

Speech-Enabled Computer-Aided Translation: A Satisfaction Survey with Post-Editor Trainees

Bartolomé Mesa-Lao

Center for Research and Innovation in Translation and Translation Technology
Department of International Business Communication
Copenhagen Business School, Denmark
bm.ibr@cbs.dk

Abstract

The present study has surveyed post-editor trainees' views and attitudes before and after the introduction of speech technology as a front end to a computer-aided translation workbench. The aim of the survey was (i) to identify attitudes and perceptions among post-editor trainees before performing a post-editing task using automatic speech recognition (ASR); and (ii) to assess the degree to which post-editors' attitudes and expectations to the use of speech technology changed after actually using it. The survey was based on two questionnaires: the first one administered before the participants performed with the ASR system and the second one at the end of the session, once they have actually used ASR while post-editing machine translation outputs. Overall, the results suggest that the surveyed post-editor trainees tended to report a positive view of ASR in the context of post-editing and they would consider adopting ASR as an input method for future post-editing tasks.

1 Introduction

In recent years, significant progress has been made in advancing automatic speech recognition (ASR) technology. Nowadays it can be found at the other end of customer-support hotlines, it is built into operating systems and it is offered as an alternative text-input method in many mobile devices. This technology is not only improving at a steady pace, but is also becoming increasingly usable and useful.

At the same time, the translation industry is going through a societal and technological change in its evolution. In less than ten years, the industry is considering new tools, workflows and solutions to service a steadily growing market. Given the significant improvements in machine translation (MT) quality and the increasing demand for translations, post-editing of MT is

becoming a well-accepted practice in the translation industry, since it has been shown to allow for larger volumes of translations to be produced saving time and costs.

Against this background, it seems reasonable to envisage an era of converge in the future years where speech technology can make a difference in the field of translation technologies. As post-editing services are becoming a common practice among language service providers and ASR is gaining momentum, it seems reasonable to explore the interplay between both fields to create new business solutions and workflows.

In the context of machine-aided human translation and human-aided machine translation, different scenarios have been investigated where human translators are brought into the loop interacting with a computer through a variety of input modalities to improve the efficiency and accuracy of the translation process (e.g., Dragsted et al. 2011, Toselli et al. 2011, Vidal 2006). ASR systems have the potential to improve the productivity and comfort of performing computer-based tasks for a wide variety of users, allowing them to enter both text and commands into the computer using just their voice. However, further studies need to be conducted to build up new knowledge about the way in which state-of-the-art ASR software can be applied to one of the most common tasks translators face nowadays, i.e. post-editing of MT outputs.

The present study has two related objectives: First, to report on a satisfaction survey with post-editor trainees after showing them how to use ASR in post-editing tasks. Second, based on the feedback provided by the participants, to assess the change in users' expectations and acceptance of ASR technology as an alternative input method for their daily work.

2 Method

In this study, we explore the potential of combining one of the most popular computer-aided translation workbenches in the market (i.e. memoQ) with one of the most well-known ASR packages (i.e. Dragon Naturally Speaking from Nuance).

2.1 Overview

Two questionnaires were developed and deployed as a survey. The survey was divided into two phases, a prospective phase in which we surveyed post-editor trainees' views and expectations toward ASR and a subsequent retrospective phase in which actual post-editor's experiences and satisfaction with the technology were surveyed. Participants had to answer a 10-item questionnaire in the prospective phase and a 7-item questionnaire in the retrospective phase. These two questionnaires partially overlapped, allowing us to compare, for each participant, the answers given before and after the introduction and use of the target technology.

2.2 Participants profile

Participants were recruited through the Universitat Autònoma de Barcelona (Spain). The group included 11 females and 4 males, ranging in age from 22 to 35. All 15 participants had a full degree in Translation and Interpreting Studies and were regular users of computer-aided translation software (mainly memoQ and SDL Trados Studio). All of them had already performed MT post-editing tasks as part of their previous training as translators and, at the moment of the data collection, they were also taking a 12-hour course on post-editing as part of their master's degree in Translation. None of the participants had ever user Dragon Naturally Speaking, but four participants declared to have tried the speech input options in their mobile phones to dictate text messages.

2.3 Procedure

Individual sessions occurred at a university office. In the first part of the session, each participant had to complete an on-line questionnaire. This initial survey covered the following topics:

1. General information about their profile as translators; including education, years of experience and employment status.

2. Background in computer-aided translation software in their daily life as professional translators.
3. Experience in the field of post-editing MT outputs and training received.
4. Information about their usage of ASR as compared to other input methods and, if applicable, likes and dislike about it.

In the second part of the session, after the initial questionnaire was completed, all participants performed two post-editing tasks under the following two input conditions (one each):

- Condition 1: non-ASR input modality, i.e. keyboard and mouse.
- Condition 2: ASR input modality combined with other non-ASR modalities, i.e. keyboard and mouse.

The language pair involved in the tasks was Spanish to English¹. Two different texts from the domain of mobile phone marketing were used to perform the post-editing tasks under condition 1 and 2. These two texts were imported to a memoQ project and then fully pre-translated using MT coming from the Google API plug-in in memoQ. The order of the two input conditions and the two texts in each condition were counterbalanced across participants.

In an attempt to unify post-editing criteria among participants, all of them were instructed to follow the same post-editing guidelines aiming at a final high-quality target text². In the ASR input condition, participants also read in hard copy the most frequent commands in Dragon Naturally Speaking v.10 that they could use to post-edit using ASR (*Select <w>*, *Scratch that*, *Cut that*, etc.). All of them had to do the basic training tutorial included in the software (5 minutes training on average per participant) in order to improve the recognition accuracy. Following the training, participants also had the chance to practice the dictation of text and commands before actually performing the two post-editing tasks.

¹ Participants performed from L1 to L2.

² The post-editing guidelines distributed in hard copy were: i) Retain as much raw MT as possible; ii) Do not introduce stylistic changes; iii) Make corrections only where absolutely necessary, i.e. correct words and phrases that are clearly wrong, inadequate or ambiguous according to English grammar; iv) Make sure there are no mistranslations with regard to the Spanish source text; v) Publishable quality is expected.

In the third part of the session, participants completed a 7-item post-session questionnaire regarding their opinions about ASR while post-editing.

2.4 Data collection and analysis

Survey data

For questionnaires’ data, responses to quantitative items were entered into a spreadsheet and mean responses were calculated across participants. For a comparison of responses to different survey items, paired statistics were used: paired t-test for items coded as ordinal variables, and chi-square test for items coded as categorical variables. The questionnaires did not include open-ended questions or comments.

Task log files

For task performance data (which is not going to be elaborated in this paper), computer screen including audio was recorded using BB FlashBack Recorder Pro v. 2.8 from Blueberry Software. With the use of the video recordings, a time-stamped log of user actions and ASR system responses was produced for each participant. Each user action was coded for the following: (i) input method involved; (ii) for the post-editing task involving ASR, text entry rate in the form of text or commands, and (iii), for the same task, which method of error correction was used.

Satisfaction data

Responses to the post-session questionnaire were entered and averaged. We computed an overall ASR “satisfaction score” for each participant by summing the responses to the seven items that related to satisfaction with ASR. We computed a 95 percent confidence interval (CI) for the mean of the satisfaction score to create bounded estimated for the satisfaction score.

3 Survey results

3.1 Usage of speech input method

To determine why participants would decide to use ASR in the future to post-edit, we asked them to rate the importance of eight different reasons, on a scale of 1 to 7, with 7 being the highest in importance. The top reason for deciding to use ASR was that it would involve less fatigue (Table 1).

Reasons for using speech input method	Mean	95% CI
Less fatigue	5.6*	4.9, 6.4
Speed	5.5*	4.8, 6.3
Ease of use	4.9*	4.7, 5.3
Cool technology	4.7*	4.0, 4.8
Limited alternatives	3.1	2.9, 3.3
Accuracy	2.9	2.1, 3.2
Personal preference	2.7	2.3, 2.9
Others	1	1, 1.2

* Reasons with importance significantly greater than neutral rating of 4.0 ($p < 0.05$)

Table 1: Importance of reasons for using automatic speech recognition (ASR), rated on a scale from 1 to 7.

3.2 Usage of non-speech input methods

Since none of the participants had ever used ASR to perform any of their translation or post-editing assignments before, and in order to understand the relative usage data, we also asked participants about their reasons for choosing non-speech input methods (i.e. keyboard and mouse). For this end, they rated the importance of six reasons on a scale of 1 to 7, with 7 being most important. In the introductory questionnaire, most participants believed that keyboard shortcuts would be quicker and easier than using spoken commands (Table 2).

Reasons for using non-speech input methods	Mean	95% CI
They are easier	6.5*	5.7, 6.8
Less setup involved	6.1*	5.5, 6.3
Frustration with speech	5.9*	5.2, 6.1
They are faster	3.1	2.7, 3.8
Just for variety	2.0	1.3, 2.8
To rest my voice	1.3	1.1, 2.3

* Reasons with importance significantly greater than neutral rating of 4.0 ($p < 0.05$)

Table 2: Importance of reasons for choosing non-speech input methods instead of automatic speech recognition, rated on a scale from 1 to 7.

Having to train the system (setup involved) in order to improve recognition accuracy or donning a headset for dictating was initially perceived as a barrier for using ASR as the preferred input method. According to the survey, participants would also choose other input methods when ASR performed poorly or not at all, either in general or for dictating particular

commands (e.g., for some participants the command *Cut that* was consistently recognized as *Cap that*). Less important reasons were the need to rest one’s voice or to switch methods just for variety.

3.3 Opinions about speech and non-speech input methods

Participants rated their satisfaction with 10 usability indicators for both ASR and non-ASR alternatives (Tables 3 and 4).

Likes	% responding yes	
	ASR	Non-ASR
Ease	85.3	91.9
Speed	74.9	88.6
Less effort	73.9	75.3
Fun	62.3	23.6
Accuracy	52.7	85.3
Trendy	39.5	23.1

Table 3: Percentage of participants who liked particular aspects of the automatic speech recognition (ASR) system and non-speech input methods.

Dislikes	% responding yes	
	ASR	Non-ASR
Fixing recognition mistakes	74.5	–
Disturbs colleagues	45.9	–
Setup involved	36.8	–
Fatigue	17.3	12.7

Table 4: Percentage of participants who disliked particular aspects of the automatic speech recognition (ASR) system and non-speech input methods.

ASR for translator-computer interaction succeeds at easing the task (its most-liked benefit). Almost 75% liked the speed they achieved with ASR, despite being slower when compared against non-ASR input methods. Almost 74% liked the effort required to use ASR, and only 17.3% found it fatiguing. Participant’s largest complaint with ASR was related to recognition accuracy. Only 52.7% liked the recognition accuracy they achieved and fixing recognition mistakes ranked as the top dislike at 74.5%. The second most frequent dislike was potential work environment dissonance or loss of privacy during use of ASR at 45.9% of participants.

Ratings show significant differences between ASR and non-speech input methods, particularly with regard to accuracy and amusement involved (*Fun* item in the questionnaire).

3.4 Post-session questionnaire results

To further examine subjective opinions of ASR in post-editing compared to non-speech input methods, we asked participants to rate their agreement to several statements regarding learnability, ease of use, reliability and fun after performing the post-editing tasks under the two conditions. Agreement was rated on a scale of 1 to 7, from “strongly disagree” to “strongly agree”. Table 5 shows participants’ level of agreement with the seven statements in the post-session questionnaire.

Statement	Level of agreement	
	Mean	95% CI
1. I expected using ASR in post-editing to be more difficult than it actually is.	6.6*	6.5, 6.8
2. My performance with the selection of ASR commands improved by the end of the session.	6.5*	5.4, 6.9
3. The system correctly recognizes almost every command I dictate.	5.9*	5.5, 6.4
4. It is difficult to correct errors made by the ASR software.	2.9	2.3, 4.1
5. Using ASR in the context of post-editing can be a frustrating experience.	2.4	1.9, 3.8
6. I can enter text more accurately with ASR than with any other method.	2.1	1.7, 2.9
7. I was tired by the end of the session.	1.7	1.2, 2.9

* Agreement significantly greater than neutral rating of 4.0 ($p < 0.05$)

Table 5: Participants’ level of agreement to statements about ASR input method in post-editing tasks.

Ratings are on scale 1 to 7, from “strong disagree” to “strongly agree”, with 4.0 representing neutral rating.

The results of the post-session questionnaire show that participants had significantly greater than neutral agreement (positively) about ASR in the context of post-editing. Overall they agreed that it is easier to use ASR for post-editing purposes than they actually thought. They also positively agreed that the ASR software was able to recognize almost every command they dictated (i.e. *Select <w>*, *Scratch that*, etc.) and acknowledged that their performance when dictating commands was better as they became more familiar with the task.

When scores were combined for the seven statements into an overall satisfaction score, the average was 73.5 [66.3, 87.4], on a scale of 0 to

100³. Thus, this average is significantly more positive than neutral. 12 out of the 15 surveyed participants stated that they will definitely consider adopting ASR in combination with non-speech input modalities in their daily practice as professional translators.

4 Discussion

The results of the present study show that the surveyed post-editor trainees tended to report a very positive view on the use of ASR in the context of post-editing. In general, findings suggest that human translators would not regret the integration of ASR as one of the possible input methods for performing post-editing tasks.

While many questions regarding effective use of ASR remain, this study provides some basis for further efforts to better integrate ASR in the context of computer-aided translation. Some specific insights supported by the collected data are:

- Expectations about ASR were definitely more positive after having performed with speech as an input method. Participants positively agreed that it is easier and more effective than previously thought.
- Most of the challenges (dislikes) of ASR when compared to other non-input methods can be tackled if the user is provided with both ASR and non-ASR input methods for them to be used at their convenience. Participants' views seem to indicate that they would use ASR as a complement rather than a substitute for non-speech input methods.

5 Conclusions

Post-editor trainees have a positive view of ASR when combining traditional non-speech input methods (i.e. keyboard and mouse) with the use of speech. Acknowledging this up front, an interesting field for future work is to introduce proper training on correction strategies. Studies in this direction could help to investigate how training post-editors to apply optimal correction strategies can help them to increase performance and, consequently, user satisfaction.

Acknowledgments

We would like to thank all the participants in this study for their generous contributions of time, effort and insights.

References

- Dragsted, B., Mees, I. M., Gorm Hansen, I. 2011. Speaking your translation: students' first encounter with speech recognition technology, *Translation & Interpreting*, Vol 3(1).
- Dymetman, M., Brousseau, J., Foster, G., Isabelle, P., Normandin, Y., & Plamondon, P. 1994. Towards an automatic dictation system for translators: the TransTalk project. *Proceedings of the international conference on spoken language processing (ICSLP 94)*, 691–694.
- Koester, H.H. 2004. Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition. *Journal of Rehabilitation Research & Development*. Vol 41(5): 739-754.
- O'Brien, S. 2012. Translation as human-computer interaction. *Translation Spaces*, 1(1), 101-122.
- Toselli, A., Vidal, E., Casacuberta, F. 2011. *Multimodal Interactive Pattern Recognition and Applications*. Springer.
- Vidal, E., Casacuberta, F., Rodríguez, L., Civera, J., Martínez-Hinarejos, C.D. 2006. *Computer-Assisted Translation Using Speech Recognition*. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(3): 941-951.

³ A score of 100 represents a strong agreement with all positive statements and a strong disagreement with all negative statements, while a score of 50 represents a neutral response to all statements.