

Research Article

How Machine Translation Helps Foreign Language Students?

Hong Zhang^{1*}

¹College of International Studies, Yangzhou University, Yangzhou, Jiangsu Province, China

*Correspondence to: Hong Zhang, PhD, Professor, College of International Studies, Yangzhou University, No.88 South Road, Hanjiang District, Yangzhou, 225012, Jiangsu Province, China; Email: zhhong2020@126.com

Received: October 14, 2023 Revised: October 31, 2023 Accepted: December 4, 2023 Published: December 31, 2023

Abstract

Objective: The objective of this study is to examine to what extent machine translation (MT) helps Chinese students who study Spanish language.

Methods: Our study employed the following methods, Chinese students participated in an experiment consisting of two main tasks. The first was a short translation test measuring whether students can translate with better accuracy, fluency and complexity if they use MT. In the second task, those students had to translate without MT. In this article, we focus primarily on the translation task.

Results: All participants obtained higher scores in accuracy, fluency and complexity when using MT.

Conclusion: This online tool helped students complete their translation task with better spelling, terminology, mistranslation, overly literal, style guide, ambiguous translation, part of speech, addition and typography. Unfortunately, MT did not help them to avoid unit conversion errors, word order errors, functions words errors, omission errors and punctuation errors. Identifying these types of errors can help students better recognize their mistakes while learning Spanish as a second language.

Keywords: machine translation, errors using machine translation, computer-assisted language learning, learning and teaching process of Spanish

Citation: Zhang H. How Machine Translation Helps Foreign Language Students? *J Mod Educ Res*, 2023; 2: 16. DOI: 10.53964/jmer.2023016.

1 INTRODUCTION

The main purpose of this paper is to determine the differences between using machine translation (MT) and not using MT in a translation task carried out by Chinese students learning Spanish as a second language. The differences in types of error are identified.

Many researchers have examined the use of technologies

in foreign language learning. Ayres^[1] highlights the idea that computer-assisted language learning techniques “need to be used more carefully and judiciously and must be tightly integrated into the learning curriculum in order for learners to obtain maximum benefit from its use”. McAlpine and Myles^[2] took on an interesting view concerning L2 dictionary use. “The role of the dictionary, whether it is used electronically, on-line or in print form, is to help

learners expand their vocabulary and increase their awareness of common grammatical errors”. O’Neill^[3] studied electronic tools that purport to convert words or phrases from one language into another for L2 learners or users. One such resource available for those writing in a second language is the electronic dictionary. There are also authors who studied the use of MT in foreign language learning. Kliffer^[4] stated, “post-editing of MT gave students insight into the huge challenges which have confronted MT, especially the questions of how to deal with syntactic and lexical ambiguity, non-literal language, and inferencing”. García^[5] noted that MT can help the beginner and early intermediate learner to communicate more and better. Garcia and Pena^[6] noted that “MT helped beginners to communicate more, particularly, when they had a lower mastery of the language”. However, like O’Neill^[3], they also noted that while online translation does not seem to negatively affect content, it cannot be said with certainty to improve content, either.

Other studies address post-editing and the analysis of the types of errors made with MT. A study by Clifford, Clifford et al.^[7] revealed that students perceive MT to be helpful in their language learning, especially in the acquisition of vocabulary, but they are aware that MT produces errors. To this last point, it is encouraging that students are critically assessing the output of the MT tool. Niño^[8] also suggested that the use of MT for advanced language students can enhance their “awareness as to the complexity of translation and language learning”. Kaye^[9] stated that MT is a useful method for advanced learners but not for beginners. Garcia and Pena^[6], however, showed that MT helps the beginner to communicate and write more in the L2. Marsden et al.^[10] stated that students need a certain level of grammatical competence to be able to use technology such as online MT and grammar checkers. Jolley and Maimone^[11] developed a survey to reveal both students’ and instructors’ views on free online MT tools and language learning. Fredholm^[12] analyzed the errors made by students using online translation. They made fewer mistakes regarding orthography, article, noun, adjective and agreement but more mistakes regarding syntax and verb morphology. According to Fredholm^[12], as discussed in Hyland and Wong’s research^[13], there is considerable pressure on teachers to use new technologies in their teaching and a need to be able to evaluate their pedagogical affordances. The present study was designed to observe the differences between error types with and without MT. Our previous work involved the evaluation of the quality of existing raw MT output for that language combination. Zhang^[14] conducted a comparison of three MT engines, Google Translate, Baidu and Bing Translator to determine which was best suited to the language combination Chinese to Spanish. Zhang and Torres-Hostench^[15] also show that post-editing of machine translated text into the foreign

language participants are studying may help them learn the second language and describe potential improved learning outcomes in their study.

The main research question was to know what extent MT helps Chinese students who study Spanish language.

2 METHODS

Based on the B2 and C1 certificate requirement, we chose a group of 20 Chinese master’s course students (10 with B2 certificate and 10 with C1 certificate) with 21 years old and study Spanish as a foreign language of the master’s course in Translation, Interpreting and Intercultural Studies of the Universitat Autònoma de Barcelona. They were asked to do a short translation test on the Tajo River. During a period of 20 minutes, we gave Chinese sentences to our 20 Chinese students (14 females and 6 males) to translate Chinese into Spanish, all of them with the level of B2 (according to the Common European Framework of Reference, a B2 certificate being required to access the masters’ degree in translation), the first 10min consisted of using MT to translate sentence from Chinese to Spanish without MT, and we refer to this translation as Translation A. The other 10min were used to allow students to translate with the aid of MT, and we refer to this second translation as Translation B. Methodology was based on previous research by Fredholm^[12]. The study is based on descriptive methodology for these two experiments. We use screen recording to record these moments meticulously and making notice of every action done by the student.

MT is used to aid the student get an idea of how to write their thoughts. It is evident that there will be the errors, and it is up to the student to notice and correct these errors in translation. The pause between the writing of words and the erasure of words, as well as other changes in decisions and behaviors; we can find their problems in their translating process. For the analysis of the results of the participants’ translation task, three aspects were taken into account: first, an evaluation of the number of words used in the translation task, second, an evaluation of the quality of the translation task. We investigated whether the participants translate better with CAT or without it. For this reason, we combined the criteria of accuracy, fluency, and complexity from Fredholm^[12] with the criteria of accuracy and fluency from Multidimensional Quality Metrics (MQM)^[16]. In our analysis, which is based on TAUS Dynamic Quality Framework and QT LaunchPad’s MQM, errors were classified as 2 main types: accuracy and fluency^[17]. In Fredholm’s investigation^[12], errors were classified as one of 3 main types: accuracy, fluency and complexity. The following analysis will focus on three main scoring parameters: accuracy, fluency and complexity. The definitions of accuracy and fluency are: (1) Accuracy: The target text does not accurately reflect the source text, allowing for any differences authorized by specifications

Table 1. The Three Evaluation Criteria for Participants' Translation Task Score Accuracy Fluency Complexity (Source: Prepared Based on MQM)

Score	Accuracy	Fluency	Complexity
1	Contains no information	The translation is incomprehensible	Grammar and vocabulary are incorrect
2	Contains little information	The translation is non-native	Grammar and vocabulary are non-native
3	Contains a fair amount of information	The translation is good	Grammar and vocabulary are good
4	Contains all the information	The translation is excellent	Grammar and vocabulary are excellent

(<https://themqm.org/error-types-2/typology/>). (2) Fluency: Issues related to the form or content of a text, regardless of whether it is a translation or not (<https://themqm.org/error-types-2/typology/>).

Table 1 shows the criteria for evaluating participants' translation task results, developed based on the references just cited.

The results were analyzed using Excel, SPSS, and for the descriptive analysis, the following descriptive statistics have been calculated: (1) Quantitative variables: mean, standard deviation, median, minimum, and maximum were computed. Box plots were created for these variables. (2) Qualitative variables: tables of absolute and relative frequencies. Bar charts were generated for these variables.

R software v. 3.3.2 was used. Screen recording techniques were used to determine the exact problems encountered by our participants while completing the test.

3 RESULTS

3.1 Number of Words in a Translation with MT vs. without MT

As mentioned by Garcia and Pena^[6], "the number of words is an indicator of the 'amount' of the communication". The quality of the translation does not depend on the quantity of words, but that number is a relevant measure of testing the participants' command of the Spanish language.

For the first sentence, when translated with MT, the mean number of words is 28.73, and the median is 28. The minimum and maximum are 27 and 35 words, respectively. When translated without MT, the mean number of words is 27.18, and the median is 28. The minimum and maximum are 20 and 33 words, respectively. There was no significant difference regarding whether or not MT was used ($P > 0.05$).

For the second sentence, when translated with MT, the average number of words was 52.18, and the median was 50. The minimum and maximum were 49 and 59 words, respectively. When translated without MT, the mean number of words was 54.36, and the median was 56. The minimum and maximum were 42 and 77 words, respectively. There was non-significant difference regarding whether or not MT was used ($P > 0.05$).

We also noticed that without MT, there was a considerable difference between the strongest and weakest students. This indicates that the use of MT by itself did not explain differences between the strongest and weakest students. That means that there is a large gap among our participants related to their actual level of Spanish competency. It can be observed that our participants wrote fewer words without MT than when using MT. The translations varied in length from 27 to 35 words with MT and from 20 to 33 words without MT. This is because the majority of our participants thought the output of MT was correct and copied it directly without post-editing.

3.2 Results of Accuracy with and without MT

Two markers who speak Spanish as their mother tongue evaluated the test samples. Our two evaluators assessed performance on a scale of 0 to 4. "0" means "bad accuracy", "1" means "normal accuracy", "2" means "intermediate accuracy", "3" means "good accuracy", and "4" means "best accuracy".

As shown in Figure 1, when the first sentence was translated with MT, our two evaluators gave similar scores, when translated without MT, the mean score was 2.68, and the median was 2.5. The minimum and maximum were 2 and 4, respectively. There was non-significant difference ($P > 0.05$). While the second sentence was translated without MT, there was a significant difference ($P > 0.05$), which means that with MT accuracy was better. For the second sentence, when translated without MT, there was a significant difference ($P < 0.05$).

From the results, for sentence 1, when they translated a short sentence, there was non-significant difference regarding the use of MT by Marker 1, and there was a significant difference regarding the use of MT by Marker 2. When they translated a long sentence, there was a significant difference regarding the use of MT. We also noticed that in Figure 1 the mean score and median were higher without MT than with MT because our participants relied on MT, and they copied the result of MT with little post-editing.

3.3 Results of the Score of Fluency without MT and with MT

As shown in Figure 2, when the first sentence was

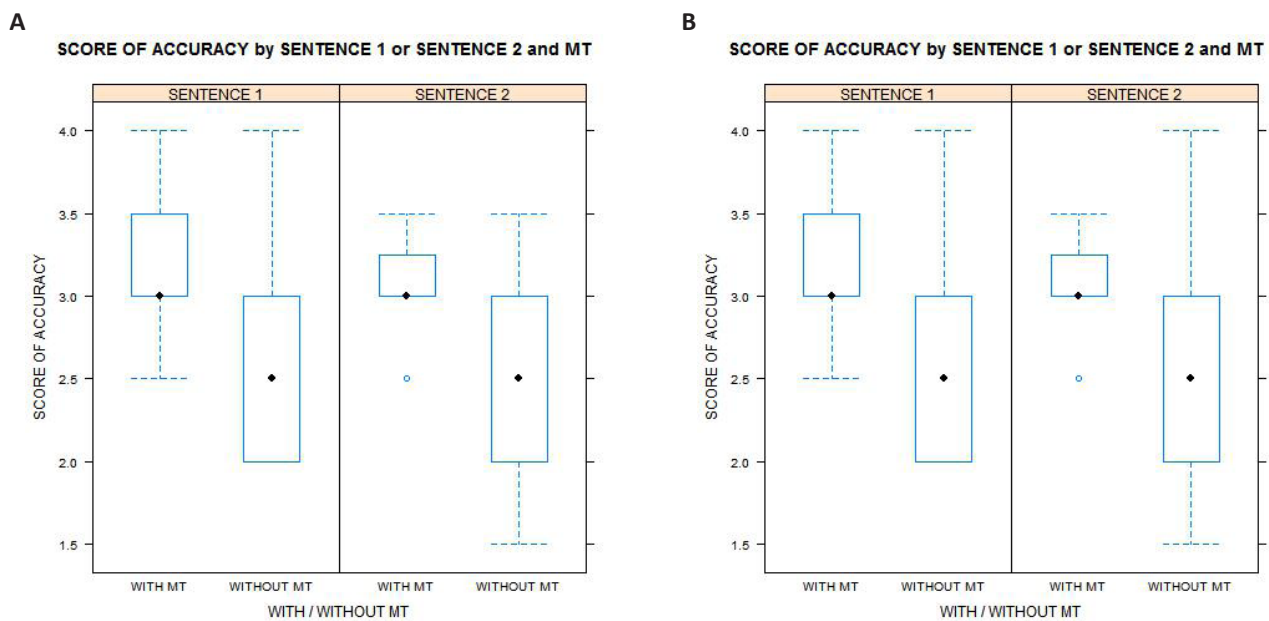


Figure 1. Score of Accuracy. A: Score of Accuracy by Marker 1; B: Score of Accuracy by Marker 2.

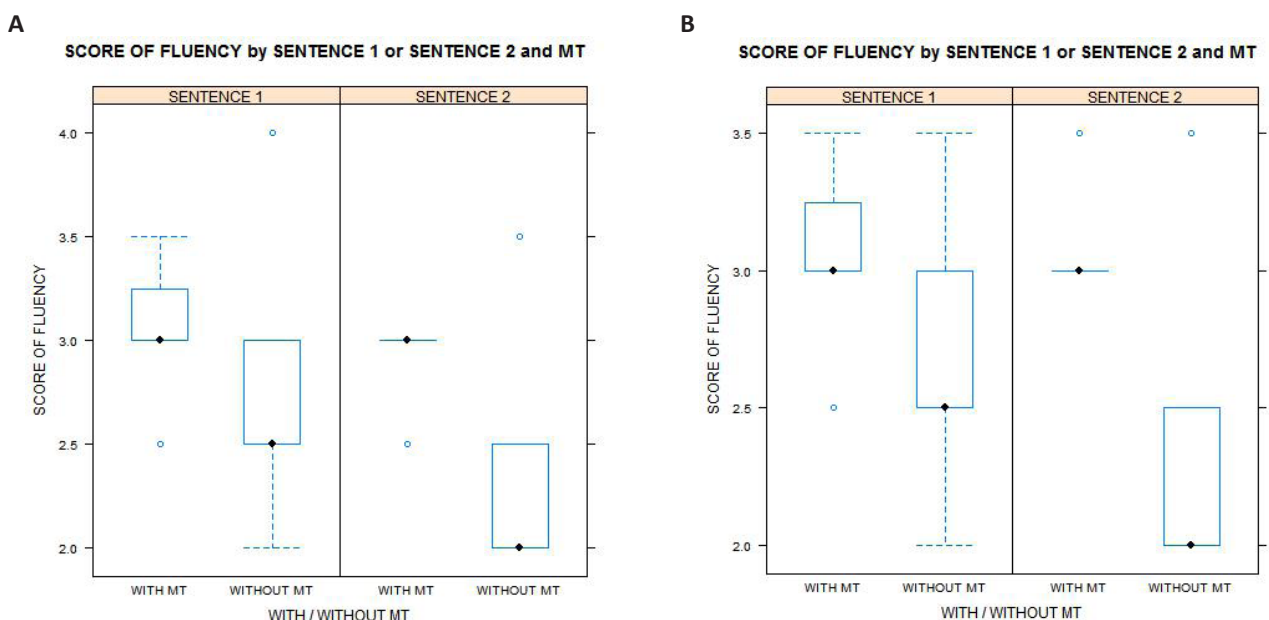


Figure 2. Score of Fluency. A: Score of Fluency of Marker 1; B: Score of Fluency by Marker 2.

translated with MT, from the mean score, median score, the minimum and maximum score, our two evaluators gave the similar score. There was non-significant difference ($P>0.05$). When the second sentence translated without MT, the mean score was 2.36, and the median was 2. The minimum and maximum were 2 and 3.5, respectively. There was a significant difference ($P<0.05$).

The results revealed that our two markers gave comparable scores for our participants regarding fluency. When they translated a short sentence, there was non-significant difference regarding the use of MT, and when they translated a long sentence, there was a significant difference

regarding the use of MT.

3.4 Results of the Score of Complexity without MT and with MT

In accordance with previous research on measures of complexity^[12,18-21], our investigation focuses on two aspects: grammatical complexity and lexical complexity. The majority of participants use present tense indicative, the infinitive and the future indicative two markers who speak Spanish as their mother language evaluated the test samples. Our two evaluators assessed performance on a scale of 0 to 4. “0” means “bad complexity”, “1” means “normal complexity”, “2” means “intermediate

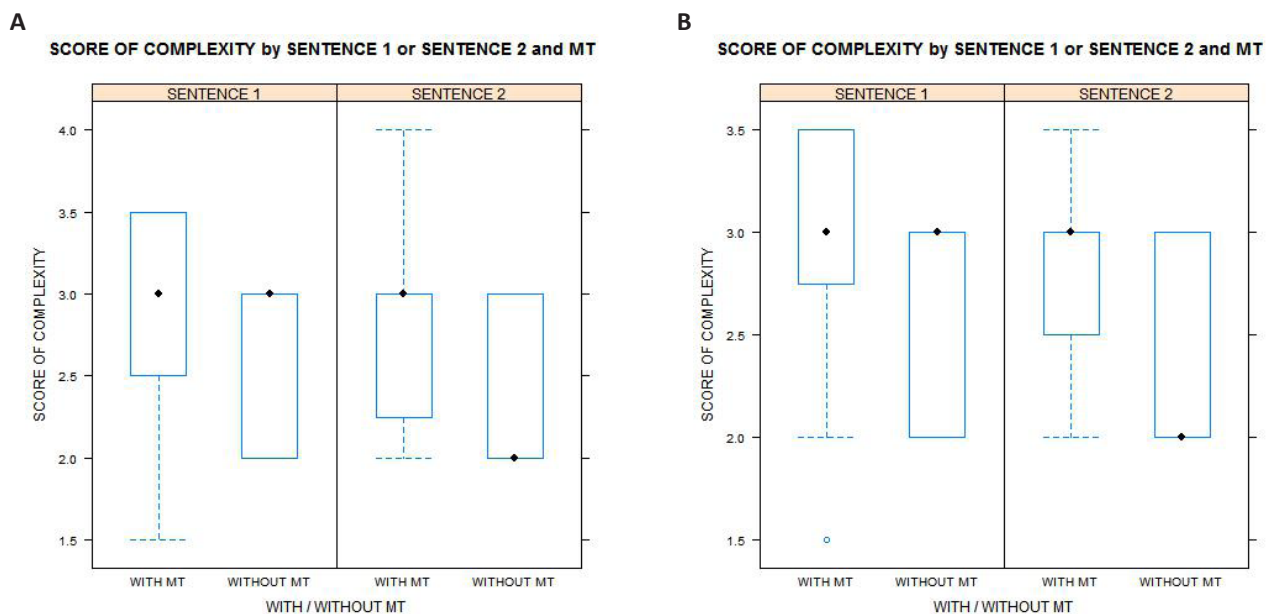


Figure 3. Scores of Complexity. A: Scores of Complexity by Marker 1; B: Score of Complexity by Marker 2.

complexity”, “3” means “good complexity” and “4” means “best complexity”.

As shown in Figure 3, when the first sentence was translated without MT, there was non-significant difference ($P>0.05$). When the second sentence was translated without MT, the mean score was 2.45, and the median was 2. The minimum and maximum were 2 and 3, respectively. There was non-significant difference ($P>0.05$).

The results showed that when our participants translated a long sentence, there was a significant difference regarding the use of MT. The results of testing accuracy, fluency and complexity showed that our two markers gave roughly the same scores. We also noticed that without MT, there was considerable difference between the strongest students and the weakest students, proving that MT is useful and helpful for advanced students. For sentence 2, there was a significant difference, showing that the use of MT can explain differences between students when translating long sentences.

We also analyzed the scores given by Marker 1 and Marker 2. For sentence 1 and 2, there was a non-significant difference between our two markers and the use of MT ($P>0.05$) or without MT ($P<0.05$), this indicates that the two judges awarded similar scores to the translations provided by the students, and the reasons for the variability in experimental results are related to the length of the translated sentences and the students' language proficiency.

3.5 To What Extent does MT Help Chinese Students Who Study Spanish Language?

They used MT, our participants performed better

in terminology, overly literal phrases, mistranslation, ambiguous translation, style guide, part of speech, addition and typology than without MT. We also observed that participants with MT cannot resolve unit conversion errors.

From Table 2, we can see that the use of MT has benefited participants E02, E05, E11 and E12, but we do not observe a significant improvement in the case of participants E05 and E12. On the other hand, it has been observed that participants E05 and E12, who already had many problems in the translation task without MT, they have not been able to take advantage of MT to solve their problems and doubts.

In the pilot study, in the PE test, MT made 19 mistakes in total, which were divided into two categories following the types of mistakes proposed by MQM^[16]: accuracy and fluency. Regarding the mistakes related to accuracy, the participants corrected better four types of specific mistakes (“N05. Mistranslation”, “N08. Accuracy in general”, “N09. Accuracy in general”, and “N19. Omission”). However, mistakes related to accuracy in general, to omissions, to mistranslations and to the names of entities were frequently repeated among the participants, since they could not detect neither correct them. Regarding the mistakes corresponding to fluency, the following five mistakes were easy to identify and correct for the participants: “N03. Word order”, “N10. Punctuation”, “N11. Spelling”, “N14. Word order” and “N16. Word order”. On the contrary, mistakes related to prepositions and the name were difficult for participants. Most participants made mistakes due to the lack of vocabularies and their insufficient knowledge of Spanish. The Table 3 shows the total number of errors made by our participants when completing the translation with and without MT.

Table 2. Results of Participants

Computer Number	Without or With MT	Sentence 1or 2	Nmuber of Words	Maker 1			Maker 2		
				Accuracy	Fluency	Complexity	Accuracy	Fluency	Complexity
E02	without MT	sentence 1	30	4	3	3	4	3,5	3
		sentence 2	46	2	2	3	2	2	3
	with MT	sentence 1	27	2,5	2,5	2	2,5	2,5	2,5
		sentence 2	49	3	2,5	3,5	3	3,5	
E05	without MT	sentence 1	24	2	3	3	2	3	3
		sentence 2	61	2	2	2	2	2	2
	with MT	sentence 1	27	3	3,5	3,5	3	3,5	3,5
		sentence 2	49	3	3	4	3	3	3,5
		sentence 2	49	3	3	2	3	3	2
E11	without MT	sentence 1	32	2	2,5	2	2	2,5	2
		sentence 2	56	3	2	2	3	2	2
	with MT	sentence 1	35	3,5	3	3,5	3,5	3	3,5
		sentence 2	58	3,5	3	3	3,5	3	3
E12	without MT	sentence 1	33	3	2	2,5	2,5	2	2
		sentence 2	77	3,5	3,5	3	3,5	3,5	3
	with MT	sentence 1	30	3,5	3	3	3,5	3	3
		sentence 2	50	3	3	2	3	3	2

Table 3. Total Number of Errors With and Without MT

Total Types of Errors	With MT	Without MT
Terminology	49	60
Overly Literal	31	35
Mistranslation	23	38
Unit Conversion	22	21
Function Words	20	9
Omission	17	11
Word order	9	2
Punctuation	4	0
Ambiguous translation	1	2
Style guide	1	0
Spelling	0	21
Part of speech	0	2
Addition	0	2
Typography	0	1

In short, the MT proved to be helpful in solving the following types of mistakes related to accuracy: mistranslation, omissions, and official names, among others. MT helped participants in terms of fluency: regarding word order, prepositions, typography and other spelling mistakes.

4 DISCUSSION AND CONCLUSION

We also discovered that our students had difficulties in correcting long and complicated sentences, probably because of the differences between Chinese and Spanish in terms of sentence structure and grammar rules.

All participants received higher scores in accuracy, fluency, and complexity when they translated a text into Spanish with the help of MT. When they translated a short sentence, there was non-significant difference regarding the use of MT, and when they translated a long sentence, there was a significant difference regarding the use of MT. From the results of the complexity score, we observed that there was non-significant difference regarding the use of MT for sentence 1 and sentence 2 because the grammar and sentence structure of Spanish are quite different from Chinese.

The most obvious difference between the translations with and without MT is the number of spelling errors. Participants with MT made 0 errors in spelling, whereas participants without MT made 21 errors in spelling. The reason for this result is as follows^[22,23]: lack of spell checker, limited time of the test and reliance on their incorrect knowledge of Spanish, and carelessness of keyboard use. However, our participants made fewer errors in word choice, omissions, word order and punctuation without MT than with MT. The reason for this result is twofold: first, the MT system cannot understand the meaning of the text, and the majority of the MT was overly literal; second, our participants had strong command of Spanish grammar. When lexicon are very difficult for both Chinese students and MT.

MT helped intermediate and advanced Chinese students who study Spanish as a foreign language to translate Chinese into Spanish. For example, from our data, we

have a percentage of translations in which our students performed better with the aid of MT. In the future, we can design a course for these students covering, for example, instruction in types of online translation tools, offline tools, linguistic resources, dictionaries, terminology glossaries, and corpora. After comparing the results of translation done with MT aid and translations made without MT aid. It can be seen that both experiments can do better with the help of MT, level C1 do better than B2.

When translating into Spanish, MT helped participants to better correct errors in spelling, terminology, mistranslation, overly literal phrasing, style guide, ambiguous translation, part of speech, addition and typography. When translating into Spanish, MT did not help participants to solve unit conversion errors, word order errors, functions words errors, omission errors and punctuation errors.

Fredholm^[12] mentioned that the fluency measure used was the length of texts produced during a certain amount of time. However, from the results of our experiment, this measure was not useful. Based on the results of the score of accuracy, fluency and complexity, participants wrote more words but with a worse quality of translation.

The sample is too small to be statistically significant, but it is often the case that larger, more representative samples are hard to reach within classroom-based research. Our study should be useful at least as a pilot, showing that it is an area in need of research and proposing ways this research could be carried out.

For the future research, from these results we are planning to develop didactic materials to help Chinese students identify and correct the output of machine translated texts to help them to better use MT to translate in Spanish.

Acknowledgements

This work was partly supported by the Ministry of Education of the People's Republic of China (23YJC752026).

Conflicts of Interest

The author declared no conflict of interest.

Author Contribution

Hong Z conceptualized and designed the experiment, supervised the work, conducted data analysis, and drafted the manuscript. The author contributed to writing the article, read, and approved its submission.

Abbreviation List

MQM, Multidimensional Quality Metrics

MT, Machine translation

References

[1] Ayres R. Learner Attitudes towards the use of CALL. *Comput*

Assist Lang L, 2010; 15: 241-249.[DOI]

- [2] McAlpine J, Myles J. Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs. *System*, 2003; 31: 71-84.[DOI]
- [3] O'Neill EM. The Effect of Online Translators on L2 Writing in French. Champaign, USA: University of Illinois, Urbana Champaign; 2012.
- [4] Kliffer MD. Post-editing machine translation as an FSL exercise. *Porta Linguarum*, 2008; 53-68.
- [5] García I. Can Machine Translation Help the Language Learner? *ICT Lang Learn*, 2010; 4-7.
- [6] Garcia I, Pena MI. Machine translation-assisted language learning: writing for beginners. *Comput Assist Lang L*, 2011; 24: 471-487.[DOI]
- [7] Clifford J, Merschel L, Munné J. Surveying the Landscape: What is the Role of Machine Translation in Language Learning? *@Tic.Revista D'innovació Educativa*, 2013; 108-121.
- [8] Niño A. Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, 2009; 21: 241-258.[DOI]
- [9] Kaye P. Translation activities in the language classroom. *Teach Engl*, 2009.
- [10] Marsden E, Mackey A, Plonsky L. The IRIS repository of instruments for research into second languages. *Adv Methodol Pract*, 2015; 1-22.
- [11] Jolley JR, Maimone L. Free online machine translation: use and perceptions by Spanish students and instructors. In: *Learn Languages, Explore Cultures, Transform Lives*. 2015; 181-200.
- [12] Fredholm K. Online Translation Use in Spanish as a Foreign Language Essay Writing: Effects on Fluency, Complexity and Accuracy. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las lenguas*, 2015; 7-24.[DOI]
- [13] Hyland K, Wong LL. Innovation and change in English language education. Routledge: London, UK, 2013.
- [14] Zhang H. Los errores de la traducción automática Chino al Español. The 2nd International T3L Conference: Tradumatics, Translation Technologies & Localisation, Barcelona, Spain. 2016.
- [15] Zhang H, Torres-Hostench O. La efectividad de las consultas en línea en posesión de traducción automática por parte de estudiantes de lenguas extranjeras. *CIRCULO de Lingüística Aplicada a la Comunicacion*, 2023; 93: 289-303.
- [16] Lommel A, Uszkoreit H, Burchardt A. Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumática*, 2014; 12: 455-463.[DOI]
- [17] TAUS. Translation Automation User Society annual plan 2010.
- [18] Koponen M, Aziz W, Ramos L et al. Post-Editing Time as a Measure of Cognitive Effort. Proceedings of the Tenth Biennial Conference of the Association for Machine Translation of the Americas 2012 Workshop on Post-Editing Technology and Practice. San Diego, California, 2012.
- [19] Lacruz I, Shreve GM, Angelone E. Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study.

- Workshop on post-editing technology and practice. San Diego, California, USA, 2012.
- [20] Lacruz I, Shreve GM. Pauses and Cognitive Effort in Post-Editing. *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing: Newcastle upon Tyne, UK. 2014; 246-272.
- [21] Lacruz I, Denkowski M, Lavie A. Cognitive Demand and Cognitive Effort in Post-Editing. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*. 2014; 73-84.
- [22] Xu M, Wang C. Translation students' use and evaluation of on-line resources for Chinese-English translation at the word level. *Transl Interpret Stu*, 2011; 6: 62-86.[DOI]
- [23] Wang Q, Vantasin K, Xu D et al. MUFOLD-WQA: A new selective consensus method for quality assessment in protein structure prediction. *Proteins*, 2011; 79: 185-195.[DOI]