International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 13, Issue 5, 2020

Artificial Intelligence Development and Challenges (Arabic Language as a Model)

Mohammed MuzaaL Khalati[a], Tahseen Ali Hussein Al-Romany[b], [a]Wasit University-College of Humanities Arabic Language department, [b]Wasit University –College of Arts –Translation Department, Email: [a]mmiziel@uowasit.edu.iq, [b]talkubachi@uowasit.edu.iq

This paper aims to discuss the development and challenges of Artificial Intelligence using the Arabic language. Artificial Intelligence refers to the computer ability to process data and reach the results that are similar to human thinking, learning and decision making. Also, this paper highlights that the Arabic content in the online internet is only 1%, while the number of Arabic speakers is 5%.This is high difference rate is due to many reasons such as: late awareness of the need of MT and the lack of experts and scholars in the field of Arabic language. The goal of Artificial intelligence is to develop systems that are able to deal with complicated problems that are similar to human operations. In this paper, we cited some previous studies and workshops to indicate the challenges and development of Artificial Intelligence using the Arabic language, as well as the role of NLP, which is a branch of Artificial Intelligence that analyses human language using computers. This paper concludes with some important recommendations to improve Artificial Intelligence and use.

Key words: *AI, NLP, Arabic language, challenges.*

Introduction

Computer devices are involved in many fields, especially in games, competitions and others. Also, it is involved in reading, which emphasises the need for developing these techniques because artificial intelligence devices are able to process a huge amount of the Arabic vocabulary. In 1954, the world witnessed the rise of machine translation (MT), which had successfully translated more than 60 sentences from Russian to English automatically. All that created an extensive believe that MT will spread in the few next years. There are many translators who know that their Job is to replace words, sentences and tests from one language to another concerning the context, meaning, style and other important elements of

communication which are very important for the process of translation. The machine came to replace those who translated like machines. So, the translation from ST to TT, needs an efficient translator who has knowledge in both languages, experience, and an understanding of culture.

Arabic language speakers in African, Asians and Islamic countries reached 480 million according to the latest population statistics. The Arabic language came in fourth place according to its dispersal around the world. The Arabic language has great importance because it is the language of the Holy Qur'an, the language of popular poetry, and the legacy of its civilization. However, it was less used in the Artificial Intelligence field which is the topic of our paper. Our subject is to discuss the main challenges and the proper solutions to accompany evolution by relying on this technique for the benefit of the Arabic language and its speakers.

Arabic Language and Artificial Intelligence

21st Century Technology is yet to understand the Arabic language, which is considered one of the greatest in use among world languages. The Arabic language is the fourth most used language around the world, while the English language is the first. The vastness of the English vocabulary, which exceeds 12 million words, reflects its richness, flexibility, capability for expression, and its ability to detect the linguistic equation of other languages according to different studies. It is important to say that Arabic language speakers exceed 480 million people. In addition to that, 1.5 billion Muslims around the world need to use Arabic language in their practise.

Arabic Language and Computer

Regarding Arabic language and Artificial intelligence, many computer scientists and researchers seek the field of humanities to produce computer programs that read the Arabic texts accurately and translate them to a digital formula. This is a hard mission to achieve (Black,220,1996). Artificial intelligence changed by the rise of some advanced techniques to present the semantic denotation of texts, especially with the appearance of new techniques including: wordszvec, Intersent, Use and Elmo.

There have been many attempts to train the words browsers to achieve the best Arabic language tasks, but the presentation on a sentence level is still not used enough. However there are a great capabilities to use that presentation in all typical units in the natural language process field such as structure, parts of speech, parsing and accurate explanation. The importance of Arabic language process is not considered as entertainment or a subordinate issue. Rather, it is considered a great issue that the future of language relied on the even

status of Arabic in the current civilization and their scientific and Economic future (Muddwar, 30:1996).

Artificial Linguistic Intelligence in Reality and Challenges

Machines have a linguistic capability to imitate the human capability to understand and produce speech, which make it smarter and enable it to do the required tasks and sustain communication with humans. Regarding the workshop held in the King Abdulaziz University in Jadah at the centre of Research Excellence in Arabic language, speech production and understanding linguistically proceed in two stages: the first one is the production stage which generates the meaning at the brain, codes it by using specific linguistic system, and then transfers it to a sound (Midleeast Website,2019). At this workshop, Dr. Al-Aseeri said that with the application of Artificial Linguistic intelligence on producing language we have not seen a robot talk nor a machine program that is able to transfer images or videos to a text. The second stage is understanding and recognition by hearing and coding using the linguistic system, then the reaction with speech (ibid).

Challenges of the Artificial Intelligence at Arabic Text

Computers exceed human abilities in many areas such as competitions, games, and others. This includes reading tests. Thus, it is necessary to develop these technologies. Artificial intelligence devices can process huge amounts of data including Arabic expressions. Speakers of the Arabic language, including those in African, Asian, and Islamic countries, are around 480 million people, according to a statistic (Alwatant Website,2016). The Arabic language ranks fourth among other global languages. Despite the importance of the Arabic language as the language of Quran, Arabic poetry, and Arabic heritage, it remains the least used language. Thus, the present study is an attempt to keep pace with development to solve the problems facing the development of the Arabic language.

The Most Important Challenges Faced by Artificial Intelligence in the Arabic Text

1- The Arabic language includes many rules and symbols of various meanings that differ with changes of tone, character, or punctuation marks (Britannica website:2020).

The following issues should be considered when doing the translation:
• The roots of words and their derivations. For each word, there are many different and harmonious terms. This is called the phenomenon of derivation.
• The analytical basis of power in the Arabic language. It depends on understanding and interpreting the text.
• The presence of synonymy, antinomy, and metonymy.

2-Synonymy is the presence of more than one word in a different form. Whereas, antinomy means that the same word signifies two opposing meanings. Likewise, metonymy means a word with multiple meanings that are known through context and clues (Noon Websit, 2018). As mentioned in different studies, Arabic has a rich morphology, in addition to being inflected for number, gender, voice and case. Words attach to various clitics for conjunction (w+and), the definite article (Al+the), preposition (b+by/with), pronouns of processions (hm+their/them).We can also notice that the Arabic language is written usually without diacritics that denote the short vowels. All these issues create a word-level ambiguity (Zbib,154:2008).

Language Definition and Characteristics

We can define language as familiar spoken and written symbols shared between a social group to express themselves. The language function includes communication, identity expression, play, imaginative expression, and emotional release. As Henry and Crystal indicate in their paper (Web,2020), the Phonetician and language scholar Henry Sweet stated that " language is the expression of ideas by means of speech sounds combined into words are combined into sentences, this combination answering to that of ideas into thought." As we know, different languages came from different communication systems. Communication systems can be recognised as different languages if they cannot be understood without specific learning by both parties (Henry and Crystal:2020). The native speaker of any given language is called the first language speaker (native tongue) or bilingual. If an individual tries to speak any other language, he has to understand the communication system of that language, as well as its structure, grammar, vocabularies and intellectual process. They can be called Multilingual (ibid). Regarding the definition of the language and communication system we mentioned above, there are many characteristics of the Arabic language:

-Deletion and brevity
-Variety of vocabulary
-Distinction between the feminine and the masculine
-Arabisation, which is the arrangement of the word for Arabic scales
-Accuracy of expression
-Characteristics and sounds of letters, such as intensity, looseness, loudness, openness, softness, richness, and thinness, etc (Britannica Website:2020).

All these features prevent artificial intelligence from overcoming the difficulty of understanding the meaning of the expression as speakers and readers do. Despite the fact that artificial intelligence facilitates many sides of human daily practical life, it is still unable to perform the job of understanding the particular meaning of many expressions. It is unable to

interpret the poetic and Quranic meanings properly as it can only understand the literal meaning due to the diversity of positions of markers (Alatyah, 134,1983).

The Arabic language contains more than sixty-five different dialects due to the diversity of countries and regions. Each dialect has its own characteristics, sounds, and structure of expressions.

The dialects of North African countries are fast in comparison with those dialects used in the gulf. These dialects vary in the same country. In Iraq, for instance, the dialect used in the west differs from that used in the south (Ajman Website, Almahjobi,2008).

What can improve Artificial Intelligence to understand the Arabic text?
• Supporting artificial intelligence through the use of contemporary Standard Arabic data.
• Introducing Artificial Intelligence to the characteristics of the Arabic language as much as possible.
• Composing a dictionary for modern Standard Arabic, taking into account the expressions used in it, especially, medical, engineering, science, technology, or technical sciences terms, and others to serve researchers, teaching staff, and students to adopt the Arabic language in scientific and practical scientific fields in these sciences.
• Feeding Artificial Intelligence programs with linguistic features that make it capable of reading, spelling, summarising, translating, and analysing. This technology should be fed with certain programs making it able to understand the characteristics of the Arabic language.
• Upgrading Artificial Intelligence programs to be able to deal with different Arabic dialects.
• Developing Artificial Intelligence techniques to translate the text through context.
• Manufacturing devices and technologies of Artificial Intelligence.
In brief, the number of people speaking the Arabic language is around 400 millions from the middle east and Africa as it is the language of the glorious Quran. The ratio of the Arabic speakers is 6.6. In 1974, the united nations adopted the Arabic language as the sixth official language.

Arabic Language Types

Arabic language is a Semitic language similar to Acadian, Armenian and Hebrew. It is the language of Arabs who live in a very rich area and recite the language of Qur'an. Therefore millions of people rely on its translation into other language to understand the text (Alotaibi, Meftah,5,2010). Classical Arabic and Modern Arabic standards are the two kinds of Arabic language.

a-Classical Arabic

This type of language is different from Modern Standard Arabic according to the style and vocabulary, some of which is neglected. All Muslims depend on the translation to understand the text.

b-Modern Standard Arabic

This type of language is understood by all the speakers of Arabic. This type of language is used in TV shows, written materials, lectures, workshops, teaching etc.

The number of Arabic letters is 28 or 29 if we count hamza (which can be counted as a letter). The handwritten letters and the printed letters are the same with no differences. For Arabic speakers, the notion of small letters and capital letters does not exist. The letters rely on their positions in the words for shape (Alotaibi, Meftah,7,2010). We must know that the Arabic language has many distinctive features from other languages.

Problems Under-Discussion

The work of machine translation began in the late 1950s with highly predicted difficulties. Machine translations work in both linguistics and computational linguistics (Slocum 1984,1985;Tucker 1984). Translation is impossible without fundamental work on text "understanding", which led to a cutback in funding. The use of Machine Translation by Arabic language speakers will confront many obstacles due to the existence of major dialects like Iraqi, Gulf, Levantine, Maghrabi and Egyptian Arabic (Zaidan and Callison-Burch,2011,2014). In addition to these dialects, the user of MT will face differences in their morphologies, vocabularies, grammatical cases and conjugation. Most authors reached for proper solutions that lie in dialect-specific processing and modelling, as well as "classification and identification," of the Arabic texts before making it a part of and NLP system. As Haung states (2015), there are many effects that interrupt the existence of the Arabic language online compared to other languages. The most significant is the breaking through of Dialectical Arabic to replace modern Standard Arabic. The spoken language and the one used in media are assembled in DA.

Some previous studies (Zbib et al, 2012; Salloum and Habash,2012) defended the development of parallel corpora that would be kind of mixing between DA and MSA. And many similar efforts were taken by famous and known researchers to discuss, suggest and fill the gap of machine Translation of the Arabic language. In this paper, we have started with the first study of (Haung,2015) that suggests several solutions to control challenges in improving

the machine translation's quality by using a dialect classifier to build a compatible, dialect-specific Arabic English MT system.

The second study conducted by Meftouh (2015), who established parallel Arabic Dialect (PADIC), consisted of major dialects that highlighted MT experiments from MSA to DA and back. These dialects are used for conversation and not for writing as he confirmed. Another study that was conducted by Zbib etal (2012) talked about Levantine-English and Egyptian-English parallel corpora. These three studies performed a few translation experiments in order to show the variety of attributes such as MSA data limitation, inter-dialect learning, the impact of both translation, and morphological analysis from a particular dialect to MSA and then to English (Zbib,49,2012).

As Daniel states (11,1996), "all languages speakers employ a specific range of styles and sub-dialects or jargons." We have to agree that we speak a dialect, and realise that our speech patterns vary largely during the daily courses of our routines. In the morning, at the office, communication with employees may consist of optimistic fragments of a specific code (ibid). So, we continuously range between formal and informal styles of speech.

Dr. Fatimah (Alwatanvoice website: 2016) presented her paper entitled "Artificial Intelligence in the service of Arabic language's roots and fields" at the faculty of Computer and Information Technology in Abdulaziz University in Jadah. She indicated that the Arabic language is the pillar of any community and in cognitive science, communication skill development, culture industry (NLP or Tradition Computer Language), and software development. She also added that our Arabic community cannot be opened to other cultures and sciences without serious actions of translation in scientific, Technical and Modern Humanity. For building advanced programming experience, environmental development, research support, innovation to assemble language dialogue, recognition linguist applications relied on computer processing and Artificial Intelligence science to produce recognition of the Arabic languages in Arab Countries.

The NLP Researcher from Egypt, Ali Farghaly, says "Arabic doesn’t have capitalization, which is a way to indicate proper names of people, places, companies." Arabic letters also change their shape every time their position in the word changes." He added in his explanation that longer words in Arabic can be created by stringing smaller elements of language together. According to the Arabic language structure, Farghaly says, “A complex word can be analysed into a subject, a verb and an object." This kind is ambiguous because the complex word can be broken down in three or more ways. For example: The English phrase, " He ate them," has three parts, the subject ("he"), the verb (ate) and the object ("them"). But in the Arabic language, this sentence has only one word "Akalahum."

Previous Studies and Workshops

World Arabic Language Day is celebrated on December 18 of each year, and UNESCO hosted three special round tables dedicated to this year's theme, "Arabic Language and Artificial Intelligence," which was followed by a concert to be held at UNESCO headquarters in Paris on December 18 2019. Experts, academics, artists and representatives of specialised institutions met and discussed the following topics:

1-The effect of Artificial Intelligence on preserving and enhancing the Arabic language.
2-Computerisation of the Arabic language and the future of knowledge.
3-The launch of the regional report entitled, "Arabic as a Gateway to Acquisition and Transfer of Knowledge".

Artificial Intelligence and Human Translation Change

As many Researchers confirmed and predicted, NMT and Technology will play a key role in the future of translators and linguists. Augmented Translation is a concept developed by the common-sense advisory team in Massachusetts, USA. This concept will make relevant information easy to reach by translators, and they will implement various technologies (Web Journal,2018). Translators will get an advantage from adaptive NMT and automated content. The technologies include translation memory and terminology management. By using this concept by translators, they will have a rapid access to all the relevant information needed during the process of translation without returning to the different dictionaries or the previous translation of the word. The suggestions of Neural Machine Translation system will be influenced by the Translator (ibid).

In addition to all that, the output relies on the experience of the translator, his expertise in Source language and Target language and the context. In addition to what was mentioned above, the translator confronts many problematic areas during the process of translation such as: Lexical-Semantic problems, Grammatical problems, Syntactical problems, Pragmatic problems and Cultural issues. Even translation is varied according to the text title such as idioms, expressions, politics, economic, sciences and others.

Conclusion

In this research paper, segmentation is better for the Arabic language to match with English or any other languages. In this paper, we presented the definition of language, Arabic language and the types of Arabic language, as well as which type is more dominant than the others. We also presented many previous studies and workshops that used Artificial Intelligence with the Arabic language. Through these studies and workshops, it was clear that the Arabic language has many distinctive features compared to other languages. In addition to

all that, we mentioned the types of Arabic language, and which one is more dominant than the others. Also, we mentioned the translator's role and experience in supporting the AI. This paper recommends that we rely on old and new studies and techniques to support Artificial Intelligence. Finally, the researcher came up with some recommendations for cooperation between Artificial Intelligence scientist and linguists in developing new technologies and supporting linguistic technology to create easy access of users.

Recommendations

1-The first thing is to work hard and develop the used techniques to represent the text's semantic denotation, especially with the exploration of new and deep learning such as Word 2 vec, Infer Sent, Vse and Elmo techniques.
2-Train browsers to achieve the best performance in Arabic language tasks and promote the sentence level, which is still not used properly. However there are capabilities to use such representation that are relevant to other typical units in processing Natural languages such as: structure unit, parts of speech unit, parsing unit and accuracy questionnaire unit.
3-Supporting linguistic technology to spread all kinds of linguistic knowledge and facilitate their use. This can also facilitate recognition, scientific construction and linguistic theories between the Arabic language users.
4-Supporting linguistic planning by reforming the language structure, phonology system, its functions, reducing the writing rules, establishing dictionaries, protecting Arabic language vocabularies, updating them and sharing them with other spoken communities.
5-Supporting language policy by making decisions between language and science in acquisition learning.
6-Human language is a very complicated phenomena, so it is necessary to rely on language models to solve the complicated problems. We have to be open to other sciences to solve such problems. In addition, using algorithms to employ language in its services.
7-Real cooperation is needed between Artificial Intelligence scientists and linguists, because language needs great human effort.

REFERENCES

Alatiyah, K. H. (1983).Phonology system in Arabs. Baghdad, Republic of Iraq, Al-Jahth Press House.

Ali, M. A. (2016). Artificial intelligence and natural language processing: the Arabic corpora in online translation software. International Journal of Advanced and Applied Sciences, 3(9), 59-66.

Alotaibi, Y.A. and Meftah, A.H. (2010). Comparative evaluation of two Arabic speech corpora, Natural Language Processing and Knowledge Engineering International Conference, pp. 1-5.

Bower, G. H.; Black, J. B., & Turner, T. J. (1979) Scripts in text comprehension and memory. Cognition Psychology, 11(1):177-220.

Danials, Peter T. and William, B. (1996). The World's writing system. New York: Oxford UP,1996.{Comprehensive examination by eighty scholars in dictionary presenting world's writing systems).

Farghaly, A. 2008. Arabic NLP: Overview, the state of the art: Challenges and opportunities. In Proceedings of the International Arab Conference on Information Technology (ACIT'08).

Meftouh K, Harrat S, Jamoussi S, Abbas M and Smaili K (2015). Machine Translation. Experience on PADIC:A parallel Arabic dialect corpus. In the 29th Pacific Asia Conference on Language Information and Computation: 26-34.

Meftouh K, Harrat S, Jamoussi S, Abbas M and Smaili K (2015). Machine Translation. Experience on PADIC:A parallel Arabic dialect corpus. In the 29th Pacific Asia Conference on Language Information and Computation: 26-34.

Shaalan, K . and Raza, H. (2008). Arabic named entity recognition from diverse text types. In Proceedings of the 6th International Conference on Natural Language Processing (GoTAL'08).B. Nordstr¨om, and A. Ranta, Eds.

Slocum, J. (1985) A survey of machine translation: its history, current status, and future prospects. Computational Linguistics 11, 1 (Jan-Mar.1985),1-17.

Tuker, A. (1984) A perspective on machine translation, theory and practice. Comm.Assn. Computing machinery 27,4 (Apr.1984),322-9.

Weber, M. (2018). mastertcloc.unistra.fr/artificial-intelligence-human-translators/Artificial Intelligence in Translation.

Zbib, R. (2008). Aclweb.org- I Badr, R Zbib, J Glass-Proceedings of ACL-08:HLT,Short papers.

Zbib, R., Machiodi, E., Vevlin, J., Tallard, D., Mastsoukas, S., Schwartz, R. and Callison, C. (2012). Machine translation of Arabic Dialects, In Proceeding of the 2012 Conference of North American Chapter of the Association.