



Tikrit University

Journal of Al-Farahidi's Arts

DOI: <https://doi.org/10.51990/jaa.15.52.2.23>**Asst. Prof. Dr. Mohammed Barjes Salman**E-Mail: dijla1983@tu.edu.iq

Mobile: +9647701770235

Department of English Language
College of Education for Humanities
Tikrit University
Salahuddin
Iraq

Keywords:

- Arabic Language
- Homograph
- Machine Translation
- Google Translate
- Diacritics
- Drop Down

Article History:

Submitted: 04/10/2022

Accepted: 27/11/2022

Published: 08/12/2022

Homographs In Arabic as A Challenge to Machine Translation: Google Translate as A Sample

ABSTRACT

Machine translation is one of the growing up fields like all the other developing technologies which try to facilitate the matters that are related to human life. In spite of its importance, the translations of this type lack accuracy and need many modifications because they contain various kinds of shortcomings. The current research focuses on one of the problems of machine translation which is the translation of Arabic homographs because the Arabic writing system permits the existence of a number of lexical items to be written with same letters but they have different meanings and different pronunciations according to the diacritics used with letters. The research adopts Google Translate as a service of machine translation as it is the more important nowadays than all other services of this filed. The problem of translating Arabic homographs is represented in two types. The first one is that Arabs tend to type text with no diacritics which makes it difficult to guess which lexical item is used even when they are used in sentences, and the second problem is that Google Translate does not have the ability to distinguish between the homographs even when they are used with diacritics. Three samples are analysed which represent three sets of homographs that are used in sentences to check the compatibility of Google Translate to translate the homographs when they are written with diacritics and without. The results shows that the accuracy is about 30% of the whole translated homographs which is very low rate. The research suggests a solution concerning translating Arabic homographs which is to upload Google Translate with the meanings of all the versions of homographs in relation to diacritics, and when the homograph is used without diacritics, a drop down of words with diacritics should appear to determine which one is intended by the uploader of the text.



1. Introduction:

A very great interest has been directed to machine translation nowadays especially with the technological development and the wide spread of technological devices such as computers and cell phones which include various programmes, apps, or internet sites and services that are interested in machine translation. Machine translation, as illustrated by Richards and Schmidt (2010:350), is the use of programmes that are specialized in translating texts without the aid of human in such process. This field has developed greatly in the last years but still the quality of the translations varies according to the complexity of the source text.

Although there are different services that are interested in translating texts, google-translate site is, nowadays, the most widely used site for this purpose as it can be easily accessed and it can translate long texts in contrast to some other services that can translate limited texts. Great efforts are exerted to develop this site but still there are many problems that appear in its translations. This may result from the complexity of the source language or the unlimited expressions in each language which makes it impossible to upload all the possible expressions in the source language. One of the challenges that machine translation may face is the translation of homographs in Arabic into other languages as the homography phenomenon is very familiar in Arabic and the machine translation services can usually discriminate the different meanings that these homographs can have.

This research is interested mainly in the translation of homographs in Arabic but a clear distinction should be made first between homography and some other relevant lexical relations which are: synonymy, antonymy, homonymy, homophony and polysemy. Usually, no clear demarcation is presented to each of these lexical relations, and sometimes more than one of these lexical relations may be given similar definitions which may lead to confusion. In the next a few lines, the distinction will be made clear throughout giving precise definitions, examples and a table that summarizes everything. All the lexical relations, mentioned above, should be defined according to three factors or variables which are meaning, spelling (written form) and pronunciation (spoken form). The presence or absence of one or more of these factors results in different lexical relations as can be summarized in the following table:

Table (1): Lexical Relations Differences according to Meaning, Spelling and Pronunciation

Lexical Relations	Meaning	Spelling	Pronunciation
Synonymy	+	-	-
Antonymy	-	-	-
Homonymy	-	+	+
Homophony	-	-	+
Homography	-	+	-
Polysemy	≈	+	+

* ≈ means partially equal

Concerning synonymy, it is a lexical relation that means two or more lexical items that share a similar meaning but they have different spellings and pronunciations such as 'big-large' (Lyons, 1995:60-62). The word

'similar' is selected other than 'same' which is used by some scholars because it is illogical to find two or more lexical items in the same language that are replaceable in all possible contexts without any difference. At least, one difference can be found to make these lexical items distinct but generally they share a similar meaning.

Antonymy refers to two lexical items which have opposite meanings such as: 'tall-short', 'old-young', and 'high-low' (Cowie, 2013:37-38). Thus, antonyms have different meanings, spellings and pronunciations.

Homonymy means two or more lexical items which have the same spelling and pronunciation but with unrelated meanings such as 'ear' as an organ of hearing and 'ear' of corn (Leech, 1974:228). So, although homonyms have the same spelling and pronunciation, they have different meanings that make them distinct from each other to the extent of being mentioned as different entries in the dictionary as if they have no similarity concerning spoken and written forms.

Polysemy refers to a lexical item which has multiple senses but these senses are related to each other as opposed to homonyms whose meanings are irrelevant to each other. An example of polysemy is the lexical item 'hook' which has multiple meanings that are related to the same general meaning such as: (1) a piece of material, usually metal, curved or bent and used to suspend, catch, hold or pull something, (2) short for fish-hook, (3) a trap or snare, etc. In dictionaries, lexicographers tend to deal with homonyms as different entries, as mentioned above, to indicate that they are considered different words which have different historical origins but they come accidentally to have the same spelling and pronunciation. Concerning polysemes, they present the related meanings under the same entry to indicate that these multiple meanings are related to each other in a way or another (Saeed, 2009:63-65). Although the meanings of polysemes are related to each other, they cannot be considered as having the same meaning because they are used in different contexts and they refer to different entities but they have the same general meaning that all of them share. Native speakers can easily conclude this similarity of this shared general meaning. For example, the lexical item 'head' can be used to refer to the highest organ of the human beings or animals, the top of the tree or mountain, the head of the department, i.e., the chief or responsible person, etc. All these meanings are related to the same general meaning which is the highest part or position of something but still each meaning is used in different context as it is clear in this example. This is the reason of using the mathematical sign \approx (partially equal) under the heading of 'meaning' for polysemy in Table (1) above because the meanings of polysemy cannot be considered 100° the same although they are related to the same general meaning. This sign means, in mathematics, anything that is intentionally similar but not exactly equal to something else.

Homophony is a lexical relation that means lexical items having the same pronunciation but they have different meanings and spellings such as: 'lead' /led/ (metal) and 'led' /led/ (the past form of the verbs 'lead') (Cruse, 2006:80).

After presenting the definitions of the lexical relations above, it is time now to define ‘homography’ which is the concept under study. Cruse (2006:80) indicates that it means two lexical items which have the same written form but they have different meanings and different spoken forms such as ‘lead’ /led/ (metal) and ‘lead’ /li:d/ (to guide). This lexical relation appears rarely in English. That’s why, it does not represent a real problem that may lead to ambiguity as few pairs can be found in English that are related to this lexical relation.

In Arabic, homography represents a critical problem that readers of Arabic texts or the programmes may face when dealing with homographs, as it will be shown below. In English, the homographs are pairs, i.e., two lexical items having the same spelling but they have different meanings and pronunciations. In Arabic, one can find more than five homographs which share only the spelling and they differ in their meanings and pronunciations. The Arabic graphological system consists of components. The first is related to letters of the Arabic alphabet which are 29 letters, three of them are vowels and the rest are consonants. The second component is called ‘harakat’ or diacritics which are signs put over or under the letters to indicate short vowels, duplication or sometimes they mean the absence of any vowel. The following figure illustrates these two components:

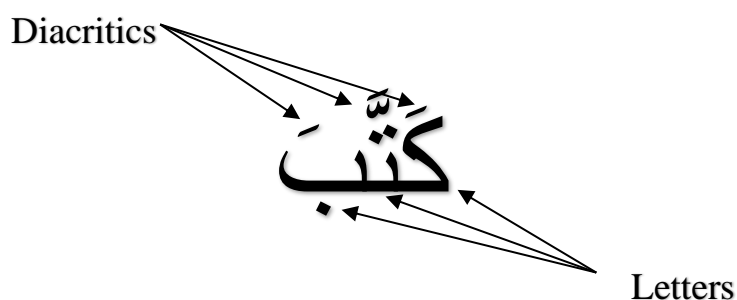


Fig (1): Components of Arabic Alphabetical System

The main diacritics and those that will be the concern of the current study are the following (َ ِ ُ). The first is short /ʊ/, the second is either /ə/ or /ʌ /, the third is short / I / the fourth means the letter is duplicated, and the last one means that the letter is not followed by any vowel. Nowadays, Arabs tend to write letters only and they neglect writing diacritics in both handwriting and typing because they know that the native speaker of Arabic can guess the meaning and pronunciations by his/her intuition and according to the context which the words are used. That’s why, the words which have the same letters but different diacritics and consequently different pronunciations come to be homographs when letters are used alone without diacritics. An example of homography in Arabic is the three-letter word (فعل) which can have a number of readings and meanings according to the diacritics used with it as follows:

- 1- فَعَلَ /fIʕl/ (an action or a verb)
- 2- فَعِلَ /fəʕəl/ (did)
- 3- فَعَّلَ /fəʕʕəl/ (activated)

4- فُعل /fʊʕl/ (was done)

As it is seen all the above examples are written with same letters but they have different pronunciations and meanings and when they are written without diacritics, they result in having four possible words. Such homographs represent a critical problem that machine translation may face because the correct meaning cannot always be guessed even the word is put in a sentence or a context as sometimes the same words which are used in a sentence can be pronounced differently to give different meanings as it will be shown in the data analysis below.

2. Research Methodology:

This research is concerned with the machine translation of Arabic homographs into English. 'Google Translate' (henceforth GT) is selected to test the compatibility of machine translation services to translate homographs in Arabic as this service or site is nowadays used widely to translate texts from one language into another. Throughout the career of evaluating the translations of English texts translated from Arabic, it was known that most of them are translated by GT as it is easy to access, free of charge, available to all, simple, etc.

The data selected in this paper are Arabic homographs put in sentences to check whether GT can guess the meaning of the homographs or not. The test consists of two stages. The first one includes feeding GT with Arabic homographs that are devoid of diacritics to check the ability of giving accurate translations which is the usual way of typing nowadays, i.e., without diacritics. The context of the sentence is only available aid that GT may use to determine the intended homograph which is a very complex process as several possible readings of the homograph are possible. The second stage includes repeating the same sentences to be translated by GT but they are written with diacritics to check whether GT can recognize the meanings of the homographs when they are differentiated by the diacritics or not. The data will be homographs that are used widely in everyday language and that one can rarely find a text that doesn't include a number of them.

In the analysis, the transcription of the homographs will be presented to show the pronunciation of each homograph in each sentence. The meaning of the homograph and the whole sentence will also be given to show how the lexical item may have different meanings and may have different parts of speech according to the diacritics used with it. When the translation of the sentence is given, the translation of the homograph will be underlined to determine the lexical item under analysis. A table will be used with each sample in which the first column presents the Arabic text, the second is the transcription of the homograph, the third is its meaning in English, and the last includes the translation of GT. The first row is for the sentence without diacritics which may have all the readings in the rows below it in the table. Because this sentence can be pronounced differently and may give different meanings, its transcription and meaning in Arabic is not given. Later, statistics will be given to show the rate of accuracy in translating homographs by GT.

The results of this test leads to suggest a solution to the problem of translating homographs in Arabic which can be useful in reducing the incorrect translations.

3. Data Analysis:

Sample (1):

The first homograph is the lexical item (فعل) which is put with same elements in four sentences. In the first, it is without diacritics and the other three are different readings for this lexical item as it is shown in the following table with the GT translation of all sentences:

Table (2): Sample 1 (فعل)

Source Text	Transcription	Target Text	GT
فعل الرجل ذلك	-	-	The man did that
فَعَلَ الرجل ذلك	/fəʕəl/	The man <u>did</u> that	The man did that
فَعَّلَ الرجل ذلك	/fəʕʕəl/	The man <u>activated</u> that	And the man did that
فِعْلُ الرجل ذلك	/fiʕl/	The man's <u>doing</u> of that	the man did that

In the second sentence, the underlined lexical item is the past form of the verb (يفعل) 'to do', in the third it has a different meaning which is 'activated' in which the second letter is duplicated, and in the last sentence it is nominal which refers to the action of doing something.

The homograph (فعل) without diacritics in the first sentence which can be any of the three sentences below it is translated as the past form in the second sentence neglecting the other possible meanings. The second homograph is translated accurately as the past form 'did' is used and the whole sentence is correctly formed which equals the Arabic text. The third sentence including the used homograph is translated inaccurately because the translation starts with the conjunction 'and' followed the same translation of the previous sentence and this does not give an equivalent translation. The last one also represents inaccurate translation because in the Arabic text the homograph is a noun but the translation gives the same translation of the second sentence.

Sample (2):

In this sample, the lexical item (كتب) is analyzed which includes five readings in addition to its form without diacritics as follows:

Table (3): Sample 2 (كتب)

Source Text	Transcription	Target Text	GT
كتب الولد	-	-	boy books
كَتَبَ الولدُ	/kətəbə/	The boy <u>wrote</u> (something)	the boy wrote
كُتِبَ الولدُ	/kʊtʊbʊ/	The boy's <u>books</u>	boy books
كَتَبَ الولدَ	/kəttəbə/	Somebody <u>made</u> the boy <u>write</u>	the boy wrote
كُتِبَ الولدُ	/kʊtlbə/	The boy (as a word) <u>was written</u>	write the boy
كَتَبِ الولدَ	/kəttlbi/	<u>Make</u> the boy <u>write</u> .	write the boy

It is clear that all the meanings of the lexical item (كتب) are related to writing but they vary from being the past form of the verb (يكتب) 'to write', the plural form of the noun (كتاب) 'book', making somebody write, the passive form, and the imperative form which means ordering somebody to make somebody else write.

The first sentence is translated as a noun phrase which means the boy is the possessor of the books and this represents only one of the possible

five meanings of the homograph (كتب). The translation of the second homograph is accurate as it uses the past form of the verb 'write' and the whole sentence is formed accurately which equals the Arabic one. The third homograph is a noun and the whole sentence is a noun phrase and GT makes an accurate translation in which the underlined homograph is translated as the plural form 'books'. The fourth one is an inaccurate translation because the meaning of the Arabic sentence is that somebody made the boy write something while GT translate it as 'The boy wrote' which doesn't equal the Arabic text. The fifth and sixth ones are translated as 'write the boy' and this translation does not equal the Arabic texts.

Sample (3):

This sample is related to the homograph (كبر) which may have five forms as follows:

Table (4): Sample 3 (كبر)

Source Text	Transcription	Target Text	GT
كبر المكان	-	-	get bigger
كَبِرَ المكان	/kəbʊrə/	The place <u>became big</u>	enlarge the place
كَبِّرَ المكان	/kəbbərə/	Somebody <u>made the place big</u>	enlarge the place
كُبِّرَ المكان	/kʊbbɪrə/	The place <u>was made big</u>	enlarge the place
كِبْرُ المكان	/kɪbrʊ/	<u>The size</u> of the place	big place
كَبِّرِ المكان	/kəbbɪrɪ/	<u>Make the place big.</u>	enlarge the place

In this sample, the homograph (كبر) has five meanings and pronunciations according to the type and place of diacritics used with it. They are as follows: the past form of the verb (يكبر) 'to become big', the action of making something big, the passive voice of this previous meaning, a noun which means the size or how big something is, and an imperative verb that means ordering somebody to make something big.

The first sentence which is without diacritics is not translated accurately and the given translation does not represent any of the possible readings of this sentence and the underlined homograph. The second one includes the homograph (كَبِرَ) which means that something became bigger but the translation is an imperative form which includes an order to somebody to enlarge the place and this is an inaccurate translation. The imperative sentence 'Enlarge the place' is also used to translate the third and fourth sentences and here the meaning of the two homographs in these two sentences are not translated accurately as they mean that somebody enlarged the place as a statement in the past tense and the passive voice of this respectively. The fifth sentence includes the underlined homograph (كِبْرُ) which is a noun that refers to the size of something or how big something is but the translation is 'big place' which does not represent an equivalent translation to the Arabic text. The last one is equivalent as the Arabic text is an imperative sentence in which there is an order to somebody to enlarge the place and the translation conveys exactly this meaning.

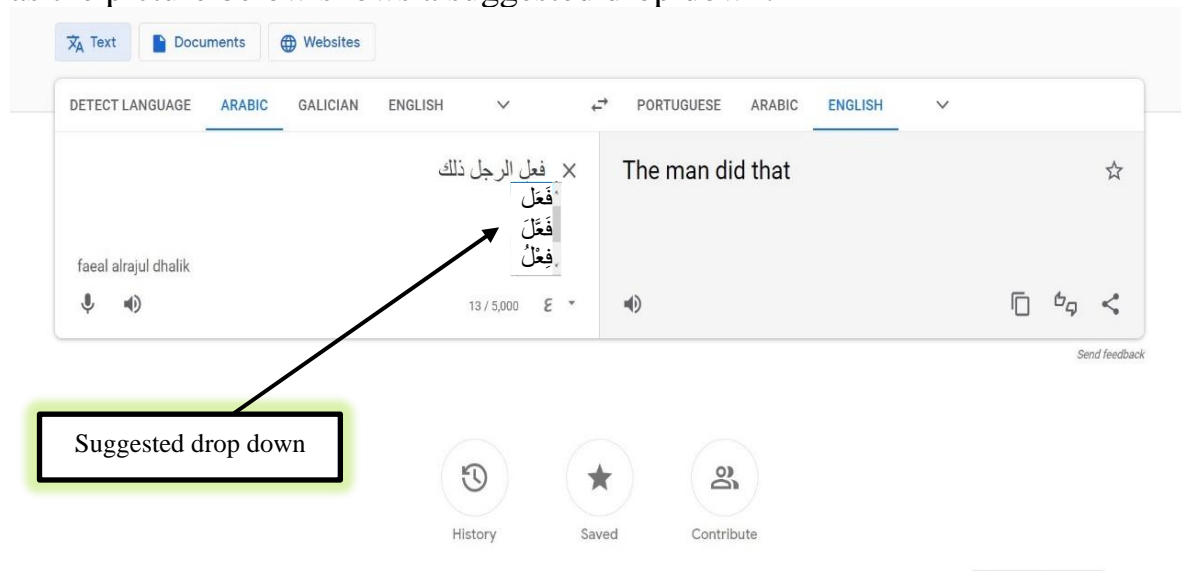
4. Findings and Discussion:

Concerning the three sentences in the first rows in which the homographs are without diacritics, the translations in the first two samples are accurate but each one conveys only one possible translation while there are two other

possible translations for the first sample and other four translations for the second. This means that all the other possible meanings are neglected and this is not even hinted by GT which means that the accuracy of these two translations represents only 25% as the possible meanings are (8) and only (2) of them are reflected in the two translations. Concerning the third sample, the translation of the homograph without diacritics is not accurate and it does not represent any of the possible meanings of this homograph.

The analyzed homographs with diacritics are (13). Only four of them are accurate translations which equal the homographs in Arabic and this represents only (30,76%) of the total number of the analyzed homograph.

One the solutions that can be suggested to avoid inaccuracy of translating Arabic homographs is that GT has to be supported with the data that make it have the property of distinguishing among the homographs in Arabic when they are written with diacritics. When the homograph is written without diacritics, a drop down should appear under the homograph which includes all the possible forms of the lexical item which have the same letters but they are differentiated by the diacritics used with the letters as the picture below shows a suggested drop down:



it is concluded that:

1. Homography is a phenomenon that is widely available in Arabic compared with English which has a limited number of pairs that are related to this phenomenon.
2. GT does not have the property that the homograph can have other translations when it is written without diacritics which presents inaccurate translations.
3. When the homographs are written with their diacritics in Arabic, they are easily read and their meanings are identified by Arab readers but GT does not always have the ability to differentiate among the forms of homographs when they are written in their full forms concerning the diacritics.
4. Drop-down is one of the solutions that GT can adopt to avoid inaccurate translations when translating Arabic lexical items which are related to homography.

References

- [1] Cowie, A. P. (2013). Semantics. Oxford: Oxford University Press.
- [2] Cruse, Alan (2006). A glossary of Semantics and Pragmatics. Edinburgh: Edinburgh University Press.
- [3] Leech, Geoffrey (1974). Semantics. London: Penguin Books.
- [4] Lyons, John (1995). Linguistic Semantics. Cambridge: Cambridge University Press.
- [5] Richards, Jack and Schmidt, Richards (2010). Longman Dictionary of Language Teaching and Applied Linguistics. London: Longman.
- [6] Saeed, John (2009). Semantics. London: Wiley-Blackwell.
- [7] <https://translate.google.com>.