

2016

ANALYSIS OF HUMAN VERSUS MACHINE TRANSLATION ACCURACY

Shihua Brazill

Montana Tech of the University of Montana

Michael Masters

Montana Tech of the University of Montana

Pat Munday

Montana Tech of the University of Montana

Follow this and additional works at: https://digitalcommons.mtech.edu/grad_rsch

Recommended Citation

Brazill, Shihua; Masters, Michael; and Munday, Pat, "ANALYSIS OF HUMAN VERSUS MACHINE TRANSLATION ACCURACY" (2016). *Graduate Theses & Non-Theses*. 223.
https://digitalcommons.mtech.edu/grad_rsch/223

This Non-Thesis Project is brought to you for free and open access by the Student Scholarship at Digital Commons @ Montana Tech. It has been accepted for inclusion in Graduate Theses & Non-Theses by an authorized administrator of Digital Commons @ Montana Tech. For more information, please contact sjuskiewicz@mtech.edu.

TRANSLATOLOGIA

[ABOUT](#) [EDITORIAL BOARD](#) [SUBMISSION GUIDELINES](#) [DOWNLOAD ISSUES](#)

[CONTACT](#)

ANALYSIS OF HUMAN VERSUS MACHINE TRANSLATION ACCURACY

[Home](#) > [2017](#) > [January](#) > [Analysis of human versus machine translation accuracy](#)

Analysis of human versus machine translation accuracy

By [Shihua Brazill](#) | [TRANSLATOLOGIA 1/2016](#)

Shihua Chen Brazill¹, Michael Masters², Pat Munday¹

¹ Professional and Technical Communication, Montana Tech, Butte (USA)

² Anthropology, Montana Tech, Butte (USA)

Corresponding author: Shihua Chen Brazill, Montana Tech

1300 West Park Street, Butte, MT 59701-1419

Search

Call for Papers
(prolonged till 20-
SEP)

TRANSLATOLOGIA
1/2019

Reviews 1/2019

TRANSLATOLOGIA
1/2018

Reviews 1/2018

TRANSLATOLOGIA

E-mail: shihuabrazill@gmail.com

Telephone: 011-1-406-548-7481

The purpose of this study was to determine whether significant differences exist in Chinese-to-English translation accuracy between moderate to higher-level human translators and commonly employed freely available Machine Translation (MT) tools. A Chinese-to-English language proficiency structure test and a Chinese-to-English phrase and sentence translation test were given to a large sample of machine (n=10) and human translators (n=133) who are native Chinese speakers with at least 15 years of familiarity with the English language. Results demonstrated that native Chinese speakers with this minimum level of English proficiency were significantly better at translating sentences and phrases from Chinese to English, compared to the ten freely available online MT applications, which unexpectedly showed a considerable degree of variation in translation accuracy among them. These results indicated that humans with at least a moderate level of exposure to a non-native language make far fewer translation errors compared to MT tools. This outcome is understandable, given the unique human ability to take into account subtle linguistic variants, context, and capricious meaning associated with the language and culture of different groups.

Key words: human translation, machine translation, translation error, Chinese to English translation

Introduction

Machine translation (MT) is largely domain-limited and generated for a specific purpose. The literature includes a number of research studies that examine existing online MT services. This research describes various domains of MT evaluations, and shows that MT is not generally intended for a literary translation, but rather for a specific purpose.

For some years, MT – especially online translation systems – have been studied in comparison to expert human translation. Aiken et al. (2006) made an early contribution with an evaluation of Spanish-to-English translations using Yahoo SYSTRAN. More recently, Aiken et al. (2009) compared four free online MT systems including Google Translate, Yahoo SYSTRAN, AppliedLanguage, and x10 for the domain of common tourist phrases and some complex phrases from Spanish and German to English. They concluded that Google Translate was the most accurate of MT tested, and was especially useful for gisting, i.e. yielding an understandable meaning even if the grammar was garbled.

Seljan et al. (2011) conducted graded evaluations of texts from four domains (city description, law, football, monitors) translated from Croatian into English by four free online translation services (Google Translate, Stars21, InterTran and Translation Guide) and text translated from English into Croatian by Google Translate. They pointed out that Google Translate is a statistical MT based on a large number of corpora that support many languages. Machine-translated texts were evaluated by inter-raters judging fluency and adequacy, with the inter-rater agreement measured using Pearson's correlation and Fleiss kappa. Results indicated that the quality of free online MT differed for specific language pairs, domains, terminology, and corpus size. Some tools performed better at translating specific language pairs, and the fluency and adequacy of different tools was highly domain dependent. For example, the domain of city description resulted in the lowest grades for all free online MT services because city description has the most freedom in its style. Error analysis indicated that untranslatable words were the biggest factor resulting in low grades, and that Google Translate was better for translating frequent expressions but not for translating

language information such as gender agreement.

Seljan et al. (2015) used human evaluators to score results of machine translated texts for one non closely-related language pair, English-Croatian, and for one closely-related language pair, Russian-Croatian. Four hundred sentences from the domain of city descriptions were analyzed, i.e. 100 sentences for each language pair and for two online statistical MT systems, Google Translate and Yandex.Translate. Analysis was carried out based on the criteria of fluency and adequacy, and enriched by error analysis. In this study fluency referred to style and adequacy referred to meaning, and Cronbach's alpha was used to measure internal consistency. Results demonstrated that Google Translate and Yandex.Translate scores varied for adequacy and fluency depending on whether the translation was English-to-Croatian or Croatian-to-English. Based on these results the authors concluded that when using MT tools, realistic expectations and using appropriate text genre (i.e. domain) will influence the perception of the translation quality. For instance, using the correct domain, similar language pairs, and regular word order results in higher scores. Also, machine translators proved better at translating simple sentences and subject-verb-object order than translating complex sentences. Morphological errors/wrong word endings were the most common error, followed by untranslatable/omitted words and lexical errors/wrong translations.

Kit and Wong (2008) evaluated six free online MT tools, including Babel Fish, Google Translate, ProMT, SDL free translator, Systran, and WorldLingo for the domain-specific translation of legal text. Using reliable, objective, and consistent methods such as BLEU (BiLingual Evaluation Understudy) and NIST (National Institute of Standards and Technology), they translated text from 13 languages into

English. The domain consisted of a large corpus of legal texts of importance to law librarians and law library users. Users with MT tool experience were able to identify limitations of MT, which was generally not able to identify language exceptions and ambiguities (for example lexical ambiguity and structural ambiguity) of the linguistic features compared to translations performed by expert human translators. However, these types of translations were difficult both for MT and for humans without subject knowledge. Both MT and experts made frequent errors and often repeated the same errors. Additionally, texts that included slang, misspelled words, complex sentences, and uncorrected punctuations also caused incorrect translations. While MT could be considered a good solution for understanding information when translation quality is not the first priority, MT quality varied widely from language to language and domain to domain. The authors also pointed out that using back translation or “round-trip translation” was not an effective approach for evaluating MT quality because some words can be translated in different ways. A more effective way to evaluate machine technology was to compare a specific MT with a human-performed translation; generally speaking, the closer the MT outcome was to the human translation, the better the tool. Additionally, the degree of linguistic diversity between two different languages resulted in less accurate MT results. For example, the accuracy of MT systems with Asian-European language pairs was much poorer than with European pairs such as Spanish-English.

Garcia-Santiago and Olvera-Lobo (2010) analyzed MT translations from German and French into Spanish using Google Translate, ProMT and WorldLingo. These tools were selected because they were capable of performing translations between these language pairs, had a wide diffusion, and could perform translations quickly. The results indicated that MT accuracy differed for various

language pairs, with ProMT best for translating from German to Spanish, followed by Google Translator; and with WordLingo the best for translating from French into Spanish, followed by ProMT.

In a broad historical overview, Gaspari & Hutchins (2007) stated that Babel Fish, which was launched on December 9, 1997, was the first free online MT. Since then, many free MT online services became widely available and are regarded as a fast tool available for all internet users. Despite the MT services' poor output quality, they can be useful for obtaining the gist of a corpus. MT vendors offer free MT online services to promote their sales of full MT systems. Though professional translators appeared less likely to use online MT system in their work, people with limited English knowledge were more likely to use free online MT services. Users who have some knowledge of the target language tend to use MT online services as electronic dictionaries to look up or check vocabulary. In 1996, two speakers at AMTA (Association for MT in the Americas) conferences argued that online MT was the ideal solution to real communication problems. However, the use of MT online services might lead to safety violations in certain domains and the resulting legal issues remain unresolved. MT online providers are working on adding a wider range of language pairs, becoming more domain specific, and enhancing translation quality by creating more lexical entries and powerful rules.

Advanced Translation technology applications include MT and Computer-assisted Translation (CAT). CAT uses Translation Memory (TM) as a component. TM requires human translators to populate and grow a translation database. According to Somers (2003), MT tools and CAT tools are different from each other. MT can accomplish the tasks to a certain extent like translators do. CAT tools avoid repetitive tasks in order to help translators to be

more efficiency. The author also states that we should not rely merely on MT, even though it sometimes produces reasonable results, as post-editing is vital after MT. Post-editing refers to correction of MT translation output by linguists. Newton (2002) defines MT as translation that is performed by a computer, with or without human assistance. Newton (2002) states that MT outputs are low quality without human translators to rewrite and edit the translation. However, Arnold (1994) points out that in some cases, MT can produce good results. Even where the quality is lower, it is often easier and cheaper to revise 'draft quality' MT output than to translate entirely by hand. (Arnold, 1994, p.11).

According to Drummer (1996), MT uses computerized systems to translate the source language texts to the target language texts. The need for MT is obvious because documents, such as books, articles, and other educational tools, must be translated to various languages to satisfy diverse readers. Tripathi and Sarkhel (2010) stated that language and translation are vital tools in intercultural communication, and for providing access to diversity. Free MT tools, which included Microsoft's Bing Translator, and Google Inc.'s Google Translate, were easy to access by Internet users to bridge the diverse language gaps. However, as shown above, the literature also argues that the quality of MT is inferior to professional human translators. MT translates word-for-word and fails to convey complete meanings between the source and target texts. The reason that MT tools are not as accurate as professional translators may be due to the linguistic irregularities, ambiguities, and lack of universal grammar and vocabulary.

Kazemzadeh and Fard (2013) defined CAT as "an alternative approach to computer translation that integrates human expertise into the automatic translation

process” (p. 23). The authors briefly reviewed the history of translation and technology. At the beginning of the 1950s Cold War, because of the strife between the United States and the Soviet Union, many documents were translated between Russian and English. The inefficient and slow traditional human translation process could not meet such a high demand in all subject matters in a short amount of time; therefore, computer technologies were developed to help complete the translations quickly and cheaply. Translation technologies have continued to grow rapidly in recent years because many users do not realize that computers do not achieve the nuances of professional translators. Kazemzadeh and Fard (2013) pointed out that human translation is far more expensive compared to machines because of the limitations of human productivity. For example, they demonstrated that a professional translator can only translate up to 2,000 words each day while assuring translation quality in technical subject areas (Kazemzadeh and Fard, 2013).

Olohan (2011) classified TM as a type of CAT tool that could speed up the translation process. By using translation memory, translators can avoid translating repetitive words or phrases and maintain the consistency of the translation, saving both the translator’s time and the client’s budget. In other words, TM enables you to translate new texts while reusing specific elements from previous translations many times. SDL Studio is a popular TM that is used worldwide by professional translators. However, feedback indicates that SDL Studio is complicated to use. Therefore, some choose easier software as a substitute. Similarly, Walmer (1999) found that even though setting up and maintaining a TM database is difficult, in the long run it will save money. A TM will tremendously improve the speed, quality, and consistency of translation. The professional translators then need only focus on sections that the TM database cannot translate.

Alcina (2008) argued that translation technologies will speed up the translation process and lower costs, but not all translators will use them. For instance, TM did not work well for audiovisual or literary translation because the language requires a more elegant and flexible vocabulary. However, people who translate specialized fields, such as legal, technical, and localization, find that TM is worthwhile and helpful.

Using translation technology without human editing often causes mistranslation. Vilar et al. (2006) identified classification errors in MT output. These errors include missing words, reversed word order, incorrect words, unknown words, and errors in punctuation. Public signs that were mistranslated in China, often referred to as “Chinglish,” were likely the result of mistranslation by machines without professional human translation post-editing, as these are common MT errors. Cui and Zhao (2014) provided some practical guidance for Chinese to English translation, such as adding, deleting, rewriting, and reorganizing the message to improve translation quality. In this case, MT might fail to fulfill these creative, flexible, and aesthetic functions.

Translation is a highly sophisticated task, which includes knowledge of both the source language and the target language, understanding the content of the texts translated, and knowing how to integrate translation experience to progressively increase the translation quality. An experienced professional translator understands the significance of cultural differences, audiences’ needs, and research gaps regarding what is required to produce high quality translation.

Human translation increases translation quality through cultural sensitivity. Buden and Nowotny (2009) noted that translation applied not only to the words of different languages, but also from one culture to another. MT was

languages, but also from one culture to another. MT was limited in its ability to properly and effectively translate cultural perceptions. Sun (2003) indicated translators must be attentive to the target culture's expectations. Information or practices that were valuable in one culture could be regarded as offensive in another culture. In some Asian cultures, consuming dog meat is acceptable, but this practice is considered bizarre and culturally unacceptable in America. In this case, showing respect and consideration for the target culture is vital. MT is unable to comprehend such cultural differences; if the cultural context is mistranslated, potential misunderstandings or even offense can result.

Cultural awareness is a major lacuna of MT. Sun (2003) emphasized the importance of cultural sensitivity when translating a language, and cultural growth often depends on the ability to gain a fresh appreciation of other cultures through translation. Translation, in this sense, is not just about words and ideas. Objects, for example, are also translated. For instance, European cultures translated Chinese gunpowder as something to be used for weapons as well as for celebration. It is difficult to understand a text without cultural context. He Sanning (2009) illustrated how cultural context affects translation. For example, "sexy" is a positive term in western culture; however, when "sexy" is translated into Chinese, *xinggan* (sexy) indicates loose sexual morality, a negative term in Chinese culture. Cultural gaps between English and Chinese mean that Chinese terms such as "kezhang, chuzhang, juzhang, shifu" are virtually untranslatable (p. 95). Understanding cultural differences can minimize communication gaps, but because MT is incapable of incorporating cultural differences, it often produces low translation quality. From the example above, if a person traveled to China and used a machine translator to say that someone was attractive or "sexy," it could be perceived as insulting.

In certain situations, such as healthcare and international business, it is especially vital to have correct translation. For these situations, professional human translators are better equipped than machines at understanding the context of the meaning. In healthcare settings, mistranslation can result in lawsuits, potential injury, or even death of patients. Jacobs, Shepard, Suaya and Stone (2004) found that those patients who have a deficiency in English receive better health care quality when they communicate through interpretation. Sathe (2006) synthesized a case study about the effects of mistranslation.

A Spanish-speaking teenager told his girlfriend that he was "intoxicado" before fainting. When his girlfriend called the paramedics, they understood the word to denote "intoxicated", while the intended meaning was "nauseated." As a result, the patient was treated for drug overdose before being re-diagnosed with intracerebral hematoma with brain stem compression and a subdural hematoma secondary to a ruptured artery. (p. 7)

With the translation of European languages, "false friends" such as "intoxicado" are a common problem. As another example, the German word "das Gift" means poison. This case specifies how dangerous mistranslation can be; professional human translation produces more accurate results and is vital in high-risk situations. Since professional human translation is more accurate than MT, MT cannot be trusted when people's lives are involved.

Healthcare is not the only area in which poor translation can have adverse effects; incorrect MT also affects customers' perceptions of product quality. The European Commission (2012) reported that poor MT quality could result in customer-incurred cost, customer dissatisfaction cost, and company reputation cost. This shows translation

is an important part of international business. Cultural differences reflect communication gaps between the source materials and target materials. MT fails to meet the needs of intercultural communication, while professional human translation will increase customers' satisfaction of the company's products.

Translation is certainly an important factor in intercultural communication. As Ye and Shi (2009) explained, translation is a process of conveying meanings between different languages and cultures. MT has been used in a broad number of fields; however, a machine translates word-for-word, which produces poor translation quality compared to human translation.

From Munday's perspective (2009), a word-for-word method should be revised using a sense-for-sense method. A sense-for-sense translation method focuses on the whole meaning of the contexts. The need for this revision is clear when looking at an example such as the Chinese sign that might be freely translated by human translation to English as "Be careful! Do not slip and fall." When translated literally by machine, it translates to "Slip and fall down carefully." This kind of translation error is so common on public signs in mainland China that it has given rise to the term, "Chinglish."

Translation services affect people's lives in numerous ways, and are an increasingly important mechanism for cross-cultural communication and the social, political, and economic integration of different cultures and linguistic groups throughout the world. The aim of the current study is to investigate whether significant differences exist in the accuracy of Chinese to English human and MT using a large number of machine translators and human individuals with a broad range of experience with the Chinese and English languages.

This research is particularly relevant because many

THIS RESEARCH IS PARTICULARLY RELEVANT BECAUSE MANY individuals, institutions, and companies rely on these MT tools as a result of their broad availability and ease of use. However, little is known about how they actually compare to human translational ability, which is also able to take into account symbolism, context, and others elements of a cultural group, which may be important components of accurate translation and in conveying a meaningful message between different languages.

Methods

Sample and Procedure

The total sample included 10 different freely available MT applications and 133 human translators. The researcher conducted the MT tests with an expectation of a basic translation or gisting. A basic translation would present the overall meaning, but the word order, sentences structures, etc. might be incorrect. The students translated Chinese to English without using MT tools. The translated sentences came from the domain of educational textbooks, ranged in length from 2 to 8 words, and comprised ordinary vocabulary from everyday life.

Additionally, a more complex Chinese sentence was tested from the domain "Idiom and Culture". The sentence "虚心使人进步, 骄傲使人落后 (Modesty helps a person to make progress whereas conceit makes a person lag behind)" was entered into each of the 10 MT sites listed in Appendix B in order to test the accuracy of each MT engine. This complex sentence is from a Chinese idiomatic expression or set phrase, which consists of 12 words without using any specific terminology. The webpage information and results indicate that <https://www.translate.com/> and <http://www.bing.com/translator/> use Microsoft translator as an engine. The application <http://www.freetranslation.com/en/translate-english-chinese-use-GDI-as-an-engine>, <http://www.netat.net/> is

Chinese uses SDL as an engine; <http://www.netat.net/> is powered by Discuz; and www.dict.cn, fanyi.youdao.com, <http://tran.httpcn.com/> and fanyi.baidu.com/translate appear to use their own proprietary translation engines because they produce different results than either the Google or Bing engines; fy.iciba.com uses Kingsoft Corp as an engine. This test of the different machine translators used in the study and the variable translation output generated by them indicates that although some of the MT tools use the same engine, they appear to have tweaked their specific algorithms, as each machine produced different numbers of errors despite some of them sharing a common engine.

The human translator group included both males (30) and females (102) who are native Chinese speakers and who had studied English for at least 15 years. These subjects' ages ranged from 19 to 45, and included individuals with variable levels of proficiency beyond the minimum 15 years of English language experience (table 1).

Table 1: Participant Groups

Group	Number
Students of second-year English audio/visual studies from a college in China	23
Students of third-year English majors from a university in China	40
Students of third-year English translation majors from a university in China	36
Native Chinese speakers participating through social media	28

Professional translators from China	3
English teachers from China	3
Free online MT applications	10

Data Collection

Data collection for the current study utilized the following instruments:

1. Chinese-to-English language proficiency structure test for human subjects

The test was used to examine different participants' English proficiency and to divide them into five levels of Chinese-to-English translation skills for the study. The researcher compared the individuals' English language levels to their stated translation levels. This reliable and valid test consisted of 25 vocabulary word translations that evaluated the translators' levels of English language proficiency.

The test included five sections and each section included five words. The first section was selected from first-grade vocabulary words. The second section was selected from third-grade vocabulary words. The third section was selected from fifth-grade vocabulary words. The fourth section was selected from seventh-grade vocabulary words. Finally, the last section was selected from ninth-grade vocabulary words.

2. Chinese-to-English phrase and sentence translation test for human subjects

To identify the translators' translation ability, a phrase and sentence translation task was administered. The translation task was divided into three categories. The first category included four beginner phrases, which are commonly used in both the Chinese and English languages. The second category included six intermediate sentences. These sentences are also common expressions. The third category included five advanced Chinese-to-English translations, which require the participants to have a solid English language foundation, as well as knowledge of both cultures.

3. Chinese-to-English language proficiency structure test for machine translators

The investigator used the vocabulary word test to determine different machine tools' accuracy in translating vocabulary words.

4. Chinese-to-English phrase and sentence translation test for machine translators

The same phrase and sentence translation test was used to measure different machine tools' accuracy in translation.

Data Analysis

This study classifies as a quasi-experiment, because individuals were actively recruited to participate in the study, rather than being randomly selected from the population. Additionally, because a certain minimum level of English language proficiency was required to be considered for the study, this further limited the ability of the researchers to randomly draw from a larger sample of native Chinese speakers.

To assess the accuracy of responses to the Chinese-to-English language proficiency structure test, the number of mistakes made by each human and machine in the sample

were simply counted by the examiner, who had the correct answers prepared before the tests were distributed. For the Chinese-to-English phrase and sentence translation test, the examiner counted the number of mistakes in each sentence and then summed the total number of mistakes in each category.

Analysis of mean differences in Chinese-to-English phrase and sentence translation accuracy between the human and machine groups was carried out using Analysis of Variance (ANOVA). Two-sample t-tests were used to evaluate the primary research question relating to overall differences in translation error between the broader human and machine groups, as well as to test for individual differences between specific groups of translators. Generalized linear regression analysis was also used to examine whether age, years of education, and group affiliation were related to the number of translation mistakes in sentences, as well as to control for these variables in assessing differences in translation errors between the sexes.

Results

A test of the primary research question concerning whether significant differences exist between human and machine translators shows that these two groups indeed differ by a substantial margin. In translating the 15 phrases and sentences from Chinese to English, mistakes made by humans ($=3.2$, $SD=4.8$, $n=133$) were far fewer than those made by the machines ($=18.4$, $SD=9.07$, $n=10$), and a two-sample t-test indicates that this difference is highly significant ($t = -5.24$, $p = 0.00$).

Although the overall difference between human and MT accuracy is considerable, it is clear that a great deal of variation exists among the human group, and that a number of individuals show up as outliers, which acts to

skew the sample mean more toward that of the machine translator average (figure 1).



Figure 1. Boxplot showing results of a two-sample t-test of human vs. MT accuracy of phrase and sentence translation tests (outliers represented by an asterisk).

Separating out the human translators into their respective groups, with different levels of English language proficiency beyond the minimum 15 years of experience as designated in table 1, shows that the outliers in figure 1 are exclusively second-year English audio/visual studies students from a college in China (figure 2). The remaining groups can be seen to cluster near the bottom of figure 2, with these four averaging only 1.45 translation errors among them, which is more than half of the mean error rate of 3.2 when the audio/visual students were included among them in the above analysis.



Figure 2. Difference in translation errors between machines and each of the different human translation groups sampled.

It is also clear from figure 2, that even this lowest ranking human group with the highest number of translation errors ($=10.5$, $SD=7.15$, $n=23$) are still markedly more accurate than the average MT ($=18.4$, $SD=9.07$, $n=10$). And a two-sample t-test between the ten machine translators and this lowest ranking group in the analysis shows that this difference is significant at $\alpha=0.05$ ($t=2.44$, $p=0.029$).

This is also indicated by an analysis of the personal self-evaluation results, in which subjects were asked to evaluate their translation ability prior to completing the

language ability tests and translations. In the survey, subjects were asked to rank their translation ability from 1-5, and because all those recruited to participate in the study had at least some understanding of the English language, only 3, 4, and 5 were selected, which are represented here as low, moderate, and high. A one-way ANOVA test of differences in mean translation errors among these self-evaluated groups shows marked variation among them ($F=14.72$, $p=0.000$), and naturally with the biggest difference in group means between the high, and MT groups (figure 3). However, a significant difference (two-sample t-test- $t=3.02$, $p=0.011$) also exists between the MT group ($=18.4$, $SD=9.07$, $n=10$) and those individuals with the lowest stated ability level ($=8.91$, $SD=6.83$, $n=31$). Here it can be seen that those individuals who ranked themselves as being of a lower translation ability level, were again far more accurate in interpreting the 15 different phrases and sentences compared to the MT tools.



Figure 3. Differences in phrase and sentence errors between machine translators and subjects who stated their translation abilities to be low, moderate, or high.

It is also useful to evaluate differences in translation accuracy among the 10 machine tools used in this study. This information could be valuable for individuals and groups who lack access to human resources for translating sentences from Chinese to English, which have been shown here to be far superior to using these types of freely available machine translators. A comparison of these resources shows that the human group is again the most accurate, and that considerable variation exists among the 10 machine translators, ranging from eight mistakes up to

34 total mistakes across the 15 sentences and phrases used in this study (table 2).

Table 2. Sentence and phrase translation errors across the ten freely available MT tools, and in relation to the average number of errors among the human groups.

Translator	Translation Errors
Human	3.20
bing.com/translator	12.00
dict.cn	10.00
fanyi.baidu.com/translate	9.000
fanyi.youdao.com	18.00
freetranslation.com	29.00
fy.iciba.com	8.000
tran.httpcn.com	22.00
netat.net	26.00
translate.com	16.00
translate.google.cn	34.00

It can be seen from table 2 that the least accurate, and likely the most widely used, is the Google Translate tool. While the most accurate resource, at least for Chinese to English translation, is the Chinese website fy.iciba.com, which only made eight errors across the 15 phrases and sentences in this study. This indicates that not only is it preferable to utilize human translators with at least a minimum level of proficiency in the language, but also that if these resources are not available, it is very important to carefully choose which MT tool will be used.

An initial test of sex differences revealed that males and females showed significant differences in translation accuracy in which mistakes made by males ($=1.97$, $SD=3.27$, $n=30$) were significantly fewer than females ($=3.56$, $SD=5.12$, $n=103$) as indicated by a two-sample t-test ($t=-2.04$, $p=0.0275$). However, this result was actually due to the fact that the second-year English audio/visual studies students, who had the highest number of translation errors of any human group, were entirely female with the exception of one male. Upon further examination, a generalized least squares regression analysis shows that no difference exists between males and females after accounting for this group affiliation effect ($F=0.01$, $p=0.992$).

The same is also true of age as a predictor of sentence and phrase translation error. For example, these two variables are weakly correlated in a bivariate regression analysis ($F=5.95$, $p=0.016$), however, age accounts for only a very small amount of the variations in translation accuracy ($R^2=4.3\%$). And when years of education, which is obviously highly correlated with age ($F=109.94$, $p=0.000$, $R^2=45.6\%$), is controlled for in a least squares regression analysis, age is no longer a significant predictor ($F=0.61$, $p=0.435$) of errors made in translating phrases and sentences from Chinese to English among individuals in the study sample.

Discussion

The above analysis of human versus MT accuracy reveals that humans are far better at translating Chinese to English phrases. Across a broad range of individuals with varying degrees of education and experience in Chinese to English translation beyond the minimum 15 years of association with the English language, humans were consistently better at translating the 15 different sentences and phrases used in this study.

and phrases used in this study.

Comparisons of the mean number of mistakes between the human group and the machine group indicated that humans generate significantly fewer mistakes in translation compared to machines. Additionally, comparing the machine average to that of the human group with the most mistakes, showed that even this worst performing group in the sample was still significantly more accurate than the average of the ten different freely available MT tools used in the analysis. Also, a subsequent analysis of differences in mean translation accuracy among the machine group and those who ranked their translation ability in three different hierarchical categories, revealed that even people at a low stated translation level were still far more accurate than the MT tool average. Results of this study also indicate that considerable variation exists among the freely available MT tools, and that caution is warranted in choosing the best one for translating words, phrases, and sentences between any two languages.

While some difference was initially observed between the sexes with regard to translation ability, these results were actually due to other factors, and specifically differences in sex composition and years of education among the participant groups. Moreover, the latter of these is certainly a logical cofactor, since as age and years of education go up, translators naturally have more experience to draw from, and therefore the quality of translation accuracy is improved.

Taken together, the current study reveals that MT is not as accurate with regard to comprehending and interpreting phrases and sentences, which likely relates to a machines inability to recognize subtleties in meaning, and cultural differences between linguistic groups. If the cultural context is mistranslated, potential misunderstandings and even offense can result. Human translation increases the

translation quality through cultural sensitivity, while MT is limited in its ability to properly and effectively translate cultural perceptions based only on how it was coded to do so.

The benefits of human over MT are apparent, and particularly in a functional and logistic capacity across consumer-focused industries as diverse as healthcare, business, and manufacturing. Additionally, in high-risk situations, precise instruction is needed to avoid misinterpretation. For example, instructions used in the healthcare industry require accurate translation for proper use. Brach, Fraser, and Paez (2005) suggested using professional human translators in a patient's language to improve healthcare quality. From the poorly translated sign example given in the introduction, MT could easily mistranslate the meanings. The sign's translated errors, made by MT, could be dangerous to a person's safety. Machine mistranslation, at an extreme, could result in lawsuits, potential injury, harm or even death for patients.

Incorrect MT can also affect customers' perception about product quality. Liu (2010) found that MT remains a risky proposition that can damage business relations between different cultures. Mistranslations can be seen as dishonest or unethical and can lead to customers doubting the products and a company's ethics. If customers find it difficult to read and understand translated instruction, it will cause them to doubt the product quality. These doubts could lead customers to have a bad impression of and to mistrust the company. Long-term effects of using MT on the company may result in customer dissatisfaction, product rejection, and financial loss.

Limitations of the Study

First, the number of languages examined was obviously limited in the current study. Here, only mistakes made by

MT tools and native speakers of Chinese were recorded for words, phrases, and sentences that were translated from Chinese to English. It is possible that if this same study was carried out translating Spanish to English, or French to Chinese, that the results may be different. However, given the marked number of mistakes made by the machine translators, and consistently significant differences between this and each of the human groups, it is expected that similar results would be found regardless of the languages used.

Secondly, the researchers did not examine patterns of mistranslation made by machines and humans. For example, common errors were not investigated across the sample in order to identify which words or groups of words may have been consistently translated in error, which could potentially add to a better understanding of why some mistakes are made for both humans and machines. Lastly, the sample size for males and females was not balanced, where there were 103 female subjects, but only 30 males in this study. However, because sex was not found to be correlated with mean translation error after controlling for age, years of education, and translation category, this sex disparity is not expected to have affected the results of the study in any way.

Conclusions

MT is a fast way to obtain a domain-specific translation but is not recommended for literary translation. The domain used in this study were non-literary Chinese vocabularies, phrases, and sentences (see Appendix A). These vocabularies, phrases, and sentences range from beginner to advanced levels. Based on the results of this study it is clear that human translators, with at least a moderate exposure to a non-native language, are more accurate than translation carried out using freely available MT tools. Even though numerous MT tools are now available they are

restricted to word-for-word translation; machines are unable to grasp the subtle differences in meaning associated with different cultures. For example, a sentence mistranslated by machines in the current study was “Customers come first”, which when translated literally from Chinese-to-English by machines without cultural input, it turned out to be “The customer is god.” The translation is obviously meaningfully inaccurate, and may not be appropriate for much of the English-speaking language group in this case, as it could even be considered offensive to some. An additional example from this study that was “Watch your head”, which was mistranslated by one of the MT tools as “be careful to meet”, which again is highly inaccurate and could even result in physical injury among English speakers, as it doesn’t even come close to conveying the original warning.

In the long run, mistranslation produced by machines could continue to be culturally offensive; cause harm in high-risk situations; and negatively affect a company’s brand, reputation, and revenues. The results of this research strongly indicate that international companies should adopt human translation resources as opposed to relying on cheaper, but considerably less effective and less accurate MT. Additionally, it is recommended that companies employ team translation where more than one person is involved in the translation process, which would further increase the accuracy of translation as a result of involving multiple individuals with variable interpretations of linguistic connotation and cultural meaning.

References

Aiken, M., Vanjani, M. B., & Wong, Z. (2006). Measuring the accuracy of Spanish-to-English

translations. *Issues in Information Systems*, 7(2), 125-129.

TRANSLATIONS. ISSUES IN INFORMATION SYSTEMS, 7(2), 125-128.

Aiken, M., Ghosh, K., Wee, J., & Vanjani, M. (2009). An evaluation of the accuracy of online

translation systems. *Communications of the IIMA*, 9(4), 67.

Alcina, A. (2008). Translation technologies scope, tools and resources. *Target*, 20(1), 79-102.

Arnold, D. (1994). *MT: an introductory guide*. Blackwell Pub.

Buden, B., Nowotny, S., Simon, S., Bery, A., & Cronin, M. (2009). Cultural translation: An

introduction to the problem, and responses. *Translation Studies*, 2(2), 196-219. doi:

10.1080/14781700902937730

Brach, C., Fraser, I., & Paez, K. (2005). Crossing the language chasm. *Health Affairs*,

24(2), 424- 434. doi: 10.1377/hlthaff.24.2.424

Cui, Y., & Zhao, Y. (2014). The use of second-person reference in advertisement translation with

reference to translation between Chinese and English.

International Journal of Society,

Culture & Language. 26-36.

Drummer, A. (1996) Literature review: MT. *MT for South African Languages*. Retrieved August 28, 2015, from

<http://people.cs.uct.ac.za/~bsharwood/downloads.html>

European Commission. (2012). Quantifying quality costs and the cost of poor quality in

translation. doi:10.2782/44381

García-Santiago, Lola, and María-Dolores Olvera-Lobo. "Automatic Web Translators as Part of a Multilingual Question-Answering (QA) System." (2010).

Gaspari, F., & Hutchins, J. (2007). Online and free! Ten years of online MT:

Origins, developments, current use and future prospects. *Proceedings of the MT Summit XI*, 199-206.

He, Sanning. (2009). Basic approaches to improve translation quality between English and

Chinese. *Asian Social Science*, 4(7), 92.

Jacobs, E. A., Shepard, D. S., Suaya, J. A., & Stone, E. L. (2004). Overcoming language

barriers in health care: Costs and benefits of interpreter services. *American Journal of*

Public Health, 94(5), 866-869.

Kazemzadeh, A. A., & Fard Kashani, A. (2013). The effect of computer-assisted translation on

L2 learners' mastery of writing. *International Journal of Research Studies in Language*

Learning, 3(3).

Kit, C., & Wong, T. M. (2008). Comparative evaluation of online machine translation systems

with legal texts. *Law Libr. J.*, 100, 299.

Liu, Y. (2010). The dangers and risks of using MT for your English to Chinese translation project. Retrieved from Nanjing Shanglong Communications Co. Ltd.: <http://www.chinatranslations.com/risks-of-machine->

translation-in-Chinese-to-English-and-English-to-Chinese-translation.html

Munday, J. (2009). *Introducing translation studies: Theories and applications*. Routledge.

Newton, J. (Ed.). (2002). *Computers in translation: a practical appraisal*. Routledge.

Olohan, M. (2011). Translators and translation technology: The dance of agency. *Translation Studies*, 4(3), 342-357.

Sathe, N. (2006). Interpreting the language of healthcare. *Letter from the Editors*, 7.

Seljan, S., Brkić, M., & Kučič, V. (2011, January). Evaluation of free online machine translations

for Croatian-English and English-Croatian language pairs. In Proceedings of the 3rd International Conference on the Future of Information Sciences: INFUTURE2011-Information Sciences and e-Society (pp. 331-345).

Seljan, S., Tucaković, M., & Dunder, I. (2015). Human Evaluation of Online Machine

Translation Services for English/Russian-Croatian. In *New Contributions in Information Systems and Technologies* (pp. 1089-1098). Springer International Publishing.

Somers, H. (Ed.). (2003). *Computers and translation: a translator's guide* (Vol. 35). John

Benjamins Publishing.

Sun, Y. (2003). Translating cultural differences. *Perspectives: Studies in Translatology*,

11(1), 25-36. doi: 10.1080/0907676X.2003.9961459

Iriratni, S., & Sarknel, J. K. (2010). Approaches to MT.

Annals of library and

information studies, 57, 388-393.

Vilar, D., Xu, J., d'Haro, L. F., & Ney, H. (2006, May). Error analysis of statistical machine

translation output. *In Proceedings of LREC* (p. 697-702).

Walmer, D. (1999). One company's efforts to improve translation and localization. *Technical*

Communication, 46(2), 230-23.

Ye, Z., & Shi, L. X. (2009). *Introduction to Chinese-English Translation*. Hippocrene Books.

Appendix A. Tests of Language Proficiency

The principle aim of this study is to collect information about the accuracy of human translation as opposed to MT. There are no perceived risks associated with taking part in this experiment. This study is completely anonymous. Participation is voluntary, and subjects' consent will be implied by their proceeding into the study. If you have any questions, comments, or concerns, please contact Shihua Chen Brazill by phone at 406-548-7481 or email at shihuabrazill@gmail.com

Mandarin Chinese-to-English Proficiency Vocabulary Test

- Gender:
- Age:
- Year of education:
- Translation level: 1 2 3 4 5

(Please circle one of the numbers above, 1 is for beginner, 5 is for expert.)

Number	Chinese Word	English Translation
1	我	I
2	车	Car
3	九	Nine
4	你	You
5	上	Up
6	草	Grass
7	抓住	Catch
8	梦想	Dream
9	六月	June
10	主意	Idea
11	杰出的	Outstanding
12	解决方案	Solution
13	援助	Aid
14	问题	Issue
15	昏昏欲睡的	Drowsy
16	悬崖	Cliff
17	教堂	Church
18	开始	Start

18	已知	已知
19	偷偷的做	Sneak
20	领先	Ahead
21	摘要	Abstract
22	无自信	Diffidence
23	统一的	Uniform
24	节俭的	Frugal
25	对话	Dialogue

Chinese-to-English Phrase and Sentence Translation Test

Number	Chinese Sentences	English Translation
1	红茶	Black tea
2	加油	Cheer up!
3	高考	College entrance examination
4	吃药	Take medicine
5	我很喜欢它。	I like it very much.
6	他的身体很健康。	He is very healthy. / He is in good health.

7	我没有英文名。	I do not have an English name.
8	我的学习很忙。	I am very busy studying.
9	该房子正在建造中。	The house is under construction.
10	你应当把你的物品保管好。	You should take good care of your things.
11	不准拍照。	No photography.
12	顾客是上帝。	Customers come first.
13	自觉排队。	Stand in line.
14	小心地滑。	Wet floor.
15	小心碰头。	Watch your head. / Lower your head.

Appendix B. MT Applications Information

MT Applications Information
translate.google.cn
fanyi.youdao.com
fy.iciba.com
dict.cn

fanyi.baidu.com/translate
http://tran.httpcn.com/
http://www.netat.net/
http://www.freetranslation.com/en/translate-english-chinese
http://www.bing.com/translator/
https://www.translate.com/

« Film translation in the People's Republic of Poland: cultural problems and reception

The “translation handbook” method applied to primary sources »

Copyright © Department of Translation Studies, Faculty of Arts, Constantine the Philosopher University in Nitra, Slovakia. All rights reserved. Theme: **Education Care**