Amazigh Verb Conjugator

Fadoua Ataa Allah, Siham Boulaknadel
CEISIC, IRCAM
Avenue Allal El Fassi, Madinat Al Irfane, Rabat, Morocco
E-mail: {ataaallah, boulaknadel}@ircam.ma

Abstract

With the aim of preserving the Amazigh heritage from being threatened with disappearance, it seems suitable to provide Amazigh with required resources to confront the stakes of access to the domain of New Information and Communication Technologies (ICT). In this context and in the perspective to build linguistic resources and natural language processing tools for this language, we have undertaken to develop an online conjugating tool that generates the inflectional forms of the Amazigh verbs. This tool is based on novel linguistically motivated morphological rules describing the verbal paradigm for all the Moroccan Amazigh varieties. Furthermore, it is based on the notion of morphological tree structure and uses transformational rules which are attached to the leaf nodes. Each rule may have numerous mutually exclusive clauses, where each part of a clause is a regular expression pattern that is matched against the radical pattern.
This tool is an interactive conjugator that provides exhaustive coverage of linguistically accurate conjugation paradigms for over 3584 Armazigh verbs. It has been made simple and easy to use and designed from the ground up to be a highly effective learning aid that stimulates a desire to learn.

Keywords: Amazigh language, Verbal morphology generation, less resourced language

1. Introduction

Amazigh language, also referred to as Berber in western literature, is a branch of the Afro-Asiatic (Hamito-Semitic) languages (Greenberg, 1966; Ouakrim, 1995). It is spoken over the Northern part of Africa which extends from the Red Sea to the Canary Isles and from Niger in the Sahara to the Mediterranean Sea. In Morocco, according to the geographic area, there are three main varieties of Amazigh: Tarifit in the North, Tamazight in the Center, and Tashelhit in the South. It is the mother tongue of approximately half of the population. However for many decades, it was only oral exclusively reserved for family and informal domains (Boukouss, 1995). While by the creation of the Royal Institute of Amazigh Culture (IRCAM) in 2001 and the constitution update of July 2011, the status of Amazigh has progressively changed to institutional then to official status beside Arabic.

These changes have strengthened the possibility of promoting the Amazigh language, and enabled it to get an official spelling (Ameur et al., 2004), proper coding in Unicode Standard (Andries, 2008; Zenkouar, 2008), appropriate standards for keyboard realization (Ait Ouguengay, 2007), and linguistic structures (Ameur et al., 2004; Boukhris et al., 2008). Nevertheless, these processes are not sufficient for a less-resourced language such as Amazigh to join the well-resourced ones.

In this context, many scientific research studies are undertaken at national level to improve the current situation. Primarily, they focus on optical character recognition (Amrouch et al., 2010; Es Saady et al., 2010; Fakir et al., 2009). But those concentrated on natural language processing are limited to resources building (Iazzi & Outahajala, 2008; El Azrak & El Hamdaoui, 2011; Boulaknadel & Ataa Allah, 2011), morphosyntactic annotation (Ataa Allah & Jaa, 2009; Outahajala et al., 2010), and basic tools (Ataa Allah & Boulaknadel, 2010-a; Ataa Allah & Boulaknadel, 2010-b; Ataa Allah & Boulaknadel, 2010-c). However to the best of our knowledge there is not any computational treatment dealing with verbal morphology for the standard Amazigh language. In this aim, this paper proposes a computational approach that applies nonconcatenative treatment to Amazigh verbal morphology generation.

In the remainder of this paper we describe, in Section 2, some Amazigh language characteristics by focusing on the verbal morphology, and we present, in Section 3, the data and the methodology used in our conjugator system in addition to an example of generating. Finally, in Section 4, conclusions are drawn.

2. Amazigh Language Characteristics

The characteristics of the Moroccan Amazigh language was explained in detail in the linguistic books of IRCAM, especially in the one named “La nouvelle grammaire de l’amazighe” (Boukhris et al., 2008). Nevertheless, in this section, we highlight some points.

2.1 Tifinaghe-IRCAM graphical system

Since the ancient time, the Amazigh language has its own writing that was adapted by IRCAM in 2003, to provide an adequate and usable standard alphabetic system, called Tifinaghe-IRCAM. This system contains:

- 27 consonants including: the labials (ⵀ, ⵃ, ⴼ), dentals (ⵜ, ⴷ, ⵟ, ⴹ, ⵍ, ⵏ, ⵇ, ⵅ), the alveolars (ⵣ, ⵙ, ⵥ, ⵚ), the palatals (ⵛ, ⵊ), the velar (ⴽ, ⵅ), the labiovelars (ⴽⵯ, ⵅⵯ), the uvulars (ⵖ, ⵅ, ⵀ), the pharyngeals (ⵄ, ⵃ) and the laryngeal (ⴰ);
- 2 semi-consonants: ⵢ and ⵡ;
- 4 vowels: three full vowels ⴰ, ⵉ, ⵓ and neutral vowel (or schwa) ⴻ which has a rather special status in Amazigh phonology.

Furthermore, the IRCAM has recommended the use of the International symbols for punctuation markers: “ ” (space), “.”, “,”, “;”, “:”, “?”, “!”, “…”; the standard numeral used in Morocco (0, 1, 2, 3, 4, 5, 6, 7, 8, 9); and the horizontal direction from left to right for Tifinaghe writing (Ameur et al., 2004).

2.2 Amazigh Verbal Morphology

Language morphology is a knowledge of the ways in which the language's words can have different surface representations. Hence, the Amazigh morphology is considered rich and complex in terms of its inflections involving infixation, prefixation and suffixation.

The Amazigh morphology covers five main lexical categories, which are noun, verb, adverb, pronoun, and preposition (Boukhris et al., 2008). The focus of this work is practically on verb morphology.

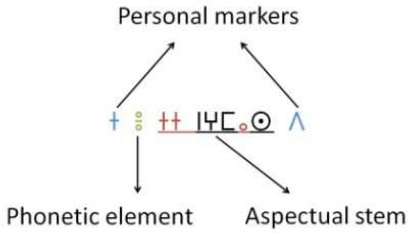
The verb, in Amazigh, has two forms: basic and derived forms. The basic form (radical) is formed through the amalgamation of a root and a pattern. A root is a sequence of one or many consonants and the pattern is a template of vowels (V) and consonants (C) (Root: "ⵏⵏⵢⵛⵙ" *nγms*, Pattern: C'C'CCC[1], Radical: "ⵏⵏⵢⵛⵙ" *nnγms* to inquire). While, the derived one is based on the combination of a basic form and one of the following prefixes morphemes: ⵙ/ⵙⵙ *s/ss* indicating the factitive form, ⵜⵜ *tt* marking the passive form and ⵎ/ⵎⵎ *m/mm* designating the reciprocal form. Whether basic or derived, the verb inflects in three moods (indicative, imperative and participial), where in each mood the same personal markers are used (cf. Table 1); and four aspects (aorist, perfective, negative perfective and imperfective) that are marked with vocalic alternations, prefixation or consonant gemination/degimination. The indicative and the participial moods are based on the four aspects, while the imperative mood has two forms simple and intensive that are based respectively on the aorist and the imperfective aspects (Boukhris et al., 2008).

Table 2 shows an example of the inflectional forms of the verb "ⵏⵏⵢⵛⵙ" *nnγms* in the indicative and the imperative moods for the 2nd person singular, in addition to their singular forms in the participial mood.

Indicative mood	
Aorist	ⵜⵏⵏⵢⵛⵙⴰ *tnnγmsd*
Perfective	ⵜⵏⵏⵢⵛⵙⴰ *tnnγmsd*
Negative perfective	ⵜⵏⵏⵢⵛⵣⵙⴰ *tnnγmisd*
Imperfective	ⵜⵜⵏⵏⵢⵛⴰⵙⴰ *tettnγmasd*
Imperative mood	
Simple	ⵏⵏⵢⵛⵙ *nnγms*
Intensive	ⵜⵜⵏⵏⵢⵛⴰⵙ *ttnγmas*
Participial mood	
Aorist	ⵣⵏⵏⵢⵛⵙⵉ *innγmsn*
Perfective	ⵣⵏⵏⵢⵛⵙⵉ *innγmsn*
Negative perfective	ⵣⵏⵏⵢⵛⵣⵙⵉ *innγmisn*
Imperfective	ⵣⵜⵜⵏⵏⵢⵛⴰⵙⵉ *ittnγmasn*

Table 2: An example of the inflectional forms of the verb "ⵏⵏⵢⵛⵙ" *nnγms*

Whereas, the following template describes the verbal form for the 2nd person singular of the imperfective aspect of the indicative mood "ⵜⵜⵏⵏⵢⵛⴰⵙⴰ" *tettnγmasd* (you are inquiring):



	Indicative mood			Imperative mood			Participial mood
		Masculine	Feminine		Masculine	Feminine	Masculine / Feminine
Singular	1st pers. 2nd pers. 3rd pers.	… ⵢ "... γ" ⵜ … ⴰ "t ... d" ⵣ … "i ... "	… ⵢ "... γ" ⵜ … ⴰ "t ... d" ⵜ … " t ..."	2nd pers.	… Ø	… Ø	ⵣ … ⵏ "i ... n"
Plural	1st pers. 2nd pers. 3rd pers.	ⵏ … "n ... " ⵜ … ⵎ "t ... m" … ⵏ "... n"	ⵏ … "n ... " ⵜ … ⵎⵜ "t ... mt" … ⵏⵜ "... nt"	2nd pers.	… ⵙⵜ/ⵎ "... at/m"	… ⵎⵜ "... mt"	… ⵏⵣⵏ "... nin"

Table 1: Personal markers for the indicative, imperative and participial moods

1 C' is used when a consonant is reduplicated

Where:

ⵜ – ⴰ: the 2nd singular personal marker (*t* – *d*) in the indicative mood;

ⵓ: the phonetic element schwa *e* that is added;

ⵜⵜⵎⵙⴰⵙ: the aspectual stem of the imperfective aspect (*ttnɣmas*) that is formed in this example by the addition of the morpheme "ⵜⵜ" *tt*, the insertion of the vowel "ⴰ" *a*, and the deletion of the first consonant "ⵏ" *n*;

ⵏⵖⵎⵙ: the radical (*nnɣms*) that represents the lexical entry;

and

ⵖⵎⵙ: the root (*nɣms*).

3. Conjugator System

To the best of our knowledge, online Amazigh conjugators are scarce and the most of the undertaken works related to verbal morphology did not focus on the creation of an end-user tool such as a conjugator. To overcome this scarce, we decided to develop an online Amazigh conjugator system, which has been implemented in a way to deal with the challenge of generating the verbal morphology forms according to all the Moroccan Amazigh regional varieties. This system is based on a list of classified verbs, a rule-based approach and a discrimination trees' network.

In general, Finite-state approach has been successful in the description and computational implementation of a wide variety of natural languages especially in morphology generation (Koskenniemi, 1983; Beesley, 1989; Beesley & Karttunenauri, 2003). However, the finite state transducer requires well defined rules and established irregularities. Since it is not the case for the Amazigh verbal morphology which is still under study, we have opted to apply first a discrimination trees' network that will help in determining the exceptions and validating the morphological rules.

This classification–based approach has also proved its efficiency and convenience in the treatment of morphology phenomena such as generation, especially in defining a fine–grained word–class–specific subclassification (Finkler & Neumann, 1988; Leavitt, 1994; Cavalli-Sforza, 2000). It organizes the transformational rules depending on the values of the features in a feature structure and acts as a discrimination network for retrieving the rules appropriate to a given feature structure.

3.1 Data and methodology

In the context of Amazigh standardization process, 3584 verbs have been collected from various written sources such as textbooks, specialized vocabularies, and dictionaries. The choice of these verbs was made on the basis of the representativeness of the three Moroccan varieties (Tarifit, Tamazight, Tashelhit). Actually, the verbs are divided into thirty one conjugation classes, according to the aorist/perfective and the aorist/imperfective oppositions. The last class includes irregular verbs (Laabdelaoui et al., 2012).

Known that the personal markers are depending on moods, we have proceeded by a distinction between morphological operations at the level of personal markers and those at the level of the aspectual stem. The changes occurring at the personal markers' level are determined by mood (indicative, imperative, participial), gender (masculine, feminine), number (singular, plural), and person (first, second, third). Whereas, the operations occurring at the stem level are related to intrinsic features of the inflectional class to which the inflected form belongs, and relatively depending on the verb patterns to deal with the verb conjugation in the negative perfective aspect, and some exceptions due to consonant gemination/degemination (i.e. "ⵣⵔⵔⵛⵜⵅ" *immctg* (he moves) -> "ⵣⵜⵜⵎⵛⵜⴰⵅ" *ittmctag* (he is moving)) and phonological alternation (i.e. "ⵣⵉⵒⵓ" *inḍu* (he jumps) -> "ⵣⵉⵒⵒⵓ" *inṭṭu* (he is jumping)).

3.2 Architecture System

The Amazigh online conjugator system is composed, as illustrated in Figure 1, of three modules: the first one is destined to ensure the user-system interaction. It enables the user to type the verb to be conjugated, and to display the conjugation results.

The second module recognizes the verb classes and extracts the information needed for its conjugation. This information concerns the verb pattern, and the linguistic rules that help to generate the aspectual stem for each class, known that a specific verb could belong to many conjugation classes.

The third one is designed to generate all the verb forms in different aspects and moods. It is composed of three conjugation steps.

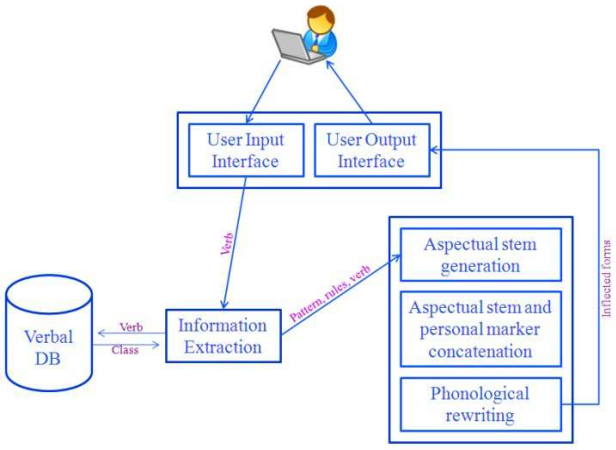


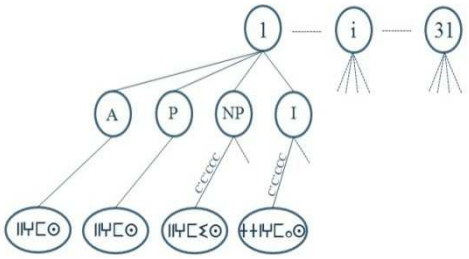
Figure 1: Amazigh conjugator architecture

To implement the first step, that generates the aspectual stem according to the verb class and the irregularity fixed by the pattern, the system uses a discrimination trees' network composed of thirty one trees representing the verb classes. Each tree branches into four sub-trees that correspond to the aspect forms, in which the rule attached to a leaf node effects the desired morphological transformations for that node. The leaf node consists of one

or more mutually exclusive clauses, where each part of a clause is a regular expression pattern that is matched against the radical pattern. The morphological transformation, is nonconcatenative, includes addition (ⵙⴰⵅⵛ *sagm* (to wait) -> ⵙⴰⵅⴰⵛ *sagam* (be waiting)), deletion (ⴱⴱⵇⴱⵖⵇ *bbḍbuṛ* (to boil) -> ⵜⵜⴱⵇⴱⵖⵇ *ttbḍbuṛ* (be boiling)), and replacement (ⵅⵅⴰⵎⵎ *ggall* (to swear) -> ⵅⵅⵖⵎⵎ *ggull* (be swearing)). In the second step, the module conjugates the verb forms by concatenating the aspectual stem with the corresponding inflections depending on mood, gender, person and number. The third concerns applying the phonological rules to deal with the concatenation at the morpheme boundaries.

3.3 An example of generation

Consider the example verb form "ⵏⵏⵢⵛⵙ" *nnɣms* (to inquire), and the extracted information (class, pattern, aspects' rule) given by the second module. In the generation phase, the feature value pairs (rule, aspect) = {(aorist=radical, aorist); (perfective=radical, perfective); (pre-final insertion of the vowel "ⵉ" *i*, negative perfective); (prefixation of "ⵜⵜ" *tt* + pre-final insertion of the vowel "ⴰ" *a*, imperfective)} are applied to the radical in order to produce the aspectual stems. Furthermore, the pattern is also used to fix specific rules according to the verb irregularities. In the case of this example, the pattern is used to apply the degemination to the first radical consonant in the imperfective stem as illustrated in Figure 2.



A: Aorist P: Perfective

NP: Negative Perfective I: Imperfective

Figure 2: First step generation results

Once the aspectual stems are generated, the system proceeds to a concatenation step of the aspectual stems and the personal markers corresponding to each mood according to gender, person and number (cf. Table 3).

Aspect	**2nd person singular**
Aorist	ⵜⵏⵏⵢⵛⵙⵏ *tnnɣmsd*
Perfective	ⵜⵏⵏⵢⵛⵙⵏ *tnnɣmsd*
Negative perfective	ⵜⵏⵏⵢⵛⵉⵙⵏ *tnnɣmisd*
Imperfective	ⵜⵜⵏⵏⵢⵛⴰⵙⵏ *tttnɣmasd*

Table 3: Second step generation results for the second person singular in the indicative mood

The last step of generation module concerns the application of phonological rules to deal with the concatenation at the morpheme boundaries. In this example, the concatenation of the imperfective aspectual stem with the second singular personal marker of the indicative mood presents a spelling mistake at the phonological level. To lead with this problem, the vowel schwa "ⴻ" *e* is added at the morpheme boundaries (cf. Table 4).

Aspect	**2nd person singular**
Aorist	ⵜⵏⵏⵢⵛⵙⵏ *tnnɣmsd*
Perfective	ⵜⵏⵏⵢⵛⵙⵏ *tnnɣmsd*
Negative perfective	ⵜⵏⵏⵢⵛⵉⵙⵏ *tnnɣmisd*
Imperfective	ⵜⴻⵜⵜⵏⵏⵢⵛⴰⵙⵏ *tettnɣmasd*

Table 4: Third step generation results for the second person singular in the indicative mood

4. Conclusions

Despite of the significant efforts made, through the work of researchers at IRCAM and other actors at the national level, to promote the Amazigh language and culture and integrate them into information and communication technology field, the computational processing resources and tools of the Amazigh language are still scarce especially those related to morphology generation.

In this context, we have proposed, in this paper, to present an online computational system that handles Amazigh verbal morphology generation. Our approach was motivated by practical concerns. Thereby, the current implementation has been used to conjugate 3584 verbs. The results are given to linguistic searchers to help in refining the existed morphological rules, and establishing new ones.

Conjugation tool is linguistically motivated model for verb paradigms. It is novel and has great potential for pedagogic applications in teaching the intricacies of the Amazigh verb conjugation system. As the conjugator system is implemented on a database management system, it is simple to add new attested verbs independently of the rest of the tool.

With regard to future work, there are a number of paths which we would like to pursue:

- The first involves updating the system regarding the linguist's recommendation in order for the system to have better coverage.
- The second requires evaluating the system performance.
- The third path of research involves carrying out larger case studies in terms of including other lexical categories.
- While our approach is a perfect fit for web due to the ease of database update's consultation and exploitation, we are confident that the approach can also be applied on a desktop and mobile application where the exploitation of the recent database updates can be enabled throw a particular system that will inform the user.

- Finally, we wish to deploy the finite state model in building an Amazigh generator/analyser, once the morphological rules are well defined and the irregularities are established.

5. Acknowledgements

We would like to express our gratitude to the colleagues of the Grammar's Studies Research Unit at the Royal Institute of Amazigh Culture for their precious collaboration.

6. References

Ait Ouguengay, Y. (2007). Quelques aspects de la numérisation des polices de caractères : Cas de Tifinaghe. *La typographie entre les domaines de l'art et de l'informatique*, pp.159- -181.

Amrouch, M., Rachidi A., El Yassa, M. and Mammass, D. (2010). Handwritten Amazigh Character Recognition Based On Hidden Markov Models. *International Journal on Graphics, Vision and Image Processing*, 10(5), pp. 11- -18.

Andries, P. (2008). *Unicode 5.0 en pratique, Codage des caractères et internationalisation des logiciels et des documents*. Dunod, France, Collection InfoPro.

Ameur, M., Bouhjar, A., Boukhris, F., Boukous, A., Boumalk, A., Elmedlaoui, M., Iazzi , E..M. and Souifi, H. (2004). *Initiation à la langue amazighe*. IRCAM, Rabat, Morocco.

Ataa Allah, F., Boulaknadel, S. (2010). Amazigh Search Engine: Tifinaghe Character Based Approach. *International Conference on Information and Knowledge Engineering*. Las Vegas, Nevada, USA, pp. 255--259.

Ataa Allah, F., Boulaknadel, S. (2010). Pseudo-racinisation de la langue amazighe. *Traitement Automatique des Langues Naturelles*. Montréal, Canada.

Ataa Allah, F., Boulaknadel, S. (2010). Online Amazigh Concordancer. *International Symposium on Image Video Communications and Mobile Networks*. Rabat, Morocco.

Ataa Allah, F., Jaa. H. (2009). Etiquetage morphosyntaxique : Outil d'assistance dédié à la langue amazighe. *1er Symposium international sur le traitement automatique de la culture amazighe*. Agadir, Morocco, pp. 110 --119.

Beesley, K.R, Karttunenauri, L. (2003). Finite State Morphology. CSLI Publications, Stanford, CA.

Beesley, K.R. (1989). Computer analysis of Arabic morphology: A two-level approach with detours, Proceedings of the 3rd Annual Symposium on Arabic Linguistics. Salt Lake City, Utah, USA, March 1989.

Boukhris. F., Boumalk. A., Elmoujahid, E.H., and Souifi, H. (2008). *La nouvelle grammaire de l'amazighe*. IRCAM, Rabat, Morocco.

Boukous, A. (1995). Société, langues et cultures au Morocco : Enjeux symboliques. Najah El Jadida.

Boulaknadel, S., Ataa Allah, F. (2011). Building a standard Amazigh corpus. *International Conference on Intelligent Human Computer Interaction*. Prague, Tchec.

Cavalli-Sforza, V., Soudi, A., Mitamura, T. (2000). Arabic morphology generation using a concatenative strategy. *The 6th Applied Natural Language Processing Conference*. Seattle, Washington, USA, pp. 86--93.

EL Azrak, N., EL Hamdaoui, A. (2011). Référentiel de la Terminologie Amazighe : Outil d'aide à l'aménagement linguistique. *4ème atelier international sur l'amazighe et les TICs*. Rabat, Morocco.

Es Saady, Y., Rachidi, A., El Yassa, M. and Mammas, D. (2010). Printed Amazigh Character Recognition by a Syntactic Approach using Finite Automata. *International Journal on Graphics, Vision and Image Processing*, 10(2), pp. 1--8.

Fakir, M., Bouikhalene, B. and Moro, K. (2009). Skeletonization methods evaluation for the recognition of printed tifinaghe characters. *1er Symposium International sur le Traitement Automatique de la Culture Amazighe*. Agadir, Morocco, pp. 33--47.

Finkler, W., Neumann, G. (1988). MORPHIX - A Fast Realization of a Classification-based Approach to Morphology. 4th Österreichische Artificial Intelligence Tagung, Springer. pp. 11--19.

Greenberg, J.H. (1966). *The Languages of Africa*. The Hague.

Iazzi, E.M, Outahajala, M. (2008). Amazigh Data Base. *HLT & NLP Workshop within the Arabic world: Arabic language and local languages processing status updates and prospects*. Marrakech, Morocco, pp.36--39.

Koskenniemi, K. (1983). Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production. Ph.D. thesis, Helsinki University.

Laabdelaoui, R., Boumalk, A., Iazzi, E.M., Souifi, H. and Ansar, K. (2012). *Manuel de conjugaison de l'amazighe*. IRCAM, Rabat, Morocco.

Ouakrim, O. (1995). Fonética y fonología del Bereber. Survey at the University of Autònoma de Barcelona.

Outahajala, M., Zekouar, L., Rosso, P. and Martí,M.A. (2010). Tagging Amazigh with AnCoraPipe. *Workshop on Language Resources and Human Language Technology for Semitic Languages*. Valletta, Malta, pp. 52--56.

Leavitt, J.R. (1994). MORPHE: A Morphological Rule Compiler. Technical Report, CMU-CMT-94-MEMO.

Zenkouar, L. (2008). Normes des technologies de l'information pour l'ancrage de l'écriture amazighe. *Etudes et documents berbères*, 27, pp.159--172.