**European Journal of Humanities and Educational Advancements (EJHEA)**
**Available Online at:** https://www.scholarzest.com
Vol. 3 No. 07, July 2022
**ISSN:** 2660-5589

# THE AMBIGUITY IN THE ARABIC LANGUAGE WITH MACHINE TRANSLATION

**Zainab Ayad Mahmood**
zainabayadmahmood@gmail.com
Teacher of English
University of Kufa

**Article history:**
**Received:** 20th May 2022
**Accepted:** 20th June 2022
**Published:** 28th July 2022

**Abstract:**
While many people believe that machine translation has assisted translators' jobs and facilitated the transfer of meaning from one language to another for non-translators also, it is a process that is not without flaws and complexities. For instance, one of the most common issues in the Arabic language is ambiguity, which is due to a variety of factors, including complicated morphology and the wide range of vocabulary.

**Keywords:** Machine translation, language, Arabic, ambiguity, morphology

For a long time translating using machines has been the humanities greatest aspiration, this has become an actuality in the last century, through computer programs that can translate a variety of texts from one specific natural language to another. As soon as electronic computers appeared, attempts started to create translation systems. Computers were used in Britain to solve German Enigma codes during World War II and deciphering language began as a metaphor for machine translation but the absolute reality is not ideal, as always, there are no machines that can instantly translate any text into any language with a few keystrokes to make an excellent translation with no human assistance or interaction with the machine. As has always been the case, the main difficulties to machine translation are not arithmetic but linguistic, they are difficulties related to lexical ambiguity and grammatical complexity from ellipses, non-grammatical structures, and differences in vocabulary between languages, in short, to create equal meaning from analyzing written symbols and output texts in other linguistic symbols. The difficulty of natural language processing with computers can be summed up in one word: ambiguity. Word meaning, morphology, syntactic features and roles, and interactions between distinct portions of a text are all uncertain in natural language. Humans can deal with ambiguity to some extent by considering the larger context and background information, but there is still a lot of misunderstanding even among humans. The speaker may purposefully be vague to avoid committing to a specific interpretation. In that instance, the uncertainty must be preserved in the translation.

## FACTORS OF AMBIGUITY IN THE ARABIC LANGUAGE

Arabic is the mother tongue of more than 330 million people. It is one of the main languages in the world and among the five languages of the United Nations. It is spoken by about 260 million people in 22 countries in the Middle East and North Africa, where it is used as the primary language in their oral and written communication. The Arabic language has some unique difficulties in the scope of language technology research, as the countries that speak Arabic are distinguished by the phenomenon of bilingualism, and the reason for this is because they traditionally speak different dialects that are not included in writing and do not always circulate. But they have a language they share called Modern Standard Arabic (MSA).

One of the most important features of this language is that it has complex grammatical and morphological structures, a high level of ambiguity that accompanies the meanings of vocabulary, and a lot of ambiguity caused by the writing system, which in turn ignores diacritics (such as inflection marks, short vowels and double consonants). So Arabic is one of the morphologically complex languages with a wide range of morphological properties. These characteristics are assimilated using both the sequence (affixes and stems) and the morphological (root and patterns) model with its formation of a morphological and phonetic organization that displays the spelling of words that interact with orthographical differences. We will look at the set of attachable function words (or clitics) in Arabic, as opposed to inflectional traits like gender, number, person, and voice. These clitics are put alongside the word, which adds to its ambiguity. Three levels of cliticization can be applied to a word basis in exact order: [CONJ+ [PART+ [Al+ BASE +PRON]] The BASE can have a definite article Al+ 'the' or a member of the +PRON class of pronominal enclitics (e.g. +hm 'their/them') at the deepest level. Possessive pronominal enclitics can be attached to nouns, verbs, and prepositions (as objects). Verbs and prepositions do not require the use of a definite article. On nouns, +PRON and Al+ do not coexist. The particle proclitics (PART+) are the next group. b+ 'by/with' and s+ 'will' are examples of single-letter prepositions and other functional particles. The conjunctions (CONJ+) w+ 'and' f+'so' are found at the flattest stage of attachment. They can link themselves to any word. Although the archaic question particle proclitic >+ can be used before conjunctions, it has nearly vanished in modern standard Arabic. Clitics can interact with the phonological and orthographic shapes of the word to which they are linked. The feminine ending +p (Ta Marbuta), for example, can only be used at the end of a word. It becomes the letter +t in the medial position.

These are three issues we have highlighted in the examples above related to preprocessing. The first one, it is a very important issue to address the ambiguity in the Arabic vocabulary to decide if it is possible to describe your clitic, you must specify that your clitic is present in the vocabulary that we take in the context. Secondly, once a specific analysis has been identified, the process of splitting or extracting a word feature must be clear about the shape of the resulting word to be clear in such a way that no undue ambiguity is added. Third, empirically taking into account the degree of division, which in turn causes a difference from one system to another, and this depends on the extent of the obtainable corpora, there is a scattering under defection, whereas excessive defection adds ambiguity, in addition, it requires pre-processing tools with high-quality because therefore, mistakes and clutter will occur when it fails.

The Arabic language is characterized by having a large vocabulary that exceeds the number of vocabulary in the English language, and this is seen as a problem due to the complex morphology. The training corpus of the vocabulary will be huge if we consider each group of morphological and stem suffixes as an individual word. The vocabulary quantity is large, which means the chance of repeating each word is very low if we compare, for example, the average length of the English sentence, which is 33 words compared to 25 words from the Arabic side. The large size of this vocabulary would create problems of variance in the Arabic side and difference in the translation report forms because most of the time the Arabic vocabulary appears in the data of the text as their counterparts in the English language. When applying statistical machine translation systems to the language pair, this is a common trouble with the extent of difference in this huge amount of vocabulary. For example, if we have two languages F and E, F is a language that contains more vocabulary than the E language, in this case, we will have problems with analysis during translation, while in the case of translation from E to F we will have problems with generation. The main challenge is the demand to integrate linguistic knowledge into Arabic data-based machine translation.

Recent specialized attempts to build data-driven systems for translation from and into Arabic have proven that the problem of words and complex grammatical structures promotes resorting to incorporating some linguistic knowledge into these machine translation systems. Standardizing this knowledge at minimal cost poses a problem in how to do this since there are consequences for the number of language resources added to computational complexity and portability.

## CONCLUSION

We can improve the quality of machine translation systems more undoubtedly through three points, the first of which is by imposing restrictions or developing complex ways of inserting the file, or by limiting the scope of the vocabulary by formatting it in an exact language format, avoiding the multiplicity of meanings, complex sentence structures, and forms. The third point is to have prefixes, suffixes, portions of words and phrases, sentence boundaries, or different kinds of grammatical indicators added to the entered (pre-edited) texts. Finally, in an interactive mode, the algorithm may indicate issues of ambiguity and options for human operators (typically translators) to resolve during the translation process.

## REFERENCES:

1. Abdelhadi, S., Farghaly, A., Neumann, G., & Zbib, R. (2012). Challenges for arabic machine translation. John Benjamins Pub. Co.
2. Habash, N. Y. (2010). Introduction to arabic natural language processing. Morgan & Claypool.
3. Hutchins, W. J., & Somers, H. L. (1992). An introduction to machine translation. Academic.
4. K., K. E. F., & Asher, R. E. (1995). Concise history of the Language Sciences: From the Sumerians to the Cognitivists. Pergamon.
5. Koehn, P. (2012). Statistical Machine Translation (3rd ed.). Cambridge University Press.
6. Olive, J. P., Christianson, C., & McCary, J. (2011). Handbook of Natural Language Processing and machine translation: Darpa Global Autonomous Language Exploitation. Springer.
7. Roux, J. C., Ndinga-Koumba-Binza, H. S., & Bosch, S. E. (2012). Language science and language technology in Africa: A Festschrift for Justus C. Roux (1st ed.). Sun Press.