

Losses and Gains in Computer-Assisted Translation: Some Remarks on Online Translation of English to Malay

Mozhgan Ghassemiazghandi

University Sains Malaysia

Tengku Sepora Tengku Mahadi

University Sains Malaysia

ABSTRACT

The article begins with a concise study of the significance of the translation technology in modern life as well as machine and computer-assisted translation. It then describes the technology accessible to translators and examines the losses and gains of the used tools in computer-assisted translation such as the electronic dictionaries, and specifically Google translate. The paper studies the influence of the online dictionaries on the professional translator with a view to determining the extent translation done using such dictionaries can be accurate. Loss in machine translation is inevitable due to the differences between English and Malay as they are entirely two different languages and not-related language pairs for translation. The online dictionary and translation software cannot replace the human translator and guarantee high-quality translations, despite their efficiency and outlooks. Online dictionaries and other translation means accelerate and facilitate the translation process only by minimizing the expected time for translation. Combination of electronic technologies with comprehensive knowledge of the translator and translation theory may result in a high-quality translation. Translation software and programs nonetheless, will not replace humans even in the future. As mentioned, the main aim of the paper is to investigate the new technologies in machine translation tools to discover the losses and gains in translation of English to Malay by using online dictionaries. Machine translations employing online dictionaries are compared with the translation done by a human translator to analyze the probable errors in machine-translated texts.

1. Introduction

Nowadays, machine translation (MT) is a significant technology that represents an essential part of natural language processing systems. The quality and the number of specialized dictionaries of the software define the efficiency of machine translation (González-Rubio & Casacuberta, 2014). According to European Commission, one of the world's largest translation services, the EU contracts 1,750 translators on a full-time basis due to global multilingualism. The EU employs external translation providers that produce almost one-fourth of its output

translation to cope with demand variations. The EU translation services translated more than 1,800,000 pages, which cost about one billion Euros in 2008 (EC, 2009). Given this volume, human translation alone without computer assistance would logically not be practical.

For the purpose of this paper, a human translator assisted only by a computer for specific tasks such as typing and looking up specialized terms and expression in the dictionary is referred to as human translation. Based on the data given above, human translating will be a high-priced and time-consuming method. There are not adequate qualified translators, while the need for superior translation has undeviatingly increased. To minimize this problem, computer and computer technology need to be relied upon to facilitate the computerized translation of extensive numbers of documents. Professional translators should post-edit the automatic translation for an accurate translation (Alabau, Sanchis, & Casacuberta, 2014).

Modern machine translation systems are not capable of delivering high-quality translation without human translator intervention. The human translator must edit the output of machine translation and translation software to improve the quality of translation. Computer-assisted translation (CAT) is an interactive translation process between human and computer. Human translator uses computer software just in order to facilitate and accelerate the translation process (Barrachina et al., 2009).

2. Losses and Gains in Computer Translating

Even the most advanced software does not have the skills of a professional translator and fluency of native speakers. Machine translation is not a straightforward task, as each word might have different meanings according to context. An accurate translation needs a good understanding of context as well as language structure and rules. Computer needs some capabilities to deal with translation difficulties in the same way as human.

A computer can only translate non-ambiguous texts that exist in the computer's dictionary with the same meaning. The output translation qualifies for personal knowledge and not academic aims. Computers translate technical texts quite well, as technical documents have a restricted topic and monotonous style (Precup-Stiegelbauer & Laura-Rebeca, 2013). Météo, a translation system for weather forecasting from English into French is an example of a successful computer translation in the domain-specific and controlled environment (Hutchins, 2001). However, computers are not able to produce a high-quality translation for other texts, which are more interesting and appealing to readers. On the other hand, professional human translator can translate all ranges of text.

Because many words have multiple meanings, the process of translation is made more difficult. Computers translate words based on a one-to-one substitution without considering other possible meanings, whereas humans choose the proper words according to the contexts. A word with different meanings may have various translations based on how the word is being used in the context. For example, the word 'book' has different meanings, even though the spelling is the same. Computer just translates it to one word '*buku*' in Malay and ignores other existing meanings like 'reservation'. It demonstrates that sense should be taken into account in the translation. Human translators can readily distinguish which meaning best fits the context among the multiple uses of the word. Although, it may still be difficult for them to choose the

best translation, it is still possible with effort. A precise translation necessitates a perception of the text, situation, and various uses of the word in the language to determine the appropriate substitution.

Computers need to distinguish between general vocabulary and specialized words, whereby the former might be culturally influential. Overused words should be avoided in general language, as the variety is highly valued. On the other hand, overused words are allowed in technical translation and specialized terminology, as the consistency is highly valued. Unfortunately, computers are not able to distinguish between general and specialized words (Nitta, 1986; Precup-Stiegelbauer & Laura-Rebeca, 2013). Humans can distinguish between general and specialized use of the words. There are many different terminology databases for specialized terms that help the translator to choose an appropriate word. Computers have an amazing memory as compared to humans, but computers cannot decide and choose the best meaning of the words based on the situation. Computers would be desperately disorganized to deal with general and specialized domains while humans can easily differentiate these two types of texts (Precup-Stiegelbauer & Laura-Rebeca, 2013; Şahin, 2013).

Computers and even novice translators might overlook the differences in meaning and therefore produce poor translation by using inappropriate words. Translators must be familiar not only with the source language, but also with target language and culture in order to produce a useful and reliable translation and deal with any translation difficulties appropriately. An accurate translation also needs to consider the intended audience, regionalism and culture in the total context. Computer is not a native speaker of any language and presumably the cultural knowledge in computer is not comparable with humans. Accordingly, it can be concluded that computers cannot translate like humans, as they are not equipped with the learning potential like humans. The computer is, however, a lifesaver as time is scarce in modern life, and people want to be productive in the shortest period. Computers aid people to accelerate translation work faster than a human translator (Dimitriu & European, 2006; Precup-Stiegelbauer & Laura-Rebeca, 2013).

Google translate is a multilingual translation service to translate written source text to target text that supports 80 languages. Google Translate is the most famous and easily accessible machine translation. Nowadays, Internet increasingly develops over recent years, and Google translate, for example, quickly helps people to get an idea of the foreign language contexts. Google Translate is only applicable to standard and coherent texts, which exist in the computer's dictionary with the same meanings. (Precup-Stiegelbauer & Laura-Rebeca, 2013).

3. Discussion

Several examples of machine translations, from English into Malay are selected to support the aforementioned ideas. The English texts are extracted from *Doctor in the House*, *The Memoirs of Tun Dr. Mahathir Mohamad*. Both the English (Mohamad, 2011) and Malay (Mohamad, 2012) versions of the book are used in this study. The original text is being compared with machine translation and human translation to evaluate the efficiency of translation. Then it will be concluded to what extent the translation is trustable and how much human intervention is necessitated to achieve an accurate translation. Here is the Google Malay translation of some

English paragraphs of Tun Dr. Mahathir Mohammad's A Doctor in the House. Although the hard copy of the Malay translation was used for comparison, comments made in this paper are general, and the translation texts are not presented here due to space constraint.

Original Text	Google Translation
<p>Though my mother had warned me many times that a doctor's life meant having to forgo sleep and work odd hours, she and my father seemed silently pleased that I was going to go to college. Moreover, when I was on contract with the office, I had been paid \$80 a month. Now that I had become full-time and pensionable, I was being paid RM60 monthly.</p>	<p>Walaupun ibu saya telah memberi amaran kepada saya banyak kali bahawa kehidupan yang doktor bermakna perlu melepaskan tidur dan bekerja jam ganjil, dia dan bapa saya seolah-olah senyap gembira bahawa saya akan pergi ke kolej. Lebih-lebih lagi, apabila saya berada di atas kontrak dengan pejabat, saya telah dibayar \$ 80 sebulan. Sekarang bahawa saya telah menjadi sepenuh masa dan berpencen, saya dibayar RM60 setiap bulan.</p>

Example 1: P. 128 Para. 3, A doctor in the house

Reading both the original text and the translation, one would get an understanding about the meaning of the text, but the text structure is grammatically poor. A long process is required to edit the text and to achieve an excellent translation. The sense of a few words has not been captured. For example, the word 'odd' has been translated to '*ganjil*' by computer, which is a literal translation, while the correct translation is '*bekerja tidak tentu masa*'. The literal translation, or one by one substitution of words does not seem enough to give an idea to the reader as more explanation is needed to clarify the text. The preposition used in the paragraph has been translated to '*di atas kontrak*' by computer instead of '*secara kontrak*'. The linking words are missed in Google translation. As discussed earlier, context of culture affects the specific meaning of the language. Computers are not able to analyze the cultural context of Malay language. For example, the word '*ibu*' has been translated to '*dia*' that is correct in terms of general meaning but the cultural aspect of respect and politeness is missed in computer translation.

Original Text	Google Translation
<p>I went to the lab where the second-year students were being introduced to anatomy and asked one of them to uncover one of the corpses. It was a Chinese man and he smelled strongly of formalin. I told myself it was dead and could do me no harm. Slowly, by facing my fears, I managed to reach a point where dead bodies no longer frightened me.</p>	<p>Saya pergi ke makmal di mana pelajar-pelajar tahun kedua sedang diperkenalkan kepada anatomi dan bertanya salah seorang daripada mereka untuk mendedahkan salah satu daripada mayat-mayat. Ia adalah seorang lelaki Cina dan dia berbau kuat formalin. Saya memberitahu diri saya itu telah mati dan saya boleh melakukannya tidak membahayakan. Perlahan-lahan, dengan menghadapi ketakutan saya, saya berjaya sampai ke titik di mana mayat tidak lagi menakutkan aku.</p>

Example 2: P. 292 Para. 5, A doctor in the house

This translation is somehow senseless. It is just disordered words that are translated word-by-word by machine. The second sentence has a vague meaning. The structure and formation of the sentences are not natural, e.g. the phrase '*dia berbau kuat formalin*' should be '*berbau*

formalin yang kuat (smelt of formalin). The computer cannot distinguish the different meanings of the words as some words occur in more than one meaning. For example, the verb *'menyuruh'* has been translated to *'bertanya'* by computer, while it is not about asking question but it is asking to instruct. The pronoun used in the paragraph has been translated to *'itu'* (that) by computer instead of *'ia'* (it). The literal translation of *'boleh melakukannya'* does not seem enough in the last sentence. More explanation, *'dan ia tidak mungkin membahayakan saya'* (it is not likely to harm me), is required for transferring the message appropriately. As can be seen, computer translation is not perfect without post-editing of human.

Original Text	Google Translation
<p>I did not think we could face too many problems establishing these industries if we gave them enough careful thought. But I overestimated the Malaysian capacity to learn how to operate a major industry. At the time it was built, the cement factory was the best in the country. It was located by the sea on the island of Langkawi, off the northwestern coast of Peninsula, close to the limestone hills, which were to be quarried for the plant.</p>	<p>Saya tidak fikir kita boleh menghadapi terlalu banyak masalah mewujudkan industri-industri ini jika kita memberi mereka pemikiran yang cukup berhati-hati. Tetapi saya di luar tafsiran keupayaan Malaysia untuk belajar bagaimana untuk mengendalikan industri utama. Pada masa yang ia dibina, kilang simen adalah yang terbaik di negara ini. Ia terletak di tepi laut di Pulau Langkawi, di luar pantai barat laut Semenanjung, berhampiran dengan bukit-bukit batu kapur yang tidak akan dikuari untuk kilang.</p>

Example 3: P. 329 Para. 3, A doctor in the house

The computer translation of the selected text did not translate most of the words appropriately. The computer cannot make a distinction between the functions of the words. Thus, the first sentence has been translated by *'memberi mereka pemikiran yang cukup berhati-hati'* instead of *'memikirkannya dengan cukup teliti'* and *'di luar tafsiran'* instead of *'terlebih menjangka'*. Moreover, the translation even omitted words such as *'orang'*, which refer to Malaysian people whereas other words have not been translated correctly. A word can change the whole meaning, for example, *'untuk mempelajari cara mengendalikan'* has been translated to *'untuk belajar bagaimana untuk mengendalikan'* rendering it a stilted translation. Another obvious mistranslation is the affirmative sentence that has been translated with a contrary meaning. The phrase *'yang akan dikuarikan untuk'* (will be quarried) has been translated to *'tidak akan'* (will not) in which the original meaning is lost in computer translation.

Original Text	Google Translation
<p>Their armed forces are well organized with groups of increasing sizes under officers of different ranges wielding ascending orders of power and authority. Their huge armies move and act with precision.</p>	<p>Angkatan tentera mereka dengan baik dengan kumpulan-kumpulan yang semakin meningkat saiz di bawah pegawai julat yang berbeza memegang tampuk perintah menaik kuasa. Tentera besar mereka bergerak dan bertindak dengan tepat.</p>

Example 4: P. 391 Para. 4, A doctor in the house

The first sentence is a string of ineffective words with inappropriate associations. The phrase *'pegawai julat yang berbeza'* has been translated by computer to *'kumpulan menjadi*

bertambah besar, which is a mistranslation and meaningless. Moreover, the Malay translation even omitted verbs, for example, *'diuruskan dengan baik dalam kumpulan'* has been translated to *'dengan baik'*. Other words have not been translated appropriately as computers cannot differentiate between the senses of the words. The word *'julat'* is has been used wrongly from the point of syntax and vocabulary. It is an inappropriate translation for *'pegawai-pegawai dari bahagian yang berbeza'* that emphasizes the power and authority of the army in the context.

Original Text	Google Translation
<p>Malaysia is a developing country with a small domestic market, which means nothing that we produce enjoys economies of scale. To industrialise, we needed to protect our infant industries. Accordingly the national car was protected by a lower excise duty than import duty on foreign car. No one really complained because they were used to paying high prices for imported car. The government did not lose out either as higher sales of Proton and an expanding market earned it increased revenue. At the time it was built, the cement factory was the best in the country. It was located by the sea on the island of Langkawi, off the northwestern coast of Peninsular, close to the limestone hills which were to be quarried for the plant.</p>	<p>Malaysia adalah sebuah negara membangun dengan pasaran domestik yang kecil, yang bermaksud apa-apa yang kita menghasilkan menikmati ekonomi bidangan. Untuk perindustrian, kita diperlukan untuk melindungi industri baru kami. Itu kereta nasional telah dilindungi oleh duti eksais yang lebih rendah daripada duti import ke atas kereta asing. Tiada siapa yang benar-benar mengadu kerana mereka digunakan untuk membayar harga yang tinggi untuk kereta import. Kerajaan tidak ketinggalan sama ada sebagai jualan yang lebih tinggi daripada Proton dan pasaran yang lebih luas yang diperolehi ia meningkat. Pada masa yang ia dibina, kilang simen adalah yang terbaik di negara ini. Ia terletak di tepi laut di Pulau Langkawi, di luar pantai barat laut Semenanjung, berhampiran dengan bukit-bukit batu kapur yang tidak akan dikuari untuk kilang.</p>

Example 5: P. 517 Para. 2

Reading the original text and Malay translation, the meaning of the text is dropped. The original text means the current activity did not benefit from economies of scale, while Google has been translated with the opposite meaning that every production benefits from economies of scale. It is an obvious mistranslation when the text includes *'apa-apa'* (anything) for 'nothing' which should be *'tiada yang kita hasilkan'*. A long process is required to achieve an accurate translation. Some words such as *'meningkatkan'* (to increase) are missing in translation. Moreover, the sense of many words has not been taken into account. Another example is a passive form *'digunakan'* (used to) which describes an action that was normal for them, while it has been translated as 'apply' or 'use' in the Malay translation that is totally different in meaning. Computers are not able to put the words into the coherent and relevant order. The words *'juga tidak rugi kerana jualan Proton lebih tinggi dan pasaran lebih luas'* has been translated by computer to *'tidak ketinggalan sama...'* and thereby, reflected a mistranslation. Other near-mistranslations include *'kita diperlukan'* for 'we needed it' which should mean *'kita perlu'*. Some words are mistranslated in Malay translation as well; *'kita'* (teller and listener both are involve in the activity) has been translated to 'kami' (while listener is not involved) by computer.

4. Result

From the small number of selected examples, we can see that there are many mistranslations and near-mistranslations and potentially ambiguous translations. Computer-

assisted translation or Google Translate in this case study is a useful tool restricted only to the aspect of memory and time. The product is not yet high-quality translation without the intervention of the human mind. Computers undoubtedly do not translate like humans despite the increasingly advanced specialized dictionaries added to the software. Sometimes, computers skip some words, as they cannot recognize them all. Computer software is unable to recognize the vagueness, complex sentence compositions, allegories, metaphors, and unknown sentences. While computer-assisted translation is time and effort saving, machine translations are of little value without human intervention. Machine translation cannot capture the hidden details and the mood of the source text because they do not resolve like the human mind does. To sum up, computer translation technology, specifically Google translate is not accurate. Imperfections of computer translation must be compensated by the intervention of intelligent humans with a computer-assisted translation.

References

- Alabau, Vicent, Sanchis, Alberto, & Casacuberta, Francisco. (2014). Improving on-line handwritten recognition in interactive machine translation. *Pattern Recognition*, 47(3), 1217-1228. doi: <http://dx.doi.org/10.1016/j.patcog.2013.09.035>
- Barrachina, Sergio, Casacuberta, Francisco, Cubel, Elsa, Lagarda, Antonio, as, Jesus Tom ´, Vilar, Juan-Miguel, . . . Vidal, Enrique. (2009). Statistical Approaches to Computer-Assisted Translation. *Computational Linguistics*, 35(1), 3-28.
- Dimitriu, R., & European, Institutul. (2006). *The Cultural Turn in Translation Studies*: Institutul European.
- EC. (2009). *Translating for a multilingual community*. European Commission, Directorate General for Translation Retrieved from http://bookshop.europa.eu/is-bin/INTERSHOP.enfinity/WFS/EU-Bookshop-Site/en_GB/-/EUR/ViewPublication-Start?PublicationKey=HC3210532
- González-Rubio, Jesús, & Casacuberta, Francisco. (2014). Cost-sensitive active learning for computer-assisted translation. *Pattern Recognition Letters*, 37(0), 124-134. doi: <http://dx.doi.org/10.1016/j.patrec.2013.06.007>
- Hutchins, John. (2001). Machine Translation and Human Translation: In Competition or in Complementation? *International Journal of Translation*, 13(1-2).
- Mohamad, M. (2011). *A Doctor in the House: The Memoirs of Tun Dr. Mahathir Mohamad*: MPH Group Publ.
- Mohamad, M. (2012). *Doktor umum: memoir Tun Dr. Mahathir Mohamad*: MPH Group Publishing.
- Nitta, Yoshihiko. (1986). Problems of machine translation systems: Effect of cultural differences on sentence structure. *Future Generation Computer Systems*, 2(2), 101-115. doi: [http://dx.doi.org/10.1016/0167-739X\(86\)90004-X](http://dx.doi.org/10.1016/0167-739X(86)90004-X)
- Precup-Stiegelbauer, & Laura-Rebeca. (2013). Automatic Translations Versus Human Translations in Nowadays World. *Procedia - Social and Behavioral Sciences*, 70(0), 1768-1777. doi: <http://dx.doi.org/10.1016/j.sbspro.2013.01.252>

Şahin, Sedat. (2013). Computer (Mechanical) Translations in Future. Procedia - Social and Behavioral Sciences, 70(0), 1312-1315. doi: <http://dx.doi.org/10.1016/j.sbspro.2013.01.192>