

Performance of Neural Machine Translation tools in translating Arabic non-diacritised terms: a case study of Algerian Family Code

Anes Stiti
University of Algiers 2
anes.stiti@univ-alger2.dz

Recibido: 10.12. 2024
Revisado: 17.05. 2025
Aceptado: 13. 06. 2025

Abstract This paper aims to evaluate the output of three machine translation tools that rely on the neural approach and artificial intelligence, it seeks to measure the performance of these tools in rendering Arabic non-diacritised legal words extracted from the Algerian Family code written originally in Arabic. To achieve this, a sample of 10 passages containing Arabic non-diacritised terms was selected, the criteria of selecting the passages was based on containing words having multiple meanings according to varying diacritics put on their letters. The samples were given to three different neural machine translation tools to be translated from Arabic to English, the results were then compared according to the error-rate of each tool. The case study showed varying results for the three machine translation tools, it demonstrated the ability of machine translation in disambiguating the correct meaning of Arabic terms from context even with the absence of diacritical marks.

Keywords: Neural Machine translation; Artificial intelligence; Legal texts; Algerian family code; Diacritics; MT evaluation.

Efectividad de las herramientas de traducción automática neuronal en la traducción de términos árabes sin diacríticos: estudio de caso del Código de Familia argelino

Resumen: Este estudio tiene como objetivo evaluar el rendimiento de tres herramientas de traducción automática neuronal (NMT), las cuales se basan en la inteligencia artificial y el enfoque neuronal, en la interpretación de términos jurídicos árabes no diacríticos al inglés. Los términos fueron extraídos del Código de Familia argelino, originalmente redactado en árabe. Para lograr este objetivo, se seleccionó una muestra de diez pasajes basada en la presencia de palabras árabes no diacríticas con múltiples significados posibles, dependiendo de las diacríticas aplicadas a sus letras. Los pasajes seleccionados fueron traducidos por tres

herramientas NMT diferentes del árabe al inglés, y sus resultados fueron analizados y comparados calculando la tasa de error de cada herramienta. El estudio revela resultados variables entre las herramientas y destaca sus respectivas capacidades para desambiguar el significado correcto de los términos árabes utilizando el contexto, a pesar de la ausencia de marcas diacríticas

Palabras clave: Traducción automática neuronal; Inteligencia artificial; Textos legales; Código de Familia argelino; Diacríticas; Evaluación de la traducción automática.

Sumario: 1. Introduction. 2. Literature review. 3. Arabic Diacritics. 3.1. The Fathah. 3.2. The Dammah. 3.3. The Kassrah. 3.4. The Tanwin Nunation. 3.5. The Shaddah. 3.6. The Sukun. 4. Challenges of translating legal texts. 5. Neural machine translation. 6. Methodology. 7. Data analysis. 8. Discussion. 9. Conclusion.

1. Introduction

In recent years, the field of machine translation has witnessed great improvements and changes especially with the emergence of Neural machine translation, thus changing preconceived opinions and beliefs of the failure of this mechanism.

Nowadays, machine translation is widely used among translators and professionals due to its efficiency and rapidity in translating different types of texts, in fact, machine translation is said to translate 143 billion words per day which is 150 times more than what all human translators combined can achieve in one single day (Pym & Torres-Simon, 2021). Besides, neural machine translation tools can translate from and between many of the world's languages, such as google translate who can provide translation for 243 languages (Whitney, 2024), Bing translator with over 100 languages (Microsoft, 2021), and Systran translator with 55 languages (systransoft, 2024).

The Arabic language being one of the languages covered by neural machine translation tools, is getting more and more attention by MT engineers due to the large number of its speakers and users reaching nearly 400 million worldwide (Alnosairee, 2021). However, the system of the Arabic language is rather unique by its use of *Tashkil* i.e. diacritics which serve as guiding marks to show the exact pronunciation of words as well as defining the grammatical function of Arabic words. (Hallberg, 2022)

The presence of Arabic diacritics is highly important for different types of texts, especially sacred texts like the Quran, or sensitive texts like legal texts whose misinterpretation may engender legal liability, or pose legal

disputes, yet, a high percentage of Arabic writing is written without any diacritic due the possibility of disambiguation through context (Mubarak, et al., 2019). However, machine translation tools may face challenges when translating Arabic terms that have the same form but whose meaning may vary according to the diacritic below or above each one, in fact, they may struggle choosing the exact equivalent especially for words that may have more than four meaning; for instance, the Arabic term "قَدَم" literally meaning foot, may have several meanings without diacritics according to Almaany (Almaany, 2024); it may mean: "قَدَمَ" *to come*, "قَدَمَ" *offered*, "قَدَمَ" *getting old*, and "قَدَمَ" *foot*.

This study aims to explore the performance of neural machine translation tools in translating non-diacritised Arabic legal terms extracted from the Algerian Family code, furthermore, it aims to compare three different neural machine translation tools and explores the difference between them by asking the following research questions:

- Are neural machine translation tools able to translate non-diacritised Arabic legal terms correctly?
- To what extent can neural machine translation tools simulate human brain and guess the meaning of non-diacritised Arabic legal terms from context?
- Is there any difference between NMT tools efficiency in translating non-diacritised Arabic legal terms?

To help find answers to the above-mentioned research questions, the following hypotheses were suggested;

- Neural machine translation tools may translate some non-diacritised Arabic legal terms and fail to translate others.
- Neural machine translation tools may guess the meaning of non-diacritised Arabic legal terms from context through Deep learning structures and algorithms that provide solutions such as attention mechanism.
- The outcome of neural machine translation tools may vary in translating non-diacritised Arabic legal terms according to the training and data base of each tool.

2. Literature review

The performance and efficiency of machine translation has attracted lot of attention in recent years, this tool can translate thousands of words in just few seconds and assist linguists and translators in their daily tasks. However, MT needs to be evaluated regularly by experts, whether through human or automatic metrics (Freitag, et al., 2021) in order to enhance its performance and efficiency, more specifically, several studies have dealt with the evaluation of machine translation translating from Arabic to English and vice versa.

In the study of (Hadla, Hailat, & Al-Kabi, 2014), the researchers evaluated the efficiency of two online machine translation tools; Google and Babylon translators using the BLEU metric; which is a widely used and famous algorithm for evaluating machine translation outcome (Reiter, 2018). The research was conducted on a corpus of 1000 Arabic sentences and their translations and concluded that Arabic sentences posed serious challenges to MT tools, however, a slight difference between the two programs was reported on the level of precision in favour of Google translation. Furthermore, the tools evaluated back then in 2014, were using Statistical Machine Translation approach which is no longer adopted by the programs in the time being, in fact, the majority of MT systems migrated to Neural Machine Translation approach for the promising results it yielded.

Furthermore, In the study of Stiti & Benaouda (2023), a human evaluation was conducted on Google Translation, to demonstrate its efficiency in translating four Algerian codes to English, the study focused on spotting errors according to levels of language (Lexical, Semantic, Syntactic, Pragmatic), the study concluded that even though Google translation was able to translate effectively on the lexical and syntactic level, it still faced challenges at the semantic and pragmatic levels when translating culture and religious-specific terms of Algerian codes. Furthermore, the study reported that post-editing required for the outcome of google translate is minimized compared to previous studies.

In another study related to measuring and evaluating the performance of machine translation tools, Saleh & Tagadeen (2023) focused on the aspect of Diacritics and its effects on the performance of machine translation, the researchers chose Yemeni proverbs as a corpus for their study, the experiment was then conducted by selecting passages containing Arabic diacritised homographs which were translated by three different machine translation tools; Systran translate, Reverso translate, and Free Translation Online, they were then compared according to error rate, and the findings revealed that Systran translate mistranslated all the samples

provided, meanwhile, Reverso's and Free Translation's error rates were 75% and 66.6% respectively. The results show a high error rate among all machine translators even though the passages provided by the authors were diacritised, this could be interpreted as an incapacity of machine translation tools in recognising or identifying Arabic Diacritical marks.

In essence, the previously mentioned studies reveal a multitude of challenges related to using machine translation tools for translating Arabic texts to English, the results did not report results similar to human like translation, but we hope to achieve better results through our study which will be focused on a slightly different scope, for we intend to assess the performance of machine translation tools in disambiguating Arabic non-diacritised words through context, the study will be focused on legal texts, more specifically on Algerian Family code.

3. Arabic Diacritics

Diacritics are marks or symbols put above or below letters acting as guides to the right spelling and reading of words (Layla, 2018), they may change the whole meaning and pronunciation of words, those signs are used in several languages; such as the *ê, ç, î* in French, *ö, ş* in Turkish, or the *ü* in German (Trask, 1997), in the Arabic language this characteristic is called *tashkil* meaning literally *taking shape*, and the difference between the use of diacritics in Arabic and other languages is that in Arabic they may change the whole grammatical function of a word from Nominative to Accusative. besides, they may change the voice from active to passive: for instance, the passive in Arabic is formulated just by changing the Diacritical mark from *Fathah* *◌َ* above the first letter of the word to a *Dammah*.

Arabic has eight types of diacritics (Hallberg, 2022) acting as indicators to the real pronunciation of a word, they are called: *Al fathah, Al kassrah, Al dammah, Tanwin*; which has three types (*Tanwin Fath, Tanwin Kassr, Tanwin Damm*), the *Shaddah*, and *Sukun*.

3.1 The Fathah

The *Fathah* in Arabic is a small diagonal stroke *◌َ* put above the letter to indicate that the sound of the vowel is similar to the English /ə/, for instance if we add a *Fathah* to a *b* sound, it will be pronounced as [bə].

3.2 The Dammah

The *Dammah* in Arabic is a small curl ُ put above the letter to indicate the /u/ sound, for instance, if we add a *Dammah* to the *b* letter, it will be pronounced as [bu].

3.3 The Kassrah:

The *Kassrah* in Arabic is a small diagonal stroke ِ put below the letter to indicate that the sound of the vowel similar to the English /i/, for instance if we add a *Kassrah* to a /b/ sound, it will be pronounced as [bi].

However, the three diacritical marks do not change just the pronunciation of words, but the whole meaning as well, to understand more we give the example of the word "كتب" literally meaning *wrote* in English may have several meanings according to the three above-mentioned marks; it may be written with a *Fathah* on all its three letters "كُتِبَ" to refer the past tense of the verb *to write*, it may be written with a *Kassrah* on its second letter "كُتِبَ" to refer to the passive voice of the verb *to write* which is *written*, and may be written as well with a *Dammah* on its first and second letter "كُتِبَ" to refer to the plural tense of the word book which is *books*.

3.4 The Tanwin Nunation

The *Tanwin* in Arabic has three types as it follows the above-mentioned diacritical marks, it equals doubling the *Fathah*, *Kassrah*, or the *Dammah*, and serves initially to differentiate between nouns and other parts of speech (Ibnulyemen Arabic, 2024), phonetically it produces the sounds [ən] with the *Fathah* Tanwin, [in] with the *Kassrah* Tanwin, and [un] with the *Dammah* Tanwin.

3.5 The Shaddah

The *Shaddah* in Arabic has the shape of a small w letter (Arabic Path, 2021) "ّ" put above Arabic letters to indicate doubling the consonant (Hallberg, 2022), it equals putting stress in English, in other words, it is like combining two similar letters and writing only one letter plus the *Shaddah* symbol, for in Arabic to put stress on a letter is to double it, for instance: the

Arabic word "الطَّعام" literally meaning food, is originally "الططعام", however, due to aesthetic considerations, the doubled consonant is merged into only one letter plus adding the *shaddah*.

3.6 The Sukun

The *Sukun* in Arabic is a small circle ◌ put above a letter to indicate that it is *Sakin* literally *vowelless*, which means that the letter which has *sukun* above it does not take any vowel and is pronounced by taking a pause on it (Greer, 2024). The *sukun* plays a crucial role in the pronunciation and interpretation of Arabic words, for instance, the word "نفس" when diacritised with a *Sukun* above its second letter means *soul* in English, and when diacritised with a *Fathah* above the same letter becomes the equivalent of *Breath* in English, this may be a real challenge for novice learners of Arabic, and for machine translation as well as they have to consider the possibility of having several meanings for only one word.

4. Challenges of translating legal texts

As the scope of this study will be focused on legal texts, it is crucial to mention the challenges encountered by human translators and machine translation tools in rendering these types of texts. Legal texts deal primarily with the subject of law (Rotman, 1996) which requires precision and clarity in order to convey rules and regulations and protect and people's rights and lives, legal texts are characterized by the performative nature of their language (WIECŁAWSKA, 2015, p. 2) because drafting legal texts is not a mere stating of facts or describing events but conveying obligations, permissions, or expressing prohibitions.

One of the challenges of translating legal texts is the translation of culture-specific terms (El-Farahaty, 2016, p. 482) which may lack the equivalent in the target language, such as the religious terms that belong to the Islamic Sharia abundant in the Algerian family code, or laws of Arab countries that derive their rulings from Islamic Sharia. The term "خطبة" which shall be mentioned in our case study, refers to the promise made by the bridegroom to marry the bride translated to English by the term *engagement*, however, the Arabic term may have an Islamic background that requires the approval of the bride's parent to conclude the engagement and marriage contract.

The legal translator may also face challenge of different legal systems in translating legal documents (COSMULESCU, 2021, p. 141), legal texts follow the structure and drafting style of the legal system of their country. we can list five different legal systems around the world; ``They are civil law, common law, customary law, religious law, and hybrid or mixed systems.`` (openstax, 2024), these legal systems can be found in different parts of the world, for example: the legal system of English speaking countries like the United States and Britain is Common Law, which is based on the system of precedents, it relies on past judgments and cases to decide new ones, on the other hand, the Civil law system relies on written codes and laws and decrees that define the rights and obligations of persons.

In addition, there are cases where a mixture of more than one legal system can be found in a country, such as the Algerian legal system that adopts the legal system but relies on the Islamic Sharia in drafting its codes (legalcommunitymena, 2023), another example can be found in the Philippine, where more than one legal system is adopted; The Sharia law, civil Law and Common law (openstax, 2024).

Consequently, this must all be taken into consideration when translating legal texts and evaluating machine translation outcome, especially with the rapid growing of neural machine translation tools that became able to translate between lots of languages, and assist translators in different types of texts.

5. Neural machine translation

Machine translation is gaining more and more interest by translators, linguists and laypeople because of its free and easy access, in addition to its rapidity and the good results it achieves. However, machine translation has gone through various stages to reach its current status that relies on the Neural approach. Most machine translation tools and language service providers steered their attention to neural machine translation just few months after its emergence thus surpassing the previous approach statistical machine translation (SMT) which dominated the market of machine translation for long (Álvarez, Badia, & Oliver, 2021, p. 33). In fact, neural machine translation was able to achieve better results by relying on Artificial Neural Networks that mimic the human brain, Alvarez et al mentioned that both automatic metrics and human evaluation of MT confirmed neural machine(NMT) translation achieved better results than statistical machine translation (SMT) (Álvarez, Badia, & Oliver, 2021, p. 36).

Moreover, Escribe (2019, p. 37) conducted a human evaluation of Neural machine translation output, the study focused on the first book translated wholly by an MT tool. “Deep Learning” was translated by DeepL translator from English to French in 2018 and published later after post-editing its raw output (Escribe, 2019, p. 38). The study compared the post-edited version of the translated book with the raw output provided by DeepL translator to determine the efficiency of machine translation. The results showed that the machine translation tool produced some errors, but the raw output could be understood and was satisfying to some extent. This could be interpreted as a big step forward for machine translation tools in achieving long translation projects in a record time.

6. Methodology

To answer the questions of this research, a comparative analytic study was conducted by extracting ten passages from the Algerian family code originally drafted in Arabic. The passages were chosen specifically for containing words and terms that may have several interpretations and meanings according to diacritical marks put above or below their letters, the passages were then translated to English by three different Machine Translation tools adopting the Neural approach; Google Translator, Bing Translator, and Systran Translator.

The output of the three machine translation tools was then analysed and compared through an error-analysis basis in order to determine the efficiency and performance of each translator. We selected only non diacritised samples to evaluate the efficiency of MT tools in disambiguating Arabic words through context.

7. Data analysis

—Example 01

Source text	المادة 5: الخطبة وعد بالزواج. Literally: Engagement is a promise of marriage.	
Machine translation	Google translator	Article 5: <i>Engagement</i> is a promise to marry.
	Bing translator	Article 5: <i>The engagement</i> is a promise of marriage.
	Systran translator	Article 5: <i>Engagement</i> is a promise of marriage

The example above defines engagement using the Arabic word "خطبة" which literally means engagement when diacritised with the *Kassrah* below the letter "خ". however, the Arabic word "خطبة" may have another equivalent in the English language when diacritised with the diacritical mark *Dammah* "ُ" put above the letter "خ", meaning *sermon* (Almaany, 2024), like the sermon given by the Imam on Friday.

The word "خطبة" may constitute a challenge for machine translation tools, as it comes without any diacritical mark in the Algerian family code, supposing the human reader may disambiguate the real meaning through context.

The three neural machine translation tools were able to disambiguate the true meaning of the word "خطبة" and gave the exact equivalent English which is *Engagement*, even though the word given was free of any diacritical mark.

—Example 02

Source text	الصداق هو ما يدفع نحلة للزوجة من نقود أو غيرها المادة 14: literally : Dowry is the money or other things given to the wife with good heart.	
Machine translation	Google translator	Article 14: The dowry is <i>what is paid</i> to the wife in cash or other things.
	Bing translator	Article 14: Dowry is what a <i>bee</i> pays the wife money or other things.
	Systran translator	Article 14: Alms pays a <i>bee</i> to a wife for money or other legal things

The example above shows the importance of the dowry given to the wife as a gift (Mian Zafar, 2020, p. 1), and the way it is given; *with good heart* i.e. graciously or kindly, the Algerian family Code used a culture-specific word "نحلة" that belong to the Islamic Sharia to demonstrate that.

However, it may have two equivalents in English given the Diacritical mark put above or below it: As it may mean *graciously* or *with god heart* when diacritised with the Kassrah put below the first letter "ن", and it can mean *bee* which is the insect that gives honey when diacritised with the Fathah put above the first letter "ن" (Almaany, 2024).

The three neural machine translation tools were not able to disambiguate to word "نحلة" as it came without any diacritical mark in the Algerian Family Code, we notice that Google translator omitted it altogether, and the other translators (Bing and Systran) gave the most common equivalent *bee* which led to a semantic error affecting the whole meaning of the sentence.

—Example 03

Source text	المادة 62 : الحضانة هي رعاية الولد وتعليمه والقيام بتربيته على دين أبيه والسهر على حمايته وحفظه صحة وخلقاً. Literally : Custody is the maintenance and education of the child, his upbringing upon the religion of his father, preserving and protecting his health and morals.	
Machine translation	Google translator	Article 62: Custody is the care and education of the child, raising him according to his father's religion, and ensuring his protection and preservation of his health and <i>morals</i> .
	Bing translator	Article 62: Custody is the care of the child, his education, raising him according to his father's religion, and ensuring his protection and preservation of health and <i>morals</i> .
	Systran translator	Article 62: Custody is the care and education of a child, the upbringing of a child to his or her father's religion, the protection of the child and the preservation of his or her health and <i>morals</i> .

The example above demonstrates the importance of preserving the child's morals by using the Arabic word "خلقاً" which can have several equivalents in English according to the diacritics put above or below it. It is clear from the context of the family code article above that the author refers to the meaning of "خلقاً" with the diacritical mark *Dammah* put above the first letter which is rendered in English with the word "Morals" (Almaany, 2024).

However, the Arabic word may create ambiguity to machine translation tools as it was given without any diacritical mark, which can lead those tools to render it in English as "Creation"; the equivalent of the word "خلقاً" diacritised with the "Fathah" put above the first letter (Surah quran, 2024).

We notice that the three machine translation tools were able to discern the real meaning of the Arabic word from context and provided the exact equivalent in English, even though it was given clear from any diacritic.

—Example 04

Source text	المادة (70) : تسقط حضانة الجدة أو الخالة إذا سكنت بمحضونها مع أم المحضون المتزوجة بغير قريب محرم. Literally : The maternal grandmother or maternal aunt is deprived of her right of custody if she comes to live with the mother of the child in care who has remarried a non-mahram relative..	
Machine translation	Google translator	Article (70): The custody of the grandmother or aunt shall be forfeited if she lives with her child in her custody with the mother of the child in her custody who is married to someone other than a <i>mahram relative</i> .
	Bing translator	Article 70: The custody of a grandmother or aunt shall be forfeited if she lives with the mother of the child married to a <i>non-mahram relative</i> .
	Systran translator	Article 70: The nursery of a grandmother or aunt who lives with her mother who is married to a <i>close relative</i> is usually out of <i>place</i> .

The example above demonstrates the case where the custody of a grandmother or an aunt is forfeited or dropped, it states that marrying a non-mahram relative may drop the right of custody, using a culture specific term related to Islamic Law "محرم" which means the person or persons that a woman is not allowed to marry due to their blood relationship, relationship acquired through breastfeeding, or through marriage (GIWPS, 2022).

The Arabic word "محرم" may have several equivalents in English according to its diacritisation, it may be translated as: *muharram* which is the first month of the Islamic calendar when diacritised with the *Dammah* on the first letter, and the *shaddah* on the third letter of the word, it may also be translated as *forbidden*, or *Mahram* when diacritised with the *Fathah* on the first letter, and the *Fathah* on the fourth letter of the word to designate the meaning intended In the context of this law article.

Both Google and Bing translators were able to disambiguate the non-diacritised Arabic term and rendered it correctly through "borrowing" by the word *Mahram* widely used in Islamic texts written in English (Oxford reference, 2024), while Systran Translator could not discern the meaning of the term and translated it by *out-of-place*.

—Example 05

Source text	المادة (79) : يراعي القاضي في تقدير النفقة حال الطرفين وظروف المعاش ولا يراجع تقديره قبل مضي سنة من الحكم. Literally : In assessing maintenance, the judge takes into account the spouses' situation and living conditions. This assessment may not be called into question before one year has passed from the pronouncement of the judgment.	
Machine translation	Google translator	Article (79): When estimating the alimony, the judge shall take into account the circumstances of both parties and the living conditions, <i>and shall not review his estimation</i> before one year has passed from the ruling.
	Bing translator	Article 79: In assessing alimony, the judge shall take into account the situation of both parties and the circumstances of the pension <i>and shall not review his assessment</i> before the lapse of one year from the judgment.
	Systran translator	Article (79): The judge shall take into account the circumstances of the parties and the pension <i>and shall not review his assessment</i> before the completion of a year of the judgment.

The Arabic word "يراجع" in the example above may be translated by two equivalents to English according to its diacritical mark, the first possibility is when it is diacritised with a *Dammah* above its first letter and a *Kassrah* below its fourth letter to designate the active voice, and It can designate the passive voice when it is diacritised with *Dammah* above its first letter and a *Fathah* below its fourth letter.

It is clear from the context that the author of the article uses the passive voice to demonstrate that the judgment of the judge shall not be revised before one year has passed from its pronouncement, this can be discerned by the absence of the subject of the sentence, however, the three machine translator tools were not able to disambiguate it, and instead, used the active voice as the whole passage was free on any diacritic.

—Example 06

Source text	المادة 87: وفي حالة الطلاق، يمنح القاضي <i>الولاية</i> لمن أسندت له حضانة الأولاد. Literally: In the event of divorce, the judge entrusts the exercise of <i>guardianship</i> to the parent to whom custody of the children has been entrusted.	
Machine translation	Google translator	Article 87: In the event of divorce, the judge grants <i>guardianship</i> to the person who has been assigned custody of the children.
	Bing translator	Article 87: In the event of divorce, the judge grants <i>guardianship</i> to the person who has been entrusted with custody of the children.
	Systran translator	Article 87: In the event of divorce, the judge shall grant jurisdiction to the person who has custody of the children.

The example above demonstrates how guardianship of the child is granted to one of his parents using the Arabic word "الولاية" without any diacritical mark, this may lead to a sort of confusion for machine translation tools, for the Arabic word has two possible meanings; the first one is the English equivalent *Guardianship* when the diacritics on the word are: a *Fathah* above the third letter, and a "Fathah" above its sixth letter, and the second one is the English equivalent of *province* or *state* when the diacritics on the word are: a *Kassrah* above the third letter, and a *Fathah* above its sixth letter.

It is clear from the context of the Law article that the author is referring to the meaning of guardianship as the word "الولاية" is surrounded by terms as child, judge and custody, besides, all three machine translation tools were able to disambiguate the meaning of the term without the need of diacritical marks and translated it by the *Guardianship*.

—Example 07

Source text	المادة (90) : إذا تعارضت مصالح الولي ومصالح القاصر يعين القاضي متصرفا خاصا تلقائيا أو بناء على طلب من له مصلحة. Literally: In the event of a conflict between the interests of the guardian and those of his ward, an ad hoc administrator <i>is appointed</i> ex officio or at the request of a person with an interest, by the judge.
-------------	--

Machine translation	Google translator	Article (90): If the interests of the guardian and the interests of the minor conflict, the judge <i>shall appoint</i> a special administrator automatically or upon the request of the person with an interest.
	Bing translator	Article 90: If the interests of the guardian and the minor's interests' conflict, the judge <i>shall appoint</i> a special administrator automatically or at the request of an interested party.
	Systran translator	Article 90: If the interests of the guardian and the minor are in conflict, the judge shall automatically <i>appoint</i> a private person or at the request of the person having an interest.

The Arabic word "يعين" may have two interpretations when not diacritised, the first one is to *assist* or *help* when it has a *Dammah* above its first letter, and a *Kassrah* below its third letter, and the second meaning is to *appoint* when it is diacritised with a *Dammah* above its first letter, and a *Kassrah* below its third letter plus a double consonant or a *Shaddah*.

The three machine translation tools were able to translate the Arabic term with the right equivalent in English although the author introduced it without any diacritic.

—Example 08

Source text	لمادة 144: الأخت الشقيقة بشرط انفرادها، وعدم وجود الشقيق، والأب، وولد أصلب، وولد الابن ذكرا أو أنثى، وعدم الجد الذي يعصمها Literally: Article 144: A full sister, on the condition that she is alone, and that there is no full brother, father, a biological son, or a male or female grandson, and the absence of a grandfather who allows her to inherit through agnation	
Machine translation	Google translator	Article 144: A full sister, provided that she is alone, and there is no brother, father, <i>biological child</i> , son's child, male or female, and there is no grandfather who is her agnate.
	Bing translator	Article 144: The sister provided that she is alone, and there is no brother, father, <i>crucifixion</i> , son male or female, and not grandfather who incapacitates her
	Systran translator	Article 144: a sister on condition of her own, the absence of a brother, the father, <i>the son of a steel</i> , the son of a male or female, and the lack of a grandfather who makes her nervous

The example above lists the category of people who have the right the inherit fifth of the heritage, and among them is the biological son, who is referred to in Arabic as "ولد الصلب", a compound word of "ولد" referring to *child* + "الصلب" referring to *backbone* (Surah quran, 2024), however, diacritical marks may give the second part of the compound word different interpretations.

The first meaning being the original one with a *Dammah* on the third letter, the second meaning being *steel* with a *Fathah* on the third letter, and the third meaning being *crucifixion* with the same diacritics as the previous one.

Due to the absence of diacritical marks in the text, only Google Translate was able to disambiguate the Arabic term from the context and gave the right translation *biological son*, unlike Bing translator who translated it as *crucifixion son*, and Systran translator who translated it the *son of steel* thus, leading to a semantic error affecting the whole meaning of the Law article.

—Example 09

Source text	المادة 200: تصح الوصية مع اختلاف الدين. Literally: The will shall be valid even with different religions.	
Machine translation	Google translator	Article 200: A will is valid even if the <i>religion</i> is different.
	Bing translator	Article 200: A will is valid with a difference of <i>religion</i> .
	Systran translator	Article 200: The will is valid with different <i>religions</i> .

The example above affirms that the will of the deceased shall be valid even when the religion of the inheritors is different, using the Arabic term "الدين" to refer to religion without any diacritical mark.

The term "الدين" may refer to several equivalents according to the diacritical mark put above or below it; it may refer to *religion* when diacritised with a *Kassrah* below its third letter + a *Shaddah* above the third letter, and it may refer to *debt* when diacritised with a *Fathah* plus a *Shaddah* above its third letter.

Although the example was free of any diacritic, the three machine translation tools were able to discern the right equivalent from the context of the law article and translated it as *religion*.

—Example 10

Source text	المادة 32: يفسخ النكاح إذا اختل أحد أركانه، أو اشتمل على مانع أو شرط يتنافى ومقتضيات العقد أو ثبتت ردة الزوج. Literally: Marriage shall be annulled if one of its pillars is vitiated, or if it includes an impediment or a condition inconsistent with the requirements of the contract, or the <i>apostasy</i> of the husband has been approved.	
Machine translation	Google translator	Article 32: The marriage contract shall be annulled if one of its pillars is missing, or if it includes an impediment or condition that is inconsistent with the requirements of the contract, or if the <i>husband's apostasy</i> is proven.
	Bing translator	Article 32: The marriage shall be annulled if one of its elements is disturbed, or if it contains an impediment or condition that contradicts the requirements of the contract or if the <i>husband's apostasy</i> is proven.
	Systran translator	Article 32: The marriage shall be annulled if one of the pillars of the marriage fails, or if it includes an impediment or condition that contravenes the requirements of the contract or proves the <i>spouse's return</i> .

The example above lists the cases where the marriage contract shall be annulled, and includes the husband's apostasy which is defined as disbelieving or quitting the religion of Islam (Alalwani, 2011), the Author of the Law article chose for that the term "ردة" without any diacritical mark.

The term may be rendered by several English words as it has multiple meanings according to the diacritic above or below its letters, the Arabic term may mean *apostasy* when diacritised with a *Kassrah* below its first letter, and may mean *return* when diacritised with a *Fathah* above its first letter.

Google translator and Bing translator were able to disambiguate the meaning of the Arabic term "ردة" from the context, even though it was provided without any diacritic, unlike Systran translator who opted for the

second meaning of the word, thus leading to a semantic error affecting the meaning of the whole passage.

Table 1: number of errors of each machine translation tool and the rate error in total

Examples	Google translator		Bing translator		Systran translator	
	Number of errors	Rate	Number of errors	Rate	Number of errors	Rate
	2	20 %	3	30 %	4	40 %

8. Discussion

We find that Machine Translation tools adopting the neural approach achieved good results in translating Arabic legal words, in fact, the three tools were able to disambiguate the correct meaning of the words even though they were given without any diacritic mark. Besides, the samples were selected for having multiple meanings according to each diacritic put below or above their letters.

However, the three machine translation tools gave varying results in rendering these Arabic non-diacritised words, we notice that Google Translate achieved the best scores among all translators with an error rate of 20% i.e. succeeding in rendering 80% of the samples which is satisfying taking in account the type of text and the lack of diacritics that play a crucial role in disambiguating meaning in the Arabic language.

In contrast, the error rates of Bing and Systran translators were 30% and 40% respectively which is nearly twice the error rate of Google translator. This reflects the leading role Google plays in developing machine translation tools and the domain of language services even translating languages that belong to different families.

9. Conclusion

In conclusion, we found that Machine translation has gone through major developments by introducing neural approaches that rely on artificial intelligence in order to improve its output and translate between lots of language pairs. However, we saw that some languages may have some characteristics that other languages don't, such as Arabic which relies on diacritics to disambiguate the meaning of its words. The native speaker may

rely on context to disambiguate the Arabic words, and this is reflected in a big portion of Arabic texts that dispense with diacritical marks.

This study demonstrated how the lack of this diacritics can affect the output of machine translation in rendering legal Arabic texts and found that the output of MT tools varies. Some tools could achieve a success rate of 80%, while others achieved 70% and 60% success rate. In fact, the results of this study are satisfying when compared to the results of the previous studies mentioned in the beginning of our research.

In addition, we conclude that machine translation tools rely on context to disambiguate non-diacritised Arabic legal terms through the attention mechanism introduced to the Neural approach which can weigh all the components of each sentence, unlike the previous approaches that found difficulties in doing that.

References

- Alalwani, T. J. (2011). *Apostasy in Islam A Historical and Scriptural Analysis*. London: The International Institute of Islamic Thought.
- Almaany. (2024). Retrieved 12 03, 2024, from Almaany: <https://www.almaany.com/ar/dict/ar-en/خطبة/>
- Almaany. (2024). Retrieved 12 02, 2024, from Almaany.com: <https://www.almaany.com/ar/dict/ar-en/نحلة/>
- Almaany. (2024). Retrieved 12 02, 2024, from www.almaany.com: <https://www.almaany.com/ar/dict/ar-en/قدم/>
- Almaany. (2024). Retrieved 11 27, 2024, from Almaany.com: <https://www.almaany.com/ar/dict/ar-en/خلق/>
- Alnosairee, A. A. (2021). A Sociolinguistics Study in Arabic Dialects. *Journal of linguistics*, 1-17.
- Álvarez, S., Badia, T., & Oliver, A. (2021). Comparing NMT and PBSMT for post-editing in-domain formal texts: A case study. In *Translation, interpreting, cognition* (pp. 33-46). Language Science Press.
- Arabic Path. (2021). Retrieved 11 21, 2024, from arabicpath.com: <https://arabicpath.com/read-arabic/shadda#>
- Cosmulescu, A. M. (2021). Views on the complexity of legal translation. *Romanian Journal of English Studies*, 137-144.
- El-Farahaty, H. (2016). Translating Lexical Legal Terms Between English and Arabic. *International Journal for the Semiotics of Law* , 473–493.
- Escribe, M. (2019). Human Evaluation of Neural Machine Translation: The Case of Deep Learning. *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology*, (pp. 36-46). Varna.

- Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation. *Transactions of the Association for Computational Linguistics*, 1460–1474.
- GIWPS. (2022, 08). Retrieved 12 02, 2024, from giwps.georgetown.edu: <https://giwps.georgetown.edu/wp-content/uploads/2022/08/Mahram-Womens-Mobility-in-Islam.pdf>
- Greer, G. (2024, 02 21). *Kaleela*. Retrieved 11 21, 2024, from kaleela.com: <https://kaleela.com/id/the-arabic-sukoon-what-it-is-and-how-to-pronounce-it>
- Hadla, L. S., Hailat, T. M., & Al-Kabi, M. N. (2014). Evaluating Arabic to English Machine Translation. *International Journal of Advanced Computer Science and Applications*, 68-73.
- Hallberg, A. (2022). Principles of variation in the use of diacritics (taskīl) in Arabic. *Language Sciences*, 1-15.
- Ibnulyemen Arabic*. (2024). Retrieved 11 20, 2024, from www.ibnulyemenarabic.com: <https://www.ibnulyemenarabic.com/arabic-language/tanween-in-arabic/>
- Layla , L. M. (2018). Phonetics and Phonology: Diacritics in English. *legalcommunitymena*. (2023, 01 16). Retrieved 11 23, 2024, from legalcommunitymena.com: <https://legalcommunitymena.com/algeria-legal-system-history/#:~:text=The%20legal%20system%20of%20Algeria%20is%20based%20on%20a%20combination,for%20all%20members%20of%20society.>
- Mian Zafar, K. I. (2020, 05). *Islamic Law and Tradition of Dowry and Recovery through Family Courts*. Retrieved 11 10, 2024, from <https://zafarkalanauri.com/wp-content/uploads/2020/05/Islamic-Law-and-Tradition-of-Dowry-and-Mahr.pdf>
- Microsoft. (2021, 10 11). Retrieved 11 20, 2024, from [microsoft.com](https://www.microsoft.com/en-us/translator/blog/2021/10/11/translator-now-translates-more-than-100-languages/): <https://www.microsoft.com/en-us/translator/blog/2021/10/11/translator-now-translates-more-than-100-languages/>
- Mubarak, H., Abdelali, A., Darwish, K., Eldesouki, M., Samih, Y., & Sajjad, H. (2019). A system for Diacritizing Four Varieties of Arabic. *EMNLP* (pp. 217-222). Hong Kong: Association for Computational Linguistics.
- openstax. (2024). Retrieved 11 23, 2024, from openstax.org: <https://openstax.org/books/introduction-political-science/pages/11-3-types-of-legal-systems-around-the-world>
- Oxford reference. (2024). Retrieved 11 21, 2024, from www.oxfordreference.com:

- <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100126871>
- Pym, A., & Torres-Simon, E. (2021). Is automation changing the translation profession. *International Journal of the Sociology of Language*, 1-19.
- Reiter, E. (2018). A Structured Review of the Validity of BLEU. *Computational Linguistics*, 393-401.
- Rotman, E. (1996). The inherent problems of legal translation : theoretical aspects. *Indiana International & Comparative Law Review*, 1-10.
- Saleh, N. A., & Tagadeen, I. N. (2023). Effect of Diacritics on Machine Translation Performance: A case Study of Yemeni Literature. *International Journal of Language and Literary Studies*, 324-342.
- Stiti, A., & Benaouda, A. (2023). Translating Algerian laws using machine translation. *Cahiers de Traduction*, 421-445.
- Surah quran*. (2024). Retrieved from surahquran.com: <https://surahquran.com/english-arabic-aya-7-sora-86.html>
- Surah quran*. (2024). Retrieved 11 23, 2024, from surahquran.com: <https://surahquran.com/english-arabic-aya-6-sora-39>
- systransoft*. (2024). Retrieved 11 29, 2024, from www.systransoft.com: <https://www.systransoft.com/resources/language-catalog/>
- Trask, L. (1997). *University of Sussex*. Retrieved 11 20, 2024, from <https://www.sussex.ac.uk/informatics/punctuation/misc/diacritics#:~:text=Diacritics%2C%20often%20loosely%20called%20%60accents,indicate%20something%20about%20their%20pronunciation>.
- Whitney, L. (2024, 06 27). *Zdnet*. Retrieved 11 27, 2024, from www.zdnet.com: <https://www.zdnet.com/article/google-translate-gets-110-new-languages-with-ais-help-bringing-the-total-to-243/>
- Wieclawska, E. (2015). Textual features of legal texts in the domain of commercial law. *Studia Anglica Resoviensia*, 82-90.