



Translation Quality Assessment of Translating Machines: the Arabic geminate Lexemes as examples

A M Fathi¹   and O H Ebraheem²  

¹Directorate General of Education in Nineveh, ²Department of Translation, College of Arts, University of Mosul

Article information

Article history:

Received: 12 June 2025

Revised: 12 July 2025

Accepted: 22 July 2025

Keywords:

Machine translation,
translation quality assessment, geminate
lexemes,
Arabic,
English

Correspondence:

ameera.23arp169@student.uomosul.edu.iq

osama.ibrahim@uomosul.edu.iq


DOI: <https://doi.org/10.69513/jnfh.v4.i2.a2> ©Authors, 2025, College of Education, Alnoor University.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract

Machine translation, a recent development in the field, requires continuous evaluation and development to achieve the highest levels of translation, as free as possible from serious errors. One challenge for Arabic-English machine translation is the system's ability to recognize words that double, which can change meaning because Arabic morphology affects semantics. This study selected ten of the best-rated machine translation systems. Four pairs of Arabic words, differing only in a doubling sound, were translated by these machines. The machines compared and evaluated the translation to determine the machines' accuracy in distinguishing between words with different doubling sounds. The results revealed significant inconsistency in the translation of these word pairs.

تقييم جودة الترجمة الآلية: أمثلة على الكلمات العربية المضعفة

اميرة محمد فتحي¹   اسامة حميد ابراهيم²  

¹المديرية العامة للتربية في نينوى، الموصل، العراق، ²قسم الترجمة، كلية الآداب، جامعة الموصل، الموصل، العراق.

المستخلص

يُعدّ الترجمة الآلية تطورًا حديثًا في مجال الترجمة، ويتطلب تقييمًا وتطويرًا مستمرين للوصول إلى أعلى مستويات الدقة، مع الحرص على خلوها قدر الإمكان من الأخطاء الجسيمة. ومن التحديات التي تواجه مستخدمي الترجمة الآلية من العربية إلى الإنجليزية قدرة الآلة على تمييز الكلمات التي تحتوي على تكرار صوتي، مما يؤدي إلى تغيير في المعنى نظرًا لتأثير الصرف العربي على الدلالة. اختارت هذه الدراسة عشرة من أفضل أنظمة الترجمة الآلية تقييمًا. وقامت هذه الأنظمة بترجمة أربعة أزواج من الكلمات العربية، تختلف فقط في صوت التكرار الصوتي. ثم قارنت الأنظمة الترجمات وقيمتها لتحديد دقتها في التمييز بين الكلمات ذات أصوات التكرار المختلفة. وكشفت النتائج عن تباين كبير في ترجمة هذه الأزواج من الكلمات. الكلمات المفتاحية: الترجمة الآلية، تقييم جودة الترجمة، الكلمات المكررة، العربية، الإنجليزية.

1. Introduction

The rapid development of artificial intelligence and the proliferation of machine translation software make it crucial to continually evaluate the performance of these programs to achieve the highest levels of accuracy. Individuals and institutions have increasingly relied on this type of software because it saves time, effort, and cost. The future of translation is increasingly leaning toward relying on these applications. This does not mean that these technologies are perfect and without problems. Therefore, human oversight remains important in examining and evaluating the performance of these machines. One problem in translation is the shift between discrete and undecidable languages. One example of this is the translation between Arabic and English. Arabic is a discretized language, i.e., diacritics have grammatical and semantic functions. Depending on our experience in this field, we noticed that most, if not all, machines are not sensitive to diacritics. This will necessarily lead the machines not to distinguish between discretized and undiacritized lexemes, resulting in inaccurate translations. Some text types in certain fields require the differentiation between these two types of lexemes, such as legal and religious texts.

Habash (2010, p. 121) (1) states that Arabic is the most morphologically complex language. One manifestation of this complexity is Arabic's diacritical system. Habash (2010, p. 120) (1) adds that the absence of Arabic diacritics adds to the ambiguity of translating from Arabic to English.

Diacritics in Arabic - known as /tashkīl/ (تشكيل) - are graphical symbols added above or below consonants to indicate short vowels, vowel absence, or other phonological and morphological features. Diacritics are crucial in formal, religious, or pedagogical texts to prevent lexical and grammatical ambiguity. They include:

- a. Short vowels: fatha (َ), damma (ُ), kasra (ِ)
- b. sukūn (ْ): absence of vowel
- c. tanwīn (ً, ٍ, ٍ): markers of indefiniteness
- d. Shadda (ّ): indicates gemination or consonant doubling

Ryding (2005, p. 49) (2) states that Arabic morphology is semantically driven; the various verb forms (derived stems) carry particular semantic values

such as causative, reflexive, intensive, and reciprocal meanings. Diacritics are part of the morphological structure of Arabic words. Therefore, translators should pay special attention to the unclear differences between geminate and non-geminate words. In machine translation, treating Arabic diacritics as inflectional elements will certainly degrade translation quality. Some diacritics in Arabic can result in grammatical or semantic changes, such as changing the transitivity of verbs or intensifying meaning.

This study aims to examine the limitations of current MT systems in handling Arabic discretized lexemes, with a focus on shadda highlighting semantic loss when shadda is ignored.

Simply put, this study will assess the quality of machine translation of geminate lexemes using the following framework:

1. Listing the pairs of lexemes that differ only in shadda (gemination marker) which are expected to receive identical translation in spite of their morphological differences.
2. Putting the lexemes in meaningful sentences
3. The selected translation machines will be fed with each sentence at a time and the images of the resulting translations will be inserted below each lexeme.
4. The translations will be assessed for quality.

The ultimate aim is to show the accuracy of the available machines in translating the diacritized Arabic geminate lexemes into English.

This study addresses the following research questions:

1. How do current MT systems handle Arabic geminate words with or without diacritics, especially shadda?
2. What are the semantic consequences of ignoring such diacritics during translation?

The study depends on the fact that Arabic morphology is semantically based and that diacritics are integral to this structure. By analyzing the behavior of MT systems when faced with discretized lexemes.

1. Shadda (ّ): The Gemination Marker

Arabic shadda (or tashdeed) ('intensification') indicates gemination. Ryding (2005, p. 688) (2) defines gemination as a term applied to lexical roots wherein the second and third root consonants are identical. He adds (2005, p. 24) that gemination refers to the doubling of a consonant so that it is pronounced

longer in duration than a single (non-geminate) consonant. He emphasizes that shadda does not normally appear in written text, but it is necessary to know when it does. This point is very important when machine translation (MT) is involved. Shadda is one of the most important diacritics in Arabic because it is mainly derivational since it changes meaning. Examples such as كَتَبَ /kataba/ (he wrote) and كَتَّبَ /kattaba/ (he made somebody to write) show that shadda is contrastive. He adds that gemination results from a derivational process, that is, it can change word meaning and create words. Another example is the verb stem /darasa/ (to study) and /darrasa/, (with doubled /...raa.../ meaning to teach). He states that the meanings are related but not identical. There are types of shadda, but this study is interested in shadda that results in geminate lexemes with non-geminate counterparts, as in the examples above. Ryding (2005, p. 251) (2) states that, in Arabic, this lengthening is contrastive because it changes the meaning of a word. Therefore, the existence of shadda is supported to be accompanied by a difference in meaning in translation.

2. Morphological and Semantic Impact

Shadda often marks a derived verb form with a different meaning from the base form. It serves as a semantic operator, adding meanings such as:

- a. Causatively (e.g. /kattaba/ “he caused to write” vs. /kataba/ “he wrote”)
- b. Intensiveness (e.g., /qaṭṭa‘a/ “he cut to pieces”)
- c. Reciprocity or reflexivity (e.g. /takassarat/)

McCarthy (1981, p. 375) (3) states that some roots change, such as the gemination of the middle radical, yielding derivatives such as causatives or agents. This means that it is contrastive and must be treated as such.

4. Computational and Translational Implications

The persistent and largely overlooked challenge remains: the accurate translation of diacritized Arabic words—particularly those containing shadda (ّ), which signals gemination or consonant doubling. Diacritics in Arabic are not mere pronunciation aids; they frequently carry essential grammatical and semantic information (Watson, 2002, p. 139) (4). For example, the verb كَتَبَ /kataba/ means “he wrote,” while كَتَّبَ /kattaba/ - differ in a shadda - means “he caused (someone) to write.” It is hypothesized here that most, if not all, MT systems will translate them identically, ignoring the morphological distinction. The problem is that even diacritized words are treated by translating machines (TMs) as undiacritized, which means that such machines are not trained to deal with diacritized words as separate words that differ from their undiacritized counterparts. Belinkov & Glass (2015) (5) developed a model for a recurrent neural network with long short-term memory layers for predicting

diacritics in languages like Arabic. The problem in this study is not the translation machines’ inability to predict diacritics, but the problem lies in the treating both forms (diacritized and undiacritized words) as one thing.

4. Data analysis and findings

In this section, available translation machines will be assessed for the translation of Arabic geminate lexemes that are part of pairs of lexemes that differ only in shadda. The analysis is based on the degree of each machine’s sensitivity to diacritized lexemes. The study uses a website that combines many (TMs) in one page (Tomedes, 2025) (6). The idea is that a non-geminate lexeme is entered into the space allotted for translation, and by clicking “translate,” all the selected machines’ translations will appear below. After that, the geminate lexeme of the pair is entered, and the translations of the two lexemes are compared to see which machine is sensitive to diacritics, i.e., which one provides a different translation for the two members of the pair. The aim is to show the machines that translate from Arabic to other languages, focusing on their sensitivity to diacritics. There are different classifications of the available (TMs), but this study will not rely on prior assessments of these machines. The analysis will classify machines according to their ability to handle the subtle differences in meaning between diacritized geminate lexemes and their non-geminate counterparts.

Aziz (1989, pp. 30-31) states that geminate lexemes derived from a non-geminate base express intensification. He presents pairs such as قطع /qaṭa‘a/ (to cut) and قَطَعَ /qaṭṭa‘a/ (to cut into pieces); and كسر /kasara/ (to break) and كَسَّرَ /kassara/ (to break into pieces) as examples of intensification by gemination. He also presents examples of causation or transferring an intransitive verb into a transitive verb: كَتَبَ /kataba/ (he wrote) and كَتَّبَ /kattaba/ (made somebody write). These pairs show the grammatical and semantic effect of shadda (gemination). Therefore, when translating such lexemes using (MTs), the machines should demonstrate differences in meaning depending on whether the lexeme is geminate or not. Now let us see how different (TMs) render such pairs when translating from Arabic to English.

Pair (1) كَتَبَ /kataba/ (he wrote) vs. كَتَّبَ /kattaba/ (he dictated or made someone to write)

1. كَتَبَ /kataba/ (wrote the teacher the lesson)
كُتِبَ المعلمَ الدرسَ

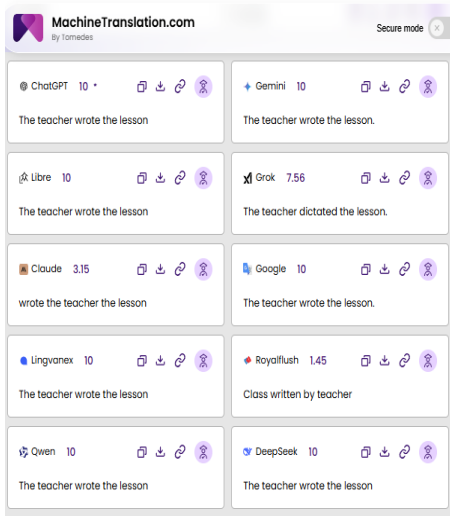


Image (1) translations of //kataba/

2. كَتَبَ /kattaba/ (dictated the teacher the lesson) كَتَبَ المعلمَ الدرسَ

Translation Quality Assessment of /kataba/ (he wrote) vs. كَتَبَ /kattaba/

The two verbs /kataba/ and /kattaba/ make a pair that share the meaning of (write), and both are grammatically similar in being transitive. However, the latter has an inherent prepositional phrase (to somebody). The important thing is that the gemination in the latter requires that the MT treat the two forms of the same base as distinct lexemes when translating from Arabic to English. In the translation of the former, one machine (Royalflush) failed to translate the sentence accurately. The interest of the present study is in the translation of the diacritized lexemes, that is to say, the second member of the pair, and in this case, it is /kattaba/. This is supposed to be different from the undiacritized non-geminate lexeme. Image (2) shows that all machines failed to translate the sentence accurately except one (Grok). Once again Royalflush failed to translate the sentence. This shows clearly that there is a serious problem in translating this diacritized lexeme by 90% of the sample machines, which are considered the best machines according to ranking by AI machines.

Pair (2) قَطَعَ /qaṭa‘a/ (cut) vs. قَطَّعَ /qaṭṭa‘a/ (cut to pieces)

1. قَطَعَ /qaṭa‘a/ قطع الرجل الحبل (cut the man the rope)

Image (3) translations of قَطَعَ /qaṭa‘a/

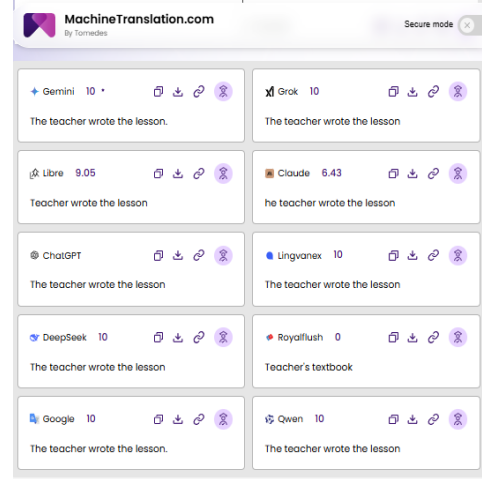
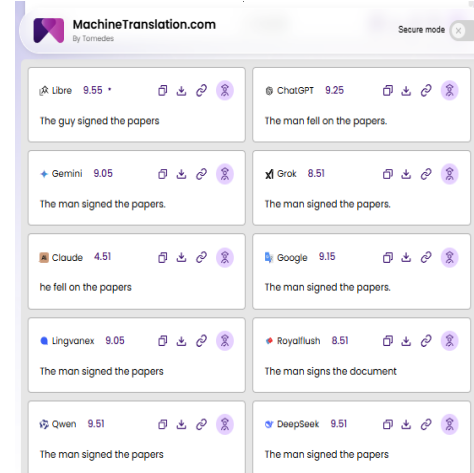
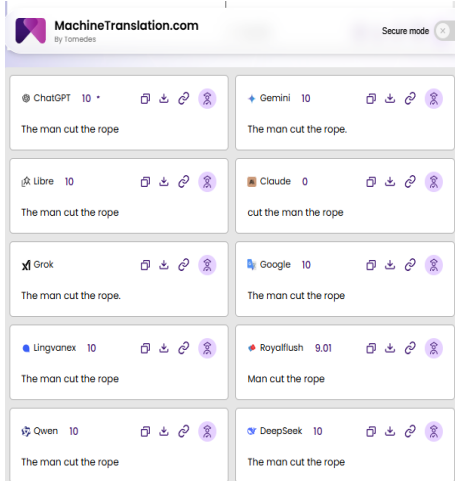


Image (2) translations of //kattaba/

2. قَطَّعَ /qaṭṭa‘a/ قطع الرجل الحبل (cut to pieces the man the rope)

Image (4) translations of قَطَّعَ /qaṭṭa‘a/



Translation Quality Assessment of قَطَعَ /qaṭa'a/ (cut) vs. قَطَّعَ /qaṭṭa'a/ (cut to pieces)

The equivalent of the first member of the pair /qaṭa'a/ is (cut) and it is expected that most, if not all, machines will render it accurately. It is clear from image (3) that as expected all machines rendered it as (cut). The second member of the pair with gemination diacritic قَطَّعَ /qaṭṭa'a/ should be translated as (cut to pieces). This difference can have serious consequences in legal and political discourse. Therefore, mistranslating this lexeme is not a simple mistake. All the TMs translated this lexeme into “cut” ignoring the intensification accompanying the shadda (gemination). There is a big difference between cutting something into two pieces and cutting something into pieces. This means that these machines are not sensitive to diacritics, or the programmer did not provide enough information for the machine to notice the difference. The correct rendering of the latter sentence should be something like:

- *The man cut the rope to pieces.*

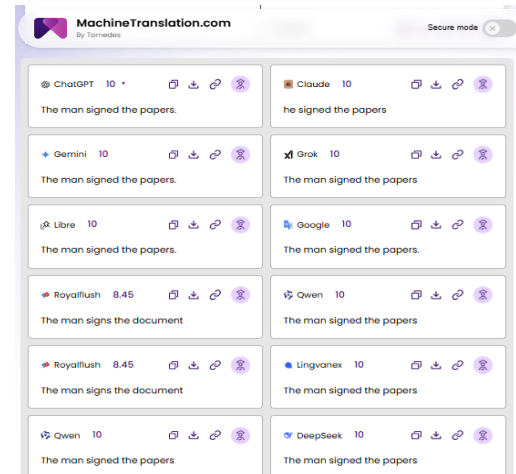
Pair (3) وَقَّعَ /waqqa'a/ (signed) vs. وَقَعَ /waqa'a/ (fell)

1. وَقَعَ /waqa'a/ (fell) وَقَّعَ /waqqa'a/ (signed) (The man fell on the papers)

Image (5) translations of وَقَّعَ /waqqa'a/

2. وَقَّعَ الرجل على الأوراق /waqqa'a/ (signed) وقع الرجل على الأوراق (The man signed the papers)

Image (6) translations of وَقَّعَ /waqqa'a/

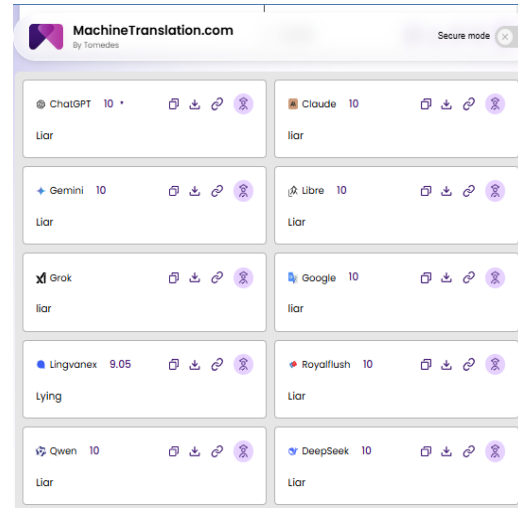


Translation Quality Assessment of وَقَّعَ /waqqa'a/ (signed) vs. وَقَعَ /waqa'a/ (fell)

Some pairs of Arabic lexemes that differ only in one having the shadda (gemination) do not share a common basic meaning. They can pose a problem for the TM if not diacritized or if the machine lacks an efficient mechanism to infer meaning from context. One example of this type of pair is /waqa'a/ (fell) and /waqqa'a/ (signed). The non-geminate form, /waqa'a/, means (fell) while the geminate form means (signed). We know that some contexts can be misleading i.e. the

Context does not help the machine guess the meaning because of the similarity of the elements surrounding the lexemes.

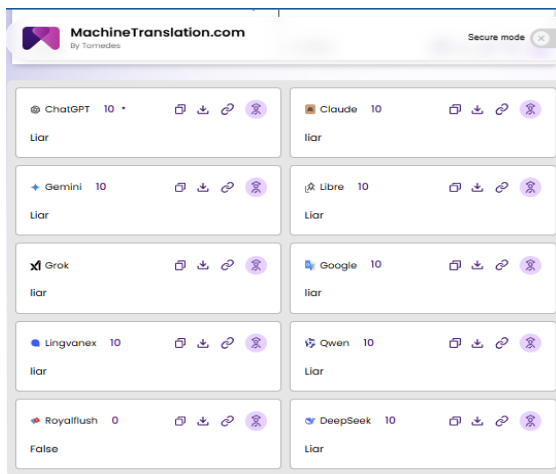
Image (5) shows the translations of /waqa‘a/ by the ten TMs. We can see that (8) of the machines mistranslated the verb into (signed) while only (2) chose the correct equivalent (fell). This reveals the confusion created by the context that misleads the machines into making wrong choices. Image (6), on the other hand, lists the translations of /waqqa‘a/ (signed). Although all machines translated the lexeme accurately, they still mistranslated the undiacritized form. The problem here is that the machine should consider the undiacritized form as having the meaning (fell) even if the context is misleading. This reveals a serious programming defect in these machines. Only the diacritized form should be interpreted as meaning (sign).



Pair (4) كاذب /kādib/ (liar) vs. كَذَاب /kaḏḏāb/ (habitual liar or big liar)

1. كاذب /kādib/ (liar)

Image (5) translations of كاذب /kādib/



2. كَذَاب /kaḏḏāb/ (habitual liar or big liar)

Image (6) translations of كَذَاب /kaḏḏāb/

Translation Quality Assessment of كاذب /kādib/ (liar) vs. كَذَاب /kaḏḏāb/ (habitual liar or big liar)

This pair presents an attribute of lying. In Arabic, as is previously stated, morphology is semantically based. Therefore, these two lexemes which share the basic meaning of (liar), differ in having the intensified meaning of “habitual” or “big” added by the gemination of the phoneme /ḏḏ/ in the second lexeme. This additional meaning should be considered by the machine as a difference and hence the equivalent should not be one and the same. The former usually indicates that the person being described as /kādib/, (without gemination) lies only this time not always. However, the latter attaches to the person the characteristic of repeated lying. This has social, pragmatic and psychological consequences. Image (5) shows that all machines used the accurate equivalent of the lexeme, /kādib/. However, image (6) which presents the machine translations of the geminate intensified lexeme /kaḏḏāb/ also rendered the lexeme as (liar), which shows that the machines do not differentiate between the two different lexemes. The accurate translation of this lexeme should be (habitual liar or big liar) as mentioned earlier.

Conclusion

The machine translations of the four pairs by the selected (10) machines which are the top ranked machines show a serious problem in either misinterpreting the diacritics if they exist or they ignore the diacritics as non-functional elements. The increasing dependence on machine translation by different elements of society dictates on all specialists in the fields of computational linguistics and people

working in the field of translation to look for solutions to ensuing problems in order to reach the best levels of MT.

As far as the present problem is concerned, we suggest that machine translation systems should be fed with a separate entry for each member of a pair that differs in

shadda and, consequently, in meaning or grammatical function.

This study demonstrates the need for diacritic-aware design in future Arabic MT development. It is recommended that Arabic diacritics be integrated into machine code.

References

1. Habash N. *Introduction to Arabic Natural Language Processing*. Morgan & Claypool. 2010. doi: 10.1007/s10590-011-9087-
2. Ryding K. *A Reference Grammar of Modern Standard Arabic*. Cambridge University. 2005 Press. https://imamhamzatcoed.edu.ng/library/ebooks/resources/Modern_Standard_Arabic_Reference_Grammar.pdf
3. McCarthy J J. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*. 1981;12(3), 373–418.
4. Watson J C. *The Phonology and Morphology of Arabic*. Oxford University Press. 2002.
5. Belinkov Y, & Glass J. Arabic Diacritization with Recurrent Neural Networks. *Belinkov, Y., & Glass, J.R. (2015). Arabi Conference on Empirical Methods in Natural Language Processing*; pp:2281–2285. Association for Computational Linguistics.
6. Tomedes *MachineTranslation.com*. Retrieved from. 2025 <https://www.machinetranslation.com/>
7. Aziz Y Y. *A Contrastive Grammar of English and Arabic*. University of Mosul. 1989