

A Pilot Study Exploring User Acceptance of Machine Translation

Lynne Bowker

School of Translation and Interpretation
University of Ottawa
Ottawa, Canada
Lynne.Bowker@uottawa.ca

Melissa Ehgoetz

School of Translation and Interpretation
University of Ottawa
Ottawa, Canada
mehgoetz@rogers.com

Abstract

The volume of text to be translated keeps rising and translators are under pressure. Where texts are translated to meet the requirements of official language policies, some may be intended for “information only” and may not require high-quality translation. If machine translation (MT) could be used in such cases, translators would be able to focus on texts that do require higher quality. This article reports on a pilot study that investigates the reception of MT output among university professors, where an MT system was used to translate internal memos. Participants were asked to rate 3 different target texts for the same source text, taking into account quality, cost and speed with which each was produced. The first target text was a human translation; the second a raw machine translation; the third a post-edited version of the same machine translation. Two thirds of the respondents indicated a willingness to accept post-edited MT output, while the remaining third preferred to receive human translations.

1 Introduction

The volume of text to be translated has increased radically, and this has been coupled with a shortage of translators. Settings that have a policy of official multilingualism create another challenge: the translation of many texts is legislated, which may mean that texts that would not normally be translated elsewhere (e.g. memos) must be translated, which can create a large volume of work.

One way to ease the burden may be to consider making a greater use of MT for certain tasks. MT systems can produce translations more quickly and often more cheaply than human translators; however, in the majority of cases, the quality of MT is inferior to the quality of human translation (HT).

To test the hypothesis that people will accept a lower quality of translation for some types of texts

if the price is cheaper and the turnaround time faster, we conducted a pilot study at the University of Ottawa, which is a fully bilingual educational institution offering all services in both French and English. This policy of official bilingualism generates a considerable amount of translation work.

2 Objective

The pilot study aims to determine whether recipients of certain types of translated texts would accept lower quality translations in exchange for savings in time and cost, or whether they would prefer to receive higher quality translations, even if this means a higher price and slower turnaround.

3 General Methodology

There are many different ways to evaluate translation in general and MT in particular. The method selected depends on the goal of the evaluation and on which stakeholders are involved (e.g. researchers, sponsors, developers, purchasers, translators, recipients). For example, Chesterman and Wagner (2002:80) note that one approach is to view translation as a service, intangible but wholly dependent on customer satisfaction; therefore, to measure translation quality, one need only measure customer satisfaction. This approach is referred to as a *recipient evaluation*, described by Trujillo (1999: 255) as an evaluation performed by the translation recipients in order to evaluate quality, cost and speed. According to Loffler-Laurian (1996:69), the recipients of MT output are in the best position to judge whether or not it satisfies their requirements.

The recipient evaluation that is outlined here differs from other well-known recipient evaluations (e.g. Henisz-Dostert 1979; Vasconcellos and Bostad 1992) in a number of ways. Firstly, the participants in our pilot study were given multiple versions of a translated text (a human translation, a raw MT text, and a post-edited MT text) to compare. Moreover, they were aware of which mode of translation had

been used to produce each version. In addition, recipients received information about the time and cost required to produce each of these versions. To the best of our knowledge, no other recipient evaluation has made this type of information available to the participants.

4 Source Texts

It is well known that not all types of texts are equally amenable to being translated by MT systems. High profile texts destined for public dissemination are best left to human translators. However, more mundane texts typically require a fast turnaround time because the information they contain must be distributed to employees in a timely fashion (e.g. there is no point announcing a meeting after the fact). Therefore, these low prestige jobs cannot be given a low priority. Rather, they must be dealt with swiftly, but maybe clients would accept a lower quality in return for speed.

We selected the text type *administrative memos*, which typically contain information such as announcements of meetings, invitations to events, calls for nominations for committees, results of elections, notification of maintenance work.

These seemed to be a good choice for MT for the following reasons: 1) they are internal documents intended for “information only” purposes, 2) they are short (200 words or less), 3) they contain simple sentence structures, 4) they are relatively unambiguous, 5) the specialized terms can be entered into the MT system dictionary, 6) they need to be translated quickly.

In total, we gathered 17 different memos to use as source texts. In their original form, these memos were bilingual in nature; however, for the purpose of this pilot study, the French-language portion of each memo was retained for use as a source text.

5 Participants

We identified 121 professors in the Faculty of Arts who use English as their primary language at work.

6 Pilot Study

For each source text, three different English-language translations were procured:

- (A) a human translation, which consisted of the English version of the original bilingual memo;

- (B) a raw (i.e. unedited) machine translation produced using the desktop version of the Systran MT system (v4.0);

- (C) a post-edited version of the MT output where a human translator corrected mistakes that were found in the raw MT output generated in (B).

The time and cost for producing the three versions of each source text is summarized in Table 1. Note that in all cases, raw MT output was by far the fastest and cheapest, while HTs always took the longest time and had the highest price. The post-edited MT always fell between these two extremes, usually having a production time that was 3 to 4 times lower than that required to produce a HT, and typically costing 3 to 4 times less as well.

Text	Word count	HT (A)		Raw MT (B)		PE MT (C)	
		Time (min)	Cost (CD\$)	Time (min)	Cost (CD\$)	Time (min)	Cost (CD\$)
1	73	17.51	8.76	2	0.84	5	2.94
2	108	25.90	12.96	2	0.84	7	3.78
3	89	21.34	10.68	2	0.84	5	2.94
4	65	15.59	7.80	2	0.84	5	2.94
5	160	38.37	19.20	2	0.84	10	5.04
6	63	15.11	7.56	2	0.84	7	3.78
7	77	18.47	9.24	2	0.84	5	2.94
8	46	11.03	5.52	2	0.84	5	2.94
9	158	37.89	18.96	2	0.84	9	4.62
10	162	38.85	19.44	2	0.84	8	4.20
11	76	18.23	9.12	2	0.84	5	2.94
12	74	17.75	8.88	2	0.84	4	2.52
13	114	27.34	13.68	2	0.84	11	5.46
14	200	47.96	24.00	2	0.84	9	4.62
15	189	45.32	22.68	2	0.84	10	5.04
16	57	13.67	6.84	2	0.84	4	2.52
17	150	35.97	18.00	2	0.84	7	3.78

Table 1. Summary of production times and costs.

6.1 Human translations

It is generally accepted that a translator can translate about 2000 words per day assuming an 8-hour (or a 460-minute) work day. Dividing the total number of words by the total number of minutes produces a rate of 4.17 words per minute. To determine the amount of time required to produce each text, the total number of words was divided by 4.17 to arrive at the number of minutes.

To calculate the cost of producing the texts, the conservative rate of \$0.12/word was used. This is much lower than the average rate of \$0.21 charged by translators in Canada (OTTIAQ 2002). Nevertheless, the decision was made to charge a lower rate because of the straightforward nature of the texts (i.e. short; unambiguous; simple syntax).

6.2 Raw MT output

For MT, tests were conducted which revealed that it took approximately 2 minutes to open the software package, import the source text, and run it through the translation engine. Since all the texts were less than 200 words, it meant that in practical terms, there was no difference in processing time.

To calculate the cost of producing raw MT output, we used the conservative hourly rate¹ of \$25/hour, which is considerably lower than the average rate of \$50.41 (OTTIAQ 2002). We felt it was fair to charge a lower rate because running the MT software is relatively straightforward. This works out to a rate of \$0.42/minute. Since the total processing time for each text was two minutes, it cost \$0.84 to generate raw MT output.

6.3 Post-edited MT output

To obtain the post-edited texts, a human translator took the raw MT output and, in consultation with the source text, revised it. Different levels of post-editing can be applied to MT output depending on the degree of quality required for the final text (Allen 2003). The type used in this pilot study was rapid post-editing (RPE), which seeks to correct blatant and significant errors but not stylistic ones.

The post-editor kept track of the time required to post-edit each text, which varied depending on the length and on the number of problems the MT system encountered. This time was added to the amount of time required to produce the raw MT output (i.e. 2 minutes) to arrive at the total time required to produce a post-edited text.

To calculate the cost of producing the post-edited versions, we again used the conservative rate of \$25/hour. We felt this was justified because the type of editing required was RPE, rather than full post-editing, which aims to produce a level of quality comparable to that of a high-quality HT.

7 Experiment

Once the three versions of each text had been produced, packages were prepared for the participants. The instructions given were as follows:

1. Please read the three different translated versions (A, B, C) of the short text.
2. Please consult the information regarding the time/cost of producing each version.

3. Balancing overall quality alongside the time/cost of producing the texts, please rank the three versions in order from best (1) to worst (3) in terms of which you would be most satisfied to receive.
4. Comments: Please feel free to add any comments that you feel are relevant.

Note that participants were *not* given the original French-language source text as a reference.

8 Results and Discussion

The survey was distributed to 121 professors in the Faculty of Arts; however, only 31 responses were received, which represented a 25.6% response rate. The responses can be summarized as follows:

- HT as preferred choice: 10 (32.3%)
- Raw MT as preferred choice: 0
- PE MT as preferred choice: 21 (67.7%)

Of the responses received, 67.7% chose the post-edited MT version as the text that they would be most satisfied to receive when balancing the quality of the text against the time and cost required to produce it. Note that although professors are not required to pay for such translations directly, they still have a vested interest in saving the university time and money with regard to translation. If the translations are produced more quickly, then they will receive the relevant information in a more timely fashion, and if the university is able to save money on translation, these savings can be re-directed to other needs at the university. A number of participants suggested, however, that if post-edited MT were to be used to translate memos, a note should be added to the bottom of the memo so that readers were aware of the fact that the text they were reading was MT-generated.

Meanwhile, 32.3% of the respondents indicated that they would prefer to receive the HTs even though they were more expensive and took longer to produce. They felt that even internal-use texts should be high-quality translations. One possible reason for this may be the fact that a considerable number of the participants are language specialists of one type or another. Both Church and Hovy (1993:253) and Wagner *et al.* (2002:90-91) observe that translators seem to object to the use of MT more than clients do. Of the 14 departments that make up the Faculty of Arts, there are at least 8 that are language-oriented to some degree: Classics, Communications, English, Lettres françaises, Linguistics, Modern Languages and Literatures, Second

¹ In Canada, it is normal for translators to charge by the word for translation but by the hour for editing or revision.

Language Institute, and Translation and Interpretation. It is therefore possible that the nature of their work led some participants to be particularly sensitive to issues of language quality, whereas non-language professionals may have been more willing to accept the lower quality texts.

Finally, no one felt that the raw MT output was acceptable and it was consistently rated as the worst option. In fact, in their comments, many participants indicated that the raw MT output was “completely unacceptable” because it contained some errors and it took too long for readers to work out the intended message. Therefore, any time saved by the translator in producing the target text was lost many times over by each reader.

9 Conclusions and Future Research

While raw MT output did not appear to meet the basic needs of the participants, slightly more than two thirds of them felt that the quality of the post-edited MT output was sufficient given the purpose of the text, and they felt it represented the best value for money in terms of production cost and time. In addition, translators at the university’s translation service indicated they would be happier to spend less time working on routine texts and devote more energy to texts that were more interesting and challenging. Thus, this pilot study suggests that the application of MT to “information only” texts in an environment where translation is legislated to meet the needs of a policy of official bilingualism is worthy of further exploration.

As a next step, it would be very interesting to conduct a similar experiment in the opposite language direction (i.e. translating from English into French) to see if those recipients would react differently to receiving texts generated using MT. It is well known that language is a sensitive issue, particularly in cases where so-called majority and minority languages co-exist. Even though the University of Ottawa has a policy of bilingualism, the university is located in the province of Ontario, which is predominantly English speaking, and within the country of Canada, which is also predominantly English speaking. It is therefore possible that Francophones would react differently than Anglophones when faced with the choice of receiving human translations vs MT output. It could be the case, for example, that if a group is seeking to preserve or strengthen a minority language, the group members would prefer to have high-quality translations even if the production price is higher. Further research

would need to be conducted to determine whether or not this is a factor, and we are in the process of setting up additional experiments to explore this in more detail.

References

- Allen, J. (2003) ‘Post-editing’, in Harold Somers (ed) *Computers and Translation: A Translator’s Guide*, Amsterdam/Philadelphia, John Benjamins, 297-317.
- Chesterman, A. and E. Wagner (2002) *Can Theory Help Translators? A Dialogue Between the Ivory Tower and the Wordface*, Manchester: St. Jerome Publishing.
- Church, K.W. and E.H. Hovy (1993) ‘Good Applications for Crummy Machine Translation’, *Machine Translation*, 8: 239-258.
- Henisz-Dostert, B., R.R Macdonald and M. Zarechnak (1979) *Machine Translation*, Mouton Publishers, The Hague.
- Loffler-Laurian, A.-M. (1996) *La traduction automatique*, Villeneuve d’Ascq: Presses Universitaires du Septentrion.
- Ordre des traducteurs, terminologues et interprètes agréés du Québec (OTTIAQ) (2002) ‘2002 Survey on Rates and Salaries’ [distributed to OTTIAQ members]
- Trujillo, A. (1999) *Translation Engines: Techniques for Machine Translation*, London: Springer.
- Vasconcellos, M. and D.A. Bostad (1992) ‘Machine Translation in a High-Volume Translation Environment’, in John Newton (ed) *Computers in Translation: A Practical Appraisal*. London/New York, Routledge, 58-77.
- Wagner, E., S. Bech and J.M. Martínez (2002) *Translating for the European Union Institutions*, Manchester: St. Jerome Publishing.