Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DESIGN AND OPTIMIZATION OF AN XML-BASED GRAMMATICAL STRUCTURE OF THE HOLY QURAN

[1]AYOUB KADIM, [2]AZZEDDINE LAZREK
Department of Computer Science, Faculty of Science, Cadi Ayyad University
P.O. Box 2390, 40000, Marrakech – Morocco
E-mail: [1]a.kadim@uca.ma, [2]lazrek@ucam.ac.ma

ABSTRACT

The Holy Quran is considered one of the most studied books along the human history. In fact, until now, scholars and researchers continue to treat it from various specialties: Quranic jurisprudence, rhetoric, grammar, etc., what is called the Quranic sciences. In the same way, we aim through this work to add a new contribution by designing a model that can absorb and organize all the information relating to the Quranic grammar. So, after studying Arabic grammar and Quranic structure, we proposed a methodology to classify Arabic grammar units and index Quranic elements to have an optimal and rich structure. The proposed model is based on XML language for more flexibility and compatibility with various systems and to more benefit from it in Quranic and Natural Language Processing researches. And finally, we discuss contributions, limitations, applications and prospects of this work.

Keywords: *Holy Quran, Quranic sciences, Arabic Language, Grammatical analysis, Natural Language Processing, XML language*

1. INTRODUCTION

1.1 Arabic Language

Arabic is one of the oldest human languages [1] and most consistent and organized ones. Indeed, the ancient Arabs did not have a material civilization as known in other civilizations as Babylonians, Romans and others. In fact, Arabs civilization was manifested in their language [2] which had a very important influence in their lives. This was due to the isolated desert environment in which they lived [3] and their simple Bedouin life which made them away from the effects of other civilizations [4]. This has been a major factor in protecting the Arab tongue from deviation and influence by non-Arabic languages. So their creative energy was focused on poetry and literature, and were known by eloquence, rhetoric and powerful memorization; the majority of them did not know reading and writing. With the coming of Islam, scholars and specialists in Arabic sciences had put several laws and rules to regulate the Arabic language and facilitate its learning.

1.2 Quranic Grammar

First of all, the Holy Quran was destined to the eloquent Arabs. So, to better understand the meaning of the Quranic texts it is necessary to have a good knowledge of the Arabic language [5]. Hence came the importance of Arabic language studies for Quranic studies especially the traditional Arabic grammar which processes the states of the Arabic word extremities in the composition case [6]. There are several features and characteristics that distinguish the Quranic text from the other Arabic texts, especially in the grammatical level; as *stopping* and *starting* (الوقف والابتداء) [7], using a variety of Arabic *versions* (لغات العرب) other than *Quraish*'s one[1], using *turning style* (أسلوب الالتفات) [9], etc.

From these characteristics, we can touch the contributions and additions of the Holy Quran for the Arabic language in rhetoric, grammar [10] and other language issues. By dropping this on digital processing level, we can say that processing the Quranic grammar digitally can have an

[1]The Arabic language version dominant on the Holy Quran is the language version of *Quraish* (قريش). Tribe of *Quraish*, due to their proximity of the *Kaaba*, receives the other Arab tribes coming for Pilgrimage, having different language versions. So, they select and choose the most eloquent expressions to adopt it and use it [8].

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

important contribution in the Arabic Natural language Processing (NLP).

1.3 Related Works

By tracking the works already done in the Quranic grammar processing, we can classify it according to its depth in the digital grammatical analysis of the Quranic texts.

1.3.1 Electronic books

The shallowest works in digital Quranic grammar processing are the electronic books talking about Quranic grammar, which are electronic copies manually entered and organized according to the author index. Hence, they do not necessarily respect the classification of verses. For example, the book *Ierab Al Qor'an* (إعراب القرآن للزجاج): Grammatical Analysis Of The Quran (by *Azzajjaj*), in the Shamela[2] library, gives a variety of grammatical cases as titles, and then for each case gives the corresponding Quranic texts; as in section XV: What came in the Quran with deleting the preposition and the genitive (ما جاء في التنزيل من حذف الجار والمجرور). There are also separate e-books as light software to launch in the computer. For example, *Ierab Al Qor'an* (by *Ibn Sidah*) is an e-book classified and arranged according to the verses classification. Within the same classification: *Mushkil al Qur'an* (مشكل القرآن لمكي): which confused in the Holy Quran (by *Makky*), and with the same title but more detailed (by *Al Kharrat*), but these last two e-books focus just on the confusing grammatical cases.

1.3.2 Websites

Then, we find some web interfaces providing easier navigation, to access directly to the desired verse and view the related grammatical analysis. For example, there are the *Mosshaf* site[3] and *Altafsir* site[4] and similar sites dealing with the Quranic sciences.

1.3.3 Software

There is another category of related works which are programs to install in the computer, providing to the user browsing and selecting the desired verse and word; as the software *Ierab Al Qor'an Al Karim* (إعراب القرآن الكريم) of the RDI Company. And here we move from the verse level to a deeper level: the words. Indeed, the program addresses the grammatical analysis of the Quran word by word, and displays the grammatical analysis of a word by clicking on it.

1.3.4 Using Syntactic Treebank

The deeper work in the digital Quranic grammar processing is the Quranic Arabic Corpus[5] project where the analysis comes down to the level of word parts and the grammatical relations linking word parts, words and sentences basing on a Syntactic Treebank [11].

2. CONCEPTION

2.1 Classification

When we want to computerize the Quranic grammar, first thing, we started looking at the traditional Arabic grammar through the computer systems design. We found that we can classify the grammatical information in such a way to have a smaller and more organized model. We analyzed the expressions of traditional Arabic grammar and its elements in order to divide them into units. Each unit will be classified and coded according to its possible values.

Let's take the following example:

سعى :فعل ماض مبني على الفتح المقدر على الألف تعذرا

s'a (try): past verb Indeclinable in the accusative marked by the vowel *a* hidden at the end because of the impossibility of pronunciation[6]

We notice so, that we can consider several units (underlined) such that each unit can have different values. And that will be the basis of the adopted classifications.

2.1.1 Nature of the word

In the expression of traditional Arabic grammar, we start with the description of the nature of the concerned word (noun, verb or particle) and then we describe its grammatical state; for example: a preposition (حرف جر) i.e. that puts the following word in genitive; a conditional noun (اسم شرط) i.e. it expresses the condition; an imperative verb (فعل أمر) i.e. verb in the imperative mood, etc. For coding the elements, we have created a code composed of three numbers: the first for the nature of the word, the second for its grammatical state and the third to distinguish the category of the word between the categories having the same nature and the same grammatical state (See table 1).

Table 1: Classification Of The Word Nature And Its Grammatical Class

Word nature طبيعة	Grammatical state الحالة الإعرابية	Grammatical class الصنف الإعرابي	Code الرمز

[2]http://shamela.ws/index.php
[3]http://www.mosshaf.com
[4]http://www.altafsir.com
[5]http://corpus.quran.com
[6]The translation of the Arabic grammatical terms was based on Pierre Cachia's dictionary [12] and the William Wright's book [13].

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

الكلمة			
Noun اسم	Nominative رفع	First name in the nominal clause مبتدأ	111
		Second name in the nominal clause خبر	211
		Subject فاعل	311
		…	…
	Accusative صب	Object مفعول به	121
		…	…
	…	…	...
Verb فعل	Past ماض …	Family of "Kana" (was) أخوات كان	112
		Family of "Thanna" (think) أخوات ظن	212
		…	…
	Present مضارع	Family of "Kana" (was) أخوات كان	122
		…	…
	…	…	...
Particle حرف	Genitive جر	Preposition حرف جر	113
		Jurative particle حرف قسم	213
		…	...
	Accusative صب	Subjunctival particle حرف صب	123
		Subjunctival and negative particle حرف صب ولم	223
		...	...
	…	...	...

2.1.2 Grammatical mark

In the traditional grammar expression, after describing the word nature and its grammatical class, we move to clarify its grammatical mark i.e. how the grammatical state is manifested. Indeed, for the Arabic words, the grammatical state is manifested either by vowels (إعراب بالحركات) or letters (إعراب بالحروف); and sometimes does not manifest (إعراب مقدر). And we distinguish between Declinable words (الكلمات المعربة) and Indeclinable words (الكلمات المبنية) in the expression. For example, we say for Declinable words: word in nominative (مرفوع) and we say for Indeclinable ones: *fixed in nominative* (مبني على الرفع) (See table 2).

Table 2: Classification Of The Grammatical Marks

Grammatical mark علامة الإعراب	Inflection الإعراب	Grammatical state الحالة الإعرابية	Vowel الحركة	Code الرمز
Word in nominative marked by the vowel *u* at the end مرفوع بالضمة الظاهرة على آخره e.g. وَقَالَ **الرَّسُولُ**	Declinable معرب	Nominative رفع	An apparent **vowel**: *u* حركة ظاهرة: الضمة	111
word in genitive marked by the letter *Yā* (ي) مجرور بالياء e.g. قُل **لِّلْمُؤْمِنِينَ**	Declinable معرب	Genitive جر	A **letter** replacing the vowel: the *Yā* replacing the vowel *i* حرف نائب: الياء النائبة عن الكسرة	631
Word fixed in accusative marked by the vowel *a* hidden at the end because of the difficulty of pronunciation مبني على الفتح المقدر على آخره استثقالا e.g. مَنْ **أَعْطَى** وَ**اتَّقَى**	Indeclinable مبني	Accusative صب	A **hidden vowel**: because of pronunciation difficulty حركة مقدرة: استثقالا	322
...	...	...	...	...

2.1.3 Grammatical relation

We clarify the grammatical relations linking the word in question with other words or sentences. For example, in the sentence:

(إن الله بكل شيء عليم) الأنفال 75

(*Verily Allah is well-acquainted with all things*)

The spoils of War 75 [14]

We say that the sacred word "الله" is the Noun of '*inna* (اسم إنَّ), and the word "كُلِّ" is annexed to (مضاف إلى) the word "شيءٍ", and all "بِكلِّ" (بـ + كلّ) is dependent to (*متعلق بـ*) the word "عليم", etc. So, all these expressions (Noun of, annexed to, dependent to, etc.) can be considered as grammatical relations that can also be classified and coded (See Table 3).

Table 3: Classification Of The Grammatical Relations

Grammatical relation العلاقة الإعرابية	Code الرمز
Agent (with the verb) فاعل (مع فعل)	1
Predicate (with the Inchoative) خبر (مع مبتدأ)	2
Condition clause (with the condition particle) جملة الشرط (مع أداة شرط)	3

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

...	...

2.1.4 Pronoun

Then we found that we can classify pronouns, given their high frequency of use in the Arabic language. So we found that they can be classified according to type, reference, number and gender (See Table 4).

Table 4: Classification Of The Pronouns

Pronoun الضمير	Type النوع	Refers to العائد عليه	Number العدد	Gender الجنس	Code الرمز
me أنا	Separated منفصل	Speaker متكلم	Singleمفرد	-	111
they: fem. هن	Separated منفصل	Absent غائب	Plural جمع	Feminine مؤنث	2331
me ـتُ	Linked متصل	Speaker متكلم	Single مفرد	-	112
you: dual أنتما	Separated منفصل	Interlocutor مخاطب	Dual مثنى	-	221
...	...	...	...	...	...

We can say that this method of coding will make the model content more semantic. Indeed, we can, for example, determine the tense of a verb by checking the second number in the word indices beginning with 2. and it also responds to the way of expressing the traditional Arabic grammar analysis, where often the expression is not very detailed; For example, in the preceding word "سعى" we can find instead of the expression: " فعل ماض مبني على الفتح" this expressions: " فعل ماض مبني على الألف تعذرا المقدر" or even "فعل ماض" , "على الفتح" (as in the word "سَعَيْنَا" it is said: "فعل وفاعل"). So the model will provide the requested information according to all the analysis level, by allowing the extraction of the desired information via its corresponding number in the code, without addressing the other details.

2.2 Indexation

When we talked about the grammatical relations, we said that the word can have a relation with a word or a sentence, and that the same word can have multiple relations. Let's take the following example:

(مَنْ يَهْدِ اللهُ فَهُوَ الْمُهْتَدِ) الكهف 17

(*He whom Allah guides is rightly guided*)

The Cave 17 [14]

The word "يَهْد" (which is a verb) has a grammatical relation with the word "مَن" (particle) and the nature of this relationship is: *the condition verb* (فعل الشرط). The same word "مَن" has a relation but this time with the sentence "هو المهتد" and the nature of this relationship is: *the Result depending upon the condition* (جواب الشرط). Thus we will need a proper indexation, from which we can locate the words and sentences in the Holy Quran in order to define the elements of grammatical relations.

We then adopted the following indexation:

Chapter , *Verse* , *Sentence* , *Word* , *Word part*

سورة ، آية ، جملة ، كلمة ، جزء كلمة

Where:

- *Chapter*: the chapter ranking relative to the other chapters of the Quran.
- *Verse*: the verse ranking compared to the other chapter verses.
- *Sentence*: the sentence ranking inside the verse.
- *Word*: the ranking word inside the sentence.
- *Word part*: the word part ranking (if exists) inside the word, relative to the other word parts.

2.2.1 Indexation levels

If we put 0 in an indexation level, the resulting index will point on all of the elements belonging to this level (See Table 5). For example, let's take this verse:

(أَفَحَسِبْتُمْ أَنَّمَا خَلَقْنَاكُمْ عَبَثًا وَأَنَّكُمْ إِلَيْنَا لَا تُرْجَعُونَ)

المؤمنون 115

(*Did ye then think that We had created you in jest, and that ye would not be brought back to Us (for account)?*) The believers 115 [14]

Table 5: Explanation Of The Indexation Levels

Index الفهرس	Significance of the index دلالة الفهرسة				
	Chapter سورة	Verse آية	Sentence ملة	Word كلمة	Word part جزء الكلمة
23,0	The believers المؤمنون	All verses جميع الآيات	-	-	-
23,115,0	The believers المؤمنون	Verse 115 الآية 115	All sentences جميع الجمل	-	-
23,115,2,0	The believers المؤمنون	Verse 115 الآية 115	Sentence 2 الجملة الثانية	All words جميع الكلمات	-
23,115,2,1,0	The believers المؤمنون	Verse 115 الآية 115	Sentence 2 الجملة الثانية	Word 1 الكلمة الأولى	The entire word الكلمة كلها
23,115,2,1,2	The believers المؤمنون	Verse 30 الآية 115	Sentence 2 الجملة الثانية	Word 1 الكلمة الأولى	word part 2 جزء الكلمة الثاني: "أَنَّ"

The figure 1 shows more the above indexation elements.

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

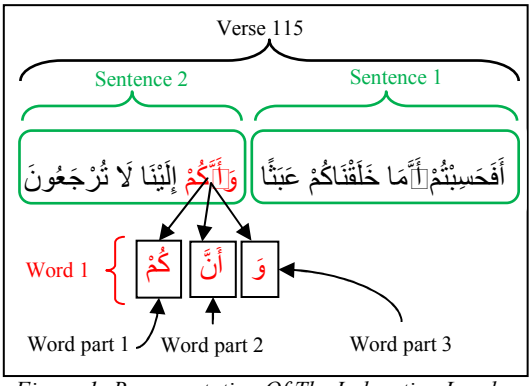

Figure 1: Representation Of The Indexation Levels

2.2.2 Indexation of the hidden elements

We often find that the specialists in Arabic grammar introduce in their grammatical analyzes some estimated words or sentences that do not appear in the text. These hidden elements are used to complete the context, so they will be considered as the other items and will have the same grammatical characteristics. Then we need to index them so as to distinguish between the original text and what is estimated by the linguist.

Hidden elements differ according to the number and the nature. For example, in the sentence " فَإِنَّ الْعِزَّةَ لِلَّهِ جَمِيعًا" *(all honour be to Allah)* is estimated a verb, such as: استقرت (*establish*) or a name as: مستقرة (*established*) (See Figure 2).

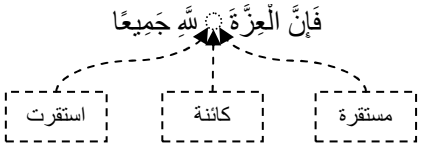

Figure 2: Example Of Diversified Hidden Elements

They may also differ according to the location, as the estimation of أبتدئ (*I start*) or ابتدائي (*my beginning*) in البسملة (the first verse of the Holy Quran) (See Figure 3).

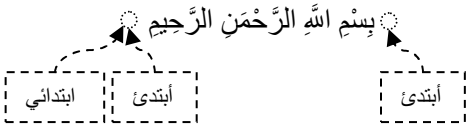

Figure 3: Example Of Hidden Elements In Different Locations

They may also be included inside a word, as the estimation of the hidden word أنْ (أن المضمرة) in the word وَلِيَتَذَكَّرَ (See Figure 4).

Figure 4: Example Of A Hidden Word Inside A Word

We can also find successive hidden elements in the same location, as in this sentence:

(قال سلام قوم منكرون) الذاريات 25

(he said: 'peace'! - people unknown)

Adh-Dhariyat 25 [15]

The estimated meaning is (See Figure 5):

قال سلام عليكم أنتم قوم منكرون

(*he said: 'peace [upon you]'! - [you are a] people unknown*)

Figure 5: Example Of Successive Hidden Items

It is then necessary to take into consideration all possible cases in order to design proper indexation to these hidden elements.

- *Indexation of the hidden elements*

To locate the place where we want integrate the hidden element, we adopted the following syntax:

The index of the preceding element "," *the hidden element ranking relative to the other hidden elements* (if there are several ones).

Thus, the indexes of the previous examples will be as followings: Sentence (جملة)

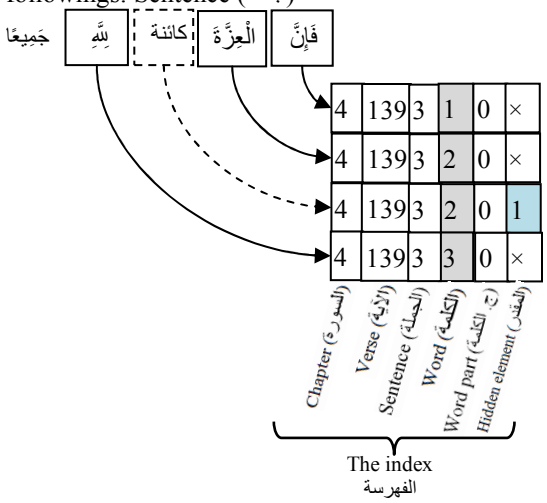

Figure 6: Indexation Of A Simple Hidden Element

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

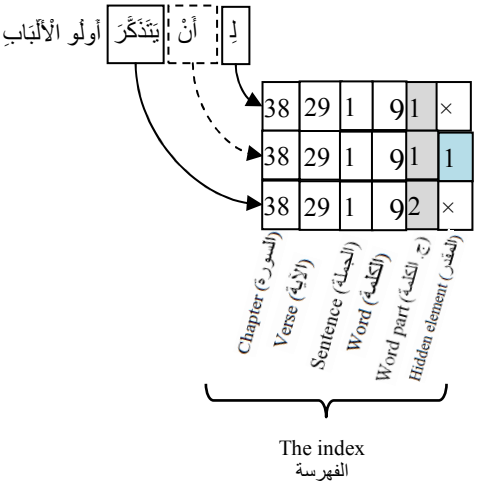

Figure 7: Indexation Of A Hidden Element Inside A Word

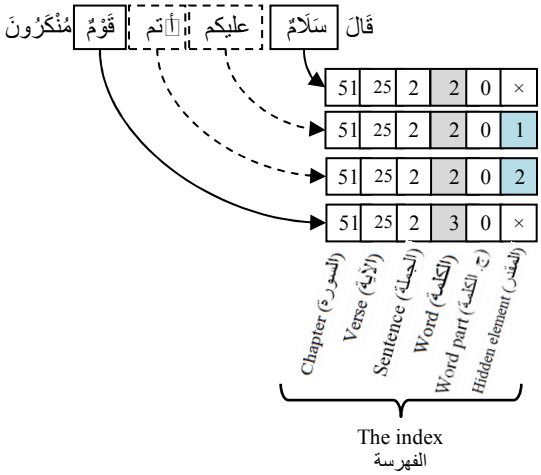

Figure 8: Indexation Of Successive Hidden Elements

3. THE XML MODEL

3.1 Overview

XML is a markup and hierarchical language. It is a kind of simplified databases where the XML file consists of tags identifying the data elements (e.g. Chapter, Verse, Word, etc.) beginning with "<item>" and ending with "</item>". Inside each element, there are some entities determining the properties of the element (for example, Number, Nature, etc.). The hierarchy appears in the fact that the parent tag contains the child tags that can also contain children, etc.

The Figure 9 shows an overview of the model.

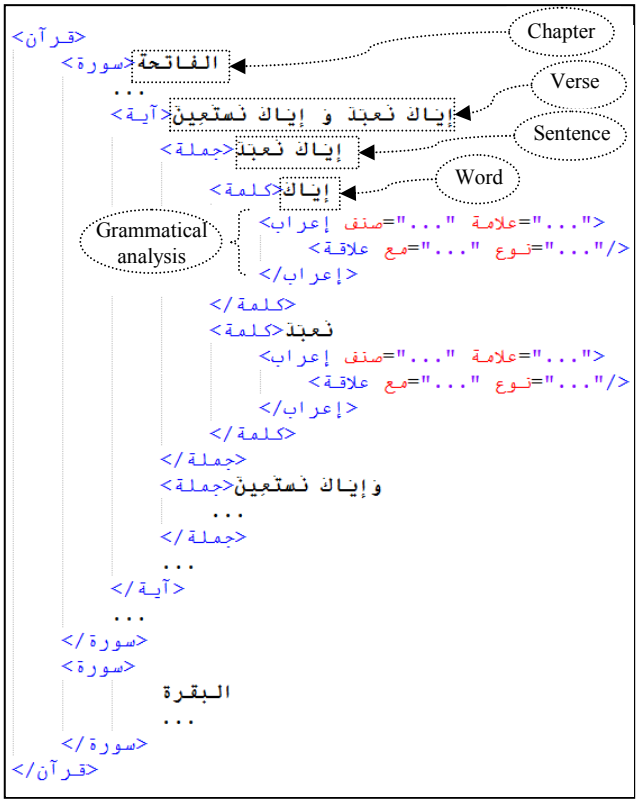

Figure 9: Overview Of The XML Model

- Tag سورة (Chapter), contains the tags آية (Verse) which concern the chapter verses.
- Tag آية (Verse) contains one or more tags جملة (Sentence), considering that at the beginning each verse consists of a single sentence.
- Tag جملة (Sentence) contains the tags كلمة (Word). Words are defined by splitting the sentence by spaces.
- Tag كلمة (Word) can contain zero or more tags جزء كلمة (Word part) considering prefixes and suffixes.

Concerning the tag attributes, they contain grammatical information that we have already classified:

- Attribute صنف (Class): contains the code concerning the word nature and its grammatical class.
- Attribute علامة (Mark): contains the code concerning the grammatical mark.

Then we define the grammatical relation(s) with the other elements. And from the fact that we can have several relations for the same element, we considered the element علاقة (relation) as an independent tag being a child of the element إعراب (grammatical analysis). The element علاقة (relation) has the attributes نوع (type) containing the code of the relation type and مع (with) containing

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the index of the other element involved in the relation.

3.2 Grammatical analysis according to the elements

In the traditional Arabic grammar we talk about the grammatical analysis of words إعراب الكلمات (or word parts) as well as the grammatical analysis of sentences إعراب الجمل, considering the sentence –grammatically- as one word (جملة لها محل من الإعراب). So, we have to foresee the tag إعراب (grammatical analysis) inside the tags: Sentence, Word and Word part.

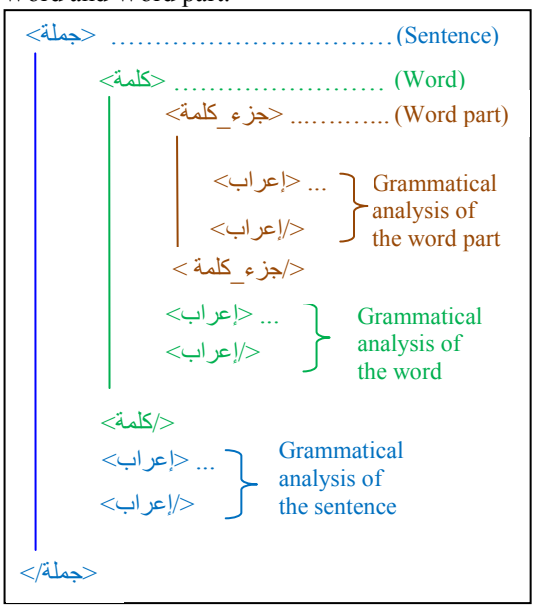

Figure 10: Grammatical analysis according to the elements

In the figure 11, we give an example of the grammatical analysis of the sentence وإياك نستعين and the grammatical analysis of the element و that is part of the word وإياك.

```
<جملة>وإياك نستعين
    <كلمة>وإياك
        <جزء_كلمة>و
            <إعراب صنف="code(حرف عطف)"></إعراب>
        </جزء_كلمة>
        <جزء_كلمة>إياك
            ...
        </جزء_كلمة>
    </كلمة>
    <كلمة>نستعين
        ...
    </كلمة>
    <إعراب صنف="code(لا محل لها من الإعراب)">
        <علاقة مع="index(إياك نعبد)"
        نوع="code(عطف جمل)" />
    </إعراب>
</جملة>
```

Figure 11: XML model of the grammatical analysis

3.3 Origins of the word parts

We also added the attribute أصل (Origin), because, in the grammatical analysis, the word parts are called as if they are independent. For example, the word بسم contains two parts: بِ and سْمِ, and in its grammatical analysis we do not say:

("***bi***": Preposition) **بِ**: حرف جر
سْمِ: اسم مجرور بـ"**بِ**"
("***smi***": The genitive governed by "***bi***")

But we say:

("***Bā***": Preposition) **الباء**: حرف جر
اسمِ: اسم مجرور ب**الباء**
("***Ismi***": The genitive governed by the "***Bā***")

Another example for the word قُلْنَا (قُلْ +نَا), we do not say:

("***qul***": past verb) **قُلْ**: فعل ما□ي

But we say:

("***qala***": past verb) **قال**: فعل ما□ي

Thus it is necessary to foresee a morphological analysis process to determine the origin of each word part. To facilitate this task, we can interface a morphological analyzer suggesting the possible origins. For example, we can use the morphological analyzer Alkhalil [16] or Buckwalter [17].

3.4 Versions of grammatical analysis

We often find in the grammatical analysis of the Quranic text that we can have several versions of grammatical analysis, and that these versions do not have the same degree.

For this, we give the opportunity to add several grammatical analysis (إعراب element) for the same entity (sentence, word or word part), giving the attribute وجه (version) indicating the degree of the current version compared to the other versions.

For example, in the grammatical analysis of the word رب in the sentence: "الحمد لله رب العالمين" we can have two versions: بدل (*replacing*) of the holy word "الله" or: نعت□(adjective) of the same word. Then the model will be as in Figure 12.

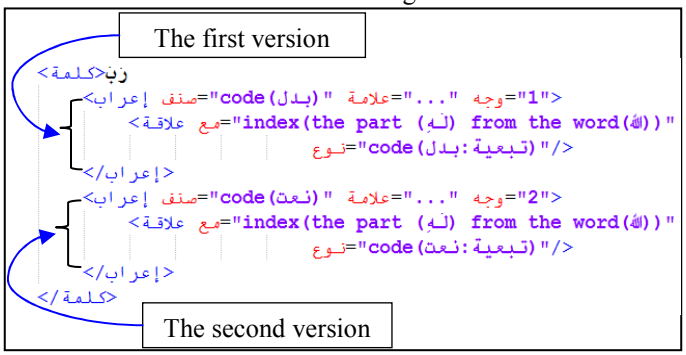

Figure 12: Example of two grammatical analysis versions

Journal of Theoretical and Applied Information Technology
30th September 2015. Vol.79. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.5 Sentences distributed over multiple verses

We can find sentences containing several parts belonging to different verses. For example, we can consider the first three verses as one sentence.

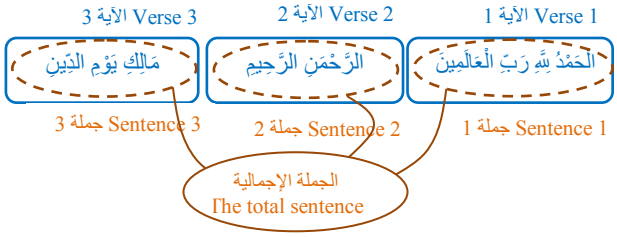

Figure 13: Example of a sentence distributed over several verses

These cases present a problem for the XML hierarchy, where the son element (here the sentence) must be completely included in the father element. For this, we have added an entity (Sequel) in the element (Sentence) that contains the index of the previous part of the total sentence. Thus, the index of the total sentence will be the index of the last part, so as to bring all parts starting with the last index up to the first.

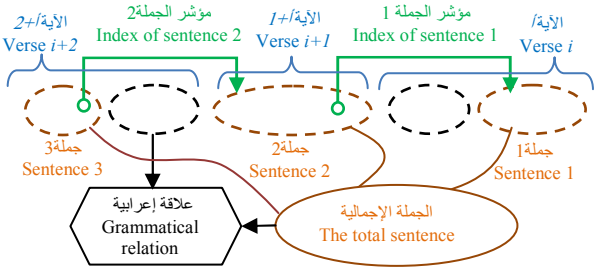

Figure 14: Indexation of a sentence distributed over several verses

To better understand this, let us take a more detailed example: in the verses 43 and 44 of Chapter النحل (Bees):

وَمَا أَرْسَلْنَا مِن قَبْلِكَ إِلَّا رِجَالًا نُّوحِي إِلَيْهِمْ فَاسْأَلُوا أَهْلَ الذِّكْرِ إِن كُنتُمْ لَا تَعْلَمُونَ (43) بِالْبَيِّنَاتِ وَالزُّبُرِ وَأَنزَلْنَا إِلَيْكَ الذِّكْرَ لِتُبَيِّنَ لِلنَّاسِ مَا نُزِّلَ إِلَيْهِمْ وَلَعَلَّهُمْ يَتَفَكَّرُونَ (44)

And before thee We sent none but men, to whom We granted inspiration: if ye realise this not, ask of those who possess the Message (43) *(We sent them) with Clear Signs and scriptures and We have sent down unto thee (also) the Message; that thou mayest explain clearly to men what is sent for them, and that they may give thought.* (44) [14]

The sentence بالبينات والزبر is considered as a sequel of the sentence فاسألوا أهل الذكر and the total sentence فاسألوا أهل الذكر + بالبينات والزبر has the relation عطف الجمل (sentence conjunction) with the first sentence: وما أرسلنا من قبلك إلا رجالا نوحي إليهم. The Figure 15 shows these details.

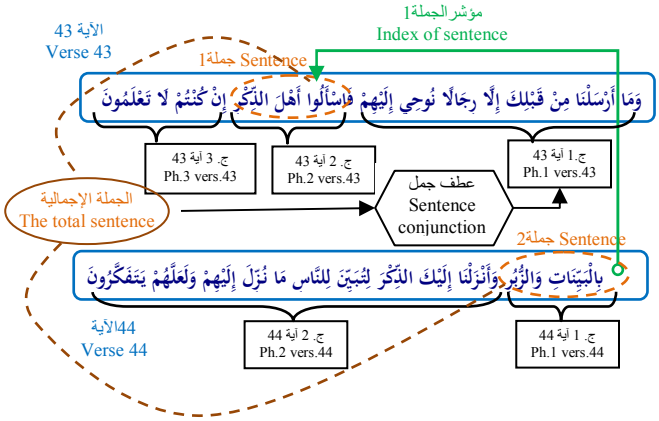

Figure 15: Detailed a sentence distributed over several verses

3.6 The reference of the grammatical analysis

The reference of the grammatical analysis is an important information that the model should also include. For this, we have two choices:

- Either we reserve for each source (book) a separate XML file.
- Or we add an attribute مصدر (Reference) in each element إعراب (Grammatical analysis)

So we adopted the second choice for these reasons:

- There are authors who have not completed the grammatical analysis of the entire Holy Quran.
- There are some grammatical analyzes that are scattered in books that are not specialized in this field (such as interpretations of the Quran)
- We will have a more global view of every word by presenting several analyses.

4. CONCLUSION

4.1 Contributions

From the presented work, we can draw the following contributions:

- It represents a new application of the XML language.
- It provides a rich resource for research related to Natural Language Processing (NLP), especially for the Arabic language whose resources are not enough numerous and available.
- Thanks to the categorization and classification, there will be an optimal data size by substituting long sentences (as in Figure 1) by small numbers; given the fact that we will have as result an XML resource that is at the end a text file that we have the interest to reduce due to the constraints of transfer and processing time. Indeed, this

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

contribution will be more important in the case of large queries that process a massive number of cases, so instead of dealing with long phrases we will have simple numbers.

- The model provides a normalization of the grammatical analysis expressions. Indeed there are often differences between the expressions of the grammarians for the same word (e.g. الما□ي / ماض، الخفض / الكسر، مبني للمجهول / للمفعول / لما لم يسم فاعله, etc.). So the unification of these expressions is a paramount phase to allow the automatic processes such as advanced search, rule extraction, etc. And this can be a first step to a universal standard of Arabic grammar.
- In addition, it will provide several grammatical analysis depth levels, so as to have the required information without browsing other details by checking directly the relevant section in the code.
- It provides the possibility of advanced search; for example, searching the cases where the predicate of كان (*Kana*) is a verbal sentence having a present verb.
- It supports adding multiple versions of grammatical analyzes of the same person.
- In addition to the benefits of XML structure: flexibility, extensibility and compatibility with various systems, etc.

4.2 Limits

Like every new work, we can see the following limitations of our work:

- It requires a community of specialists in Quranic sciences, the Arabic language and XML language to ensure and validate the proposed structure.
- It also requires specialists to delimit the sentences and word parts, and to enter the information related to the grammatical analysis.
- We also mention the limitations of XML format: security, complexity, etc.

4.3 Applications

We can have diverse applications of this work related to the creation of a reference for everything related to the grammatical analysis of the Holy Quran that will be accessible and used by researchers in various entities: Arabic sciences, Quranic sciences, NLP, XML ... As well as developers of web and mobile applications.

4.4 Perspectives

Regarding the present work, we think to introduce future enhancements:

- Create more specialist groups to discuss the validation and improvement of the structure and its components.
- Extend the model to cover the Arabic language.
- Combine this structure with other XML structures concerning the other Quranic sciences (as *Tajweed* التجويد: pronunciation of the Quranic text, *Tafseer* التفسير: explanation of the Quran, etc.).
- For the Quranic versions (Qiraat), we work currently on the available version (Hafs). However, currently, we have the vision to work on other Quranic versions separately, because there are often differences between versions that affect the division of words and phrases which will complicate the model if we want to assemble all the versions into one file.
- We are also working on the design of an advanced web interface containing all the features required to edit the XML model and view its content adequately.
-

REFERENCES:

[1] Thomas Chenery, "The Arabic language, a lecture", 1869.

[2] Nezar Najjar, "The civilization of Arabs is based on their language", 2010.
(□زار□جار، "قامت حضارة العرب على لغتهم"، 2010)
http://fedaa.alwehda.gov.sy/_archive.asp?FileName=5915214832010031422405 6

[3] Dakhil Hamad Ben Nasser, "The impact of the desert in the genesis and evolution of Arabic poetry until the end of the Abbasid era", 2013.
(دخيل حمد بن□اصر. "أثر الصحراء في□شأة الشعر العربي وتطوره حتى□هاية العصر العباسي"، 2013.)

[4] Gustave Le Bon, "La civilisation des Arabes", *Librairie de Firmin-Didot*, 1884.

[5] Roubhi Lakhdar, "The importance of the grammatical example in the Holy Qur'an interpretation, the interpretation (inclusive statement of Ibn Jarir al-Tabari) as a model", *Al-Athar - Journal of Arts and Languages*, University of KasdiMarbah- Ouarqalla, Algeria, Vol. 6, 2007.
(روبحي لخضر، "أهمية الشاهد النحوي في تفسير القرآن الكريم، تفسير (جامع البيان لابن جرير الطبري)□موذجا"، الأثر- مجلة الآداب واللغات، جامعة قاصدي مرباح- ورقلة- الجزائر، العدد السادس، 2007.)

[6] Al-Sabban, "Sabban's Footnotes on Al-Ashmouni explanation of Alfiat Ibn Malik, arranged by Mustapha Houssain Ahmed", *Dar Al-Fikr*, Vol. 1, pp. 18.
(حاشية الصبان على شرح الأشمو□ى لألفية ابن مالك، □بط وترتيب مصطفى حسين أحمد، دار الفكر، المجلد الأول، ص. 18.)

[7] Ibn Al-Jazary, "Al-nachr Fi Al-Qiraeat Al-achr, corrected and revised by Ali Mohammed Dabbae", *Dar Al-Koutoub Al-ilmiah*, Vol. 2, pp. 224.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(ابن الجزري، "النشر في القراءات العشر، تصحيح ومراجعة علي محمد الضباع"، دار الكتب العلمية، الجزء الأول، ص. 224.)

[8] Hanna Al-Fakhouri, "Inclusive in the history of Arabic literature", *Dar Al-Jalil*, Beirut, 1st Edition, 1986, pp. 52.
(حنا الفاخوري، "الجامع في تاريخ الأدب العربي"، دار الجليل، بيروت، الطبعة الأولى، ص. 52، 1986)

[9] Ghalib Ben MohammedAl-Hamidi, "Method of Changing Context, and its types, in the Holy Quran", *Conference of University of Al-Minia*, Egypt, 1429 AH.
(غالب بن محمد أبو القاسم الحامضي، "أسلوب الالتفات وأقسامه في القرآن الكريم"، مؤتمر جامعة المنيا، مصر، 1429 هـ)
http://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/4050050/gh4.doc

[10] Abdeleal Salem Makram, "The Holy Quran and its impact on the grammatical studies", *Ali Al-Jarrah institution*, 2nd Edition, 1978, pp. 45.
(عبد العال سالم مكرم، "القرآن الكريم وأثره في الدراسات النحوية"، مؤسسة علي الجراح الصباح، الطبعة الثانية، 1978، ص. 45.)

[11] Kais Dukes and Tim Buckwalter, "A Dependency Treebank of the Quran using Traditional Arabic Grammar", *The 7th International Conference on Informatics and Systems (INFOS)*, 2010.

[12] Pierre Cachia, "The monitor A Dictionary of Arabic Grammatical Terms Arabic-English English-Arabic", *Librairie du Liban & Longman London*, 1973.
(بيير كاكيا، "العريف معجم في مصطلحات النحو العربي عربي-كليزي كليزي-عربي"، مكتبة لبنان، بيروت ولنجمان، لندن، 1973.)

[13] William Wright, "A grammar of the Arabic language", *Cambridge university press*, 3rd Edition, 1967.

[14] Abdullah Yusuf Ali, "The Holy Qu'ran with English Translation of Its meaning, Roman Transliteration and full Arabic Text", *Dar Al-Kitab Al-Aziz*.

[15] Daryabadi, Abdul Maajid, "The Glorious Qur'an: Text, Translation & Commentary (Koran)", *The Islamic Foundation*, 2010.

[16] Abderrahim Boudlal, Abdelhak Lakhouaja, Azzeddine Mazroui, et al. "Alkhalil Morpho SYS1: A Morphosyntactic Analysis System for Arabic Texts". *International Arab Conference on Information Technology*, 2010.

[17] Tim Buckwalter, "Buckwalter Arabic Morphological Analyzer Version 1.0", 2002.