IJHER International Journal of Humanities and Educational Research
ISSN: 2757-5403
Volume 6, Issue 5, October 2024

Received: 14/07/2024 **Accepted:** 10/08/2024 **Published:** 01/10/2024

CONSISTENCIES IN MACHINE TRANSLATION: THE CASE OF ARABIC-ENGLISH TRANSLATION OF MEDIA AND LEGAL TERMS BY BING TRANSLATOR AND GOOGLE TRANSLATE

Dr. Mujahed Hossien Tahir ZAYED[1]

Arab American University, Palestine

Abstract

Machine translation has recently been noticed to be used by translators as they translate media and legal terms. The present study aims at investigating the appropriateness of the translations provided by Google Translate (GT) and Bing Translator (BT) for media and legal terms from Arabic into English. The researcher employs an analytical qualitative approach to achieve the objectives of the study. Thus, the researcher relied on comparing the translations provided by GT and BT to the meanings of the given terms in specialized dictionaries to decide on the appropriateness of machine translation and the translations of the programs to each other for the sake of identifying consistency. The results of the study showed that the investigated translation programs provide more appropriate translations for legal terms with a little preference for Bing Translator over Google Translate. The results of this study also reveal that Bing Translator and Google Translate were consistent in their translations of media and legal terms. The two investigated translation programs were more consistent in translating media terms than legal terms. In light of the results, the researcher recommends that GT and BT be carefully used as they provided inconsistent translations with the meanings of the terms in specialized dictionaries in a few cases.

Key words: machine translation, media terms, legal terms, Google Translate, Bing Translator.

http://dx.doi.org/10.47832/2757-5403.28.51

[1] mujaheed.zayed@aaup.edu

Introduction

Machine Translation (MT) has become strongly present in translation industry and online platforms, such as social media networks. Moreover, it is generally used without any human intervention or post-editing in social media networks. As such, the target text obtained through machine translation has to be investigated and assessed for the sake of finding out whether these translation tools are consistent or not in terms of their translation products. Many researchers maintain that MT needs to be evaluated through developing a standardized set of formulas (Ulitkin et al., 2021; García, 2014). This need increased with the use of machine translation in translating technical terms from various fields such as medicine, law and media. Al Sharou et al. (2021) claims that the translation provided by MT is generally grammatically ill-formed. This way, human users cannot always rely on machine translation as some translations were proven to contain critical errors. In fact, it hard for MT to deal with technical terms and provide accurate or acceptable translations. To lessen such a problem, recent research has looked into automatic methods to detect critical errors in machine translation, with a view to inform users of such errors. This was framed as a track in a Shared Task on Quality Estimation (Specia et al., 2021).

Pitfalls of Machine Translation

Despite the fact that machine translation has significantly increased translation efficiency, the translation quality is still subpar due to some unavoidable issues in the original output. First of all, it is discovered that machine translation produces inconsistent terminology. As such, a phrase is employed to convey the precise idea of a specific thing. The term "inconsistency of terms" refers to the fact that during the translation process i.e. when a term is converted from the source language to the target language, it may take on various expressions, but a machine will translate these multiple expressions into various iterations of the same text in the target text. Because Arabic uses the different parts of speech for many grammatical components with morphological alterations, Arabic sentences are essentially formed using different rules and it is difficult for computers to translate them. The aforementioned Arabic formula never applies to Chinese as it uses the same part of speech for different grammatical components which also makes it difficult for a machine to analyze it (Guo & Wang, 2017). Because of the context or the various collocations of the same phrase, machine translation is prone to provide inconsistent terms, especially for lengthy documents.

In addition, machine translation provides inappropriate segmentation of punctuation. Punctuation is seen as being a crucial component of written language. Chinese punctuation currently uses symbols derived from the English punctuation system. They exhibit both Chinese language traits as well as the majority of the key features of English punctuation. As a result, there are certain discrepancies between Arabic and English punctuation marks, which account for a large portion of the disparity between Arabic and English in terms of sentence structure and expression. But in machine translation, punctuation is frequently

disregarded. For instance, Arabic uses commas and other marks as punctuation following the meaning of the stretch of a language. Machine translation will copy them into the target text during the conversion of Arabic to English, which could lead to certain translation issues.

Moreover, redundancy has been noticed in machine translation. It alludes to redundant, overlapping, or functionally repeated language in the translation (Cui & Li, 2015). Chinese is known for its repetition, and using distinct terms as synonyms in the form of four-character words to emphasize a point is particularly prevalent. However, English avoids repetition and frequently substitutes pronouns and prepositions for the repeated portion. The goal of machine translation is to fully translate the source language's text, which is difficult given the conventions of English expression.

Finally, the machine translated texts have some kind of lexical gap. Lexical vacancy is the term used to describe the difficulty of translating words completely into the target language from the source language. Due to the cultural and socioeconomic differences between China and other nations, there are lexical gaps that result in cultural defaults when translating words with cultural connotations. Aside from new words coined to reflect the advancement of the times, many ancient words have also acquired new meanings in the ever-evolving modern civilization. As a result, the quality of the translation cannot be guaranteed if the lexical void cannot be filled. Because of how quickly society is changing and the machine's initial database cannot keep up with the rate of change, or because there is insufficient research on cultural differences, machine translation cannot correctly detect the meaning of the source text, which will result in these translation issues.

Related Work

One of the key difficulties in MT research is creating high-quality automatic evaluation metrics for translation. The majority of the currently used metrics heavily rely on parallel corpora for aligned texts as references (Papineni et al., 2002). To gauge the effectiveness of MT systems, one might assess translated outputs against references. The string-based measures are of crucial importance, including BLEU (Popovi'c, 2015), taking into account the lexical matching rate for translation quality (Snover et al., 2006). BERT Score (Sellam et al., 2020), for example, is a metric that uses pre-trained language models to estimate the semantic importance of texts and has been shown to match human evaluation of machine performance (Kocmi et al., 2021). Some reference-based evaluation criteria, however, require supervised training in order to be effective (Rei et al., 2020a; Mathur et al., 2019). Although these automatic evaluation criteria are frequently used in MT evaluation, they are ineffective in low-resource language translation situations because there are no analogous references that are grounded in reality (Mathur et al., 2020). In order to identify MT quality, researchers have recently focused on quality evaluation (QE), mostly using direct assessment (DA) and post-editing (PE), which are based on human evaluation methodologies. Despite a few early attempts at automatic evaluation meter prediction that were unsuccessful (Blatz et al., 2004),

most modern QE metrics require human-annotated DA and PE data at the sentence level for training on the target language pairings. Recent developments have shown the usefulness of COMET-QE-MQM (Rei et al., 2020) on WMT shared tasks. Round-trip Translation for Quality Estimation Many NLP actions have made extensive use of round-trip translation for data augmentation (Edunov et al., 2018). There isn't a resounding consensus, nevertheless, if it were to be used to assess translation quality. There are two different evaluation paradigms, automatic evaluation and human assessments, that can be used to determine the quality of machine translation. Metrics like the BLEU and BERTScore are commonly taken into account in the experiments because the attempts on QE using RTT fall within the category of automatic evaluation. Somers (2005) first came to the conclusion that there was no correlation between RTT and forward translation quality by comparing the RTT BLEU ratings of various online statistical MT systems on two language pairings. With the studies evaluating one MT system for 10 English-centric language pairs, Koehn (2005) later corroborated earlier findings. Our effort clears up the confusion that was transmitted from SMT to NMT.

Problem Statement

The use of machine in translation industry has increased and become very common over the past few years. As such, the increasing use of machine translation can be attributed to translators' need to meet the needs of globalization in terms of providing multilingual translations for huge projects presented by multinational companied or organizations. Gambier (2014) maintains that the logical professional balance between translation supplies and demands is profoundly changing. Yet, the amount of work is not the only changing thing. Floran (2010), claims that the way in which translations are accomplished has changed over the past years. This leads us to consider the use of machine translation in terms of translation quality, especially whether machine translation is consistent or not. The researcher has noticed that media texts and legal terminology occupy a major part of translation industry and thus machine translation is widely used in translating such materials from Arabic into English. The researcher has also noticed that machine translation, especially Google Translate and Bing Translator, provide different English translations for the same Arabic media term. This inconsistency in translating media and legal terms from Arabic into English usually leads to confusing translators and thus ends up with deviated target texts. Therefore, the researcher conducts this study to investigate the consistency of Google Translate and Bing Translator in translating media and legal terms from Arabic into English.

Research Objectives

The present study aims to achieve the following objectives

1. To explore the translations provided by Google Translate and Bing Translator in translating media and legal terms from Arabic into English.

2. To investigate the consistency of Google Translate and Bing Translator in translating media and legal terms from Arabic into English.

Research Questions

The present study aims to answer the following questions

1. Do Google Translate and Bing Translator provide comprehensible translation for media and legal terms from Arabic into English?

2. Are Google Translate and Bing Translator consistent in translating media and legal terms from Arabic into English?

Research Methodology

The present study adopts a qualitative approach in which the case of Arabic media and legal terms and expressions are translated into English using Google Translate and Bing Translator. As such, the researcher will select certain number of media and legal terms and translate them using the aforementioned translation programs. Once the terms are translated, the researcher will investigate whether Google Translate and Bing Translator provide comprehensible renditions to the terms or not. Then, the researcher will examine whether each one of the two translation programs provide the consistent translations for the media and legal term. This will be achieved through comparing the translation of each term and then utilizing the strategy of parallel text analysis. This way, the researcher will determine whether the translations provided by the translation programs are consistent or not.

Data and Procedures

Based on the objectives of the study, forty media and legal Arabic terms will be selected from specialized glossaries. Then, the selected terms will be translated into English using Bing Translator and Google Translate. As such, the translations obtained from the two translation programs will be considered to find out whether they are comprehensible or not. To make sure that the translations are comprehensible, the researcher will present them to two translation experts. In the next procedure, the researcher will compare the translations provided by the programs to find out whether they are consistent in translation or not. This procedure requires the researcher to use the strategy of parallel text analysis as the process of comparing the texts to each other makes is the most appropriate to decide on the consistency of the translations.

Data Analysis

No.	Arabic Term	Bing Translator	Google Translate	Meaning in Specialized Dictionaries	Consistency in Programs	Matching with Specialized Dictionaries
1	السلبية الاجتماعية	Social negativity	social negativity	Social passive	Consistent	None
2	السلطة الرابعة	Fourth Estate	The fourth power	Fourth estate	Inconsistent	BT
3	السلك السياسي	Political Corps	Political wire	Political corps	Inconsistent	BT
4	السياسة غير المعلنة	Unspoken Policy	Unspoken policy	Backdoor policy	Consistent	None
5	السينما الناطقة	Talking Cinema	Talking cinema	Sound film	Consistent	None
6	الشخص الذي يستطلع الآراء	The person polling	The person who polls	Polltaker	Inconsistent	None
7	الصحافة السينمائية	Film Journalism	Film journalism	Filmdom	Consistent	None
8	الصحف الرخيصة	Cheap newspapers	Cheap newspapers	Dreadful papers	Consistent	None
9	الصحفي الرسمي	Official Press	Official press	Gazetteer	Consistent	None
10	الصوت الحاد	Sharp sound	The sharp sound	All -top sound	Consistent	None
11	الصندوق الأسود	Black Box	The black box	Black box	Consistent	BT
12	الصورة الأولى في الصحافة	First photo in the press	First photo in the press	Wood cuts	Consistent	None
13	الصورة التفصيلية	Detailed picture	The detailed picture	Detail image	Consistent	None
14	الصورة السلكية	Wired image	Wired photo	Wire photo	Inconsistent	None
15	الضوء الأخضر	Green Light	Green light	Green light	Consistent	BT GT
16	العقل الجماهيري	The Mass mind	The mass mind	The Mass mind	Consistent	BT GT
17	الغالبية العظمى	The vast majority	The vast majority	The vast majority	Consistent	BT GT
18	الفترة الانتقالية	Transition period	Transitional period	Lead time	Inconsistent	None
19	القدرة على الإقناع	Persuasion	Persuasion	Cogency	Consistent	None

20	القوى العظمى	Great Powers	Great powers	Superpower	Consistent	None
21	الموكل إليه	Assignee	Entrusted	Assignee	Inconsistent	BT
22	محضر المحكمة	Court Minutes	Court report	Court minutes	Inconsistent	GT
23	عقوبة الإعدام	Death penalty	The death penalty	Capital punishment	Consistent	None
24	تقاضي	Charging	litigation	Adjudication	Inconsistent	None
25	إدعاء	Claim	Claim	Allegation	Consistent	None
26	إعتقال	Arrest	Arrest	Arrest	Consistent	BT GT
27	الأهلية القانونيّة للتقاضي	Legal capacity to litigate	Legal capacity to litigate	Competency	Consistent	None
28	حكم غيابي	Judgment in absentia	Sentenced in absentia	Default judgment	Inconsistent	None
29	بيّنة	Evidence	Evidence	Evidence	Consistent	BT GT
30	المدعى عليه	Defendant	Defendant	Defendant	Consistent	BT GT
31	إحتيال	Fraud	Scam	Fraud	Inconsistent	BT
32	إعتراف	Recognition	Recognition	Confession	Consistent	None
33	التحقيق	Investigation	Investigation	Investigation	Consistent	BT GT
34	اختصاص المحكمة	Jurisdiction of the Court	Jurisdiction of the court	Jurisdiction of the court	Consistent	BT GT
35	التزييف	Counterfeiting	Fake	Counterfeiting	Inconsistent	BT
36	هيئة محلّفين	Jury	Jury	Jury	Consistent	BT GT
37	دعوى قضائية	Lawsuit	Lawsuit	Lawsuit	Consistent	BT GT
38	القتل العمد	Murder	Murder	Murder	Consistent	BT GT
39	لاغ وباطل	null and void	Null and void	Null and void	Consistent	BT GT
40	اليمين الكاذبة	False oath	False oath	Perjury	Consistent	None

Table (1) Consistencies in BT, GT and Specialized Dictionaries translations

Consistency	Occurrences	Media Terms	Legal Terms	Percentage	Media Terms	Legal Terms
Consistencies in programs	29	15	14	72.5 %	37.5 %	35 %
Inconsistencies in programs	11	5	6	27.5 %	12.5 %	15 %
BT and GT Mismatching with Specialized Dictionaries	Occurrences	Media Terms	Legal Terms	Percentage	Media Terms	Legal Terms
	21	15	6	52.5 %	37.5 %	15 %
Matching with Specialized Dictionaries	Occurrences Google Translate	Percentage		Occurrences Bing Translator	Percentage	
	13/40	32.5 %		18/40	45 %	

Table (2) Occurrences and Matching with Specialized Dictionaries

Results and Discussion

The present study aims to find out whether Google Translate and Bing Translator provide appropriate translation for media and legal terms as well as investigate the consistency in the translations provided by the two translation programs. As far as the appropriateness of the translation is concerned, the data analysis showed that Bing Translator provides a relatively more appropriate translation for both the media and legal terms. This can be figured out from the percentage of matching the translation provided by BT and the meaning of the terms in the specialized dictionaries. As such, many terms were translated by BT the providing an identical translation with the meaning in the specialized dictionaries, such as the term "إحتيال" which was translated as "fraud", "التزييف" translated as "Counterfeiting" and "الموكل اليه" "Assignee" among others. Bing Translator provided identical translation with the specialized dictionaries at a percentage of 45 %. On the other hand, Google Translate provided a lower degree of matching with meanings in the specialized dictionaries at a percentage of 32.5 %. These results also revealed that Bing Translator provides translations that match more with meanings of the specialized dictionaries in the legal terms. This could be attributed to the facts that legal terms are more fixed than the media terms and that Bing Translator is better fed with legal terminologies. Despite the fact that Bing Translator provided more appropriate translations, the occurrences were not greatly more than those of Google translate. These results clearly indicate that both Bing Translator and Google Translate provide appropriate translations for media and legal terms even though Bing Translator showed more matches with the meanings of the terms in specialized dictionaries.

In terms of consistency, the data analysis showed that the two translation programs were consistent to a certain extent when translating media and legal terms. Hence, both BT and GT provided similar translation for 29 out of 40 terms at a percentage of 72.5 %. For example, both programs translated the terms "هيئة محلّفين", "بينة" and "الغالبية العظمى" providing the meanings "jury", "evidence" and "the vast majority". These translations are also similar to the meanings of the given terms in the specialized dictionaries. It is of crucial importance to mention that the two translation programs were more consistent in translating media terms than legal terms. This is clear in the occurrence of consistent translations at 15 for media terms and 14 for legal terms at percentages of 37.5 % and 35 % in sequence.

However, Bing Translator and Google Translate were consistent in providing translations for 11 terms out of 40. The notable thing in the inconsistency is that the two programs showed more inconsistency in translating legal terms than media terms. This result is clear from the 6 occurrences of inconsistent translation for legal terms and 5 occurrences for inconsistencies in media terms at 15 % and 12.5 % in sequence.

The data analysis also showed that the consistent translations provided by Google Translate and Bing Translator for media terms are less similar to the meanings of the terms in the specialized dictionaries. On the other hand, the translations provided by the two translation programs for legal terms were more similar to the meanings of the terms in specialized dictionaries. This can be considered as an indication Bing Translator and Google Translate can be better used when it comes to translating legal terms that media terms.

Conclusions

When translating media and legal terms, translators often use machine translation, such as Bing Translator and Google Translate. The present study aimed to investigate whether the two translation programs provide acceptable translation for the aforementioned terms or not. The results of the data analysis revealed that BT and GT provide a relatively appropriate translation for media and legal terms. It is concluded that the investigated translation programs provide more appropriate translations for legal terms as these translations are similar to the meanings of the terms in specialized dictionaries with a bit higher preference for Bing Translator over Google Translate. The results of this study also indicate that Bing Translator and Google Translate were consistent in terms of the translations they provided for the media and legal terms used for the data analysis. The two investigated translation programs were more consistent in translating media terms than legal terms. Based on the results, Bing Translator and Google Translate can be used by translators in translating media and legal terms from Arabic into English as they showed consistency in translating these terms and the translations provided matched in most of the occurrences the meanings of the terms in specialized dictionaries. Yet, translators should use these translation programs carefully as they were proven to provide inconsistent translations which do not match with the meanings of the terms in specialized dictionaries in a few cases.

References

Blatz, J., Erin Fitzgerald, George Foster, Simona Gandrabur, Cyril Goutte, Alex Kulesza, Alberto Sanchis, and Nicola Ueffing. (2004). Confidence estimation for machine translation. In Coling 2004: Proceedings of the 20th international conference on computational linguistics, pages 315–321.

Cui, Q. L., & Li, W. (2015). The Character of Error Types of Post-editing: Perspective of Machine Translation Based on Scientific and Technology Materials. Chinese Science & Technology Translators Journal, 28(4), 19-22.

Edunov, S., Myle Ott, Michael Auli, and David Grangier. (2018). Understanding back-translation at scale. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 489–500.

Floran, D. (2010). Translation Tools. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 429–436). Amsterdam; Philadelphia: John Benjamins Pub. Co. Retrieved from http://public.eblib.com/choice/publicfullrecord.aspx?p=871816

Gambier, Y. (2014). Changing Landscape in Translation. *International Journal of Society, Culture and Language*, *2*(2), 2–12.

García, I. (2014). Training quality evaluators. *Revista Tradumàtica: tecnologies de la traducció*, (12):430–436.

Guo, G. P., & Wang, Z. Y. (2017). Research on the Pre-edit and Post-edit of Machine Translation in Science and Technology Text Translation. Journal of Zhejiang International Studies University, 3, 76-83.

Kocmi, T., Christian Federmann, Roman Grundkiewicz, Marcin Junczys-Dowmunt, Hitokazu Matsushita, and Arul Menezes. (2021). To ship or not to ship: An extensive evaluation of automatic metrics for machine translation. In Proceedings of the Sixth Conference on Machine Translation, pages 478–494.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In Proceedings of machine translation summit x: papers, pages 79–86.

Mathurm, N., Timothy Baldwin, and Trevor Cohn. (2019). Putting evaluation in context: Contextual embeddings improve machine translation evaluation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2799–2808.

Mathur, N., Timothy Baldwin, and Trevor Cohn. (2020). Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4984–4997, Online. Association for Computational Linguistics.

Papineni, K., Salim Roukos, Todd Ward, and Wei- Jing Zhu.(2002). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pages 311–318.

Popovi´c, M. (2015). chrf: character n-gram f-score for automatic mt evaluation. In Proceedings of the Tenth Workshop on Statistical Machine Translation, pages 392–395.

Rei, R. Craig Stewart, Ana C Farinha, and Alon Lavie. (2020). Unbabel's participation in the WMT20 metrics shared task. In Proceedings of the Fifth Conference on Machine Translation, pages 911–920, Online. Association for Computational Linguistics.

Rei, R., Craig Stewart, Ana C Farinha, and Alon Lavie. (2020). Comet: A neural framework for mt evaluation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 2685–2702.

Sellam, T., Dipanjan Das, and Ankur Parikh. (2020). Bleurt: Learning robust metrics for text generation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7881–7892.

Snover, M., Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. (2006). A study of translation edit rate with targeted human annotation. In Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers, pages 223–231.

Specia, L. & Frédéric B. & Marina F. & Chrysoula Z. & Zhenhao L. & Vishrav Ch. & Martins A. (2021). Findings of the wmt 2021 shared task on quality estimation. *Association for Computational Linguistics*.

Ulitkin I. & Filippova I & Ivanova N. & Poroykov A. (2021). Automatic evaluation of the quality of machine translation of a scientific text: *the results of a five-year-long experiment. In E3S Web of Conferences.*