Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 73 (2015) 234 – 241

Procedia Computer Science

The International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015)

Amazigh Representation in the UNL Framework: Resource Implementation

Imane Taghbalout[a], Fadoua Ataa Allah[b], Mohamed El Marraki[a]

[a]*LRIT, Faculty of Sciences, Mohammed V University, Rabat, Morocco*
[b]*CEISIC, Royal Institute of Amazigh Culture, Rabat, Morocco*

Abstract

This paper discusses the first steps undertaken to create necessary linguistic resources to incorporate Amazigh language within the Universal Networking Language (UNL) framework for machine translation purpose. This universal interlanguage allows to any source text to be translated into different other related languages with UNL by converting the meaning of the source text into semantic graph. This encoding is considered as a pivot interlanguage used in translation systems. Thus in this work, we focus on presenting morphological, syntactical and lexical mapping stages needed for building an "Amazigh dictionary" according to the UNL framework and the "UNL-Amazigh Dictionary" that are both taking part in enconversion and deconversion processes.

Peer-review under responsibility of organizing committee of the International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015)

Keywords: Machine Translation; Universal Networking Language; Amazigh language; UNL-Amazigh dictionary; Inflectional morphology

1. Introduction

Amazigh language, more frequently called Berber, belongs to the northern Afro-Asiatic languages' family[1] . It is the autochthon language of North Africa[2]. However, according to the census of 1994, it was spoken, in Morocco, only by about 50% of the population. While, according to the census of 2004, it is spoken by about 28.3% of the population[3].These rates indicate clearly that the Amazigh language was marginalized for many years. It was only used as an oral language until the creation of the Royal Institute of Amazigh Culture (IRCAM) in 2001. Therefore, in the aim to ensure its survival and its position in the information society, the computerization of Amazigh is highly necessary especially after it became one of the official languages of Morocco along with Arabic. To this end, a

Peer-review under responsibility of organizing committee of the International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015)
doi:10.1016/j.procs.2015.12.023

number of strategies have been proposed, including the production of a machine translation tool[4,5]. Certainly, the statistical approach is promising in the field of machine translation, but its strength depends largely on the amount of resources required in term of corpus[6,7]. Since the Amazigh language is still one of less resourced languages, it is very difficult to find an Amazigh corpus whose size could be larger than thousands of sentences. For this reason, we have opted, in a first stage, for the linguistic approach based on UNL[8,9].

The translation of any source language to any target language based on UNL interlanguage is the process of "enconverting" the source sentence to the UNL representation and then "deconverting" the target sentence from the UNL representation. The advantage of UNL-based machine translation is the ability to apply it in a multilingual environment. This approach is, actually, the best one for the Amazigh language especially that UNL offers a universal language-independent and open-source platform for multilingual applications[10,11], including Arabic, Bulgarian, Chinese, English, French, Hindi, Russian, etc.[12,13]. Thus, the realization of a single Amazigh-UNL enconverter and deconverter will allow getting the translation of the Amazigh texts into all the languages integrated in the UNL project and vice versa.

The remaining of this paper is organized as follows: In Section II, we present some morphological characteristics of the Amazigh language. In Section III, we outline the Universal Networking Language Project. In section IV, we describe the process of building required resources for integrating Amazigh language in the UNL framework. Finally in Section V, we draw the conclusions, and present potential future research directions.

2. Amazigh language morphology

Amazigh is morphologically rich language with complex inflection system. Words are classified into ten categories[14]: noun, adjective, verb, adverb, preposition, pronoun, particle, conjunction, interjection and numeral. However, in this work, we focus on noun and verb categories.

2.1. Noun morphology

In Amazigh, the noun is a grammatical category that has claimed one of the following three forms[15]:

- The simple noun is a single word (Example: 'ⴰⴷⵍⵉⵙ' [Adlis] *book*);
- The compound noun is composed at least of two elements forming a single word having its own meaning (Example: 'ⴱⵓⵜⴳⵔⴰ' [butgra] *turtle*);
- The derived noun is a noun formed by the processes of affixation of a morpheme to the base form of a simple noun (Example 'ⴰⴼⵔⴼⵔ'[afrfr] 'flight' derived from the verb 'ⴼⵔⴼⵔ'[frfr] *fly*).

The second characteristic of the Amazigh nouns is their variation in gender (female/male), number (singular/plural) and state (free/construct).

2.1.1. Gender

Most male nouns (MCL) begin with 'ⴰ' [a], 'ⵉ' [i], or 'ⵓ' [u] unless some exceptions, namely the case of kinship nouns like 'ⵉⵎⵎⴰ' [imma] *my mother*. Concerning the formulation of female nouns (FEM), we transform a singular masculine noun to feminine by affixing the morpheme 'ⵜ' [t] at the beginning and the end of the masculine noun. For example, the word 'ⵜⴰⴳⵎⴰⵔⵜ' [tagmart] *mare* is the feminine noun for the masculine noun 'ⴰⴳⵎⴰⵔ' [agmar] *horse*. Some female nouns take only the initial morpheme 'ⵜ' [t] or the final 'ⵜ' [t].

2.1.2. Number

The noun has a singular (SNG) and a plural (PLR) form, whether it is in masculine or feminine. There are four types of plural forms: the external plural, broken plural, mixed plural and plural in 'ⵉⴷ' [id] [15]:

- The external plural: is formed by an alternation of the first vowel '⵿/ⴻ' [a/i] accompanied by a suffixation of 'ⵏ' [n] or one of its variants. Example: '⵿ⴰⵎⴻⵙ' [adlis] *book* –> 'ⴻⴰⵎⴻⵙⵏ' [idlisn] *books*.
- The broken plural: involves a change in the vowels of the noun. Example: '⵿ⵣⵎⵣ' [azmz] *an epoch* –> 'ⴻⵣⵎⴰⵣ' [izmaz] *epochs*.
- The mixed plural: is formed by vowels change accompanied sometimes by the use of the suffixation of 'ⵍ' [n]. Example: 'ⴻⵉⵍⵉ' [ili] *possession* –> 'ⴻⵉⵍⴰⵏ' [ilan] *possessions*.
- The plural in 'ⴻⴰ' [id]: the noun is preceded by the morpheme 'ⴻⴰ' [id]. It is applied to a set of nouns including: nouns with an initial consonant, proper nouns, kinship nouns, compound nouns, numerals, as well as borrowed nouns. Example: 'ⵎⵕⴰⵡ' [mraw] *ten*→ 'ⴻⴰ ⵎⵕⴰⵡ' [id mraw] *the tens*.

2.1.3. State

The noun can be in two different states: Free State or annexed State (also called construct state). We talk about a noun in the Free State when it is isolated from any grammatical context. The construct state (CTS) is marked by a change affecting its initial vowel in the following grammatical contexts: the noun has a function of the subject postponed to verb; after a preposition except 'ⴰⵏ/ⴰⵔ' [al/ar] and 'ⵖⵍⴰ' [bla]; after a numeral; after 'ⴻⴰ' [id]; after the morpheme of membership and affiliation: 'ⵓ' [u], 'ⵓⵍⵜ' [ult], 'ⴰⵢⵜ' [ayt], 'ⴻⵔⵜ' [ist].

2.2. Verb morphology

The verb, in Amazigh, can be either simple (radical) or derived[16]. The simple verb is composed of a root and a pattern. The root is a sequence of one or many consonants, while the pattern is a template of vowels (V) and consonants (C)[15,16]. Whereas, the derived form is based on the concatenation of a simple verb and one of the following prefixed morphemes: 'ⵔ' [s] / 'ⵔⵔ' [ss], 'ⵜⵜ' [tt], or 'ⵎ' [m]/ 'ⵎⵎ' [mm]. The prefixation of 'ⵔ' [s]/'ⵔⵔ' [ss] gives factitive (causative) form and the morpheme 'ⵜⵜ' [tt] gives the passive form, while the third prefix gives the reciprocal form. Whether the verb is simple or derived, it inflects in three moods: indicative (IND), imperative (IMP) and participial (PTP). In each mood the same person markers are used in four aspects: aorist (AOR), perfective (PFV), negative perfective (NEG&PFV) and imperfective (NPFV) that are marked with vocalic alternations, prefixation or consonant gemination/degimination[16].

3. The Universal Networking Language

Universal Networking Language (UNL) is a computing language developed as an essential element of the UNL Project, which is an international project promoted by the Institute of Advanced Studies of the United Nations (UNU/IAS). In the UNL approach, information conveyed by natural language is expressed as a hyper graph consisting of nodes or hyper nodes inter-related by semantic relations[17]. A node is called UW 'Universal Word'. UWs are often accompanied with a set of grammatical properties called attributes. Arcs binding UNL nodes in the graph represent a relationship between the UWs. The definitions of these basic UNL elements are:

Universal Words (UWs): constitute the vocabulary of the UNL language. They are English words, accompanied with a set of linguistic and semantic restrictions. UW is the basic element to build UNL expression of a sentence.

Universal Attributes: represent the grammatical properties that can enrich the description of the universal word. For example, the UW that corresponds to the English word 'play' is 'play (icl>do)'. (icl>do) is added to describe more the UW. It means that 'play' is a verb. If the word 'play' is conjugated to the past, the attribute '@past' must be added to the UW 'play (icl>do)'. Thus, we obtain the following UW: 'play (icl>do, @past)'.

Universal Relation: is a syntactic-semantic binary relation that connects a pair of nodes in the UNL graph. The UNL system defines a set of labels for universal relations following to their roles. For instance, the relation "*agt*" (agent) defines the thing or the person who initiates an action.

The UNL framework requires a set of linguistic resources[18], including grammatical rules to produce a correct sentence in the target language, UNL Knowledge Base (UNLKB) that defines every possible relation between universal words, and three dictionaries:

- UNL dictionary which lists UWs in alphabetical order with all their corresponding linguistic and semantic properties.
- NL (Nature language) dictionary that lists entries in natural language with their linguistic properties.
- NL-UNL dictionary, called Analysis Dictionary, which is a bilingual dictionary that links UNL words (UWs) to the corresponding lexical items in the natural language. This dictionary has a generative format.
- UNL-NL dictionary, called Generation Dictionary, which is a bilingual dictionary that links UNL words (UWs) to the corresponding lexical items in the natural language. This dictionary has enumerative format.

4. Amazigh dictionaries' implementation in the UNL framework

The Amazigh dictionary and the Amazigh-UNL dictionary are designed to support morphological, syntactic analysis needed for Amazigh enconversion. To build these dictionaries, we should implement grammar rules that include both morphological and syntactic information namely inflection paradigms and subcategorization frames.

4.1. Amazigh inflection paradigms

In the UNL framework, a paradigm is a list of inflectional rules applied over the base form for generating the different inflected words to express different grammatical categories such as tense, mood, voice, aspect, person, number, gender and case[19]. For the time being, we have achieved to create eighty inflectional paradigms for the nominal category[20]. Each paradigm contains rule generation of feminine, plural and annexed forms unlike other noun classification works based only on gender and number[21,22].

Amazigh is rich morphologically and complex in terms of its inflections involving infixation, prefixation and suffixation. Hence, the implementation of the inflectional rules is very time-consuming, especially that the Amazigh language suffers from scarcity of works dealing with nominal category processing.

The inflectional rules, below (Fig. 1), represent the inflectional paradigm created for the fourth class (model noun: 'ⵓⵙⵍⵎⴰⴷ' [aslmad] *student*) of nominal category.

```
MCL&SNG&NOM: =0>"";
MCL&SNG&CTS: = "ⵓ"<1;
MCL&PLR&NOM: = "ⵉ"<1, 0>"ⵏ";
MCL&PLR&CTS: ="ⵉ"<1, 0>"ⵏ";
FEM&SNG&NOM: = "ⵜ"<0, 0>"ⵜ";
FEM&SNG&CTS: ="ⵜ"<1, 0>"ⵜ";
FEM&PLR&NOM: ="ⵜⵉ"<1, 0>"ⵉⵏ";
FEM&PLR&CTS: ="ⵜ"<1, 0>"ⵉⵏ";
```

Fig. 1. Inflectional paradigm for nouns of the fourth class.

In order to evaluate the nominal inflectional generation, we have used a corpus of test. This corpus is composed of 500 lemmas extracted from the Amazigh general dictionary[23]. By comparing the test corpus entries to the word forms generated with the implemented inflectional rules, we remarked that 28.7% of the inflected forms are dissimilar to test entries.

This dissimilarity is owing mainly to two factors: the first one is the fact that an Amazigh word could have many inflectional paradigms due to the regional varieties. For example, the noun 'ⵜⴰⴱⵔⵣⴰ' [tabrza] *tournament* has two plural forms 'ⵜⵉⴱⵔⵣⵉⵡⵉⵏ' [tibrziwin] and 'ⵜⵉⴱⵔⵣⴰⵜⵉⵏ' [tibrzatin]. The second factor is related to the misclassification of certain nouns. For instance the noun 'ⴰⴽⴰⵎ' [akal] *soil* was misclassified into a class including nouns of the pattern 'ⴰ.ⴰ.', like the word 'ⴰⴽⴰⵎ' [akam] *sting*. So, the plural form of 'ⴰⴽⴰⵎ' was associated to the

inflectional rule consisting in replacing the first 'ⵄ' by 'ⴻ' and suffixing 'ⵉ' ('ⵄⴽⵄⵎ' [akal] →'ⴻⴽⵄⵎⵉ' [ikaln]). While, it should be associated to another inflectional rule, which means the noun should belongs to another class, that consists in suffixing the morpheme 'ⴻⵍⵉ' [iwn] and replacing the first 'ⵄ' by 'ⴻ' ('ⵄⴽⵄⵎ' [akal] → 'ⴻⴽⵄⵎⴻⵍⵉ' [ikaliwn]). To remedy this problem, we are improving the nominal classification process.

Concerning Amazigh verbs, we are undertaking the implementation of their inflectional rules on the basis of the classification proposed by Laabdelaoui *et al.* [16] and Ataa Allah & Boulaknadel[24]. Thus, we are creating some inflectional paradigms. For instance: the verb 'ⵖⵄⵔⴽ' [bark] *to bless* follows the first paradigm that contains the inflectional rules listed in Fig.2 below:

```
1PS&PFV&AFM&IND:=0>"ⵢ";
2PS&PFV&AFM&IND:="ⵜ"<0,0>"ⴰ";
3PS&MCL&PFV&AFM&IND:="ⴻ"<0;
3PS&FEM&PFV&AFM&IND:="ⵜ"<0;
1PP&PFV&AFM&IND:="ⵏ"<0;
2PP&MCL&PFV&AFM&IND:="ⵜ"<0,0>"ⵛ";
2PP&FEM&PFV&AFM&IND:="ⵜ"<0,0>"ⵛⵜ";
3PP&MCL&PFV&AFM&IND:=0>"ⵏ";
3PP&FEM&PFV&AFM&IND:=0>"ⵏⵜ";
1PS&PFV&NEG&IND:= "ⴻ"<[-1], 0>"ⵢ";
2PS&PFV&NEG&IND:= "ⴻ"<[-1],"ⵜ"<0,0>"ⴰ";
3PS&MCL&PFV&NEG&IND:="ⴻ"<[-1],"ⴻ"<0;
3PS&FEM&PFV&NEG&IND:="ⴻ"<[-1],"ⵜ"<0;
1PP&PFV&NEG&IND:="ⴻ"<[-1],"ⵏ"<0;
2PP&MCL&PFV&NEG&IND:="ⴻ"<[-1],"ⵜ"<0,0>"ⵛ";
2PP&FEM&PFV&NEG&IND:="ⴻ"<[-1],"ⵜ"<0,0>"ⵛⵜ";
3PP&MCL&PFV&NEG&IND:="ⴻ"<[-1],0>"ⵏ";
3PP&FEM&PFV&NEG&IND:="ⴻ"<[-1],0>"ⵏⵜ";
1PS&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],0>"ⵢ","ⵜⵜⵜ":"ⵜⵙⵜⵜ";
2PS&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],"ⵜ"<0,0>"ⴰ","ⵜⵜⵜ":"ⵜⵙⵜⵜ";
3PS&MCL&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],"ⴻ"<0,"ⵜⵜⵜ":"ⵜⵙⵜⵜ";
3PS&FEM&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],"ⵜ"<0,"ⵜⵜⵜ":"ⵜⵙⵜⵜ";
1PP&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],"ⵏ"<0,"ⵜⵜⵜ":"ⵜⵙⵜⵜ";
2PP&MCL&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],"ⵜ"<0,0>"ⵛ","ⵜⵜⵜ":"ⵜⵙⵜⵜ";
2PP&FEM&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],"ⵜ"<0,0>"ⵛⵜ","ⵜⵜⵜ":"ⵜⵙⵜⵜ";
3PP&MCL&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],0>"ⵏ","ⵜⵜⵜ":"ⵜⵙⵜⵜ";
3PP&FEM&NPFV&IND:="ⵜⵜ"<0,"ⵄ"<[-1],0>"ⵏⵜ","ⵜⵜⵜ":"ⵜⵙⵜⵜ";
1PS&FUT&IND:=0>"ⵢ", "ⵄⴰ "<<0;
2PS&FUT&IND:="ⵜ"<0,0>"ⴰ", "ⵄⴰ "<<0;
3PS&MCL&FUT&IND:="ⴻ"<0,"ⵄⴰ "<<0;
3PS&FEM&FUT&IND:="ⵜ"<0, "ⵄⴰ "<<0;
1PP&FUT&IND:="ⵏ"<0,"ⵄⴰ "<<0;
2PP&MCL&FUT&IND:= "ⵜ"<0,0>"ⵛ", "ⵄⴰ "<<0;
2PP&FEM&FUT&IND:="ⵜ"<0,0>"ⵛⵜ", "ⵄⴰ "<<0;
3PP&MCL&FUT&IND:=0>"ⵏ","ⵄⴰ "<<0;
3PP&FEM&FUT&IND:=0>"ⵏⵜ","ⵄⴰ "<<0;
SNG&AOR&PTP:="ⴻ"<0,0>"ⵏ";
PLR&AOR&PTP:=0>"ⵏⴻⵏ";
SNG&NPFV&PTP:="ⵜⵜ"<0,"ⵄ"<[-1],"ⴻ"<0,0>"ⵏ";
PLR&NPFV&PTP:="ⵜⵜ"<0,"ⵄ"<[-1],0>"ⵏⴻⵏ";
2PS&AOR&IMP:=""<0;
1PP&MCL&AOR&IMP:= 0>"ⵄⵜ",0>" ⵄⵢ";
1PP&FEM&AOR&IMP:= 0>"ⵄⵛⵜ",0>" ⵄⵏⵢ";
2PP&MCL&AOR&IMP:= 0>"ⵄⵜ";
2PP&FEM&AOR&IMP:= 0>"ⵛⵜ";
2PS&NPFV&IMP:="ⵜⵜ"<0,"ⵄ"<[-1];
2PP&MCL&NPFV&IMP:= "ⵜⵜ"<0,"ⵄ"<[-1],0>"ⵄⵜ";
2PP&FEM&NPFV&IMP:= "ⵜⵜ"<0,"ⵄ"<[-1],0>"ⵛⵜ";
```

Fig. 2. Inflectional rules corresponding to verbs of the first class

Where:

- PFV= Perfective aspect; NPFV= Non perfective aspect; AOR= Aorist aspect;
- IND= Indicative mood; IMP= Imperative mood; PTP= Participial mood;
- NEG= Negative; AFM= Affirmative;
- MCL= Masculine; FEM= Feminine;
- XPG= person markers, where X={1,2,3} and G={masculine, feminine};
- "a"<0;= Prefixation of "a";
- 0>"a";= Suffixation of "a";
- "a"<[-1];= Insertion of "a" before the final letter.

The forty eight rules above concern just one paradigm that of the first class of Amazigh verbs among thirty others[16,24]. The implementation of this paradigm allows generating all forms of verbs belonging to the same class as 'ⴱⴰⵔⴽ' [bark] *to bless*.

4.2. Amazigh subcategorization frames

Subcategorization is the definition of the number and types of syntactic arguments that co-occur with the base form in order to form a phrase. The idea of a subcategorization is related to the concepts of valency and transitivity. Subcategorization frames are schemes that describe grammatical relations between words by defining the number and the type of specifiers, complements and adjuncts that a base form needs to constitute a phrase[25].

For the Amazigh language, we have implemented, up to now, fifteen subcategorization frames for the nominal and verbal categories. Table 1, below, gives some implemented subcategorization frames with their explanations.

Table 1. An extract of subcategorization frames defined for Amazigh language.

Subcategorization frame	Explanation	Example
VS(NP)VC(NP)VC(PH([ⵉ]))VC(PP);	Tri-transitive verbs need a Noun Phrase NP as subject, a NP as direct complement, a phrase headed (PH) by 'ⵉ' [i] as indirect complement and a Prepositional Phrase (PP) as complement	*'ⵉⵙⵙⴽⴽ ⵎⵉⵎⵓⵏ ⵜⴰⴱⵔⴰⵜ ⵉ ⵎⵎⵉⵙ ⴰⴳ ⵓⵎⴷⴷⵓⴽⵍ ⵏⵏⵙ'* [issk mimoun tabrat i mmis ag umddukl nns] *Mimoun delivered the letter to his friend for his son*
NC(PH([ⵏ]));	Nouns whose complement is introduced by the preposition 'ⵏ' [n].	'ⵉⵎⵉ ⵏ ⵜⴰⴷⴷⴰⵔⵜ' [imi n taddart] *house's entrance*
VS(NP)VC(NP)VC(PH([ⵅⴼ]));	The verb requires a noun phrase as a specifier (VS), a noun phrase as a complement (VC), and a prepositional phrase headed by "ⵅⴼ" as a complement (VC)	'ⵉⴱⴷⴰ ⵓⵙⵍⵎⴰⴷ ⵉⴷⵍⵉⵙⵏ ⵅⴼ ⵉⵏⵍⵎⴰⴷⵏ ⴳ ⵜⵉⵏⵎⵍ' [ibda uslmad idlisn khf inlmadn g tinml] *The teacher distributed books to students at the school*

4.3. Template of Amazigh dictionaries in the UNL framework

The templates of the Amazigh dictionary and the Amazigh-UNL dictionary have, respectively, the following forms[26]: *[NLW] {ID} (ATTR ...) < FLG, FRE, PRI >;* and *[NLW] {ID} "UW" (ATTR ...) < FLG, FRE, PRI >;*

where:

- ***NLW***: Head Word (Amazigh word).
- ***ID***: The unique identifier (primary-key) of the entry.
- ***UW:*** Universal Word.
- ***ATTR***: The list of grammatical, morphological and semantic features (Part of speech 'POS', Number 'NUM' …).
- ***FLG:*** ber, the Amazigh three-character language code according to ISO 639-3.
- ***FRE:*** The frequency of NLW in natural texts is used for natural language analysis. It can range from 0 (less frequent = default value) to 255 (most frequent).
- ***PRI:*** The priority of the NLW. It is used for natural language generation. It can range from 0 to 255.

The number of attributes (ATTR) varies according to the category of the entry. For instance, verbal entries must include the following attributes: POS (Part of speech), LST (Lexical structure), and TRA (Transitivity).While, the nominal entries must include POS, LST, and NUM (Number). Fig.3 presents screenshots of the Amazigh dictionary according to the grammatical category of the Head Word.

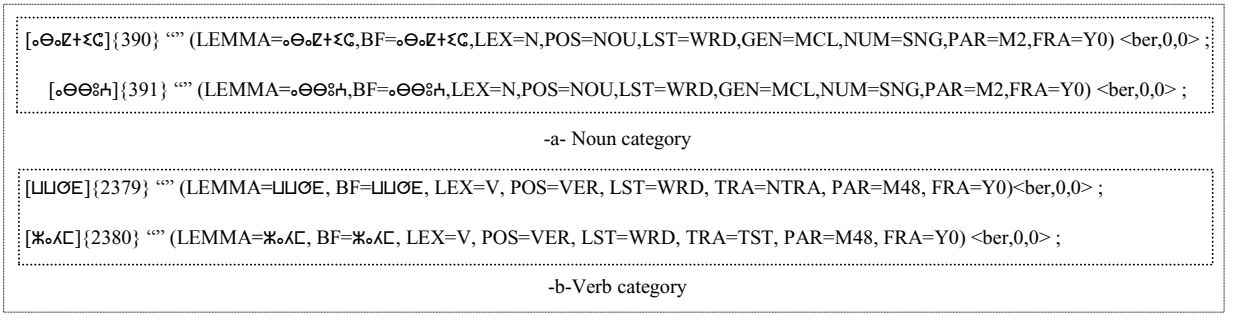

Fig. 3. Screenshots of the Amazigh dictionary according to the entry's grammatical category.

While, for building the Amazigh-UNL dictionary, we have performed the lexical mappings between Amazigh lexical items and UNL words as described in the screenshot of the Amazigh-UNL dictionary (Fig.4). Each entry is associated to the identifier of the Universal Word (UW) that is the concept carrying the meaning of the Amazigh entry. The identifier of the UW in Fig.4 is the one that is framed. For example the word "answer" is expressed by two concepts "answer (icl>state)" which is mapped with the corresponding Amazigh verb "ⵔⴰⵔ" [rar] and the concept "answer (icl>statement)" which is mapped with the corresponding Amazigh noun"ⵜⴰⵎⵔⴰⵔⵓⵜ" [tamrarut]. For the time being, we have achieved to enter 2335 mapped entries.

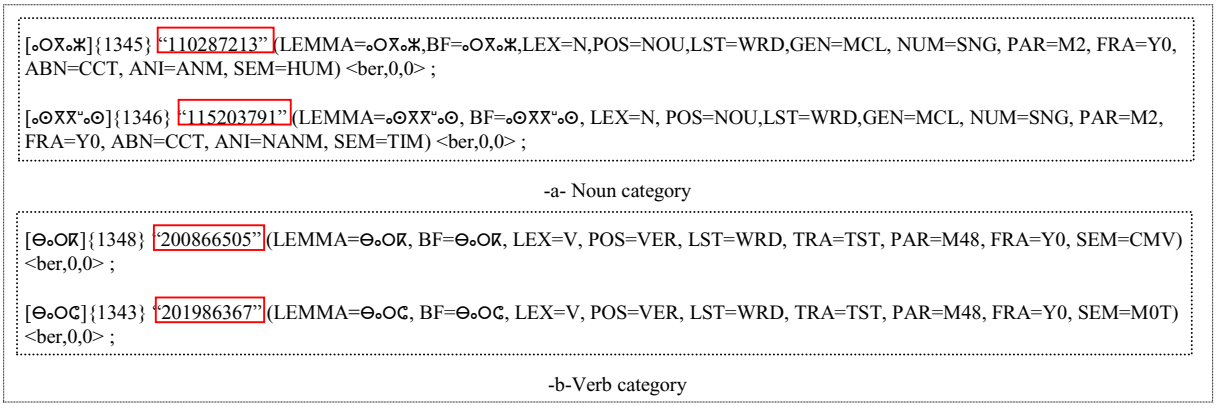

Fig. 4. Screenshots of the Amazigh-UNL dictionary according to the entry's grammatical category.

5. Conclusion and Future Perspectives

In the perspective to realize an UNL-based machine translation system for Amazigh language, we have successfully created inflectional paradigms for the nominal category and subcategorization frames. We are also undertaking the implementation of the verbal inflectional rules and the lexical mapping between Amazigh and UNL words. Our future plan will carry on the creation of the verbal inflectional rules, and the lexical mapping stage. Then, developing transformation rules used to convert the UNL graphs into Amazigh text, according to the UNL Grammar Specifications, and disambiguation rules that improve the results of the tokenization, and prevent wrong lexical choices.

References

1. Kirsty R. Meroitic – an Afroasiatic language?. SOAS Working Papers in Linguistics, vol. 14,pp.169-206; 2006
2. Francis M. Berbères, berbérité : noms, territoires, identités. Cahiers de sociolinguistique, pp.175- 214;2006
3. Boukous A. Dynamique d'une situation linguistique: le marché linguistique au Maroc. Morocco ; 2005
4. Ataa Allah F, Boulaknadel S. Toward Computational Processing of less Resourced Languages: Primarily Experiments for Moroccan Amazigh Language. Theory and Applications for Advanced Text Mining, Prof. Shigeaki Sakurai (Ed.), pp. 197-218; 2012
5. Ataa Allah F, Boulaknadel S. La promotion de l'amazighe à la lumière des technologies de l'information et de communication. Asinag n°9, pp. 33-48, ISSN : 2028-5663; 2014
6. Sonja N, Hermann N. Statistical Machine Translation with Scarce Resources Using Morpho-syntactic Information. Association for Computational Linguistics, vol. 30, No. 2, pp.181-204 ;2004
7. Maučec M.S, Kačič Z, Verdonik, D. Statistical machine translation of subtitles for highly inflected language pair. Pattern Recognition Letters, Vol. 46. Elsevier Science Inc. New York, NY, USA, pp. 96-103; 2014
8. Uchida, H, Zhu M. The Universal Networking Language beyond Machine Translation. International Symposium on Language in Cyberspace, Seoul of Korea; 2001
9. Noha, A., Sameh A. Evaluation of Arabic Machine Translation System Based on the Universal Networking Language. Natural Language Processing and Information Systems ,Lecture Notes in Computer Science vol. 5723, pp. 243-257;2010
10. Boguslavsky I, Cardeñosa J, Gallardo C, Iraola L. The UNL initiative. An overview. Computational Linguistics and Intelligent Text Processing, 6th International Conference CICLing, Mexico; 2005
11. Boitet C, Boguslavskij I.M, Cardeñosa J. An Evaluation of UNL Usability for High Quality Multilingualization and Projections for a Future UNL++ Language. In A. Gelbukh (ed.) Proceedings of CICLing, Lecture Notes in Computer Sciences, vol. 4394, Springer-Verlag, pp. 361-373; 2007
12. Stoykova V. Lexical Information for Bulgarian in Universal Networking Language. In: Proceedings of the 8th WSEAS International Conference on Computer Engineering and Applications, Tenerife, Spain, pp. 414-417; 2014
13. Kumar P. UNL Based Machine Translation System for Punjabi Language. Ph.D. thesis, School of Mathematics and Computer Applications, Thapar University, Patiala, India; 2012
14. Ataa Allah F, Boulaknadel S, Souifi H. Jeu d'Etiquettes Morphosyntaxiques de la Langue Amazighe, Asinag, n°9, pp. 171-184, ISSN : 2028-5663; 2014
15. Boukhris F, Boumalk A, El Moujahid E, Souifi H. La nouvelle grammaire de l'amazighe. IRCAM, Rabat, Morocco; 2008
16. Laabdelaoui R, Boumalk A, Iazzi, E.M, Souifi H, Ansar K. Manuel de conjugaison de l'amazighe. IRCAM, Rabat, Morocco; 2012
17. Uchida H, Zhu M. The Universal Networking Language (UNL) Specification Version 3.0.Edition 3, Technical Report, UNU, (2005/6-UNDL Foundation, International Environment House, Tokyo; 2004
18. UNDL. http://www.unlweb.net/wiki/Lex; (Viewed 23/04/2015)
19. UNDL. http://www.unlweb.net/wiki/Inflectional_paradigm ; (Viewed 23/04/2015)
20. Taghbalout I, Ataa Allah F, El Marraki M. Amazigh noun inflection in the universal networking language. International Journal of Education and Information Technology. Vol. 9, pp.122-128; ISSN: 2074-1316; 2015
21. Nejme F, Bouknadel S, Aboutajdine D. Analyse Automatique de la Morphologie Nominale Amazighe. 20th TALN-Récital conference, Les Sables d'Olonne, pp.5-18; 2013
22. Raiss H, Sforza V.C. ANMorph: Amazigh Nouns Morphological Analyzer. 5th International Conference on les Technologies d'Information et de Communication pour l'Amazighe (TICAM 2012), Rabat, Morocco, November 26-27;2012
23. Ameur M, Boumalk A, El Azrak N, Laabdelaoui R. Dictionnaire général de l'amazighe. IRCAM, Rabat, Morocco; 2015
24. Ataa Allah F, Boulaknadel S. Amazigh Verb Conjugator. 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, May 26-31 2014, pp. 1051-1055; 2014
25. UNDL. http://www.unlweb.net/wiki/Subcategorization_frame; (Viewed 23/04/2015)
26. Ronaldo M, Vahan A. Generative and Enumerative Lexicons in the UNL Framework. 7th International Conference on Computer Science and Information Technologies , Yerevan, Armenia , Proceedings of CSIT,28 September - 2 October;2009