E-Assessment System Based on Natural Language Processing for Arabic Language

Dalila SOUILEM[1], Abdelkarim ABDELKADER[2], Rafik BRAHAM

National Engineering school of SOUSSE, University of SOUSSE, TUNISIA

College of Computer at Al-Gunfudah, Umm Al-Qura University, Saudi Arabia

[3] Prince Research Group, ISITCom, University of Sousse, Tunisia

Abstract— ***In this paper, we focus on the application of Natural Language Processing (NLP) in an e-assessment system for Arabic language. The development of some linguistic tools as lexicon, morphological analyzer and parser is crucial. Lexicon represents the most important resources for the system Morphological analyzer represents a decomposition, an analysis and research tool for Arab texts. But other evaluation activities require a parsing system. Parsing is a fundamental step in the process of automatic analysis of natural languages. In this paper we describe, firstly the building of a lexicon of Arabic language using the Head Driven phrase Structure Grammar (HPSG) formalism. The lexicon is represented in XML as a list of item elements, and it lists several lexical properties which are specifically targeted at morphological analysis. It also specifies morphological, syntactic and semantic features (such as gender or number), which can be used by parser and assessment system. Also we explain the ascendant approach design of parser and we give an idea on the experimentation of the realized system and its integration in e-learning environment.***

Keywords *: NLP, Parser, HPSG, XML, e-assessment, Arabic language.*

I. INTRODUCTION

An e-assessment system using NLP tools was developed to assist and evaluate learner. The assessment can occur with the learner or tutor at different levels of learning action. There are traditional exercises such as gap-fill, multiple choice questions, matching words, reordering words etc. These exercises are easily corrected by the system and they are also useful to check understanding of a text by the learners. But these types of exercises are insufficient to properly learn the language. However, it is necessary to use more advanced and robust language tools in order to integrate more complex evaluations and activities for learners to produce complete and complex sentences.

The Arabic language was considered among the most difficult languages to automatically process because of several problems. But the development of computer science and the advancement of research in the field of natural language process (NLP) are constantly inventing more and more suitable tools that have allowed removing some obstacles. But even with these advances in technology, there are still many problems in the automatic processing of Arabic. These obstacles come from the characteristics of the Arabic language itself, to the point that some researchers have claimed that there are no robust and powerful complete systems that can be integrated, for example, in a computer environment for online learning. One of the major problems, such as clumping, flexion and lack of diacritics (vowels) is the irregularity of the Arab and the presence of several linguistic phenomena in that language.

We start by presenting an overview for the HPSG formalism and some specific features of the Arabic language at the morphological and syntactic level which are the sources of difficulties and ambiguities that may arise during the automatic analysis of Arabic. Also we show the contribution of HPSG formalism for building the lexicon of simple Arabic sentences. Then, we describe some NLP tools such as the morphological and the syntactic analyzer that we are integrate in our e-assessment system.

II. BUILDING OF AN ARABIC LEXICON

A. The HPSG formalism

There are several reasons for the popularity of HPSG as linguistic formalism that provides a rigorous framework for grammar development through its unification of feature structures. Sub-categorization and semantic representation are important qualities of HPSG.

Furthermore, the founders of this theory list key components of the HPSG formalism that explain the impact it had on grammar development, such as

- Its foundation on typed feature logic;
- Using the same formalism for universal and particular generalizations;
- Allowing for language specific analyses, that are less natural described in other formalisms.

The Head-Driven component of HPSG name reveals one of the underlying principles of this formalism. The most important element of a phrase

is its lexical head (C.POLLARD & I.SAG, 1994)[6]. The lexical head incorporates both syntactic information (such as part of speech and dependency relations with other constituents) and semantic information. Lexical entities, in general, are information-rich structures, with feature values from the type signature. This lexical centered organization of HPSG eliminate redundancies, therefore the amount and complexity of phrasal rules is greatly reduced. Typed feature structures (TFS) are widely used in natural language processing and computational linguistics to enrich syntactic categories.

This formalism is based on unification. The unification of two types feature structures is an operation that produces a feature structure that is their least upper bound, with respect to the subsumption ordering. Unification fails when the two feature structures provide inconsistent information. Intuitively, unification is accomplished by first unifying the types of the root node of the two feature structures, and replacing them with the result of the type unification. By recursion, nodes that are values of the identical features are then unified, and replaced with the result of the unification. Failure occurs when an inconsistency between nodes occurs (i.e. , when two types cannot unify)

Agreement phenomena in Arabic sentence, involve morphological, syntactic, semantics and pragmatics. This is another domain in which the sign-based formalism of HPSG has yielded significant results. The central concept of the HPSG theory of agreement is the INDEX, which unifies some of the properties of constants and variables from logical formalism. In the simplest cases an INDEX is an abstract linguistic entity that is referentially linked to some object in the interpretation domain. Indices are also used with quantification, in which case they behave much like variables. Unlike constants and variables in logic, however, indices have an internal organization that reflects properties of the associated linguistic entities or referential objects. This information includes number, gender, and person. This makes it possible to straight forwardly account for a number of agreement phenomena. For instance, if we assume that person/ number/ gender information is encoded on indices and that the relation between reflexive pronouns and their antecedents involves INDEX identity, then the distribution of forms in follows immediately.

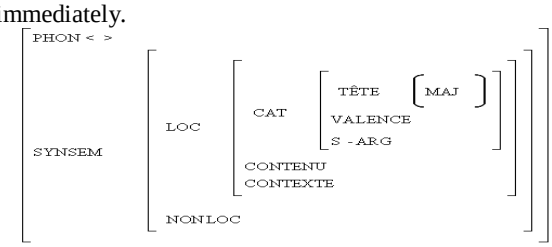

Fig 1. Feature structure for any lexical entry (A.ABDELKARIM & H.KAIS,2006)[3].

B. *Specificities of Arabic language*

The Arabic alphabet has twenty-eight letters (consonants) given that there is no difference between written and printed letters. In Arabic language, the capital letter and lowercase concepts do not exist. However, letters have, for the most part, a form somewhat different according to whether they are isolated, beginning, middle or end of the word. To this set of letters are added all the vowels that can disambiguate the phonetic words and give a sequence of consonants different syntactic functions and different meanings. Signs of vocalization are rated as diacritical marks placed above or below the letters. However the normal graphic note only long consonants and vowels. Only some special books and some children's books are printed with the vowels (we say vocalized). Much of what is written and still writes in Arabic is a non-vocalized spelling. To write this way is economic. But, to read this poses problems that cannot overcome gradually, resulting in visual memory and developing its grammatical and textual intuitions.

Also, Arabic is an inflected language in which the words change shape according to the grammatical relationships they have with other lemmas (words). So some of the words change their form (audible and / or visual), we say them to undergo the game of flexion. The different forms of the same word formed its paradigm. For this the Arab ranks as a very rich morphology language. In addition, Arabic obeys communicative and syntactic constraints. This is why it ranks as one of mixed language, neither fixed nor entirely free.

C. *The features for Arabic lexicon*

The Arabic lexicon includes essentially three categories of units: verbs, nouns and particles. As in many other Semitic languages, Arabic verb formation is based on a trilateral root, which is not a word in itself but contains the semantic core. The consonants K-T-B for example indicate “write”. Words are formed by supplying the root with a vowel structure and with affixes. Traditionally, Arabic grammarians have used the root F-A-L “do” as a template to discuss word formation. Each verb is the origin of a set of words. Verb conjugation depends on several factors:

- The time (accomplished unfulfilled).
- The number (singular, dual, plural).
- The type of the subject (male, female).
- The person (first, second and third)
- The mode (active, passive).

These factors are designated by a variety of prefixes and suffixes. For personal pronouns, the subject is included in the verb. It is therefore not necessary to precede the verb conjugated by its pronoun. The perfective form expresses a completed action, i.e. mostly past tense. The imperfective

expresses an action in progress, or incomplete, i.e. mostly present tense.

The noun is a word that refers to a person, an object or a state. The function of name is its relationship with a word or phrase; it changes with the change in this relationship without losing its linguistic meaning. Arab nouns fall into two categories, those derived from the verbal root and those who do not like proper nouns and common nouns. All nouns are either singular when there is one, dual when there are two, or plural if there are three or more. In this category, we also find the pronouns (the personal pronouns, the demonstrative pronouns, the relative pronouns and the interrogatives pronouns) and the adjectives (the analogous adjectives, the active participles, the passive participles, the verbal names). Adjectives and appositions follow the noun and agree with the preceding noun in state, gender and case.Fig2 present the lexical entry of a noun.

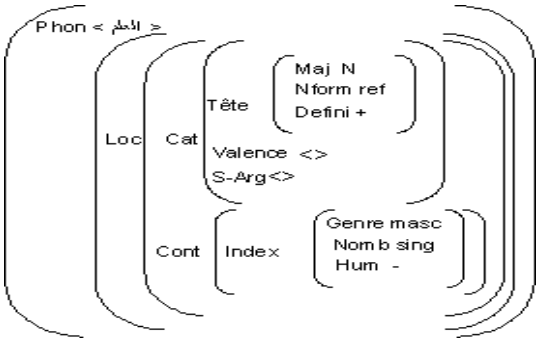

Fig2. Lexical entry of the noun «العلم ».

We notice that the noun "العلم" is a referenced name, defined, masculine, singular, and not human.

In the Arabic lexicon we have the particle. This category designates the invariable words that don't take a sense when they are not accompanied by other words. Particles are used to locate events and objects in relation to time and space, and enable a consistent text flow. It is mainly the function words such as conjunctions of coordination and subordination. We distinguishes the actual prepositions, the verbal and nominal particles, the conjunctions of proposition, the conjunctions of coordination, the conjunctions of exception, the conjunctions of call, the conjunction of negation and conjunction of condition. The particles are classified according to their semantics and their function in the sentence; They play an important role in the interpretation of the sentence. Fig3 explains the lexical entry of particle unit.

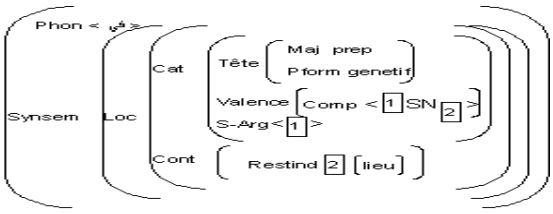

Fig3: Lexical entry for the particle "في".

We can construct two other categories that are the syntactic group (syntagm) and the sentence.

In modern grammar, between the upper limit of the syntax, consisting of the sentence, and the lower limit, constituted by the single category (base unit, inseparable), we find more or less complex syntactic group unit (syntagm). A syntagm is a set of words forming a single categorical and functional unit, but each component (unlike the compound word), retains its own meaning and its own syntax. A syntagm is, therefore, an occasional association, free, while the compound word is a permanent association.

Regardless of the level of a given syntagm, it can virtually be included in a syntagm of the upper level (grouping principle), but can also include syntagms lower level (principle of division). This dual principle of grouping and division (be, the principle of inclusion) and the structure of the syntagm identity make the of the sentence.

The syntagm possesses the categorical and functional unit of indivisible word. Analysis of syntagm will operate in three stages:

- Finding categorical unit (noun phrase, verbal phrase)
- Finding functional unit (higher unit to which depends, and its specific function);
- Finding different components as its core (which inherits its nature and function) and its lower units included in it.

Arabic syntagms can have multiple categorical units as verbal syntagm (مركب فعلي), noun syntagm (مركب اسمي) , adjective syntagm (مركب نعتي), prepositional syntagm (مركب حرفي), Annexation syntagm (مركب إضافي) and relative syntagm (مركب موصولي).In fig4 we have the syntagm representation.

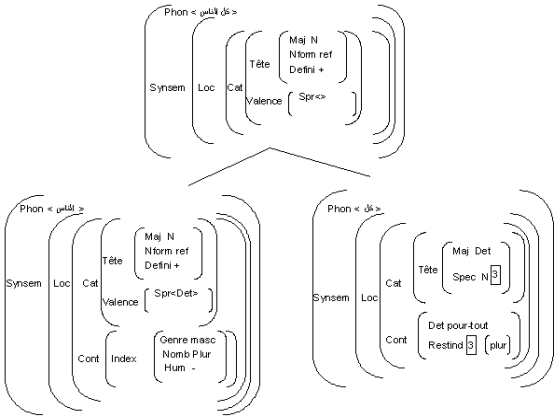

Fig4. Representation of the syntagm "كل الناس"

The sentence is defined as a self-contained unit, bringing together units organized according to various more or less complex relationships (subordination, coordination or juxtaposition). Furthermore, the sentence has a semantic unit, a communication unit, a meaning (conveyed by the message, or contained therein). A very characteristic fact, the absence or existence of a verb in mind the sentence is used to divide it into two categories: the noun phrase and verbal phrase. The tableI and tableII show the various structures that may have a simple Arabic nominal or verbal sentence.

TABLE I.
GRAMMATICAL FORMS OF THE NOMINAL SENTENCE

مبتدأ+خبر
ناسخ+اسم ناسخ+خبر ناسخ
مبتدأ+خبر(مركب فعلى)
مبتدأ(مركب موصولى)+خبر
ناسخ+اسم ناسخ+ خبر ناسخ(مركب موصولى)
ناسخ+اسم ناسخ(مركب موصولى) +خبر ناسخ

TABLE II.
GRAMMATICAL FORMS OF VERBAL SENTENCES

فعل+فاعل
فعل لازم+فاعل
فعل متعد+فاعل+مفعول به
فعل متعد+فاعل+مفعول به أول+مفعول به ثان
فعل+فاعل+مفعول به+مفعول لأجله
فعل+فاعل+حال
فعل+مفعول به+فاعل
فعل مبنى للمجهول+نائب فاعل
فعل مبنى للمجهول+نائب فاعل+مفعول به
فعل+فاعل+مفعول به(مركب موصولى)
فعل مبنى للمجهول+نائب فاعل(مركب موصولى)
فعل+ فاعل(مركب موصولى)

III. THE MORPHOLOGICAL ANALYSIS TOOL IN E-ASSESSMENT SYSTEM

Several morphological analyzers of Arabic was made including AMIA, MORPH2, ArabMorpho Sakhr, AraMorph, Xerox, Alkhalil Alsarfi ... Some analyzers are developed by non-Arab researchers as AraMorph and Xerox. Others treat each type of Arabic word in a different way such as Sakhr. Others use computational approaches whose complexity is talking such as MORPH2 and ArabMorpho. Alkhalil Alsarfi and AMIA are based on linguistic rules. In fact, efforts in this context are increasingly important to improve the performance. Comparison studies have been performed and have shown the effectiveness of "Alkhalil Alsarfi" and "AMIA" in our case. In fact they are free and open source systems. They allow to morphologically analyzing words, phrases and unvocalized or completely or partially vocalized texts. We worked on both systems by taking advantage of each. AMIA (K.NADIA & S.DALILA, 2006)[11] has the advantage of being designed as an educational tool that helps to analyze the response of the learner as the corpus studied. Alkhalil Alsarfi (ALKHALIL MORPHO SYS, 2010) [5] has advantages of language. It contains all the roots of Arabic (about 8000 roots) and all of the Arab verbs (more than 24000 verbs) and all non-canonical action names derived from verbs (14000 action names). Both systems have the capacity to generate the derived names and standard action names and all the morphological and syntactic rules relating to the process of derivation and conjugation are represented. It is the same for the morphologic transformation rules. In addition, they represent all transcription rules of the "Hamzah" and all derivatives or conjugates words.

Figures 5 and 6 show analysis and labeling of words in the sentence "كتب الولد رسالة" completely or partially vocalized using morphological analyzer.

الدخل INPUT	الخرج OUTPUT						
	الكلمة المشكولة Voweled Word	السابق Prefix	الجذع Stem	نوع الكلمة Type	الوزن Pattern	الجذر Root	الحالة الإعرابية POS Tags	اللاحق Suffix
كتب	كَتَبَ	#	كتب	مصدر أصلي	فَعْل	كتب	مفرد مذكر مرفوع في حالة الإضافة	#
	كَتَبٌ	#	كتب	مصدر أصلي	فَعْل	كتب	مفرد مذكر مرفوع نكرة	#
	كَتَبَ	#	كتب	مصدر أصلي	فَعْل	كتب	مفرد مذكر منصوب في حالة الإضافة	#
	كَتَبَ	#	كتب	فعل ماض مبني للمعلوم	فَعَلَ	كتب	ثلاثي مجرد مسند إلى الغائب(هو) متعد ولازم	#
	كَتَّبَ	#	كتب	فعل ماض مبني للمعلوم	فَعَّلَ	كتب	ثلاثي مزيد مسند إلى الغائب(هو) متعد ولازم	#
	كُتِّبَ	#	كتب	فعل ماض مبني للمجهول	فَعَّلَ	كتب	ثلاثي مزيد مسند إلى الغائب(هو) متعد ولازم	#
	كُتِبَ	#	كتب	فعل ماض مبني للمجهول	فَعِلَ	كتب	ثلاثي مجرد مسند إلى الغائب(هو) متعد ولازم	#
	كُتِّبَ	#	كتب	فعل أمر	فَعِّلْ	كتب	ثلاثي مزيد مسند إلى المخاطب أنت متعد ولازم	#
التلميذ	التلميذ	ال: التعريف	تلميذ	اسم جامد	تَفْعِيل	تلمذ	مفرد مذكر	#
	رسالة	#	رسالة	مصدر أصلي	فَعالَة	رسل	مفرد مؤنث مرفوع في حالة الإضافة	ة: تاء التأنيث
	رسالةٌ	#	رسالة	مصدر أصلي	فَعالَة	رسل	مفرد مؤنث مرفوع نكرة	ة: تاء التأنيث
	رسالة	#	رسالة	مصدر أصلي	فَعالَة	رسل	مفرد مؤنث منصوب في حالة الإضافة	ة: تاء التأنيث

Fig5. Results of analysis of words of the phrase "كتب الولد رسالة" non vocalized

الدخل INPUT	الخرج OUTPUT						
	الكلمة المشكولة Voweled Word	السابق Prefix	الجذع Stem	نوع الكلمة Type	الوزن Pattern	الجذر Root	الحالة الإعرابية POS Tags	اللاحق Suffix
كتب	كتبَ	#	كتب	فعل ماض مبني للمعلوم	فعَلَ	كتب	ثلاثي مجرد مسند إلى الغائب(هو) متعد ولازم	#
التلميذ	التِّلميذ	ال: التعريف	تلميذ	اسم جامد	فِعْلِيل	تلمذ	مفرد مذكر	#
رسالة	رِسالةُ	#	رسالة	مصدر أصلي	فِعالَة	رسل	مفرد مؤنث منصوب نكرة	ة: تاء التأنيث

Fig6. Results of analysis of words of the phrase "كتب الولد رسالة" partially vocalized

As already mentioned, analyzer can intervene with the learner or tutor at different levels of learning. It represents a decomposition tool for analysis and research in Arab texts:

- Phrases decomposition: This is the decomposition of the text into sentences.
- Partial decomposition: This is the decomposition of text graphics Arab word.
- Deep decomposition: Decomposition of each word in the text elementary constituents (proclitics, prefix, base, suffixes and enclitic).
- Decomposition by determining the formal characteristics of Arabic words. The formal characteristics represent grammatical properties such as category (verb or noun), Gender (male or female), number (singular, dual or plural), person (1st, 2nd or 3rd), the time (does, in does or imperative). Or as lexical type (derived fixed or exceptional), root and the schema.
- The conjugation of verbs.
- The use of the database to find the word definitions.

Figure 7 shows the analysis and labeling of the sentence "**كانَ المَكَّانُ يَعُجُّ بالحاضرين.**"partially vocalized using morphological Analyzer AMIA (K.NADIA & S.DALILA, 2006).

Fig 7. Results of analysis of the sentence "**كانَ المَكَّانُ يَعُجُّ بالحاضرين.**" partially vocalized

IV. PARSING USED BY E-ASSESSMENT SYSTEM

At the base of the formal specification and design of a parser is the problem of choosing a suitable formalism to represent and describe different syntactic structures. The choice of formalism is particularly important since it determines to a very precise way the architecture of the parser, and influence on the degree of coverage of the grammar to develop. The objective of grammatical formalisms is firstly to describe the natural languages to the set of possible sentences of a language, and secondly, the structural properties of these sentence (syntax) and their meaning (semantics). In the vast majority of linguistic formalisms, syntactic structures consist of hierarchical descriptions of phrases (ie group of varying size words). It is for this reason that the result of the analysis is usually a tree, where the various basic syntactic units are the nodes and hierarchical relationships are represented by branches. We are based on HPSG (Head driven Phrase Structure Grammar). In this model, the grammar is reduced to a static classification system and expresses the constraints on language information (A.Abdelkarim, D.Souilem ,2014)[2]. These constraints are represented as hierarchical structures and typed feature. They are interpreted as partial descriptions of linguistic objects

The interest of this grammar is threefold: first, as a formal system, it facilitates the definition of analysis algorithms; second, it enables the integration of heterogeneous resources, both because they use the same type of representation (feature structures) and because it allows the smooth accumulation without imposing constraints of order of application; third, it establishes strict separation between language and data processing programs.

We present the parsing used by e-assessment system to identify the different forms of simple Arabic sentences. The method we propose requires three phases: the division of the sentence into words, the loading of Structures Attributes / Value (SAV) words and parsing itself. Figure 8 shows the different steps of the parsing process continued.

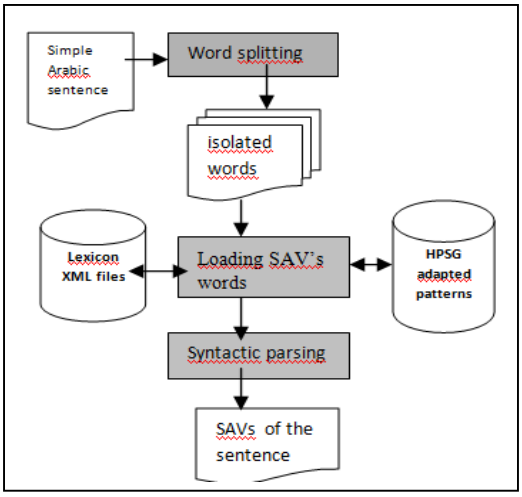

Fig8. The proposed Parsing approach

The cutting phase is to cut the sentence uses words as indicators of separation, spaces, some pronouns and some prepositions. This first phase is required to load subsequently, Structures Attributes / Value (SAV) words. So the loading phase can take all the information about each word within the sentence as SAVs. This information is stored in XML files forming the lexicon of the morphological analyzer.

We opted for XML in order to be able to structure, standardize and normalize the information used by the NLP application. The lexicon is constituted of a set of XML documentations. Every document concerns a category: Noun, verb, adjective, pronouns, particle and preposition. We extracted a corpus of sentences of Arabic grammar books through literary texts for the pupils of seventh year of the Tunisian basis teaching.

Phase parsing itself is finding the syntactic representation of a sentence using the unifying mechanism and patterns changed to Arabic. The use of these schemes is to select the appropriate scheme to build parents phrases attribute values structures of each word of the sentence. Indeed, the rule change requires a modifier always share the same features of the head, the rule specification requires that head traits (attribute) are transmitted to specifies (topical). The construction of the structure attribute / value of the sentence is based on the use of "Son-Father Heritage" principle guarantees that son features, regardless of the class will be copied to the parent phrase. Parsing a sentence will be the result of a successful unification of syntactic lexical information associated with the words in the sentence. This unification allows achieving the construction of a SAV. In addition we have defined some priority of the rules on the merger.

1) *Implementation and system testing*

Performed To evaluate the parsing system, we extracted a corpus of 200 of nominal sentences Arabic grammar through literary texts for the seventh grade students of basic education in Tunisia. These sentences belong to different types of simple sentences discussed above. Each sentence is consisting of two to six words.. Therefore, the analysis allows us to reduce the execution time as we call once the module for the actual parsing. The system retains the semantics of the sentence that may be lost during decomposition proposals. By working on a well-defined body of sentences allowed us to rule in some ways the problem of ambiguity. The results we have obtained are satisfactory for most examples.

For advanced assessment activities such as detection of gender agreement errors, construction of the syntactic structure of the sentence or syntactic categorization, the use of a parser is a requirement.

For these types of evaluation exercises ("صندقة"), we present three forms:

- The first asks the learner to determine the grammatical category of a selected word in a given sentence

Fig9. Grammatical category of selected word in a sentence

- The second type of exercise asks the learner to determine grammatical categories of all words of the sentence in the form of a box prepared by the system as shown in the following figure

Fig10. Grammatical category using prepared form of box

- Third type asks learner to determine the type of sentence (noun or verbal sentence). Then, cut the sentence syntagms (functional units) and determine the grammatical function of each syntagm. Then find the grammatical function of each constituent word of the sentence analyzed. The learner has a toolbox. Whenever it can drag and drop the tool to form the case and thus analyze the sentence. The following figure shows an example of an exercise of this type.

Fig11. Grammatical function using toolbox

Note that the failure of parsing can be explained by the following reasons: a compound word or unplanned grammatical structure proposed in the grammar or the word does not exist in the lexicon of the analyzer. Identifying the causes blockage or failure of the system will, firstly, to update the rules of grammar and, secondly, to define the new linguistic phenomena.

- It gives the sentence to be analyzed and the boxes and it is the learner to complete the holes.
- It gives the sentence to be analyzed and it is the learner to determine the type of sentence (single noun phrase or verbal sentence). Then, cut the sentence phrases (functional units) and determine the grammatical function of each phrase. Then find all the grammatical function of each constituent word phrase to analyze.

In the latter case, if the words of the sentence are completely or partially vowels, the results are correct. Note that all the types of exercises are specified according to the IMS QTI standard .(IMS, 2014) using variables and constraints, and this to ensure the interoperability and reusability.(A.ABDELKADER, D. SOUILEM, 2014)[1].

To improve coverage and robustness of syntactic system, primarily for unvocalized sentences, many difficulties are resolved. They particularly concern is the linguistic aspects, particularly with regard to the description, or aspects related to the implementation.

- Linguistic aspect: one of the fundamental problems of the syntax is that it is not possible to define a grammar covering the whole of a language. However, the increase in coverage analyzers involves taking into account most varied phenomena. The problem becomes that of the consistency of the developed grammars. The description of a particular phenomenon should not be using ad hoc mechanism. For example, the order of words in the sentence Arabic, the inversion phenomenon of topical positions and attribute, the interpolations, elliptical structures, etc. require extensive language skills.
- To carry out the implementation, we can mention two types of problems. The first concerns the generality and reusability of the analysis process. To meet this goal, the implementation should be as close as possible to the theoretical model and we need the tools used for implantation are also generic. The second problem is the treatment of categorization ambiguity. Support disambiguation mechanism by the analyzer remains a great challenge: an efficient parsing system must recognize words and assign the class or the possible syntactic categories. For this, we have implemented heuristics that reduce the categorization ambiguity. Thus, the resulting system is able sometimes to distinguish the adjective happy سعيد, the name سعيد or the verb name. The specific problem related to the treatment of names is that it is hardly possible to build a comprehensive glossary containing all occurrences: its size will be huge (all names and personal names, regions, countries, cities, villages, company names, products ...). In addition, certain proper names are homonymous with common words.

It is therefore necessary to distinguish compound words free sequences of simple words. The latter cannot be merely formal; moreover, the vast majority of compound words (and fixed expressions) are written without characteristic composition mark (hyphen or apostrophe). Their definition is syntax and semantics: their syntactic and semantic properties cannot be deduced from those of their constituents, unlike those of free sequences. The identification of compound words is similar to that of single words. For an automatic analysis system to recognize compound words or frozen expressions of a text, it is necessary to build a lexicon containing all the compound words and idioms (more simply poly lexical units) of the Arabic language. This obviously implies a working gigantic census.

V. CONCLUSION

In this paper and in the context of development and testing of language tools to integrate in an online assessment environment of the Arabic language, we identified the characteristics of simple Arabic sentence. Then, we have shown the contribution of the use of HPSG formalism for building a lexicon for Arabic and analyzing of simple sentences. Our e-assessment uses results of two morphological analyzers, "Alkhalil Alsarfi" and AMIA. Then we discussed the different types of assessment activities that can be achieved using a syntactic parser. As perspectives, we propose to extend our e-assessment system by adding other standardized types of activities and designing other NLP tools and resources.

REFERENCES

[1] A.ABDELKADER, D. SOUILEM BOUMIZA, R. BRAHAM, 2014. «E-assessment System Based on IMS QTI for the Arabic Grammar». International Journal of Advanced Computer Science and Application, IJACSA, Vol. 5, No.10, 2014. pp 92-100.

[2] A. ABDELKARIM, D.SOUILEM, R. BRAHAM, 2014. «محلل نحوي لبيئة تعلم عن بعد للغة العربية». Communications of the Arab Computer Society CACS, Vol. 7 No.1

[3] A.ABDELKARIM & H. KAIS & B.H. ABDELMAJID 2006. «Etude et analyse de la phrase nominale arabe en HPSG». Traitement Automatique des Langues Naturelles : TALN 2006, Université catholique Louvain (UCL), du 10 au 13 avril 2006 à Leuven.

[4] A. ABDELKADER, D. S. BOUMIZA, AND R. BRAHAM. 2009. «A categorization algorithm for the Arabic language». International Conference on Communication, Computer and Power (ICCCP'09), Muscat, February 2009.

[5] ALKHALIL MORPHO SYS, «A Morphosyntactic 2010 analysis system for Arabic texts».

http://www.alecso.org.tn/index.php?option=com_content&task=view&id=1302&Itemid=998&lang=ar. Retrieved February, 2014, from ALECSO.

[6] C.POLLARD & I.SAG, 1994, «Head-Driven Phrase Structure Grammars». Chicago University Press.

[7] D.KOULOUGHLI 2002. Grammaire de l'arabe d'aujourd'hui. BRODARD & TAUPIN.

[8] F.POPOWICH, C.VOGEL 1990, «CHART PARSING HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR». TECHNICAL REPORT CSS-IS TR 90-01, SIMON FRASER UNIVERSITY.

[9] INSTRUCTIONAL MANAGEMENT SYSTEM (IMS). 2014. «IMS Question & Test Best Practices and Implementation Guide».

http://www.imsproject.org/question/index.html(Accessed on Junury . 20, 2014)

[10] I.SAG & T.WASOW 1999, «A formal Introduction». Center for the Study of Language and Information. CSLI publication, Stanford University.

[11] K.NADIA & S.DALILA 2006, «Pré analyse du mot arabe basée sur une approche de filtrage pour une analyse aorphologique». Les sixièmes journées scientifiques des jeunes chercheurs en Génie Electrique et Informatique : GEI06.