AQG: Arabic Question Generator

Kheira Z. Boumsama1*, Nour H. Cherghi2, Mahfoud Sid Ali M barek3, Lamia Belguith Hadrich4

1 Department of Computer Science, FSEA, RIIR Laboratory, University Oran1, Oran 31005, Algeria
2 Department of Computer Science, FSEA, University Oran1, Oran 31005, Algeria
3 Department of Electronic, LSI Laboratory, University of Science and Technology Oran – Mohamed Boudiaf (USTO), Oran 31001, Algeria
4 Department of Computer Science, Faculty of Economics and Management (FSEGS), University of Sfax, Sfax 3018, Tunisia

Corresponding Author Email: kzboumsama@univ-orn1.dz

https://doi.org/10.18280/ria.340606

Received: 3 August 2020
Accepted: 16 December 2020

Keywords:
Arabic natural language process, question generation, semantic role labelling, semantic methods, model-based methods

ABSTRACT

The Arabic natural language process (ANLP) community does not have an automatic generator of questions for texts in the Arabic language. Our objective is to provide it one. This paper presents a novel automatic question generation approach that generates questions as a form of support for children learning through the platform QUIZZITO. Our approach combines the semantic role labelling of PropBank (SRL) and the flexibility of question models. It essentially relates to an approach of instantiation model of representation based on an analysis focused on the semantics. This allowed us to capture the maximum sense of sentence given the flexibility of the grammar of the Arabic language. This model was written in a set of Patterns and Templates based on the REGEX languages. Our goal is to enrich Quizzito's online quiz platform, which contains more than 254.5k quizzes, and to provide it with a generator of Arabic language questions for children's texts. Our Arabic Question Generator system (AQG) is functional and reaches up to 86% f-measure.

1. INTRODUCTION

There are many uses of an automatic question generator (QG): the creation of multiple-choice tests and quizzes for learning materials, human-machine dialogue systems, or interactive Question-Answer.

While this topic has been widely debated in other languages, for the Arabic language, it has never been discussed. Yet, Arabic is a language spoken by more than 300 million people in more than 22 countries.

The ANLP is considered difficult to apprehend because of its singular characteristics. His morphology has always been a challenge for the NLP specialists. Thus, its automatic processing must cope with its agglutinative nature, its inflectional richness and its diacritics bearing meaning and certain morphosyntactic traits such as declination, mode, and case; which leads to a large number of ambiguities lexical, morpho-syntactic and semantic thus curbing its treatment. In fact, the adaptation of existing models that have shown their effectiveness in treating other languages to the Arabic language is problematic [1].

Our contribution is about this theme. It focuses on an approach of instantiation of the representation model based on an analysis driven by the semantics of the sentences of a text. It allows you to automatically switch from text to generated questions. Building a question manually requires reading and understanding the text.

We have chosen to define this model using Semantic Role Labelling, i.e. the PropBank SRL [2]. SRL are based on intuitive reasoning, close to human reasoning. In addition, they provide a framework of representation and formal semantics that can guide the identification of the components of the sentence. Semantic role tagger attempts to identify the predicates (relationships and actions) and the semantic entities associated with each of these predicates. The set of semantic roles used in PropBank include both predicate-specific roles whose precise meaning is determined by their predicate (Rel) and multipurpose adjunct-type modifier roles whose meaning is consistent for all predicates (C) [2]. Our task is the identification of the components of the sentence such as: المبتدأ (mubtada, beginning), الفعل (fiel, subject), الفاعل (fial, subject), the verbal phrase (jar wamajrur, prepositional phrase), the verb (v), the agent and the beneficiary (الفاعل, الفاعل, الفعل, الجار و المجرور, prepositional phrase).

The SRL can be used to identify the appropriate question words in the absence of named entities, but the replacement of question words is not the biggest challenge to solve in QG from text. The challenge is rather the understanding of the meaning of the sentence.

We chose the semantic method based on models to capture the maximum meaning of the sentence, regardless of the syntactic order of its words. Knowing that during the syntactic annotation, it must be taken into account that the Arabic language is a pro-drop zero-subject language and that it systematically omits the morphological realization of the subject pronoun; the verb contains a pronominal index subject in its flexion. It is also important to know that the Arabic language has a mixed order of words, that is to say, neither completely free nor completely fixed. Concerning the verbal sentence, the order of the standard Arabic sentence usually follows the VSO order (that is, with a verb (V) followed by its subject (S) and object (O)). Nevertheless, one can find the
existence of other structures including the order SVO (subject-verb-object). The VOS (verb-object-subject) structure also exists, but with a lower frequency. It serves exclusively to emphasize the subject. Finally, the rare OVS (object-verb-subject) structure is sometimes used to express focus on the subject [1]. The complexity and flexibility of the grammar of the Arabic language poses a real problem in identifying the different components of the sentence.

This paper presents a new approach to generating questions from an Arabic language text combining the semantics of SRL and the flexibility of question models. We wrote a set of patterns and templates based on the REGEX language [3]. Our evaluation methodology was focused on evaluating issues from an educational perspective.

This article is composed of 4 sections. Section 2 presents a literature review on question generation methods and deals with related work. Section 3 details our contribution. Section 4 and 5 make an assessment of AQG. Finally, a conclusion of this work and some perspectives are given in section 6.

2. LITERATURE REVIEW

Efforts in QG from text can be regroup into three categories: syntax-based, semantics-based, and model-based. These three categories are not totally disjoint. Whichever approach is chosen, systems must perform at least four tasks:

1. Content Selection: selection of source text ranges (usually single sentences) from which questions can be generated.
2. Target identification: determines which specific words and/or phrases should be queried.
3. Formulation of the question: determines the appropriate question(s) based on the identified content.
4. Surface shape generation: final production of the surface shape.

These tasks are not always discrete and do not necessarily occur in that order. Identifying the target can guide the formulation of the question and vice versa.

Papasalouros and Chatzigiannakou [4] provides an overview of the existing research of the subject of applying Semantic Web technologies for automatic question generation. The review provides a classification based on technological as well as on pedagogical aspects of the works presented. According to recent research [5, 6], the majority of QG systems focus on generating questions for the purpose of assessment. The template-based approach was the most common method employed in the reviewed literature. The focus was on the generation of questions from text and for the language domain. Duan et al. [7] explored how to generate questions from given passages using neural networks. They explore two ways to generate questions for a given passage: one is a method based on the recovery using a convolutional neuron network (CNN), the other is a method based on the generation using a network of recurrent neurons (RNN). De Viron et al. [8] presented a system of automatic generation of questions for French. The generation system proceeds by transforming declarative sentences into interrogatives and is based on a prior syntactical analysis of the basic sentence. Ciguene et al. [9] focus more specifically on the automatic generation of test subjects, to assist the teacher in the development of his subjects. Their approach consists of an iterative construction of subjects "on the fly", starting from a source exercise database, built beforehand by the teacher. This work presents the design of an automatic generator based on multiple choice questions exercises called DIFAIERT-G (Different FAIR Tests Generator). Yao and Zhang [10] presented a QG approach based on minimal semantic recursion semantics (MRS), a superficial semantic analysis framework. Their method uses an eight-step pipeline to convert the input text into a set of questions. Key elements of their approach are MRS decompositions and MRS transformations. Decompositions convert complex sentences into simple sentences from which MRS transformations can generate questions.

All these researches have shown great efficiency. The question that arises is how to adapt existing methods to the Arabic language, knowing that no research work has been devoted to it in the field of automatic question generation.

3. DESCRIPTION OF AQG

The design steps for our question generation system are illustrated in Figure 1. It begins with the analysis of a text in Arabic using the online morphological analyzer MADAMIRA [11]. MADAMIRA performs segmentation into words, not sentences. To make up for that, we integrated the STAR (Arabic Texts Slicer) tool [12] as a module for the segmentation of the text into sentences. The output will be a segmented file and annotated in XML.

Then, AQG proceeds to the generation and the assignment of roles to the different actors of the sentence. We used PropBank SRLs that we needed, and we adapted them to the Arabic language, a task that was not easy given the complexity of the language. Once this phase was completed, the Pattern design modules and Template modules were designed to generate the questions. The method used is based on the semantic model method. These models, written in the form of regular expressions, required a very profound knowledge of the Arabic language grammatical rules.

3.1 Preprocessing

For the preprocessing and annotation of our texts, we have chosen MADAMIRA, a system for morphological and orthographic disambiguation of Arabic. It produces a rich feature set for each word, containing more than 14 morphological and lexical features. It includes the full diacritization of the word, identification of its stem and morphemes, part-of-speech, identification of the lemma of the
word and its English gloss as well [13]. It also provides different tokenization schemes. MADAMIRA has an implementation that is more robust, portable, and extensible and is faster than its ancestors by more than an order of magnitude [11].

The performance of MADAMIRA is quite competitive, scoring 96% accuracy for lemma and stem as over 99% accuracy for word segmentation. Diacritization is also high (96%) when excluding the case markers. MADAMIRA’s speed is close to 1000 words per second in a server-client mode [13]. MADAMIRA also provides XML and HTTP support: input and output text can be supplied as plain text or in XML [11].

3.2 Generation of roles

Semantic roles describe how words are used in sentences and the functions they fulfill. For their adaptation, we were inspired by the work [14-18]. Table 1 summarizes the set of roles we utilized and the grammatical category or categories that we assigned to each of these roles.

Table 1. Adaptation of semantic roles to the Arabic language

<table>
<thead>
<tr>
<th>ARG</th>
<th>Description</th>
<th>Grammatical category assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG/A0</td>
<td>agent</td>
<td>(faeil,subject)</td>
</tr>
<tr>
<td>ARGI/A1</td>
<td>patient</td>
<td>(mafeul-bih,object)</td>
</tr>
<tr>
<td>AM_LOC</td>
<td>locative</td>
<td>(zarf almakan, adverb of place)</td>
</tr>
<tr>
<td>AM_TMP</td>
<td>temporal</td>
<td>(zarf alzaman, temporal adverb)</td>
</tr>
<tr>
<td>AM_PRP</td>
<td>purpose</td>
<td>(jar wamajrur, prepositional phrase)</td>
</tr>
<tr>
<td>Predicate</td>
<td>verb</td>
<td>(Verb (past/present/imperative / passive voice))</td>
</tr>
</tbody>
</table>

Writing SRL must start by recognizing their grammatical categories. It is therefore important to recognize the type of sentence. In the nominal sentence, we detect its subject (مراس (mubtada,beginning)) and its predicate (الخبر (elkhabar,predicate)) and in the verbal sentence, we identify the verb (الفعل (the main verb) (sifa,adjective)) and its object (فاعل (faeil,subject) (ARGI/A0)) and this in all their possible forms:

For the other argument that the ARG={AM_LOC, AM_TMP, AM_PRP} we need to identify the locative, temporal, and the purpose expression in the sentence, respectively. We have written a set of patterns in the form of regular expressions based on the research work of linguists [19-23]. We used the syntax of Regex from Python NLTK [24].

3.2.1 SRL design of AM_TMP

This role corresponds to the presence of temporal adverb (ظرف الزمان (zarf zaman)). We tried to enumerate them in a list so that we could ask questions of temporal types. We remark that their meaning is different; it will be a duration or a precise time. Like illustrated in Table 2, we designed two sub-roles AM_TMP_K to ask a question about duration with the interrogative كم (how many) and AM_TMP_M to ask a question about a precise time with the interrogative متى (when). (_K) means كم (how long, how much time) and (_M) means متى (when).

Table 2. AM_TMP_M and AM_TMP_K

<table>
<thead>
<tr>
<th>AM_TMP_M</th>
<th>(corresponding to the question when)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(لكن) بره ندم دهرأ</td>
<td>(AM_TMP_K)</td>
</tr>
<tr>
<td>وقتم، زن</td>
<td>(AM_TMP_K)</td>
</tr>
<tr>
<td>أسبوعا هر كفم،</td>
<td>(AM_TMP_K)</td>
</tr>
<tr>
<td>يوم ثلاثية، ساعه</td>
<td>(AM_TMP_K)</td>
</tr>
<tr>
<td>(_right)</td>
<td>(AM_TMP_K)</td>
</tr>
<tr>
<td>(Duration)</td>
<td>(AM_TMP_K)</td>
</tr>
<tr>
<td>(How long)</td>
<td>(AM_TMP_K)</td>
</tr>
</tbody>
</table>

Sometimes the same word can be interpreted differently. For example, in the sentence (1) the word “ أسبوعا” (week) that indicates here a duration, is an AM_TMP_M. While in the sentence (2), the same word “ الأسبوع” (the week) must be recognized as AM_TMP_K because it signifies a moment although the adverb is listed in AM_TMP_K.

(1) (I stayed in the Levant for a week)

AM_TMP_M

(2) (I visited Baghdad last week)

AM_TMP_K

In the Arabic language, the problem is solved by introducing the morphological characteristics of the word into the sentence. Therefore, we include the part of speech in the SRL design of AM_TMP.

3.2.2 SRL design of AM_LOC

This role corresponds to the presence of adverb of place (ظرف المكان (zarf almakan)) in the sentence. We recorded the adverbs places in a list like {

شرق، غرب، شمال، جنوب، خلف،

نتAPPLE، غانو، جان، قبة، فسق، وراء، قام، بين، بار،

أين، هنا، ثم، حيث، شطر، حول، قبل، بعد، إلى، لدن، حولي، نحو، أسفل،

أين، آلي (where).} so that we could ask questions of types: أين (where).

Example:

(Maldives is located north of the western coast of the Republic of Sri Lank)

AM_LOC

At the meantime, we have distinguished a list grouping temporal and place adverbs (ظرف المكان (zarf zaman)) common to both. Common temporal and place adverbs  في (in) عنة، بين مع، قبل (before, between, with), and (attention, in). Their processing will be different because it will depend on the word categorization that follows it. This categorization can be organization, locative, Date (Temporal), Person or Numeric expression. It is extracted from English WordNet from the "gloss" tag of MADAMIRA, which gives the translation of these words into English. Therefore, we include this categorization in the SRL design of AM_LOC and AM_TMP. The following examples illustrate the categorization of some of these adverbs.
3.2.3 SRL design of AM_PRP

In Arabic, there are nearly 20 prepositions; only 8 of them are the most used, each its function. We have decomposed the AM_PRP SRL into subtypes listed in Table 3 in order to find the corresponding interrogative in writing patterns.

Table 3. Subtypes of AM_PRP

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM_PRP_M (من/From)</td>
<td>Layla's book is a gift from Master Teachers loyal to their honesty in their work, and according to this SLR, the sentence must begin with an adverb of place (one or more) followed by a name or an adjective (zero or more) followed by a prepositional sentence (zero or more).</td>
</tr>
<tr>
<td>AM_PRP_I (إلى/to)</td>
<td>Back to the country a great joy things for a great joy.</td>
</tr>
<tr>
<td>AM_PRP_N (عن/about)</td>
<td>Deep talk about studying abroad things for a great joy.</td>
</tr>
<tr>
<td>AM_PRP_AL (على/on)</td>
<td>The phone is on the chair in the room things for a great joy.</td>
</tr>
<tr>
<td>AM_PRP_F (في/in)</td>
<td>The passengers in the plane as the passengers in the plane as the passengers in the plane.</td>
</tr>
<tr>
<td>AM_PRP_B (بـ/in,with)</td>
<td>Oranges are rich in vitamins things for a great joy.</td>
</tr>
<tr>
<td>AM_PRP_K (كـ/as)</td>
<td>The girl's face is as beautiful as the moon things for a great joy.</td>
</tr>
<tr>
<td>AM_PRP_L (للـ/for)</td>
<td>Fund for poor money things for a great joy.</td>
</tr>
</tbody>
</table>

Example of a regular expression detected AM_PRP in the AQG program.

1. \(<<\text{NNP}\)+<\text{DT}+\text{NN}?>\): the sentence must begin with a proper name (one or more) followed by a name or an adjective (zero or more) followed by a prepositional sentence (zero or more). 
2. \(<<\text{DT}+\text{NN}?>\): the sentence must begin with a proper name (one or more) followed by an adjective (zero or more) followed by a prepositional sentence (zero or more).

Examples of SRL design A0

1. \(<<\text{NNP}\>+<\text{DT}+\text{NN}?>\): the sentence must begin with a proper name (one or more) followed by a name or an adjective (zero or more) followed by a prepositional sentence (zero or more).
2. \(<<\text{AM_LOC}\>+<\text{DT}+\text{NN}?>\): the sentence must begin with a proper name (one or more) followed by an adjective (zero or more) followed by a prepositional sentence (zero or more).

Examples of SRL design A1

1. \(<<\text{NNP}\>+<\text{DT}+\text{NN}?>\): the sentence must begin with a proper name (one or more) followed by a name or an adjective (zero or more) followed by a prepositional sentence (zero or more).
2. \(<<\text{AM_LOC}\>+<\text{DT}+\text{NN}?>\): the sentence must begin with a proper name (one or more) followed by an adjective (zero or more) followed by a prepositional sentence (zero or more).

Figure 2. Decomposition of the sentence "اللهو عن الدراسة أمر سيء" (Having fun from school is bad) is recognized by AQG as illustrated in Figure 2.

3.2.5 SRL Design of A0 and A1 in verbal sentence

We have designed a set of regular expressions covering all possible cases for the verbal sentence (VS) [21]. We have adopted the same procedure to those of the nominal sentence. We give some examples.

Examples of SRLs that determines the A0 and the A1 in VS.

Example of SRL for A0

1. \(<<\text{WP}\>+<\text{CC}\>+<\text{NNP}\>+<\text{PRP}\>$\): if the sentence must begin with a preposition (one or more) followed by a prepositional sentence (zero or more).
2. \(<<\text{VB}\>+<\text{PRP}\>$\): if the sentence must begin with a verb (one or more) followed by a prepositional sentence (zero or more).
3. \(<<\text{NNP}\>+<\text{PRP}\>$\): if the sentence must begin with a proper name (one or more) followed by a prepositional sentence (zero or more).
4. \(<<\text{NNP}\>+<\text{PRP}\>$\): if the sentence must begin with a proper name (one or more) followed by a prepositional sentence (zero or more).
For example the SLRs \( A0 \): 
\[
\{<DT>?,<NN>!<NN>!,<num_noun>!<num_adj>!|\langle CL\rangle\ |
\langle NN>!<PRP$/>$|<DT>?!,<NN>!|<DT>!,<NN>!;|<AM_LOC!AM_TMP!\langle PRP$/>$?
|<NNP>!|<num_noun>!|<adj_num>!|<AM_PRP>$\}
\]
detects \( A1 \) in the sentence 
"هندس برامج النظافة إلى جعل بيئتنا نظيفة " (Cleaning programs aim to make our environment clean) where \( A0 \) is underlined.

Examples of SRLs \( A1 \)

(1) \{<VB>?,<DT>?,<NN>!<NN>!,<num_noun>!,<num_adj>!,<CL>!|\langle CL\rangle\ |
\langle NN>!<PRP$/>$|<DT>!,<NN>!|<DT>!,<NN>!;|<AM_LOC!AM_TMP!\langle PRP$/>$?
|<NNP>!|<num_noun>!|<adj_num>!|<AM_PRP>$\}
\]
detected \( A1 \) in the sentence "أعط من يشاء (We give provisions to whom He wills)"

(2) \{<NN>!<NN>!,<DT>!,<NN>!,<DT>!,<NN>!;|<AM_LOC!AM_TMP!\langle PRP$/>$?
|<NNP>!|<num_noun>!|<adj_num>!|<AM_PRP>$\}
\]
detected \( A1 \) in the sentence "لا يشبه / تشبه (What), (Where), (When), (Why), (How much, How many) and (How)"

For example the SLRs \( A1 \): 
\{<DT>?,<NN>!<NN>!,<num_noun>!,<num_adj>!,<CL>!|\langle CL\rangle\ |
\langle NN>!<PRP$/>$|<DT>!,<NN>!|<DT>!,<NN>!;|<AM_LOC!AM_TMP!\langle PRP$/>$?
|<NNP>!|<num_noun>!|<adj_num>!|<AM_PRP>$\}
\]
detects \( A1 \) in the sentence "ول نصيب المسلمون على عدوهم في غزوة بدر (Muslims defeated their enemy in the Battle of Badr)"

For the generation of templates, we used the interrogatives of the Arabic language. And for each recognized pattern, we wrote a set of templates. The templates were written taking into account the type of sentence as well as the two forms of the nominal sentence and the 4 forms of the verbal sentence.

If we take the sentence "محمد كالأسد في القتال" (Muhammad is like a lion in combat). When templates are recognized the questions are asked. For this sentence, two templates are triggered.

\[
\text{Template 1: } A1 \ (A0 - \ AM \ PRP \ K) \quad \text{ما هـ} \quad A0
\]
\[
\text{ماذا يشبه/ تشبه} \quad A0 \ (A1 - \ AM \ PRP \ K)
\]

The questions posed are \( Q1 \): 
"ماذا يشبه / تشبه محمد كالأسد؟" (in \( u \) and \( Q2 \):  "ماذا يشبه/ تشبيه محمد؟"

Table 4 illustrates some examples of templates (S: sentence and Q: question).

<table>
<thead>
<tr>
<th>Example of templates in AOG program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template('من Sfi31 Sma3oul bih $jar $')</td>
</tr>
<tr>
<td>Template('ماذا $dar $target Sjar $')</td>
</tr>
<tr>
<td>Template('SPP لم $target Sma3oul bih $')</td>
</tr>
<tr>
<td>Template('من $nom $')</td>
</tr>
</tbody>
</table>

There are other questions that our AOG system may ask, like "أعط (Give a suitable title for the text)" to the question "أعط الكلمات المتصلة بالعنوان" (Give a suitable title for the text)". We choose 3 words randomly from the text.

Some examples of questions generated by AOG system are shown in Figure 4.
Table 4. AQG templates

<table>
<thead>
<tr>
<th>Template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM-PRP-M</td>
<td>من أقوال هتلر: لا تقاتل من ليس لديه شيء يوجد بدنيًا.</td>
</tr>
<tr>
<td>A0</td>
<td>تطوير الهاوبات الجميلة كورسم والخريف والموضي، وتحتضن أطعمة مثل تعاطي الهاوبات الجميلة.</td>
</tr>
<tr>
<td>A1</td>
<td>السفر شيء من الغرب لأن الرجل يشتغل فيه.</td>
</tr>
<tr>
<td>predicate</td>
<td>من الذي/التي A1 AM PRP</td>
</tr>
<tr>
<td>predicate</td>
<td>A1 AM PRP* (A1 ??) (A0 ?) ?</td>
</tr>
<tr>
<td>predicate</td>
<td>معكم A1 predicate(??) A0</td>
</tr>
<tr>
<td>predicate</td>
<td>من ? NNP من</td>
</tr>
<tr>
<td>? (A0 ??)</td>
<td>? (A0 ??) predicate(??)</td>
</tr>
<tr>
<td>predicate</td>
<td>(A1 ??) (A0 ??) predicate(??)</td>
</tr>
<tr>
<td>predicate</td>
<td>PP-AM-PRP-M ) ما هي</td>
</tr>
<tr>
<td>predicate</td>
<td>ما اللون الذي A0 استعملت الفنانة في لوحتها</td>
</tr>
</tbody>
</table>

4. EXPERIMENTS

We have integrated our system into the Quizzito online game [26, 27]. Quizzito is a platform with a set of texts in the Arabic language for children aging between 6 and 14 years old (see Figure 5). In order to evaluate the understanding of these texts, questions are asked about them and the children respond on this platform. The generation of these questions is quickly becoming automatic for other languages, but regarding the Arabic language, the designers of the platform are currently soliciting Internet workers to prepare a set of questions related to each text of their corpus.

Figure 5. Quizzito interface

Quizzito is composed by 4,712 quizzes in Arabic, each quiz having either 5 or 10 multiple-choice questions, which make 40,435 questions. The questions are derived from children books and books summaries, and have all been generated by humans. Specifics instructions are given to them, such as: “read the book before writing the questions”, and “use a simple language, that children can understand”.

For the evaluation of AQG, we selected a subset of ten documents (texts) consisting of 600 sentences taken from the corpus of Quizzito which counts 374 documents. In Figure 6, we show an example of text taken from this corpus.

Figure 6. Example of text in Quizzito

5. RESULTS AND DISCUSS

Two common types of evaluation methodologies have been put into practice [14]. The first is based on a direct human evaluation of the output of the system; the people directly involved in the research have evaluated the results of their own system. The second common methodology is based on a comparative evaluation between the output of the system (Qg) and the questions that man has generated (Qh) from the same corpus. We could consider this an indirect human evaluation. The metrics used are given in equations Eq. (1), Eq. (2) and Eq. (3).

\[
\text{Precision} = \frac{Qg \cap Qh}{Qg}
\]
\[
\text{Re call} = \frac{Qg \cap Qh}{Qh} 
\]

(2)

\[
F - \text{measure} = \frac{(2 \times \text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} 
\]

(3)

For a given document or corpus, the precision and the recall can be calculated as indicated by Eqns. (1) and (2), where the questions generated with a hyper generation are considered as the gold standard. Obviously, this type of assessment is only possible if human-generated questions are already available or if it is possible to have humans generate questions. For example, we choose a text from our corpus:

This is questions generated by our system as illustrated by the Figure 7.

\[
\text{For a given document or corpus, the precision and the recall can be calculated as indicated by Eqns. (1) and (2), where the questions generated with a hyper generation are considered as the gold standard. Obviously, this type of assessment is only possible if human-generated questions are already available or if it is possible to have humans generate questions. For example, we choose a text from our corpus:} 
\]

\[
\text{This is questions generated by our system as illustrated by the Figure 7.} 
\]

\[
(Figure \text{ 7. Questions generated by AOG from the text “الفهد” (lion))} 
\]

Below in Figure 8, we display the questions generated by the Internet users from the same text.

\[
(Figure \text{ 8. Questions writing by the Internet users from the text “الفهد” (lion))} 
\]

Therefore, in this result we have 66 (Qg) questions generate automatically from the AOG system with the text “الفهد” (lion). 47 (Qg\cap Qh) from them are linguistically correct and the others are incorrect (we put them in yellow in the Figure 7) and we have five questions manually writing by the Internet users.

The evaluation of the results obtained by our system is given by the histogram in Figure 9, which graphically illustrates the values of the f-measure obtained for each text. We note that the results obtained are satisfactory and reach up to 86%.

The unsatisfactory f-measures obtained are due to either:

1. Patterns not defined by AOG. The processing of personal pronouns is in progress.
2. Incorrect morphological analysis made by MADAMIRA. The authors speak of an error rate of 6%.

A non-fine morphological analysis made by MADAMIRA. This fine analysis will allow us to reach a better understanding of the meaning of the word especially for names. MADAMIRA does not recognize (static noun, aism jamid), (proper noun, aism eilm) (Agent noun, aism faeil), (passive participle, aism mafeul).

\[
(Figure \text{ 9. Results of f-measures obtained for evaluation of AOG)} 
\]

Another human-based evaluation of metrics examines several dimensions of the results of issue generation systems. This evaluation proves to be necessary for a correct evaluation of QG results. These metrics measure the possible deficiencies of the questions [14]. Applying these metrics to the AOG corpus, we obtained the results reported in Table 5.

\[
(Figure \text{ 9. Results of f-measures obtained for evaluation of AOG)} 
\]

\[
(\text{Table 5. Results of the evaluation of incorrect questions generated for 10 texts)} 
\]

<table>
<thead>
<tr>
<th>Category</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ungrammatical</td>
<td>77.3%</td>
</tr>
<tr>
<td>has no sense</td>
<td>8.7%</td>
</tr>
<tr>
<td>Vague</td>
<td>7.4%</td>
</tr>
<tr>
<td>Bad interrogative tool</td>
<td>6.69%</td>
</tr>
</tbody>
</table>

Ungrammatical questions are grammatically incorrect questions that do not respect the grammar rules of the Arabic language: (الصرف و النحو (grammar and conjugation/ sarf w nahw)). They represent the highest percentage of the incorrect questions in the evaluation of AQG. This is explained by the fact that we have not yet rewritten the questions generated by the templates by respecting the times and pronouns used in the patterns. The rewrite module is presented as a perspective of this work.

The 8.7% of the category that has no meaning are mainly due to the fact that the input vowels in MADAMIRA are not taken into account. The grammatical category cannot
distinguish, on its own, the A1 and the A0 in VS.

As for the use of the bad interrogative tool, 6.69% is explained by the temporal and place adverbs, which have not been recognized as such. There will, therefore, be a bad attribution of the SRL, the interrogative will be badly chosen especially between (when) and (where).

The Quizitto designers requested another evaluation of questions that is important for children learning through their platform. The evaluation of the percentage of use of the different Arabic interrogative tools by internet users and by our AQG system. The results are illustrated in Table 6.

Another comparison given in Table 7 clearly shows the large number of questions generated by our system compared to those of voluntary Internet users on the 10 previous texts.

Table 6. Use of interrogative tools by Internet users and by AQG

<table>
<thead>
<tr>
<th>Interrogatives</th>
<th>Internet users</th>
<th>AQG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inversion of verb and subject</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>هل/Do you</td>
<td>1.68%</td>
<td>6.00%</td>
</tr>
<tr>
<td>أي/who</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>ما/what</td>
<td>16.23%</td>
<td>41.00%</td>
</tr>
<tr>
<td>كم/how many (how much)</td>
<td>53.60%</td>
<td>43.00%</td>
</tr>
<tr>
<td>متى/when</td>
<td>3.82%</td>
<td>4.00%</td>
</tr>
<tr>
<td>كيف/how</td>
<td>12.67%</td>
<td>5.00%</td>
</tr>
<tr>
<td>أين/where</td>
<td>6.82%</td>
<td>3.00%</td>
</tr>
<tr>
<td>أي/which</td>
<td>0.54%</td>
<td>0.00%</td>
</tr>
<tr>
<td>أيان/(to ask about time)</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>أيان/(to ask about time)</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 7. Comparison between the numbers of questions generated by AQG and Internet users

<table>
<thead>
<tr>
<th>Text</th>
<th>Internet users</th>
<th>AQG</th>
</tr>
</thead>
<tbody>
<tr>
<td>دهاء الاعلم (hadaya aleid)</td>
<td>10</td>
<td>29</td>
</tr>
<tr>
<td>مكتبة الشاهد (kanz althaealib)</td>
<td>10</td>
<td>31</td>
</tr>
<tr>
<td>أفضل ذو الرؤوس السفيفة (altayir dhi alzuwuws alsm)</td>
<td>5</td>
<td>79</td>
</tr>
<tr>
<td>كله الوحش (kahf alwuwhsh)</td>
<td>5</td>
<td>59</td>
</tr>
<tr>
<td>هذه زهرة واحدة (timaha zahrat wahida)</td>
<td>10</td>
<td>48</td>
</tr>
<tr>
<td>يعني لحظة (yaya aleed)</td>
<td>10</td>
<td>46</td>
</tr>
<tr>
<td>الدحر (alifhah)</td>
<td>5</td>
<td>87</td>
</tr>
<tr>
<td>أربع وقفات العنزة (alraaiw wafatat alghazal)</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>نادي الألغام (sayyat alagmar)</td>
<td>5</td>
<td>56</td>
</tr>
<tr>
<td>الزواري البخيل (almizarie albakhill)</td>
<td>5</td>
<td>48</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

Generating questions can be a time consuming and effortful process. In this research, we work toward automating that process. In particular, we focus on the problem of automatically generating factual questions from individual texts. The comparison between the user-generated questions and the AQG-generated questions is according to three criteria:

1. Number of questions generated
2. The interrogatives tools used
3. Quality and rationality of the generated questions

These criteria allowed us to compare qualitatively and quantitatively the questions generated. The results obtained confirm the high quality of the questions generated by the AQG system.

At the end of this research work, we estimate to have submitted a contribution for the ANLP. Our system proves its effectiveness when the texts are long and where a human does not have the patience to read the text several times to understand it in order to generate questions. The morphological analyzer MADAMIRA has helped us a lot in linguistic preprocessing as well as in the disambiguation of words. Although we have designed all possible cases of verbal and nominal sentences, we have not been able to implement all the cases conceived for the reasons mentioned above concerning the morphological analyzer MADAMIRA.

The results achieved in this paper are very promising for the company Quizitto, as the objective is to automate the questions generation process in the Arabic language. By implementing this method, the company will be able to assist the quiz creators in the process of questions generation, letting them rather focus on the distractor generation task. Adding to that, it will reduce the costs of each quiz in terms of both time and money, allowing a rapid growth of the numbers of quizzes on Arabic content for children.

As perspectives, we plan to improve the grammatical structure of the generated questions, to filter the questions, generate answers, generate multiple-choice questions and validate the answers.

REFERENCES

textes en français. Traitement Automatique des Langues Naturelles.


