

Surveying the potential of using speech technologies for post-editing purposes in the context of international organizations: What do professional translators think?

Jeevanthi Liyanapathirana

World Trade Organization
Rue de Laussane 154
Geneva, Switzerland

jeevanthi.liyana@wto.org

Pierrette Bouillon

Fac. de traduction et d'interprétation
University of Geneva, 1211
Geneva, Switzerland

pierrette.bouillon@unige.ch

Bartolomé Mesa-Lao

Universitat Oberta de Catalunya
Av. Tibidabo, 39-43
08035 Barcelona, Spain

bmesa@uoc.edu

¹ Abstract

The present study has surveyed professional translators working in six international organizations in order to know more about their views and attitudes with regard to new translation workflows involving two different types of technologies, i.e. machine translation and speech recognition. The main aim of this survey was to identify how feasible it is to implement new post-editing workflows in an international organization using speech as an input method to edit inaccurate machine translation outputs. Overall, the results suggest that the surveyed translators do not hold a negative view on the use of ASR as part of their translation workflow, which provides a promising first step towards investigating the integration of speech based post-editing to translation workflows for productivity and ergonomic gains.

1 Introduction

Automatic speech recognition (ASR) software has quietly created a niche for itself in many situations of our daily lives (Joscelyne, 2018). 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 for smartphones. On another front, given the significant improvements in Machine Translation (MT) quality and the increasing demand for

translations, post-editing of MT has become a popular practice in the translation industry, since it has been shown to allow for larger volumes of translations to be produced saving time and costs. Workflows in the translation industry have experienced a significant transformation and it is in this new context that speech technology is likely to contribute to further innovation. With post-editing services becoming common practice among language service providers and speech recognition gaining momentum, it seems reasonable to start exploring interplays between both fields to create new business solutions and workflows (Mesa-Lao, 2014b). Most traditional international organizations with translation needs (see section 3.2) have already added a machine translation component as one of the resources offered to their human translators. However, not much has been said yet when it comes to the attitude of such professional translators using machine translation and, more specifically, using speech technology to edit machine translation outputs in the context of a large scale international organization.

In this study, we conducted qualitative research on the usage of speech and post-editing in a selected set of large scale international organizations. To our knowledge, this is the first study conducted on using post-editing and speech together in large scale international organizations. The paper is structured as follows: section 2 mentions related work for this study, followed by our method in section 3. Section 4 describes the results, leading to the discussion and conclusions in section 5.

¹ © This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CCBY-ND.

2 Related work

The use of speech as an input method to interact with computers and generate text is as old as the idea of computers themselves. 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.

In the context of translation, dictaphones were a popular tool in the context of large international organisations in the 1960s and 1970s and professional translators often collaborated with transcriptionists to dictate their translations. In the 1990s and 2000s, computational researchers began to explore ASR for translation purposes. Such developments focused mainly on reducing ASR word error rates by combining ASR and MT (Vidal et al., 2006). More recently, further efforts have been made by Translation Studies scholars in order to assess the performance of translation students and professionals when using commercial ASR systems (Dragsted et al., 2011; Zapata, 2012); to assess and analyze professional translators' needs and opinions about ASR (Ciobanu, 2014, 2016, and 2018), and to explore ASR in mobile and multimodal environments (Zapata, 2016a,b). More recently, the potential of using ASR for post-editing purposes has also been investigated (García-Martínez et al., 2014; Mesa-Lao, 2014a,b; Torres-Hostench et al., 2017, and Zapata et al., 2017). For example, it was shown in previous pilot experiments that post-editing with the aid of a speech recognition system was the fastest method for translation (Zapata et al., 2017), that voice input is more interesting than the keyboard alone for post-editing (García-Martínez et al., 2014) and that 12 out of 15 translators would welcome the integration of voice as one of the possible input modes for performing PE tasks (Mesa-Lao, 2014a,b).

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 in order 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:

- a) To understand the current situation of technology usage (specifically speech technologies) in selected international organizations with substantial translation needs.
- b) To analyze the potential of introducing speech technologies to post-edit MT within such organizations.

As a first step towards these two objectives, the following section describes our pilot survey in detail, our participants' profile and our methodology.

3 Method

In order to answer our research questions, we used two steps. As a first step, we investigated the current usage of translation technology solutions in a selected set of organizations. As a second step, we selected a set of professional translators from those organizations to gain further insights about their perceptions on using speech and other tools as part of their translation workflow.

3.1 Overview

As part of this research, two main questionnaires were designed and deployed as a survey. The first survey was distributed to a total of six organizations and contained six questions about current translation technology usage. This survey was filled up by technology managers in the respective organizations. The second survey consisted of 15 questions targeting professional translators working in the selected international organizations. Both surveys were carried out in March 2019.

3.2 Participants profile

This study involved five large scale international organizations based in Geneva and one large

scale international organization based in Luxembourg.

17 participants were selected from these six organizations using snowball sampling. The selected group included 11 females and 6 males, belonging to different age groups (3 translators between ages 20-35, 8 between 36-50, and 6 older than 50). All 17 participants are professional translators within these organizations, with multiple years of translation experience (7 translators with +20 years experience, 2 with 16-20 years, 3 with 11-15 years, 1 with 6-10 years, and 4 translators between 0-5 years experience). Their language combinations involved translating to/from English, French, Russian, and Spanish. The sample included translators working with different post-editing scenarios, i.e. post-editing via typing, translation from scratch using speech, translation from scratch using a keyboard, translation from scratch using a dictaphone, and post-editing via speech (Figure 1).

16 out of the 17 translators were familiar with standard computer-aided translation software (i.e. SDL Trados) and 14 out of the 17 translators claimed to be familiar with different categories of speech technologies which will be described in detail in the analysis section. In addition, 6 translators declared to use speech input methods in their day to day life (e.g. to dictate messages in a smartphone or to issue commands to Google Home, Amazon Alexa, etc.).

3.3 Procedure

In the first questionnaire, the managerial staff of each organization received a short questionnaire via email where they had to answer 6 simple questions on their current translation workflows. The second questionnaire was distributed internally by each international organization to their professional translators and contained 15 questions covering the following topics:

1. General information about their profile: including age, translation experience (years), employment status, and exposure to CAT tools.
2. Current translation workflow (translating from scratch, post-editing by typing,

post-editing by speech, and use of dictaphones).

3. Information about their usage of ASR as compared to other input methods (e.g. typing), and their likes and dislikes about it.
4. Their attitude towards different methods of translation, including speech based post-editing.

The first questionnaire consisted of open-ended questions. The second questionnaire was a mix of different types of questions: multiple choice questions, preference ranking questions, and dichotomous questions. These questionnaires can be found in the appendices (A and B).

3.4 Data collection and analysis

Regarding questionnaires' data, responses to quantitative items were entered into a spreadsheet, where mean responses were calculated. For binary or numeric results, the results were plotted in graphs to have a clearer overview. Open-ended questions and comments were analysed separately.

4 Survey Results

4.1 Distribution of translation technology among translators

As explained in section 3.3, the translators provided information on the translation technologies they involve in their translation processes. Figure 1 displays the results. The translators could select different technologies at the same time, since they could be competent in multiple translation workflows.

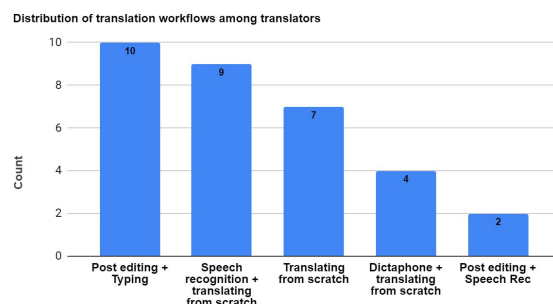


Figure 1. Translation technology usage among translators

10 out of 17 translators use typing to post-edit, 9 out of 17 already use speech recognition for translating (from scratch), and 4 out of 17 translators use a dictaphone for translation. Thus, the selected set of translators can be considered as representative of the variety of techniques used in international organizations. Also, these data show that usage of speech for the purpose of translating is not uncommon in the selected environments (at least 9 out of 17 translators are already competent in workflows involving ASR).

4.2 Usage of speech-based input methods

In order to determine why participants would decide to use ASR in the future to post-edit MT, we asked them to rate the importance of eight different reasons, on a scale from 1 to 8, being 8 the lowest in importance. The scale was 1 to 8 since there are eight reasons to be ranked by the translators - see Appendix B. The top reason for deciding to use ASR was that using speech was considered to be faster by the surveyed translators, followed by speech helping them with ergonomics. The mean value of the translator input score was neither negative nor positive with regard to the notion of speech technologies being accurate, providing a mean value of 4.0 (Table 1).

Reason	Mean
Using speech is less tiring for me	3.9
Using speech is faster for me	2.4
Using speech is easier for me	3.7
Speech is a cool technology	6
Not many other alternatives for me	7.1
Personal preference	5
Speech technologies are accurate	4
Speech helps me with ergonomics	2.6

Table 1. Ranking of reasons for using speech-based inputs in translation, rated on a scale from 1 (highest) to 8 (lowest).

4.3 Usage of non-speech input methods

Participants were also asked about their reasons for choosing non-speech input methods (i.e. keyboard and mouse). They rated the importance of six reasons on a scale from 1 (the most important) to 6 (the least important). The scale was 1 to 6 since there are six reasons to be ranked by the translators - see Appendix B. Table 2 describes the reasons why translators would not use speech input.

Reason	Mean
Not using speech is easier.	3.7
Speech requires a lot of training	4
Speech is frustrating	3.4
Speech is not faster	3.7
To rest my voice after speaking	3.5
Speech is trendy but not efficient	2.7

Table 2. Ranking of reasons for choosing non-speech input methods, rated on a scale from 1 (highest) to 6 (lowest).

The results were not very conclusive, but the main reason for their negative perception on speech technologies was their concern about its efficiency, which confirms the “neutral” attitude towards accuracy of speech recognition in Table 1.

The surveyed translators also provided open ended comments about negative views on using speech recognition. It was interesting to see how the biggest negative point of using speech recognition would be the noise factor (11 people out of 17 think speech recognition will disturb colleagues when working in an open space). This issue illustrates that using speech technologies in an organization would involuntarily depend on logistics factors. 9 out of 17 thought that using speech recognition can be tiring as well.

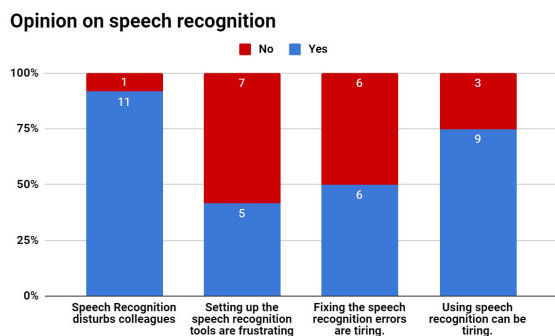


Figure 2. Translator views on using speech recognition for translation purposes.

4.4 Preferred choice of input method by translators based on requirement

The sampled translators were asked whether they would choose speech input or typing when considering the six reasons mentioned in the Figure 3 below.

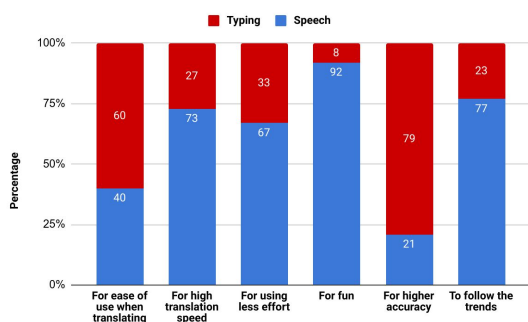


Figure 3. Choice of input method by the percentage of translators (speech/typing). Data labels illustrate percentages.

As can be seen from the results in Figure 3, a higher number of translators are open to the idea of using speech as a faster (73%) and less tiring (67%) input method when compared to typing. However, while the majority of the translators think that using speech as an input method is faster (73%), 79% do not believe that it is more accurate than typing, which agrees with our findings in Table 1 and 2.

4.5 Openness to different workflows

Since one of the main objectives of this study was to identify the potential of introducing speech input for post-editing purposes, translators were also asked about their openness to different workflows during translation. Figure 4 displays the results of our survey. 8 out of 17

translators were open to the idea of speech-based post-editing for translation and only 2 out of 17 assumed that mixing speech and post-editing together would be confusing.

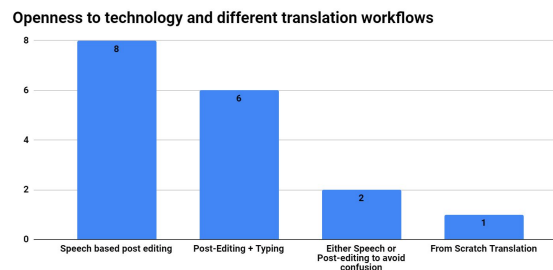


Figure 4. Translators' openness to different translation workflows.

As a second step, we further analyzed the translator's opinions in Figure 4, this time considering their current translation approaches as well. We analyzed the current translation workflow of each translator against the following three new workflows: 1) speech-based post-editing, 2) typing based post-editing, and 3) using either speech or typing post-editing but not together. Translators were divided into two categories based on their current skills: using dictaphone/any type of speech recognition tool and using typing for translation purposes. Figure 5 shows the translators' breakdown of openness to different workflows based on their translation workflow experience.

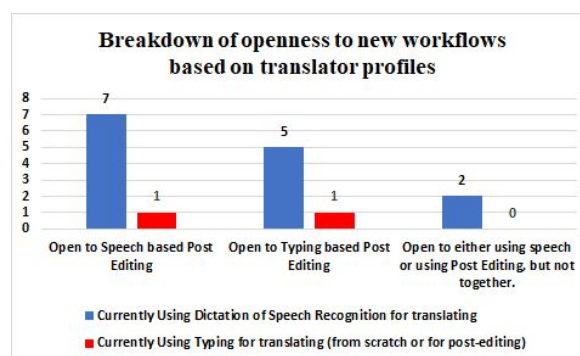


Figure 5. Breakdown of openness to new workflows based on translator profiles.

As expected, 7 out of 8 translators willing to use speech-based post-editing were already using either dictaphones or speech recognition tools,

which explains their positive attitude towards using speech along with MT post-editing.

4.6 Translator input on their choice of preferred method

As part of the second questionnaire, translators were asked to elaborate on when they would use translating from scratch, when they would prefer to post-edit via keyboard and when they would be interested in using speech as input to their translation workflow (results are displayed in Appendix D, E, and F). The main factor of choosing post-editing instead of typing from scratch was the availability of high quality machine translation or translation memory suggestions (11 out of 17 translators). Translators also mentioned that they would post-edit if “the translation does not require creativity” (2 translators), when “the translation has to be more accurate than fluent” (1 translator), and when “the translation has to be done quickly” (1 translator). Translators also mentioned that they would choose speech technologies to translate when “the text to be translated is long” (6 translators) and “when speech recognition quality is good with less errors” (4 translators). This feedback emphasizes the necessity of having not only high quality speech recognition, but also better machine translation/translation memory output, if we want to invite more translators to the idea of speech-based post-editing.

5 Discussion and Conclusion

Many other questions regarding the effective use of ASR in translation could be surveyed, but this preliminary study shows that our sample of professional translators do not hold a negative view on the use of ASR as part of their translation workflow. In general, our findings suggest that professional translators working in the context of an international organization can benefit from the integration of ASR as one of the possible input methods when translating from scratch or when editing text for post-editing purposes.

The main findings of this study are:

1. Speech as an input method (i.e. ASR or dictaphones) is mainly used by translators to translate from scratch, rather than to post-edit MT output.
2. The majority of the surveyed translators believed that speech is faster than typing and less tiresome (more ergonomic). However, they are still in doubt regarding the accuracy level of available speech recognition toolkits.
3. Along with the necessity of high-quality ASR software solutions, this survey exposed multiple other factors which make translators more inclined towards the use of speech-based post-editing. These factors include working with high-quality machine translation or translation memory suggestions, larger amounts of texts for translation, and the possibility to use private or protected workstations for translation purposes using ASR. Provided that these requirements are met, the majority of translators were open to try speech-based post-editing as a new translation workflow.

These observations thus provide a promising first step for us to continue towards a more ambitious study, where we will conduct quantitative research evaluating the productivity gains derived from speech-based post-editing. We also plan to investigate how currently available CAT tools with integrated speech support (e.g. Matecat, memoQ, and SDL Trados) can be used for this purpose.

We thus conclude this first survey on translators' perceptions on using ASR in large scale international organizations with positive results.

Acknowledgement

We would like to thank our anonymous participants from different international organizations for their valuable time and involvement in this preliminary survey.

References

- Ciobanu, D. (2014). Of Dragons and Speech Recognition Wizards and Apprentices. *Revista Tradumàtica*, 12: 524–538.
- Ciobanu, D. (2016). Automatic Speech Recognition in the Professional Translation Process. *Translation Spaces*, 5(1): 124–144.
- Ciobanu, D. (2018). Automatic Speech Recognition in the professional translation process. In Ehrensberger-Dow, M., Englund Dimitrova, B. (Eds). *Exploring the Situational Interface of Translation and Cognition* (123–143). John Benjamins Publishing Company.
- Dragsted, B., Mees, I. M., and Hansen, I. G. (2011). Speaking your translation: students' first encounter with speech recognition technology. *Translation & Interpreting*, 3(1): 10–43.
- García-Martínez, M., Singla, K., Tammewar, A., Mesa-Lao, B., Thakur, A., Anusuya, M. A., Bangalore, S., Carl, M. (2014). SEECAT: ASR & Eye-tracking Enabled Computer-Assisted Translation. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, pages 81–88, Dubrovnik, Croatia
- Joscelyne, A. (2018). Translators in the Algorithmic Age. Briefing based on the TAUS Industry Leaders Forum, Amsterdam, Netherlands.
- Matecat. (2019). *Advanced options – Matecat*. [online] Available at: <https://www.matecat.com/phrase-based-vs-neural-mt-webinar-questions/advanced-options/> [Accessed 17 Apr. 2019].
- Mesa-Lao, B. (2014a). Post-editing through Speech Recognition: A Feasibility Study with Post-editor Trainees. Abstract from 2014 *CRITT - Conference: Translation in Transition: Between Cognition, Computing and Technology* - Copenhagen Business School, Frederiksberg, Denmark.
- Mesa-Lao, B. (2014b). Speech-Enabled Computer-Aided Translation: A Satisfaction Survey with Post-Editor Trainees. In *Workshop on Humans and Computer-assisted Translation*, pages 99-103, Gothenburg, Sweden.
- SDL. (2018). *Increase your translation speed in 2018 with voice recognition software*. [online] Available at: <https://www.sdltrados.com/video/increase-your-translation-speed-in-2018-with-voice-recognition-software/128377/> [Accessed 17 Apr. 2019].
- Torres-Hostench, O., Moorkens, J., O'Brien, S., and Vreeke, J. (2017). Testing interaction with a Mobile MT post-editing app. *Translation & Interpreting*, 9(2):138-150.
- 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.
- Zapata, J. (2016a). Translating On the Go? Investigating the Potential of Multimodal Mobile Devices for Interactive Translation Dictation. *Revista Tradumàtica*, 14: 66–74.
- Zapata, J. (2016b). Translators in the Loop: Observing and Analyzing the Translator Experience with Multimodal Interfaces for Interactive Translation Dictation Environment Design. PhD thesis. University of Ottawa.
- Zapata, J., Castilho, Sh., and Moorkens, J. (2017). Translation dictation vs. post-editing with cloud-based voice recognition: a pilot experiment. In *Proceedings of MT Summit XVI 2*, Nagoya, 123–136.

Appendices

Appendix A. First questionnaire

- 1) **What type of CAT tools do you use in your organization? Please name them.**
- 2) **Do the employees of your organization use dictaphone to translate? If yes, please mention what the resources and toolkits are.**
- 3) **Do the employees of your organization use speech recognition toolkits to translate? If yes, please mention what the resources and toolkits are.**
- 4) **Do the employees of your organization use machine translation suggestions during translation? If yes, please mention what the resources and toolkits are.**
- 5) **Do the employees of your organization use machine translation suggestions to post-edit by typing? If yes, please explain.**
- 6) **Do the employees of your organization use speech recognition techniques to post-edit machine translation suggestions? If yes, please explain.**

Appendix B. Second questionnaire

Questions

- 1) **What is your age range?**
 a) 20-35 b) 35-50 c) 50 or more

- 2) **What type of translation experience do you have?**
 a) I work in an academic organization
 b) I work in an international organization
 c) I have experience in both (academic/international organizations)

- 3) **How long have you worked in the translation industry (experience can be academic or industrial)?**
 a) 0-5 years b) 6-10 years c) 11-15 years
 d) 16-20 years e) 20+ years

- 4) **Which are your language pairs during translation? e.g. English-French, etc.**

- 5) **Which of these statements is applicable to you? Multiple statements can be applicable, since you might be using different techniques for different requirements.**
 a) I use translation suggestions (MT/TM) for my translation purposes and I type to work on them.
 b) I use a dictaphone for translating from scratch.
 c) I use speech recognition toolkits (e.g. Dragon) to speak out my translations from scratch (and then correct them if necessary).
 d) I use speech recognition toolkits (e.g. Dragon) to post-edit translation suggestions (MT/TM).
 e) I prefer to type and translate from scratch.

- 6) **Do you use any computer-assisted translation tools (e.g. SDL Trados)? If so, which ones?**

- 7) **Have you used any speech recognition toolkit for other purposes, e.g. SIRI. Please explain briefly.**

- 8) **Following are some of the major reasons for using speech-based input methods (according to**

previous research). Could you please rank them according to the importance?

I would be the most important, and 8 would be the least important.

Reason	Rank
Using speech is less tiring for me.	
Using speech is faster for me.	
Using speech is easier for me.	
Speech is a cool technology.	
There are not many other alternatives for me.	
Using speech is a personal preference.	
Speech technologies are accurate.	
Speech helps me with ergonomy.	

9) Following are some of the reasons for not using speech-based input methods. Could you please rank them according to the importance?

I would be the most important, 6 would be the least important

Reason	Rank
Not using speech is much easier for me.	
Speech requires a lot of setup and training	
I get frustrated using speech.	
Using speech is not faster (at least for me).	
I don't use speech to rest my voice.	
Using speech is just trendy, but not efficient.	

10) Which technique would you use (speech recognition or typing) during the translation? Please use “yes” and “no” in each column.

Feature	Typing	Speech
For ease of use when translating		
For high translation speed		
For using less effort		
For fun		
For higher accuracy		
To follow trends		

11) Please type “yes” or “no” next to each of these statements according to your own personal views.

Feature	no/yes
Speech recognition disturbs colleagues	
Setting up speech recognition is frustrating	
Fixing speech recognition errors is tiring	
Using speech recognition can be tiring.	

12) Could you please mention reasons or situations you came across when you preferred post-editing translation suggestions rather than typing from scratch?

13) Could you please mention reasons or situations you came across when you preferred translating from scratch rather than post-editing machine translation suggestions?

14) As part of our research, we are investigating whether we can use speech technology for post-editing. In this hypothetical scenario, the users will get a machine translation suggestion or a translation memory suggestion for a given input. We would like to see if translators can use speech commands to post-edit the translation suggestion (the suggestions can come from translation memories or machine translation).

Could you please type "yes" next to the statement that is most applicable to you?

- a) Yes, I am open to the idea of speech-based post-editing.
- b) Yes, I would like to use speech for translation, but without having to work on translation suggestions coming from translation memories or machine translation. This setup would be confusing.
- c) Speech is not an option for me. I still enjoy translating from scratch via keyboard without having to work on machine translation outputs.
- d) Speech is not an option for me. I still enjoy translating from scratch via keyboard and I am happy to use machine translation outputs as a starting point.

15) Please mention situations where you would like to use speech technologies as a translation support (e.g. I would use it for long paragraphs, I would use it for short sentences, etc.).

Appendix C. Tool Usage in organizations

Category	Details
CAT tools used	Eluna, SDL Trados and Multitrans, DtSearch, MultiTerm, Groupshare, Euramis, memoQ, SmartLing.
MT tools	WipoTranslate, DeepL, eTranslate
Usage of dictaphone	2 organizations out of 5. One out of those two uses the dictaphone very rarely.
Speech recognition usage (e.g. Dragon)	4 organizations out of 5 use speech recognition.
Machine translation usage	4 organizations out of 5 use machine translation.
Post-editing using typing	4 organizations out of 5 use post-editing using typing.
Post-editing using speech	Only one translator of one organization could be found using post-editing using speech.

Appendix D. When would you choose post-editing machine translation instead of typing from scratch?

Reason	Frequency
When MT/TM quality is good	11/17
When accuracy is more important than fluency	1/17
When translation does not need creativity	2/17
To translate quickly	1/17

Appendix E. When would you choose typing from scratch instead of post-editing machine translation?

Reason	Frequency
When MT/TM quality is not good	13/17
When fluency is more important than accuracy	1/17
When text is short	2/17
When creativity is necessary	1/17

Appendix F: When would you choose speech technologies to translate?

Reason	Frequency
To translate long texts (paragraphs, articles)	6/17
I would use it anytime if the speech recognition quality is good	4/17
I would only use it to dictate long texts where post-editing is too much effort	1/17
To translate quickly	1/17