Computerising Arabic Language: Challenges and Solutions

Dr. Amel Khireddine *

M'hamed Bougara University- Boumerdes,
a.khireddine@univ-boumerdes.dz

Submission: 03/07/2020 Acceptance: 19/05/2022 Publication: 30/05/2022

Abstract:

Arabic language is characterised by a number of characteristics which make its process of computerisation a complex and a difficult one. Indeed, there are various challenges facing the computerisation of Arabic language, most importantly technical ones, due to the lack of a software computing system and the unavailability of experts specialised in it. Another challenge is the peculiar linguistic aspect of Arabic namely its syntax and semantics. This paper aims at discussing some of the problems relating to the computerisation of Arabic language- particularly in the Quran- as well as attempting to suggest some applied techniques which prove practical in settling them, and which can in turn be generalised on computerising the language.

Key words: question- answering system; the Holy Quran; Arabic language; Arabic computerisation; language processing.

1- Introduction :

* Corresponding author

 EISSN :2602-5353 / ISSN : 2170-0583

Arabic language underwent many different stages before and after Islam. It was written in its beginning without form and punctuation because the Arabs did not need these signs as they knew how to write them. However, with the advent of Islam, which concurred with the penetration of *Al-Ajmah,* a group of Arab linguists took the burden of introducing a number of rules and laws which have been since developed until they reached their current status.

The importance of Arabic language stems from the fact that it is one of the six languages used in the United Nations and that it is the basis for the Islamic religion in which the Quran, the Holly book for Muslims, was revealed. Arabic language has many characteristics. In addition of being the language of the Quran, it is rich in vocabulary and has a derivational nature. It is also characterised with commonality by which a word may have multiple meanings and synonyms, in addition to its wealth of figurative expressions including metaphors and allegories.

The Arabic language has various scientific offshoots including grammar, which is a science that examines the principles of sentence formation and syntactic rules; rhetoric; parsing, which means to explain or

EISSN :2602-5353 / ISSN : 2170-0583

clarify things; synonymy and contrast; the science of performances and rhymes which deal mainly with poetry.

Arabic language is currently written in two ways, the traditional method used in writing the Quran and in teaching Arabic in schools. The other way is Modern Standard Arabic, which does not use form and is prevalent now and is used in writing articles and newspapers. This paper discusses some of the problems relating to the computerisation of Arabic language- particularly in the Quran- as well as attempts to suggest a number of solutions to overcome such challenges.

2- The Role of the Quran in Keeping and Enriching Arabic

The Quran enriched Arabic language with its unique characteristics, both spoken and written, which contributed in enriching Arabic language. Indeed, the Quran added new meanings, methods and structures that Arabs never knew before. Such new features earned Arabic eternal richness, accuracy, beauty and stability (Khoudja, 2012). More strikingly, Quran has preserved Arabic from being lost. It has buttressed and promoted it towards perfection (Khoudja, 2012). In the same way, it has unified its dialects and made of it an international tongue, creating in the process new sciences that were not previously known (Khoudja, 2012).

The role of the Quran in keeping and saving Arabic language could be perceived in many aspects. It has contributed in the spread of the language in the Arabic and Muslim universe, given that the Arabic language is the only language for reciting the Quran and performing the prayers (Al Harash, 1999). Indeed, no holy book has sanctified a language the way the Quran does to Arabic. Furthermore, the role of the recurrence of the Quranic word between reality and metaphor preserves the meanings of Arabic language in its originality and development (Allan, 2013). In the same way, the clarity of the meaning of the word and its relationship to its family and the origin of its derivation prevented its extinction (Allan, 2013).

3- Question-Answering System

Albared et al. (2009) explain that the system for answering a question is one of the systems specialised in retrieving information and returns accurate short answers to the questions posed as a question in the natural language. For their part, Bauer and Berleant (2012) put forth that the system addresses the deficiencies resulting from Information Retrieval Systems (IR) that retrieve files and documents containing the information or part of it, while Information Extraction Systems (IE) and response systems to a question give the answer only in a concise and accurate way, if available, instead of the entire

document. Question-Answering Systems consist of three parts namely question classification, information retrieval and answer extraction (Gupta, 2013). In their study, Hamdelsayed and Atwell (2015) indicate that there are many areas used in Question-Answering Systems, such as education, biomedicine and linguistics.

There are few Question-Answering Systems in Arabic due to the fact that this field is new in relation to that language in addition to other challenges facing it. In his study, Azeldin (2012) sums up these challenges in the derivational nature of Arabic language; the non-use of large and small letters in it; in addition to the lack of computerised sources in Arabic such as dictionaries and encyclopedias.

In the same way, the use of Question-Answering Systems in the Quran are few. In addition to the challenges facing Arabic language, Abdelnasser (2014) adds the challenges of the various diacritics and the unique written and spoken characteristics of Quran.

4- Possible Solutions

To address the challenges facing the computerisation of the Quran in particular and Arabic in general, a number of general and specific solutions might prove practical. The general solutions necessitate the synergy of people,

organisations and governments interested in Arabic language. Some of these solutions can be studied as follows: first, the establishment of technical associations for the computerisation of Arabic. Second, spreading the culture of computerising Arabic language and its importance in linguistic councils and colleges interested in that language. Third, governments and organisations should show more interest in and support of researches in the field of Arabic language. Fourth, teaching methods of the computerisation of linguistics in the faculties of Arts and Computer Science.

Other solutions are mostly technical and programmatic related to the Question-Answering Systems in the Quran which can be crucial in computerising Arabic language. The existing Ottoman drawing designed by the King Fahd *Mushaf*, for King Fahd printing deals with texts as a drawing where each word gives a symbol. Thus, we need Arabic character encoding to write verses in the Ottoman script although this may lead to other problems, such as modifying some verses of the Quran and falsifying their true interpretations.

When searching, we use Modern Standard Arabic (MSA) used in writings in general and newspapers, as such we need to strip the Quranic text of the shape and the mark. This applies to all symbols, signs, shapes and all

EISSN :2602-5353 / ISSN : 2170-0583

that need to be removed. In the Quran, there are myriad repeated words that affect the search process, including letters, pronouns, adverbs, names and signs. Those common words are dealt with by deleting them and for instance by removing the *hamzas* and converting them to *Alif* without *hamza*.

In some Quranic Books and Arabic writings, some letters stick to the words. As such, when using programs to search for these words, the program will not find these words as single words, for that reason, specialists must review the Quranic books and separate such letters from words.

Based on the programs designed in the field of automatic response to a question in the field of the Quran, we need for instance to display options for the user that contain the following: show signs, symbols and shapes for those who want to search for something specific.

In the case of conjoined words, a standard Quranic Book should be available. If more than one reading is used, we need options to clarify the reading in order to achieve better results. To solve the problem of different rules, we need to present them as options and exceptions for clarification and optimal selection, as well as design different patterns of grammatical rules.

5- Conclusion :

In conclusion, in order to address the challenges facing the computerisation of the Quran and Arabic, we need the concerted efforts of scholars in the fields of Arabic language, Islamic sciences and computer science, so as to reach high levels of effective practical solutions for Arabic linguistic and Quranic computerisation. Concerned governments and linguistic associations should broaden the culture of the substantiality of computerising Arabic language and encourage researchers specialised in this area. They should also develop technical and linguistic methods for teaching this field of study in universities and colleges. With regard to the issue of the Question-Answering Systems in the Quran, which may prove vital in the computerisation of the Arabic language, a standard Quranic Book should be accessible.

6- Bibliography List:

Abdelnasser, H. (2014). "Al-Bayan: An Arabic Question Answering System for the Holy Quran". Proc. 9th Int. Work. Semant. Eval., 57-64.

Abdul-Mageed M.,& Diab M. (2012). AWATIF: A Multi-Genre corpus for moderns standard Arabic subjectivity and sentiments analysis. *Language Resources and Evaluation Conference (LREC 12)*. Istanbul. pp. 3907-3914.

Al Harash, A.H.S. (1999). computer and the learning of the Arabic language. *Human Sciences Review*, (12), 217-230.

Albared, M., Omar, N., & Aziz, M.J.A. (2009). Classifiers combination to Arabic morphosyntactic disambiguation. *Sci. Technol* ,(August), 163-171.

Allan, A.A. (2013). The role of the Holy Quran in keeping the meanings of Arabic language vocabulary in is originality and development. *Studies of Islamic Sciences and Law, 41*(2), 1191-1207.

Bauer, M., & Berleant, D. J. (2012). Usability survey of biomedical question answering systems. *Hum Genomics, (6),* 17.

Ezzeldin, A.M. (2012). "The 13th International Arab Conference on Information Technology ACIT". 2012 Dec. 10-13, A SURVEY OF ARABIC QUESTION ANSWERING: CHALLENGES, TASKS, APPROACHES, TOOLS, AND FUTURE TRENDS", 409-414.

Gupta, V. (2013). A proposed online approach of English and Punjabi question answering. *Int.J. Eng. Trends Technol, 6*(5), 292-295.

Hamdelsayed, M.A., & Atwell, E. (2015). Islamic applications of automatic question-answering. *SUST J.Eng. Comput. Sci., 1*(2), 51-57.

Kalaivani, S., & Duraiswamy, K. (2012). Comparing of question answering systems based on ontology and semantic web in different environment. *J. comput. sci., 8* (9), 1407-1413.

Khoudja, K. (2012). The advantage of the Holy Quran in keeping and enriching the Arabic language. *Majalat Kism Al Arabi* (19).

Sharaf A., & Atwell E. (2012). QuAna: corpus of the Quran annotated with pronominal anaphora. *Lrec12*, 130-137.

Sharaf, A., & Atwell, E. (2012) QurSim: A corpus for evaluation of relatedness in short texts. *Lrec*, 2295-2302.