Amazigh Noun Inflection in the Universal Networking Language

I. Taghbalout[1], F. Ataa Allah[2], M. El Marraki[1]

[1]LRIT, Faculty of Sciences, Mohammed V University, Rabat, Morocco
[2]CEISIC, Royal Institute of Amazigh Culture, Morocco
taghbalout.imane@gmail.com, ataaallah@ircam.ma, marraki@fsr.ac.ma

***Abstract*—**Amazigh is one of the official languages of Morocco. However, it is still considered as less resourced language. Thus, it is time to develop language resources and applications for processing automatically this language in order to ensure its survival and promotion by integrating it into the new information and communication technologies (NICT). In this context and in the perspective of achieving an automatic translation of the Amazigh language, we have undertaken the necessary work to incorporate this language into the UNL (Universal Networking Language) structure, which is developed inside the UNDL (Universal Networking Digital Language) foundation. The mission of the UNDL is to allow anyone to access to the information available on the Internet using her/his mother tongue, which will contribute in promoting multilingualism and reducing access constraints to information due to language barriers. In this paper, we focus on the Amazigh noun inflection paradigms to build the first Amazigh dictionary for the UNL project.

***Keywords*—**Amazigh language, UNL, Amazigh dictionary, noun morphology, Inflectional paradigms.

I. Introduction

Amazigh language or Tamazight, more frequently known as Berber, belongs to the Afro-Asiatic language (Hamito-Semitic) family [1] [2]. Actually, it is one of the official languages of Morocco along with Arabic. However, it was neglected for many years. To reduce the gap and ensure its survival and its position in the information society, the computerization of Amazigh is necessary. To this end a number of strategies have been proposed, including the production of a machine translation tool. Certainly, the statistical approach is a promising in the field of machine translation, but its strength depends largely on the amount of resources required in term of corpus [3] [4]. But, since, the Amazigh language is still one of the 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, as a first step, for the linguistic approach: translation using UNL interlanguage [5]. This choice is the result of the state of the art that we have undertaken. A part of this study will be presented in the second section of this paper.

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 [6] [7], including more than 16 official languages: Arabic, Bulgarian, Chinese, English, French, Hindi, Russian, etc. [8] [9]. 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 UNL project and vice versa.

The remaining of this paper is organized as follows: In Section II, we outline the state-of-the-art of machine translation approaches. In Section III, we present the morphological characteristics of the Amazigh language in general and the nominal category in particular. In Section IV, we introduce the UNL system. In Section V, we describe the building process of the Amazigh nominal dictionary for the UNL project. In Section V, we present our experimentation and correctness evaluation of the generated inflected forms. Finally in Section VI, we draw the conclusions, and present potential future research directions.

II. State-of-the-art

This section defines briefly the rule-based and statistical machine translation approaches by highlighting the advantages and the limitations of each approach.

A-Linguistic approaches

Rule-based machine translation is a general term that denotes machine translation systems based on linguistic information. The process of this approach is divided into three basic phases [10]:

- Analysis: Analyzing the source text in the intermediate representations;

- Transfer: Transferring these intermediate representations into intermediate representations in the target language;
- Generation: Generating the new text in the target language from the intermediate representations in the target language.

The "Vauquois triangle" proposed in [10] summarizes the three linguistic machine translation architectures [11]. Each path in the triangle "Fig.1" corresponds to a linguistic architecture.

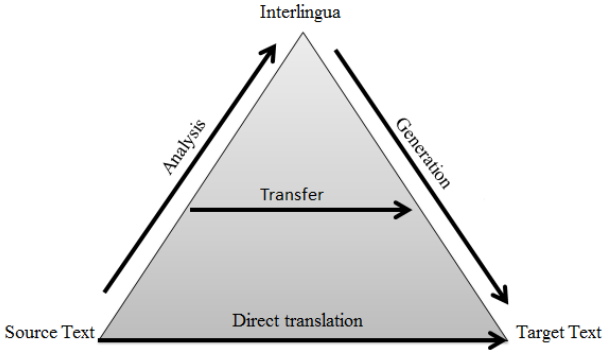

Fig. 1. Vauquois triangle

1) Direct translation

The source sentence is segmented into words. The system executes the translation word by word. The transfer step uses a bilingual table that associates for each word or sequence of source words, a set of translation rules and rearrangement for translating and reordering the words in the target sentence.

2) Transfer approach

The source sentence is analyzed with a parser and grammar. To transform the source language parse structure into a target language parse structure a bilingual table, containing the transfer rules between the source and target representations, is required to generate the target language sentence from the target parse structure.

3) Interlingua approach

The translation of any source language to any target language is the process of enconverting the source sentence to the pivot representation, and deconverting the target sentence from this pivot representation. The advantage of this method is the ability to apply it in a multilingual environment. Furthermore, the transfer of the source language representations into those of the target language is no longer needed. Thus, to cover translations between *n* languages, we only need *n* enconversion modules and *n* deconversion modules.

Table I draws some advantages and limitations of rule-based approaches.

TABLE I. ADVANTAGES AND LIMITATIONS OF RULE-BASED APPROACHES

Prerequisites	Dictionary entries and linguistic rules.
Advantages	The quality of the translation is very good, particularly when it is about a specialized translation.
Limitations	The development of linguistic rules is expensive in terms of time and effort.

B- Statistical machine translation approaches

Statistical Machine Translation (SMT) translates automatically a text from a source language to a target language by performing machine learning system using mathematical and statistics calculations on aligned bilingual data [12] [13].

SMT requires two models: a language model and a translation model. These models are based on the mathematical theory of Frederick Jelinek distribution and probabilistic estimation [14].

- **Translation model** which provides for each source segment the translation candidates in the target language
- **Language model** which is responsible for capturing the constraints imposed by the syntax of the target language using probabilistic functions.

Fig.2 illustrates the architecture of a statistical machine translation system.

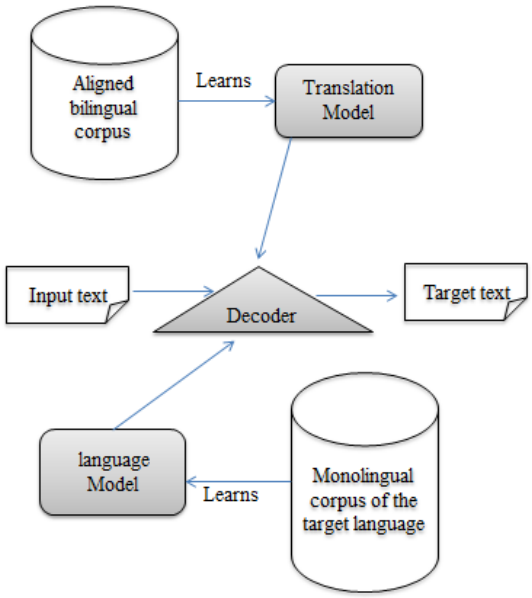

Fig. 2. An architecture example of a statistical machine translation system

Table II defines two types of statistical approaches.

TABLE II. STATISTICAL METHODS

Statistical methods	
A. ***Based on corpus***	*B.* ***Based on examples***
Learning translation rules from a bilingual corpus [10].	Reuse of existing translation examples as a basis for the new translation [15].

Statistical machine translation models were based, first, on words where the sentences are translated word by word. Then, Koehn *et al*. have improved these models and established systems based on machine translation of sequence of words [15]. Table III presents some strengths and limitations of the statistical approaches.

TABLE III. ADVANTAGES AND LIMITATIONS OF THE STATISTICAL MACHIINE TRANSLATION APPROACHES

Prerequisites	Parallel corpus
Advantages	• The quality of the translation is improved automatically and gradually as the training base is enriched; • Few linguistic resources needed; • It is faster to implement and easy to maintain; • Portability: Easy to extend toward a new language; • Possibility to evaluate experimentally practical systems and scientific hypotheses.
Limitations	Difficulty of building sufficient linguistic resources in quantity and quality.

As shown in Table III, the prerequisite for building a statistical machine translation system is the availability of large parallel corpus for training. But our language, unfortunately, doesn't own this type of corpus. Thus why, we have chosen the linguistic approach for realizing an Amazigh machine translation, especially that the interlingua approach based on UNL has proven many advantages.

III. AMAZIGH NOUN MORPHOLOGY

Amazigh morphology covers three main syntactic categories: nouns, verbs, and particles [16]. However, since the work of this paper is focused on noun category, it will be the subject matter of the rest of this section.

In Amazigh, Noun is a grammatical category that has claimed one of the following three forms:

- Simple noun is a single word (example: 'ⴰⵡⴰⵙⵙ' [Adlis] *book*).
- Compound noun is composed at least of two elements forming a single word that has its own meaning. Example: 'ⴰⵏⴱⵓⵉ ⵏ ⵔⴱⴱⵉ' [anbgui n rbbi] *guest*.
- Derived noun is a noun formed by the affixation processes of a morpheme to the base form of a simple noun.

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

A. Gender

Most male nouns begin with 'ⴰ' [a], 'ⵉ' [i], or 'ⵓ' [u] unless some exceptions, namely the case of kinship nouns like 'ⵉⵎⵎⴰ' [imma] *my mother*, which is a female name but begin with 'ⵉ' [i]. Concerning the formulation of female nouns, 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 'ⵜⵉⵙⵍⵉⵜ' [tislit] *the bride* is the feminine noun for the masculine noun 'ⵉⵙⵍⵉ' [isli] *the groom*. Some female nouns take only the initial morpheme 'ⵜ' [t] or the final 'ⵜ' [t], like 'ⵜⴰⵙⴰ' [tasa] *the liver*.

B. Number

The noun has a singular and a plural form, whatever it is in masculine or feminine. The plural forms are classified into four types: external plural, broken plural, mixed plural and plural in 'ⵉⵏ' [id] [16] [17].

- The external plural is formed by an alternation of the first vowel 'ⴰ/ⵉ' [a/i] accompanied by a suffixation of 'ⵏ' [n] or one of its variants.
- The broken plural involves a change in the vowels of the noun.
- The mixed plural is formed by vowels change accompanied sometimes by the use of the suffixation of 'ⵏ' [n].
- The plural in 'ⵉⵏ' [id] is obtained by putting the morpheme 'ⵉⵏ' [id] before the noun. It is applied to a set of nouns including nouns with an initial consonant, proper nouns, parent nouns, compound nouns, numerals, as well as borrowed nouns from foreign languages.

C. State

The noun can be in two different states: Free State or Annexed State (also called the construct state). We talk about a noun in the Free State, when it is:

- isolated from any grammatical context;
- a direct object;
- a complement of the predictive particle 'ⵏ' [d].

The construct state 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 'ⴰⵍ/ⴰⵔ' [al/ar] and 'ⴱⵍⴰ' [bla];
- after a numeral;
- after the morpheme 'ⵉⵏ' [id];
- after the morphemes of membership and affiliation: 'ⵓ' [u], 'ⵓⵍⵜ' [ult], 'ⴰⵢⵜ' [ayt], 'ⵉⵙⵜ' [ist].

The annexed state is often marked, in the case of feminine nouns, by the drop of the vowel 'ⴰ/ⵉ' [a/i] placed after the

initial '+' [t]. While for masculine nouns, usually the initial vowel 'ⴰ' [a] changes to 'ⵉ', 'ⴻ' [i] to 'ⵢ', and 'ⵓ' [u] to 'ⵡ' [w] [16].

IV. Universal Networking Language system

A. The UNL system Components

The meaning of a sentence in its natural language can be expressed in the UNL system by a graph consisting of nodes inter-related by semantic relations [18]. A node is called UW 'Universal Word'. It is often accompanied with a set of grammatical properties called attributes. Arcs binding UNL nodes in the graph represent a relationship between the UWs. The basic structure of the UNL language is:

Universal Words (UWs): they 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: they 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: it 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.

B. Basic architecture of the UNL system

The UNL system requires a set of language resources [19], 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 four 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;
- UNL-NL dictionary, called Generation Dictionary, which is a bilingual dictionary that link UNL words (UWs) to the corresponding lexical items in the natural language;
- The NL-UNL dictionary or Analysis Dictionary which is a bilingual dictionary linking Amazigh word to UNL words called Universal Words (UWs). This dictionary will be used in UNLization process (enconversion from a nature language word to UNL graphs). It is presented in the enumerative format which means that it brings all word forms and not only base forms.

In this paper, we focus, specially, on the building of the nominal NL dictionary for the Amazigh language.

V. Amazigh dictionary development

A. Template of the Amazigh dictionary

The Amazigh dictionary is presented in the enumerative format which means that it brings all word forms and not only base forms [20]. The Amazigh dictionary entry has the following form:

[NLW] {ID} (ATTR …) < FLG, FRE, PRI >;

Where:

- ***NLW***: Head Word (Berber word).
- ***ID***: The unique identifier (primary-key) of the entry.
- ***ATTR***: The list of grammatical, morphological and semantic features, subcategorization frames (FRA) and inflectional paradigms (PAR) of the NLW.
- ***FLG:*** The three-character language code according to ISO 639-3 (we use Ber for Berber).
- ***FRE:*** The frequency of NLW in natural texts is used for natural language analysis. It can range from 0 (less frequent) to 255 (most frequent).
- ***PRI:*** The priority of the NLW. It is used for natural language generation. It can range from 0 to 255.

Fig. 3 below is a screenshot of the Amazigh dictionary that we are developing.

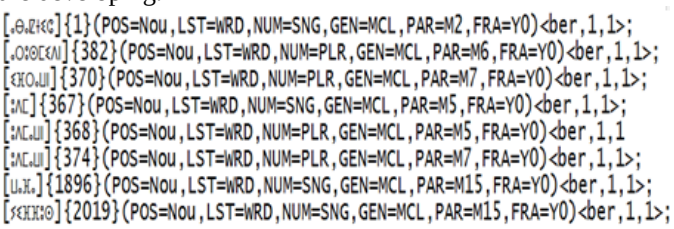

Fig. 3. Screenshot of the Amazigh dictionary

In order to perform well the development of the Amazigh dictionary, we expected to analyze and define morphological behavior of the Amazigh word forms, that by assigning linguistic features (part of speech, gender, number, transitivity, etc.) and inflectional paradigms for each entry of the dictionary.

B. Definition of inflectional paradigms

We have started feeding our dictionary by nominal entries (1400 nouns at this moment) extracted from three lexicons [21]-[23]. For this, a set of rules, allowing inflectional generation, has been defined for each entry, including inflectional forms according to gender, number and state.

The inflectional rules follow the general standard ***<CONDITION>:=<ACTION>,*** where:

- <CONDITION> is a set of features (such as MCL&PLR, i.e., masculine and plural);
- <ACTION> describes the changes to be performed over the base form (prefixation, infixation, suffixation or circumfixation).

For instance, the rule FEM&SNG:="+"<0,0>"+" means that, in order to form the feminine singular form , we have to

add the morpheme 'ⵜ' [t] at the beginning and another at the end of the base form. The grammar formalism adopted within the UNL is described in [20].

The realization of inflectional paradigms for the noun category has presented for us a real challenge concerning the formalization of the Amazigh plural forms because there are many unpredictable plural forms and there is a scarcity of documents dealing with Amazigh inflectional paradigms. According to a heuristic study, based on [16] [17] [24] [25], we have achieved, at this moment, to find out 80 classes of inflectional forms.

Table IV below shows an example for some inflectional paradigms classes

TABLE IV. AN EXTRACT OF INFLECTIONAL PARADIGMS LIST DEFINED FOR AMAZIGH NOUNS

Classes	Inflectional paradigms in UNL	Explanation	Inflected words
Class 1 ('ⵎⵔⴰⵡ' [mraw] *ten*)	MCL&SNG&NOM :=0>"";	For the masculine singular form(free state), no change	ⵎⵔⴰⵡ [mraw]
	MCL&SNG&CTS := 0>"";	For the masculine singular form (annexed state) , no change.	ⵎⵔⴰⵡ [mraw]
	MCL&PLR&NOM :="ⵉⵏ "<<0;	For the masculine plural form (free state), the rule adds ⵉⵏ [id] before the noun.	ⵉⵏ ⵎⵔⴰⵡ [id mraw]
	MCL&PLR&CTS :="ⵉⵏ "<<0;	For the masculine plural annexed state form, the rule adds ⵉⵏ [id] before the noun.	ⵉⵏ ⵎⵔⴰⵡ [id mraw]
	FEM&SNG&NOM :=0>"ⵜ";	For the feminine singular form, the rule adds 'ⵜ' at the end of the noun.	ⵎⵔⴰⵡⵜ [mrawt]
	FEM&SNG&CTS:= 0>"ⵜ";	For the feminine singular annexed state form, the rule adds 'ⵜ' at the end of the noun.	ⵎⵔⴰⵡⵜ [mrawt]
	FEM&PLR&NOM := "ⵉⵏ"<<0,0>"ⵜ";	For the feminine plural form, the rule adds ⵉⵏ [id] before the noun, and 'ⵜ' [t] at the end of the noun.	ⵉⵏ ⵎⵔⴰⵡⵜ [id mrawt]
	FEM&PLR&CTS:= "ⵉⵏ "<<0,0>"ⵜ";	For the feminine plural annexed state form, the rule adds ⵉⵏ [id] before the noun, 'ⵜ' at the end of the noun.	ⵉⵏ ⵎⵔⴰⵡⵜ [id mrawt]
Class 2	MCL&SNG&NOM := 0>"";	For the masculine singular form (free state), no change	ⴰⵎⵏⴰⵙⴽⵍ [amddukkl]
	MCL&SNG&CTS	For the masculine	ⵙⵎⵏⴰⵙⴽⵍ
('ⴰⵎⵏⴰⵙⴽⵍ' [amddukkl] *friend*)	:= "ⵙ"<1;	singular annexed state form, the rule changes the first letter by 'ⵙ' [u]	ⵓ [umddukkl]
	MCL&PLR&NOM :="ⵉ"<1,"ⴰ"<[-1];	For the masculine plural form, the rule changes the first letter by 'ⵉ' and inserts 'ⴰ' before the last letter	ⵉⵎⵏⴰⵙⴽⵍ [imddukkal]
	MCL&PLR&CTS:= "ⵉ"<1,"ⴰ"<[-1];	For the masculine plural annexed state form, the rule changes the first letter by 'ⵉ' and inserts 'ⴰ' before the last letter	ⵉⵎⵏⴰⵙⴽⵍ [imddukkal]
	FEM&SNG&NOM: = "ⵜ"<0, "ⴰ"<[-1];	For the feminine singular form, the rule adds 'ⵜ' at the begin of the noun and insert 'ⴰ' before the last letter	ⵜⴰⵎⵏⴰⵙⴽⵍ [tamddokkal]
	FEM&SNG&CTS:= "ⵜ"<1, "ⴰ"<[-1];	For the feminine singular annexed state form, the rule changes the first letter by 'ⵜ' [t] and inserts 'ⴰ' [a] before the last letter	ⵜⵎⵏⴰⵙⴽⵍ [tmddokkal]
	FEM&PLR&NOM := "ⵜⵉ"<1, "ⴰ"<[-1];	For the feminine plural form, the rule changes the first letter by 'ⵜⵉ' [ti] and insert 'ⴰ' [a] before the last letter	ⵜⵉⵎⵏⴰⵙⴽⵍ [timddokkal]
	FEM&PLR&CTS:= "ⵜ"<1, "ⴰ"<[-1];	For the feminine plural annexed state form, the rule changes the first letter by 'ⵜ' [t] and insert 'ⴰ' [a] before the last letter	ⵜⵎⵏⴰⵙⴽⵍ [tmddokkal]
Class 3 ('ⵉⵙⵍⵉ' [isli] *groom*)	MCL&SNG&NOM :=0>"";	For the masculine singular form, no change	ⵉⵙⵍⵉ [isli]
	MCL&SNG&CTS := 0>"";	For the masculine singular annexed state form, no change	ⵉⵙⵍⵉ [isli]
	MCL&PLR&NOM :=1>"ⵍ ";	For the masculine plural form, the rule adds 'ⵍ' [an] at the end of the noun	ⵉⵙⵍⵉⵍ [islan]
	MCL&PLR&CTS:= 1>"ⵍ ";	For the masculine plural annexed state form, the rule adds 'ⵍ' [an] at the end of the noun	ⵉⵙⵍⵉⵍ [islan]
	FEM&SNG&NOM: ="ⵜ"<0,0>"ⵜ";	For the feminine singular form, the rule adds 'ⵜ' [t] at the begin and at the end of the	ⵜⵉⵙⵍⵉⵜ [tislit]

		noun	
	FEM&SNG&CTS:= "ⵜ"<1, 0>"ⵜ";	For the feminine singular annexed state form, the rule replaces the first letter by 'ⵜ' [t] and adds 'ⵜ' [t] at the end of the noun	ⵜⵙⵍⵜ [tslit]
	FEM&PLR&NOM :="ⵜ"<0,1>"ⴰⵜⵏ";	For the feminine plural form, the rule adds 'ⵜ' [t] at the begin of the noun, and changes the final letter ⵜ by 'ⴰⵜⵏ'	ⵜⵜⵙⵍⴰⵜⵏ [tislatin]
	FEM&PLR&CTS:= "ⵜ"<1,1>"ⴰⵜⵏ";	For the feminine plural annexed state form, the rule changes the first letter by 'ⵜ' [t] and replaces the final letter by 'ⴰⵜⵏ'	ⵜⵙⵍⴰⵜⵏ [tslatin]
Class 4 ('ⴰⵙⵍⵎⴰⵏ' [aslmad] *Teacher*)	MCL&SNG&NOM :=0>"";	For the masculine singular form (free state), no change	ⴰⵙⵍⵎⴰⵏ [aslmad]
	MCL&SNG&CTS :="ⵓ"<1;	For the masculine singular annexed state form, the rule changes the first letter by 'ⵓ'[u]	ⵓⵙⵍⵎⴰⵏ [uslmad]
	MCL&PLR&NOM :="ⵉ"<1,0>"ⵏ"	For the masculine plural form, the rule replaces the first letter by 'ⵉ'[i] and adds 'ⵏ'[n] at the end of the noun	ⵉⵙⵍⵎⴰⵏ [islmadn]
	MCL&PLR&CTS:= "ⵉ"<1,0>"ⵏ"	For the masculine plural annexed state form, the rule replaces the first letter by 'ⵉ'[i] and adds 'ⵏ'[n] at the end of the noun	ⵉⵙⵍⵎⴰⵏ [islmadn]
	FEM&SNG&NOM: ="ⵜ"<0,0>"ⵜ";	For the feminine singular form, the rule adds 'ⵜ'[t] at the begin and at the end of the noun	ⵜⴰⵙⵍⵎⴰⵏⵜ [taslmadt]
	FEM&SNG&CTS:= "ⵜ"<1,0>"ⵜ";	For the feminine singular annexed state form, the rule replaces the first letter by 'ⵜ'[t] and adds 'ⵜ'[t] at the end of the noun	ⵜⵙⵍⵎⴰⵏⵜ [tslmadt]
	FEM&PLR&NOM :="ⵜⵉ"<1,0>"ⵉⵏ";	For the feminine plural form, the rule replaces the first letter by 'ⵜⵉ'[ti] and adds 'ⵉⵏ'[in] at the end of the noun	ⵜⵉⵙⵍⵎⴰⵏ ⵉⵏ [tislmadin]
	FEM&PLR&CTS:= "ⵜ"<1,0>"ⵉⵏ";	For the feminine plural annexed state form, the rule replaces the first letter by 'ⵜ'[t] and adds 'ⵉⵏ'[in] at the end of the noun	ⵜⵙⵍⵎⴰⵏⵉ [tslmadin]

VI. EXPERIMENTATION AND ANALYSIS

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 [26]. Thus, we compared the results of the implemented inflectional rules, applied on the basic corpus, to the list of test corpus entries. The results of the evaluation are given in the Table below:

TABLE V. RESULTS OF INFLECTIONAL RULES EVALUATION

	Similar inflected forms (%)	**Dissimilar inflected forms (%)**
Results	71.3%	28.7%

By analyzing the dissimilar plural forms, we remarked that the difference between the basic and test corpus lists are generally due to two factors, 25.3% is ascribable to the first factor while 3.4% is referred to the second:

1- The Amazigh regional varieties: In fact, in Moroccan, there are three main Amazigh language varieties: Tarifite in the North, Tamazight in the Middle Atlas and South-Eastern Morocco, and Tashelhit in the South-Western. Thus, one word could have many inflectional paradigms. For example, the noun 'ⵜⴰⵖⵔⵉⵎⵜ' [tagldit] *the kingdom* has two plural forms 'ⵜⵜⵖⵔⵉⵎⵜⵏ' and 'ⵜⵜⵖⵔⵉⵎⴰⵢ'.

2- Automatic classification problem: For instance the noun 'ⵜⵉⵜⵜⵎⵔⵉ' [ittmri] *blocking* was misclassified into a class including nouns of the pattern 'ⵜ….ⵜ', like the word 'ⵜⵉⵜⵔⵉ' [itri] *star.* So, the plural form of 'ⵜⵉⵜⵜⵎⵔⵉ' was associated to the inflectional rule consisting in replacing the final 'ⵜ' by 'ⵏ' ('ⵜⵉⵜⵔⵉ' [itri] →'ⵜⵉⵜⵔⵏ' [itran]). While, it should be associated to the inflectional rule of an irregular class, that consists in adding the morpheme 'ⵜⵏ' [tn] in the end of the word ('ⵜⵉⵜⵜⵎⵔⵉ' [ittmri] → 'ⵜⵉⵜⵜⵎⵔⵉⵜⵏ' [ittmritn]).

To reduce this dissimilarity, we are trying to be limited in the test corpus to the standardized words, and to improve our classification system.

VII. CONCLUSION & PERSPECTIVES

In the perspective to realize an UNL-based machine translation system for the Amazigh language, we successfully integrated it into the UNL framework. Then, we have proceeded to build the first Amazigh dictionary template, by defining inflectional paradigms. Currently, we have treated only nouns. It stills verbs and particles categories. Our future plan is to focus on inflectional paradigms of verbs and to undertake the mapping between items and their corresponding universal words. Thus in this stage, we will be concerned to link Amazigh words to their corresponding UWs to build the Amazigh-UNL dictionary.

REFERENCES

[1] J.H. Greenberg, "The Languages of Africa". The Hague, 1963.

[2] O. Ouakrim, "Fonética y fonología del Bereber". Survey at the University of Autònoma de Barcelona, 1995.

[3] P.F. Brown, V.J. Pietra, S.A.D Pietra, R.L. Mercer, "A statistical approach to machine translation". IBM Thomas J. Watson Research Center Yorktown Heights, NY, p. 79-85, 1990.

[4] M.S. Maučec, Z. Kačič, D. Verdonik, "Statistical machine translation of subtitles for highly inflected language pair", Pattern Recognition Letters, Vol. 46, p. 96-103, September, 2014. Elsevier Science Inc. New York, NY, USA.

[5] I. Taghbalout, F. Ataa Allah, M. Elmarraki, "Towards an Amazigh UNL dictionary", Proceedings of the 2nd WSEAS International Conference on Systems, Control and Informatics, p. 129-132, November 28-30, 2014, Athenes, Greece, WSEAS Press.

[6] I. Boguslavsky, J. Cardeñosa, C. Gallardo, L. Iraola, "The UNL initiative. An overview", Computational Linguistics and Intelligent Text Processing, 6th International Conference, CICLing 2005, Mexico City, Mexico, February 13-19, 2005.

[7] C. Boitet, I. M. Boguslavskij, J. Cardeñosa, "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, p. 361-373, 2007. Springer-Verlag.

[8] V. Stoykova, "Lexical Information for Bulgarian in Universal Networking Language", Proceedings of the 8th WSEAS International Conference on Computer Engineering and Applications, p. 414-417, January 9-11, 2014, Tenerife, Spain. WSEAS Press.

[9] P. Kumar, "UNL Based Machine Translation System for Punjabi Language", Ph.D. thesis, School of Mathematics and Computer Applications, Thapar University, February 2012, Patiala, India.

[10] B. Vauquois, "A Survey of Formal Grammars and Algorithms for Reconition and Translation", FIP Congress-68.Edinburg, 254-260, 1968.

[11] J. Hutchins, "Linguistic models in machine translation", UEA Papers in Linguistics, Vol. 9, p. 29-52, January 1979.

[12] P.Brown, J.Cocke, S.Della Pietra, V.Della Pietra, F.Jelinek, R.Mercer, P.Roossin, "A statistical approach to language translation", Proceedings of the 12th International Conference on Computational Linguistics, vol.1, p. 71-76, August 22-27, 1988, John von Neumann Society for Computing Sciences, Budapest, Hungary.

[13] S. Vetrivel, Diana Baby, "English to Tamil statistical machine translation and alignment using HMM", Proceedings of the 12th WSEAS International Conference on Networking, VLSI and signal processing, p.182-186, February 20-22, 2010, UK. WSEAS Press.

[14] P. F. Brown, V. J. Pietra, S.A.D Pietra, R.L. Mercer, "The Mathematics of Statistical Machine Translation: Parameter Estimation". IBM T.J. Watson Research Center, p. 264-311, 1993.

[15] P. Koehn, F.Josef Och, D. Marcu, "Statistical phrase-based translation". In NAACL ˝03 : Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Morristown, NJ, USA. Association for Computational Linguistics. 48-54. 2003.

[16] F. Boukhris, A. boumalk, E. El moujahid, H. Souifi, " La nouvelle grammaire de l'amazighe", Rabat (IRCAM), Maroc, 2008.

[17] L. Oulhaj, "Grammaire du Tamazight", Centre Tarik ibn Zyad center for studies and research, 2000.

[18] H. Uchida, M. Zhu. "The Universal Networking Language (UNL) Specification Version 3.0". Edition 3, Technical Report, UNU, (2005/6-UNDL Foundation, International Environment House, Tokyo, 2004)

[19] UNDLFOUNDATION, Inflectional paradigms. Available on http://www.unlweb.net/wiki/Inflectional_paradigms.

[20] UNDLFOUNDATION, Lexica. Available on http://www.unlweb.net/wiki/Lexica.

[21] M. Ameur, A. Bouhjar, A.Boumalk, N. El Azrak, R. Laabdelaoui, Vocabulaire des Médias (Français – Amazighe – Anglais - Arabe), 2009, IRCAM, Rabat.

[22] M. Ameur, A. Bouhjar, A.Boumalk, N. El Azrak, R. Laabdelaoui, معجم اللغة اللأمازيغية, 2009, IRCAM, Rabat.

[23] M. Ameur, A. Bouhjar, A.Boumalk, N. El Azrak, R. Laabdelaoui, Vocabulaire Grammatical, 2009, IRCAM, Rabat.

[24] F. Nejme, S. Boulaknadel, D. Aboutajdine,"Toward an Amazigh language processing", the 3rd Workshop on South and Southeast Asian Natural Language Processing, December 2012, Mumbai, India.

[25] H. Raiss, V. Cavalli-Sforza,."Amazigh Nouns Morphological Analyzer", 5ème Conférence internationnale sur les TIC pour l'amazighe, Rabat, Maroc 2012.

[26] M. Ameur, A.Boumalk, N. El Azrak, R. Laabdelaoui, Dictionnaire général de l'amazighe, 2015, IRCAM, Rabat.