# Amazigh-Sys: Intelligent system for recognition of amazigh words

**Rachid Ammari, Lahbib Zenkouar**
Smart Communications LAB, Mohammadia School of Engineers, Mohammed V University of Rabat, Morocco

| Article Info | ABSTRACT |
|---|---|
| | Amazigh-sys is an intelligent morphological analysis system for Amazigh language based on xerox's finite-state transducer (XFST). Our system can process simultaneously five lexical units. This paper begins with the development of Amazigh lexicon (AMAlex) for attested nouns, verbs, pronouns, prepositions, and adverbs and the characteristics relating to each lemma. A set of rules are added to define the inflectional behavior and morphosyntactic links of each entry as well as the relationship between the different lexical units. The use of finite-state technology ensures the bidirectionality of our system (analysis and generation). Amazigh-sys is the first general morphological analysis system for Amazigh based on xerox finite state able to process and recognize all lexical units and ensures a high recognition rate of input words. This contribution facilitates the implementation of other applications related to the automatic processing of the Amazigh language. |
| | |

*Corresponding Author:*

Rachid Ammari
Laboratory of Smart Communications
Mohammadia School of Engineers
Mohammed V University of Rabat, Morocco
Avenue Ibn Sina B.P 765, Agdal Rabat 10090 Morocco
Email: ammari.rachid@hotmail.fr

## 1. INTRODUCTION

The Amazigh language (Tamazight) also called Berber language is considered as a family of languages in Morocco. Nowadays the Amazigh language is considered as the axis of culture in North Africa in general and Morocco in particular [1], [2]. In the past, Tamazight in Morocco was neglected and marginalized by political decisions. The cultural, political activism, and the union of all the components of the Moroccan society made the Amazigh language known and organized. On 2011, the Moroccan Kingdom announced the constitutional reform that Tamazight became official language and will be introduced in all administrations. With the creation of the Royal Institute of Amazigh Culture (Institut Royal de la Culture Amazigh IRCAM), the Amazigh language has enjoyed an official spelling [3], has acquired in the standard Unicode [4], [5] and appropriate standards for keyboard realization with linguistic structures [6], [7]. However, this progress remains insufficient for the Amazigh language, which aims to join the group of well-resourced ones in terms of language technologies, especially those related to morphology. In this optic, many scientific researchers are undertaken, in order to strengthen its presence in the area of human language technology applications.

In terms of linguistic resources: the studies carried out on promotion of Tifinaghe: Unicode keyboards, character recognition [8], [9], Tifinaghe character-based approach [10], and standard Amazigh corpus [11]. In terms of computational linguistic tools: the main works concern online Amazighe

Concordancer [12], verb Conjugator [13], and primarily experiments for Moroccan Amazigh language [14]. In terms of morphological analysis: Despite the importance of this module as a main and essential step in several modules relating to for NLP applications, several works have been done relating to lexical units: System of analysis and generation for Amazigh nominal morphology [15]. Verbal morphological analyzer/generator for Amazigh moods [16], and finite-state morphological analyzer for Amazigh [17]. However, none of them can be considered as a general morphological analyzer for the Amazigh language. In this context, the aim of this work is to build general system able to recognize nouns, verbs, pronouns, prepositions and adverbs.

For decades, the field of computational linguistics has a long list of challenges and choices. Adopting an approach for the morphological analysis is among the strategic decisions to reach the goal. The choice between rule-based [18] and machine learning statistical methods [19] depends mainly on the availability of resources. Therefore, in the current paper, given the shortage of linguistic resources modeling Amazigh morphology by machine learning methods and these resources remain unable to satisfy the terminology of the Amazigh language we opted for rule-based approach.

The use of the XFST platform in the current work is due to several reasons. On the one hand, it has proven its efficiency in the treatment of several languages as several European languages [20], some sub American languages [21], and some Asiatic languages as Arabic language [22]. On the other hand, XFST is capable to process and manage the characteristics of each language whether inflectional, derivational concatenative, nonconcatenative, or agglutinative language. This tool also presents another strong point, the analysis and generation of all the lexical units that constitute the Amazigh language. Therefore, this work contains five main parts. We start with a basics background the Amazigh language and the specificities of Amazigh morphology was provided in section 2. The section 3 exposes technical characteristics of XFST toolkit technology. Section 4 and 5 presents respectively our intelligent system for recognition of Amazigh words based on Xerox finite-state and the last one talks about conclusions and outlines our roadmap for the near future.

## 2.    AMAZIGH LANGUAGE: HISTORICAL BACKGROUND

The Berber languages or Berber, in Berber "Tamazight" form a branch of the family of Chamito-Semitic (or Afro-Asiatic, or Afrasian) languages and cover a vast geographical area: North Africa from Morocco to Egypt, via Algeria, Tunisia and Libya, as well as part of the Sahara, and the western part of the Sahel (Mauritania, Mali and Niger) [23]. With the creation of Royal Institute of Amazigh Culture (IRCAM) in Morocco and having a main objective: the promotion and conservation of the Amazigh language and culture, this language was able to cross the teaching, administration, and media sector. After the officialization of the Amazigh language, IRCAM has defined a strategic standardization plan, which revolves around several axes: adapting a graphic system and a common basic lexicon, applying the same orthographic rules, and the same instructional for Amazigh language [24], [25]. Amazigh transliteration-in accordance with the ALA-LC system-follows, with minimal exceptions, the romanization method used by IRCAM for texts published in Amazigh language using neo-Tifinaghe writing. The transliteration used during this work is Tifinagh-Latin as shown in Table 1.

Table 1. Transliteration Tifinagh-Latin

| ⵄ (a) | ⵔ (b) | ⵅ (g) | ⵠ (d) | ⵣ (Ḍ) | ⵦ (e) | ⵋ (f) |
|-------|-------|-------|-------|-------|-------|-------|
| ⵁ (h) | ⵟ (Ḥ) | ⵂ (ɛ) | ⵜ (kh) | ⵇ (q) | ⵕ (i) | ⵊ (j) |
| ⵏ (m) | ⵉ (n) | ⵓ (u) | ⵔ (r) | ⵗ (Ṛ) | ⵖ (ɣ) | ⵙ (s) |
| ⵛ (sh) | ⵜ (t) | ⵟ (Ṭ) | ⵡ (w) | ⵢ (y) | ⵣ (z) | ⵥ (Ẓ) |
| ⵀ (bh) | ⵅ (gh) | ⵅ (dj) | ⵟ (dj) | ⴸ (d) | ⴰ (h) | ⵎ (kh) |
| ⵍ (p) | ⵯ (w) | ⵞ (č) | ⴰ (v) | ⴾ (k) | ⵍ (l) | ⵕ (S) |

## 3.    AMAZIGH MORPHOLOGY

Like most Semitic languages, the morphology of Amazigh is characterized by a complex and productive morphology and rich in terms of inflection and derivation; it has a structure multi-level and applies a non-concatenative morphotactic. This complexity is due to their process training which is based on the root (a sequence of one or more consonants (C)) which is embedded in a scheme (vowel pattern (V) and sometimes consonants (C)), to which can be added to a variable number of affixes (morpho-syntactic characteristics) and or clitic attachments [26].

### 3.1. Amazigh noun morphology

The Amazigh noun is a lexical unit that denotes an individual object, it depends in gender (feminine: ⵜⴰⴼⵔⵓⵅⵜ [tafruxt] "girl ", masculine: ⴰⴼⵔⵓⵅ [afrux] "boy"), in number (singular: ⴰⴼⵔⵓⵅ [afrux] "boy", plural: ⵉⴼⵔⴰⵅ [ifrax] "boys") and in state (free: ⴰⵏⴱⵢⵉ [anbyi] "guest", annexation: ⵜⵉⵏⴱⵢⴰ [tinobya] "Hospitality"). The peculiarity of masculine noun is their beginning, which must include one of the vowels ⴰ [a], ⵉ [i] or ⵓ [u], for example: (ⴰⵔⴳⴰⵣ [argaz] *"men"*; ⵉⵙⵍⵉ [isli] *"groom")*; ⵓⵅⵙⴰⵙ [uxsas] *"head"*). The feminine case starts with the morpheme ⵜ..., which characterizes the transition from masculine to feminine gender for example: (ⵉⵙⵍⵉ [isli] *"bride")* ➔ Feminine noun: ⵜⵉⵙⵍⵉⵜ [tislit] "bride".

The number is defined by a singular and three categories of plural: external, internal and mixed plural.

− For the external plural, we add the suffix "-- ⵏ [--n]" for masculine and "--ⵉⵏ [--in]" for Feminine form.
− The internal plural is formed by a change of internal vowels (rarely of a consonant).
− Mixed plural is marked by the alternation of an internal vowel and/or a consonant, and by the suffix -ⵏ [-n] (suffixation + internal alternation): ⵉⵣⵉ [izi] "fly" => ⵉⵣⴰⵏ [izan] "flies"

The Amazigh noun covers two states: the free state and the annexation state/constructed state. In the first one, the initial vowel remains unchanged and the second one changes the initial in certain syntactic contexts taking one of the following forms:

− Vocalic alternation ⴰ [a] ➔ masculine noun ⵓ [u]: ⴰⵔⴳⴰⵣ [argaz] "men" ➔ ⵓⵔⴳⴰⵣⵜ [**u**rgaz].
− Addition of a ⵡ [w] or ⵢ [i] to vowel nouns ⴰ [a] or ⵉ [i].
− The initial vowel ⴰ [a] remains fixed with appearance of the semi-consonant ⵡ [w] for the masculine but the feminine nouns remain unchanged.

### 3.2. Amazigh verb and moods

The verb forms a single graphical word with its obligatory marks (indices of persons, signs of appearance and derivational morphemes [causative, reciprocal or passive]) and also depends from mood (indicative, imperative or participial), aspect (aorist, perfective, negative perfective or imperfective), gender (feminine or masculine), number (singular or plural), and person (first, second or third). The Amazigh verb has two forms: basic and derived. The fusion of a root and a pattern constitutes the basic forms called also "radical". A sequence of one or more consonants forms a root and the pattern is considered a template of vowels (V) and consonants (C), for example: (Root: ⵏⵂⵍⵍⴰ [nhlla], pattern: C'C'CCC , radical: ⵍⵍⵂⵍⵍⴰ "to take care"). The derived one is formed by the combination of the basic form and the prefixes according to the context of the form, for example: [ⴼⵙⵜ [fstigh] "i stop talking"; ⴼⵙⵜⴰⵏ [fstan] "they stopped talking" (masculine); ⴼⵙⵜⴰⵏⵜ [fstant] "they stopped talking" (feminine); ⵉⴼⵙⵜⴰ [ifsta] "He stops talking"]. If the verb begins with initial vocalic, the third-person singular is indexed by the letter [y] (ⵓⵎ�ⵛ [umç] ➔ ⵢⵓⵎⵛ [y umç]).

### 3.3. Pronoun

Amazigh pronoun refers to any element that could replace a noun, nominal group or could be attached to a verb. In Amazigh language, there are two categories of pronouns: Autonomous personal pronouns, as (masculine singular nkk [nkk] "me", kyyi [kyy] "you" ntta [ntta] "him"; feminine singular nkk [nkk] "me" kmmi [kmmi] "you" , nttat [nttat] "her" plural masculine nkkni [nkkni] "us" , knni [knni] "you" , nitni [nitni] [them (masculine)]; knninti [knninti] "you (plural)"; nintnti [nintnti] "them (feminine)." The affixed pronouns to the verb and noun: depend on gender (masculine, feminine), number (singular, plural), and person (first, second, third). This dependency generates twenty possible cases for the noun and eighteen for the verb. Example of case: pronoun for direct form: ⵅⵇⵉⵞ ⵜ [zrikht] "I saw it" (masculine) or ⵅⵇⵉⵞ ⵜⵜ [zrikhtt] "I saw it" (feminine) in the third person singular.

### 3.4. Prepositions

It is a function word witch not assigned neither to the noun nor to the verb; it can be simple like gr [gr], al [al]/ ar [ar], it is a closed paradigm that regroups both simple and complex forms that express various semantic values. For the first form, consists of the following formats: gr [gr], al [al]/ ar [ar], n[n], i [i], s [s], g[g], di [di], gz [zg], xh [xh], bla [bla], r [ r], dar [dar], agd [agd], d [d]. For the second one, it is composed of two or three prepositions, which may be used or composed of two or three prepositions adverbially:' ⵉⵣⴷⴰⵔⵏ [izdarn] "below".

## 3.5. Adverbs

Adverbs are considered as elements that can change the meaning of a verb. Often, they are classified according to their semantics. The distinction of adverbs is based on the notion of place (Λₒ [da] "here"), time (ₒ⊙⊙ₒ [assa] "today"), quality ([ΛOₒ⊙] "little") and manner (ⵍⵎⵉⵃ [mlih] "well").

## 4. XEROX FINITE-STATE TOOLKIT (XFST)

A finite-state transducer (FST) is a finite state machine that contains two memory strips, according to the terminology of Turing machines: an input tape and an output tape. The role of a finite state transducer is to link a relationship. For each input string, FST creates its corresponding output. This passage between these two levels is reflected in the morphological analysis by the switch between lexical level and surface level. In the surface level or upper-side (ΛΛₒⵍ [ddant] "they left") becomes **ΛΛₒⵍ+Verb+Indi+Fem+Plr+3ʳᵈpers** in lexical level or lower-side which means that our entry is recognized as a verb in indicative mood in third-person plural feminine. The sign "+" is added by convention. Xerox finite-state toolkit (XFST) was described by Beesley and Karttunen [20]. It is considered as an integrated set of software tools able to convert a text and a binary file to the simple automata and transducers. XFST tools offers efficient programs to promote the linguistic development of traditional grammar components such as lexicons and morphotactic rules etc.

## 4.1. XFST interface

The XFST interface is based on regular expression compiler. Each compilation generates a network. Usually, the regular expressions have a simple syntax and the expressions often complex. The output of lexc grammar represents as input for this component and the labeled grammatical characteristics are added to additional rules in order to obtain the acceptable surface forms. The XFST possesses also many regular expression operators other than concatenation, union and subtraction such as cross product of two regular expressions [27].

## 4.2. Lexc grammar

Lexc grammar is a lexicons compiler designed for processing lexicons related to natural languages as well as inflectional and derivative management rules and all irregularities of morphotactic structures. Morphologically, it is considered as a standard formalism which specifies the highest lexical level. Each lexc compilation of input file generates a finite transducer. The lexicon contains roots, stems, and inflectional or derivational forms. The structure of a lexc grammar file is formed by:

– Multichar_Symbols is composed of verb, and adverb…, which should be declared under the declaration section. By convention each declaration is started with the sign "+", example to declare noun "+Noun", verb "+Verb".
– The lexicon root is obligatory, it represents the starting point for each compilation
– Other lexicon in the file represents a state, which is a continuation class.

## 5. SYSTEM ARCHITECTURE

The complex nature of Amazigh morphology requires a reliable and efficient work tool capable of handling language constraints, such as nonconcatenative, morphotactic, and morphophonological treatment. The technology chosen must allow morphological analysis and generation (bidirectional). For previous needs, the XFST technology was selected for modeling our system.

## 5.1. Amazigh-Sys: implementation and evaluation

Our aim is to conceive and build a general system (Morphological analysis system for Amazigh language based on Xerox finite state) capable of recognizing simultaneously both Amazigh nouns, verbs, pronouns, prepositions, and adverbs in a given input and identifying the component morphemes of the forms using large coverage morphological grammars through a morphological analysis. Our system architecture is articulated around two axes, the first axis defines the lexicons AMAlex that contains the five lexical units and the second axis represents a description of morphotactic rules that define the constraints on possible morpheme combinations. These two axes correspond to two levels of abstraction composing our created finite-state network. Then these levels are immediately switched to more concrete levels allowing morphological rules to convert nouns, verbs, pronouns, prepositions, and adverbs to their surface form. All the rules for each level-managed by autonomous transducers are combined by the AMAlex lexicon in order to complete the full circuit giving rise to a global and unique which will be the basis of our morphological

analysis (analysis and generation) of Amazigh words. In other words, it is a mapping between the lower side and the upper side as shown in Figure 1.

For example, the verb ⴰⴰⴲⵍ "they left" in the upper side becomes ⴰⴰⴲⵍ+amazigh_verb+past+3rdpers +masc_plr in the lower side, that means that our analysis system has recognized our lexical unit as a verb in masculine plural conjugated in the past and in third person plural.
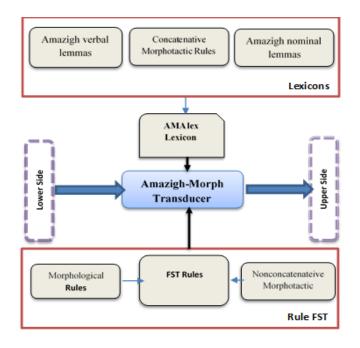


Figure 1. Intelligent system for recognition of Amazigh words

## 5.2. AMAlex: Lexicon

The lexicon is the main and preliminary step of each morphological transducer. Its quality is ensured by the quality of input words and each affects directly the processing levels. In this context, a set of lists of 3000 distinct verbs from Amazigh manual conjugation [28], 4000 distinct nouns, 1000 adverbs and prepositions from a set of resources such as Taifi dictionary [29], vocabulary resources [30], [31]. With their rules covering the morphotactic and morphological phenomena, these words lemmas represent a large part our input in Table 2. The lexicon after compilation is largely superior compared to the lexicon of entry, it is normal since the Amazigh language is inflectional (226000 words).

Table 2. An example of the input/output of upper side

| Input | Lexical unit | Output |
|---|---|---|
| ⴰⴰⴲⵍ | Verb | ⴰⴰⴲⵍ |
| | | ⴰⴰⴲⵜ |
| | | ⴳⴰⴰⴲ |
| ⴲⵅⵛⴲⵔ | Noun | ⴲⵅⵛⵔ |
| | | ⵜⵅⵛⴲⵔⵜ |
| | | ……… |
| ⴲⵛⵔⴲⵓⴲⴻ | Affixed pronoun of noun | ⴲⵛⵔⴲⵓⴲⴻⴳⵉⵙ |
| | | ⴲⵛⵔⴲⵓⴲⴻⵍⵍⴿ |
| | | ⴲⵛⵔⴲⵓⴲⴻⵍⵍⵛ |
| ⴲⵍⵍⴲⵙⵅⵙⴰ | Affixed pronoun of Verb | ⴲⵍⵍⴲⵙⵅⵙⴰⵜ |
| | | ⴲⵍⵍⴲⵙⵅⵙⴰⵜⵜ |
| | | ⴲⵍⵍⴲⵙⵅⵙⴰⴲⴲⵔ |
| | | ⴲⵍⵍⴲⵙⵅⵙⴰⴿⵙⵛ |

### 5.2.1. Rules notion

The rules of our System transducer are based on combinations of morphemes (Morphotactics) and and describes the morphology of morphemes, and describes the morphology of morphemes. Each rule implemented in the finite-state circuit reflects a phenomenon relating to the aspect of the Amazigh language. Therefore, each verbal lemmas was reclassified into 19 inflections and affixed with 18 pronoun, for example a verb in the third person masculine singular in indicative mood, is considered as an inflection. For Amazigh nominal lemmas we talk about six inflections [gender (feminine, masculine), in number (singular, plural) and in state (free, annexation)], and 20 affixed pronouns. Nevertheless, adverbs and prepositions are considered as function words, which are not attached to any other lexical entities.

### 5.3. Amazigh-Sys: analysis/generation

The main advantage of Amazigh-Sys is its bi-directionality. Thus, our system allows both the analysis (upper side) and generation (lower side) processes. The upper side direction (lookup) performs an analysis and a description of the entries (lexical units) presented as input that could be represented as a list of tokenized strings Table 2. If the operation of the analysis is successful, the transducer returns the morphology of the lexical unit (verb, noun, pronoun, preposition or adverb). The generation process. Furthermore the lower side describes the output generated based on the rules applied on each lexical unit, some examples illustrated on Table 3.

Table 3. Input/output of lower side

| Input | Output |
|---|---|
| ΛΛₒI | ΛΛₒɪ+amazigh_verb+indic_mood+inflection3+past+3$^{rd}$pers +masc_plr |
| ΛΛₒ† | ΛΛₒ†+amazigh_verb+indic_mood+inflection3+past+3$^{rd}$pers +fem_plr |
| ƐΛΛₒ | amazigh_verb+indic_mood+inflection2+past+3$^{rd}$pers +masc_plr |
| ₒꓫꓔₒO | ₒꓫꓔₒO +Amazigh_noun+variation1+masc+sing |
| †ꓫꓔₒO† | †ꓫꓔₒO†+Amazigh_noun+variation2+fem+sing |
| ₒꓝⵀₒꓢₒEƐ!§ | ₒꓝⵀₒꓢₒE +prounon_affix_noun+sgl+1pers+fem+class2 |
| ₒIIₒ𝔰ꓫⵁ† | ₒIIₒ𝔰ꓫⵁ+prounon_affix_verb+masc+direct_form+3pers+sgl+class1 |

### 5.4. Experiment and evaluation

Our transducer has been tested in both directions (upper side and lower side). Each level is tested individually. Therefore, each test result is equivalent to a level of improvement and refinement of our analyzer. Using Xerox Finite State Technology (XFST) tools, we proceeded to a collective of 8000 words before compilation) and 226 000 after compilation. After processing of 8000 words, 7434 were recognized by our analyzer with a success rate of 93% and the analysis of the total output shows that 92% of the words were recognized as shown in Table 4. Three reasons justify the unrecognized words:

− The dialect aspect of the Amazigh language (Tachlhit, Tarifit, and Tamazighte), makes some words outside the perimeters of the predefined rules.
− Some proper nouns, which are not supported by our system.
− The Amazigh language has borrowed many words from other cultures or languages takerrust [takurrost] "carrosse", takuzint, [takuzint] cuisine, astidyu [studio] "studio".

The majority of unrecognized forms is due to the standardization process, which is not completed by the linguists and does not currently cover all the categories of words. In order to perform further system evaluation, the list of the unrecognized words was, therefore, sent for normalization. Once done, the discarded list will be reinjected into our system for a new analysis and update of the results.

Table 4. Reporting of input words before and after compilation with % of success

| Root type | Total of lexical entry | Nbr of Affixed Pronoun | Nbr of Affixed Verb/noun | Nbr of inflections | Inflected Verb /noun | Total output |
|---|---|---|---|---|---|---|
| Noun | 4000 | 20 | 80000 | 6 | 24 000 | 108 000 |
| Verb | 3000 | 18 | 54 000 | 19 | 57000 | 117 000 |
| Adverb and prepositions | 1000 | | | | | 1000 |
| Total 8000 | | | | | | 226 000 |
| Total 7434 Recognized forms | | | | | | 209000 |
| Total 566 Unrecognized forms | | | | | | 17 000 |
| % of success 93% | | | | | | 92% |

## 6. CONCLUSIONS AND FUTURE WORK

The main mission of this work is to set up a first general intelligent system capable of processing and recognizing all Amazigh lexical units. 8000 words representing the different categories, were collected and tested in both directions (analysis and generation). The Amazigh Morpho-Sys is capable of achieving a motivating degree of success for analysis and generation. The 93% rate can be improved by recycling unrecognized words with more suitable rules. This improvement in percentage will allow our system to be used in the various applications related to language.

## REFERENCES

[1] R. Achab, "La néologie lexicale berbère," *Peeters, Paris/Louvain*, vol. 362, pp. 67, 1996. [Online]. Available: http://www.theses.fr/1994INAL0008

[2] M. Ameur, "Emprunt et créativité lexicale en berbère : traitement en situation d'aménagement linguistique," Thèse de Doctorat d'Etat, Université Sidi Mohamed Ben Abdellah, Fès, 2007. [Online]. Available: http://www.ircam.ma/?q=fr/node/18653

[3] A. Boumalk, "Conditions de réussite d'un aménagement efficient de l'amazighe," *Asinag*, Rabat, 2009, vol. 3, pp. 53-61. [Online]. Available: http://www.ircam.ma/sites/Default_old/files/doc/revueasing.

[4] P. Andries, "Unicode 5.0 Codage des caractères et internationalisation des logiciels et des documents," Paris, Dunod, 2008. [Online]. Available: https://www.amazon.fr/Unicode-5-0-pratique-Patrick-Andries/dp/2100511408.

[5] L. Zenkouar, "Normes des technologies de l'information pour l'ancrage de l'écriture Amazigh," *Berber Studies and Documents*, vol. 27, pp. 159-172, 2008, doi : 10.3917/edb.027.0159.

[6] M. Ameur, "Aménagement linguistique de l'amazighe : pour une approche polynomique," Rabat: *Asinag*, vol. 3, pp. 75-88, 2009. [Online]. Available: http://www.ircam.ma/sites/default/files/doc/revueasing/

[7] Ameur, Meftaha, A. Bouhjar, and F. Boukhris, "Initiation à la Langue Amazighe," Rabat: *IRCAM*, 2004. [Online]. Available: http://www.ircam.ma/sites/default/files/publications/initiation-langue-amazighe-1.pdf.

[8] IRCAM, "Conception et mise au point des polices tifinagh. Centre des Etudes Informatiques", *système d'Information et Communication,* Available: http://www. ircam.ma/fr/index.php?soc=telec&rd=1, 2003.

[9] M. Amrouch, A. Rachidi, M. El Yassa, and D. Mammass, "Handwritten Amazighe Character Recognition Based on Hidden Markov Models," *International Journal on Graphics, Vision and Image Processing*, vol. 10, no. 5, pp. 11-18, 2010. [Online]. Available: http://article.nadiapub.com/IJAST/vol33/4.pdf

[10] F. Ataa Allah, and S. Boulaknadel, "Amazigh Search Engine: Tifinaghe Character Based Approach," in *Proceeding of International Conference on Information and Knowledge Engineering*, Las Vegas, NV, Jul. 2010, pp. 255-259. [Online]. Available: https://tal.ircam.ma/talam/Articles/IKE10.pdf.

[11] S. Boulaknadel and F. Ataa Allah, "Building a standard Amazigh corpus," In *Proceedings of the Third International Conference on Intelligent Human Computer Interaction (IHCI 2011)*, Prague, Czech Republic, August, 2011, pp. 91-98, doi: /10.1007/978-3-642-31603-6_8.

[12] F. Ataa Allah, and S. Boulaknadel, "Amazigh Verb Conjugator," In *Proceedings of the 9th edition of the Language Resources and Evaluation Conference*, Reykjavik, Iceland, 2014. [Online]. Available: https://tal.ircam.ma/talam/Articles/LREC14.pdf.

[13] N. F. Zahra, S. Boulaknadel, and D. Aboutajdine, "Analyse Automatique de la Morphologie Nominale Amazighe," in *Proceedings of TALN 2013, Les Sables d'Olonne*, France, Taln, 2013, vol. 1, pp. 5-18. [Online]. Available: https://www.aclweb.org/anthology/F13-1001.pdf.

[14] F. Ataa Allah, and S. Boulaknadel, "Toward computational processing of less resourced languages: Primarily experiments for Moroccan Amazigh language," in *Text Mining. Rijeka: InTech*, pp. 197-218, 2012, doi: 105772/51069.

[15] R. Ammari, L. Zenkour, and M. Outahajala, "System of Analysis and Generation for Amazigh Nominal Morphology based on Xerox's finite-state transducer," *Procedia Computer Science*, vol. 117, pp. 217-224, 2017, doi: 10.1016/j.procs.2017.10.112.

[16] R. Ammari, L. Zenkouar and M. Outahajala, "Verbal morphological analyzer/generator for Amazigh moods based on XFST," *2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech)*, Rabat, 2017, pp. 1-8, doi: 10.1109/CloudTech.2017.8284732.

[17] F. Z. Nejme, and S. Boulaknadel, "AmAMorph: Finite State Morphological Analyzer for Amazighe," *Journal of computing and information technology*, vol. 24, no. 1, 91-110, 2016, doi: 10.20532/cit.2016.1002478.

[18] K. Koskenniemi, "Two-level morphology: A general computational model for word-form recognition and production," The Department of General Linguistics, University of Helsinki, vol. 11, p. 12, 1983.

[19] S. Goldwater, T. Griffiths and L. Johnson, "Contextual Dependencies in Unsupervised Word Segmentation," In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, 2006, pp. 673-680, doi:10.3115/1220175.1220260

[20] R. Beesley and L. Karttunenauri," Finite-state morphology: Xerox tools and techniques," *CSLI*, Stanford, 2003, [Online]. Available: http://users.itk.ppke.hu/~sikbo/nytech/gyak/05_morfo/xfst/

[21] M. Hulden and S. T. Bischoff, "An experiment in computational parsing of the Navajo verb," *Coyote Papers, Special issue on Navajo Language Studies*, vol. 16, pp. 101-118, 2010, doi: http://hdl.handle.net/10150/126389.

[22] M. A. Attia, "An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modelling Finite-State Networks," In *Challenges of Arabic for NLP/MT Conference*, The British Computer Society, London, vol. 200610, no. 1.72, p. 1482, Oct. 2006.

[23] J. Greenberg, "The Languages of Africa," *International journal of American linguistics*, vol. 29, vol. 1, 1966, doi:10.1086/464395

[24] M. Ameur *et al.*, "Graphie et orthographe de l'Amazighe," *IRCAM, Rabat, Maroco,* 2006, [Online]. Available: https://www.sudoc.fr/135364140

[25] A. Boukous, "Aménagement de l'amazighe : pour une planification stratégique," Asinag 3, Rabat, 2009, pp. 13-40, [Online]. Available: http://www.ircam.ma/sites/default/files/doc/revueasing

[26] F. Boukhris, A. Baoumalk, El H. El Moujahid and H. Souifi, "La nouvelle grammaire de l'Amazighe," *Maroc: IRCAM*, Rabat, 2008, vol. 2, [Online]. Available: http://www.ircam.ma/?q=fr/node/23419

[27] B. Megyesi and S. Rydin, "Towards a Finite-State Parser for Swedish," In *Proceedings of the 12th Nordic Conference of Computational Linguistics (NODALIDA 1999)*, Throndheim, Norway, Dec. 1999, pp. 115-123, [Online]. Available: https://www.aclweb.org/anthology/W99-1012

[28] F. Pisceldo, R. Mahendra, Manurung, R., and I. W. A. Arka, "Two-Level Morphological Analyser for the Indonesian Language," In *Proceedings of the Australasian Language Technology Association Workshop*, Hobart, Australia, December 2008, pp. 142-150, [Online]. Available: https://www.aclweb.org/anthology/U08-1018.pdf

[29] R. Laabdelaoui, A. Boumalk, E. M. Iazzi, H. Souifi and K. Ansar, "Manuel de Conjugaison de l'Amazighe," *IRCAM*, Rabat, Morocco, 2012, vol. 5, [Online]. Available: http://www.ircam.ma/

[30] M. Taifi, "Dictionnaire tamazight-français," *parlers du Maroc central Paris, L'Harmattan-Awal*, 1991, [Online]. Available: http://asegzawal.com/francais/

[31] M. Ameur *et al.*, "Vocabulaire de la langue Amazigh (Français-Amazigh)," Lexiques, no. 1, IRCAM, Rabat, Maroco, 2006, [Online]. Available: http://www.ircam.ma/?q=fr/node/27801.

## BIOGRAPHIES OF AUTHORS

**R. Ammari** Ph.D. Candidate at the Laboratory of Smart Communications, Mohammadia School of Engineers, Mohammed V University of Rabat, Morocco. My research project is about the Natural Language Processing using Machine Learning techniques. My interests lie in the fields of Computational linguistics and Natural Language Processing, Machine Learning, morphological analysis of Amazigh Language.

**L. Zenkouar** Received the Dr. Eng. degree from CEM, Université des Sciences et Techniques du Languedoc, Montpellier, France in 1983 and PhD degree from ULg (Liège) in Belgium. After working as a research assistant and an assistant professor in the Mohammadia School of Engineering in Rabat, he has been a professor degree since 1996. His research interest includes signal processing, IT and Telecommunications.