IRCAM, le périodique « Inghmisen n usinag » (bulletin d'information de l'IRCAM) et des manuels scolaires (Outahajala et *al.*, 2010).

5.5. Outils d'assistance

Dans l'objectif de faciliter et de mener à bien le processus d'élaboration de ressources langagières, des applications d'assistance ont été réalisées.

5.5.1. Base de données lexicales

Bien que de nombreux dictionnaires papiers de la langue amazighe soient édités, aucune base de données numérique n'est disponible. Pour faire face à ce manque, une application assurant la gestion des collectes et l'exploration des lexiques amazighes a été élaborée. Cette application permet de structurer les informations intra et inter-lexicales telles que la définition, les équivalents arabes et français, les synonymes et le classement par thèmes (Iazzi et Outahajala, 2008).

5.5.2. Outil d'assistance à l'étiquetage morpho-syntaxique

L'utilisation des corpus, notamment ceux annotés morpho-syntaxiquement, est devenue une étape indispensable dans un processus d'informatisation d'une langue. Cependant, cette tâche est d'une grande complexité et son coût en termes de temps et de ressources humaines limite la quantité et la disponibilité des corpus. Afin de minimiser les efforts nécessaires pour produire des corpus amazighes et d'améliorer leurs qualités en permettant des vérifications et en simplifiant les modifications et les mises à jour, un outil d'assistance à l'étiquetage morpho-syntaxique a été développé, permettant une automatisation partielle de la tâche

d'étiquetage et assurant une bonne cohérence entre les différents travaux réalisés par l'ensemble des chercheurs (Ataa Allah et Jaa, 2009).

5.5.3. Base de données littéraires

Dans une perspective de sauvegarde du patrimoine littéraire amazighe, une application visant la collecte et la mise en réseau d'un noyau d'une banque de corpus littéraires, contenant principalement des données textuelles et audiovisuelles amazighes, a été développée. Cette application permet la gestion et la consultation du contenu littéraire (Ait Ouguengay et *al.*, 2012).

5.6. Localisation logicielle

A travers la localisation de l'interface de deux cellulaires Sony Ericsson « J110i et J120i » et du système d'exploitation Microsoft « Windows 8 », une opportunité a été offerte à l'amazighe pour assurer sa présence dans le monde technologique et offrir à l'utilisateur un environnement familier, où l'expérience de l'intégration de la langue amazighe et son système d'écriture tifinaghe a été menée.

5.7. Reconnaissance optique des caractères

Dans une vision de contribuer à la sauvegarde et la numérisation du fonds documentaire amazighe, de nombreuses études ont été menées sur la reconnaissance optique de caractères amazighes dont le taux de reconnaissance est autour de 92%. Ces études se basent sur différentes approches pour la reconnaissance du caractère imprimé et manuscrit. Elles sont regroupées généralement en trois classes, à savoir les réseaux de neurones (Ait Ouguengay, 2009), (Elyachi et *al.*, 2009), une approche syntaxique fondée sur les automates à états finis (Es Saady et *al.*, 2010) et les modèles de Markov cachés (Amrouch et *al.*, 2010).

6. Conclusion

Dans un contexte général visant la pérennité et la promotion de la langue amazighe, les TICs constituent une piste prometteuse pour lutter contre la fracture numérique dont l'amazighe a souffert. Dans cette perspective, le présent article propose une méthodologie d'informatisation de l'amazighe pour agir de façon ciblée et efficace dans quatre domaines clefs, à savoir la sauvegarde du patrimoine, l'aménagement linguistique, l'éducation et les médias. Le fait de mettre l'accent sur ces domaines permettra de donner aux Marocains les moyens de s'épanouir en amazighe comme en arabe et en français et de contribuer au développement de la société. Cette méthodologie se base sur les travaux élaborés pour l'informatisation des langues peu dotées et suit un plan de développement en trois étapes partant de la réalisation d'un kit d'outils et de ressources linguistiques minimaux jusqu'à l'élaboration d'applications avancées. En outre, l'article dresse un bilan de réalisation d'outils et de ressources linguistiques amazighes permettant son traitement automatique et son intégration dans le domaine des TICs.

Références

- Ait Ouguengay, Y. *et al.* (2012), « Projet GCAM- Vers une gestion informatisée du corpus amazighe à l'IRCAM », in *TICAM 2012*, 26-27 novembre 2012, Maroc.
- Ait Ouguengay, Y. et Taalabi, M. (2009), « Elaboration d'un réseau de neurones artificiels pour la reconnaissance optique de la graphie amazighe: Phase d'apprentissage », in *Systèmes intelligents-théories et applications*. Europaia productions.
- Amrouch M. *et al.* (2010), « Handwritten Amazigh Character Recognition Based On Hidden Markov Models. International Journal on Graphics », *Vision and Image Processing*, Vol. 10, n° 5, p. 11-18.
- Andries, P. (2008), *Unicode 5.0 en pratique : Codage des caractères et internationalisation des logiciels et des documents*, France, Dunod.
- Ataa Allah, F. (2010), « Etude comparative au niveau de ressources et outils des langues peu, moyennement et bien dotées », Rapport interne, CEISIC, IRCAM.
- Ataa Allah, F. (2011), « Construction de corpus étiqueté des documents textuels de la langue amazighe », Rapport interne, CEISIC, IRCAM.
- Ataa Allah, F. (2014), «Finite-state transducer for Amazigh verbal morphology», in *Literary and Linguistic Computing*, Oxford University Press. doi: 10.1093/lc/fqu045.
- Ataa Allah, F. et Jaa, H. (2009), « Etiquetage morpho-syntaxique: outil d'assistance dédié à la langue amazighe », in *SITACAM 2009*, 12-13 decembre 2009, Maroc.
- Ataa Allah, F. et Boulaknadel, S. (2010a), « Amazigh Search Engine: Tifinaghe Character Based Approach », in *IKE 2010*, 14-16 juillet 2010, USA.
- Ataa Allah, F. et Boulaknadel, S. (2010b), « Pseudo-racinement de la langue amazighe », in *TALN 2010*, 19-23 juillet 2010, Canada.
- Ataa Allah, F. et Boulaknadel, S. (2012), « Toward computational processing of less resourced languages: Primarily experiments for Moroccan Amazigh language », in *Text Mining*, (éd) Rijeka, InTech, p. 197-218.
- Ataa Allah, F. *et al.* (2013), « Amazigh Language Desktop Converter », in *SITACAM 2013*, 2-4 mai 2013, Maroc.
- Berment, V. (2004), *Méthodes pour informatiser des langues et des groupes de langues «peu dotées»*, Thèse de Doctorat, Université Joseph Fourier, France.
- Boukhris, F. *et al.* (2008), *La nouvelle grammaire de l'amazighe*, Rabat, IRCAM.
- Boulaknadel, S. et Ataa Allah, F. (2010), « Online Amazigh Concordancer », in *ISIVC 2010*, 30 septembre - 2 octobre 2010, Maroc.
- Boulaknadel, S. et Ataa Allah, F. (2012), « Building a standard Amazigh corpus », in *IHCI 2011*, 29-31 août 2011, Tcheque.

ElYachi R. *et al.* (2010), « On the Recognition of Tifinaghe Scripts ». *Theoretical and Applied Information Technology*, vol. 20, n° 2, p. 61-66.

El Azrak, N. et Elhamdaoui, A. (2011), « Référentiel de la terminologie amazighe : outil d'aide à l'aménagement linguistique », in *NTIC 2011*, 24-25 février 2011, IRCAM.

Es Saady, Y. *et al.* (2010), « Printed Amazigh Character Recognition by a Syntactic Approach using Finite Automata ». *International Journal on Graphics Vision and Image Processing*, vol. 10, n° 2, p. 1-8.

Iazzi, E.M et Outahajala, M. (2008), « Amazigh Data Base », in *HLT & NLP Workshop*, 31 mai 2008, Maroc.

Laabdelaoui, R. *et al.* (2012), *Manuel de conjugaison amazighe*, Rabat, IRCAM.

Muhirwe, J. (2007), « Towards Human Language Technologies for Under-resourced languages », *Computing and ICT Research*, (éd) Joseph Kizza et al, Kampala.

Nejme, F.Z. *et al.* (2013), « Analyse Automatique de la Morphologie Nominale Amazighe », in *TALN 2013*, 17 - 21 juin 2013, France.

Outahajala, M. *et al.* (2010), « Tagging Amazigh with AnCoraPipe », in *HLT & NLP Workshop*, 17 mai 2010, Malta.

Scannell. K.P. *et al.* (2012), *The Irish Language in the Digital Age / An Ghaeilge sa Ré Dhigiteach*. White Paper Series, Springer-Verlag Publisher.