Translation Dictation vs. Post-editing with Cloud-based Voice Recognition: A Pilot Experiment

Julián Zapata julianz@intr.co
University of Ottawa & InTr Technologies, Ottawa-Gatineau, Canada
Sheila Castilho sheila.castilho@adaptcentre.ie
ADAPT Centre/School of Computing, Dublin City University, Dublin, Ireland
Joss Moorkens joss.moorkens@adaptcentre.ie
ADAPT Centre/School of Applied Language & Intercultural Studies, Dublin City University, Dublin, Ireland

Abstract

In this paper, we report on a pilot mixed-methods experiment investigating the effects on productivity and on the translator experience of integrating machine translation (MT) post-editing (PE) with voice recognition (VR) and translation dictation (TD). The experiment was performed with a sample of native Spanish participants. In the quantitative phase of the experiment, they performed four tasks under four different conditions, namely (1) conventional TD; (2) PE in dictation mode; (3) TD with VR; and (4) PE with VR (PEVR). In the follow-on qualitative phase, the participants filled out an online survey, providing details of their perceptions of the task and of PEVR in general. Our results suggest that PEVR may be a usable way to add MT to a translation workflow, with some caveats. When asked about their experience with the tasks, our participants preferred translation without the 'constraint' of MT, though the quantitative results show that PE tasks were generally more efficient. This paper provides a brief overview of past work exploring VR for from-scratch translation and PE purposes, describes our pilot experiment in detail, presents an overview and analysis of the data collected, and outlines avenues for future work.

1. Introduction

Machine translation (MT) post-editing (PE) and voice recognition (VR) technology are gaining ground in both translation technology research and the translation industry. Over 50% of international Language Service Providers now offer a PE service using dedicated MT engines integrated into translators' computer-aided translation environments (Lommel and DePalma, 2016). In a recent survey of 586 translators in the UK, 15% responded that they use VR technology in their work (Chartered Institute of Linguists et al., 2017). These disparate technologies tend not to be deployed in tandem, although both offer translators the potential to increase productivity and reduce the technical effort usually required to translate from scratch when using conventional word-processing hardware and software.

We carried out a pilot experiment to investigate the effects on productivity and on the translator experience (TX) (Zapata, 2016a) of integrating PE with VR and translation dictation (TD) using a sequential mixed-methods design. In the quantitative phase, four translators performed four translation tasks under four different conditions: (1) conventional TD (i.e., sight-translating using a digital dictaphone), (2) PE in dictation mode (PED) (i.e., dictating approved or amended segments into the same dictaphone), (3) TD with VR (TDVR) (using a cloud-based VR system on a tablet), and (4) PE with VR (PEVR) (using the same VR system as in task 3). The quantitative experiments consisted of three phases during which task times were measured and some input data were collected. Phase I consisted of dictating and post-editing with dictaphone or the VR system; phase II consisted of manually transcribing the recordings from tasks 1 and 2 on the researcher's laptop; and phase III consisted of revising/editing all four translations. As has been noted in a great deal of research about PE, productivity increases alone do not make a tool desirable for translators (see Teixeira, 2014; Moorkens and O'Brien, 2017). Translator attitudes and usability, the TX, are important factors in the adoption of any technology. For this reason, we have appended a follow-on qualitative phase, wherein the participants filled out an online survey, providing details of their perceptions of the task and of PEVR in general.

In this paper, we present our pilot experiment in detail. The paper is structured as follows: First, we provide a brief overview of past work exploring VR for from-scratch translation and PE purposes. Then, we describe the experimental setup, and present an overview and analysis of the quantitative and qualitative results. In the conclusion, we describe avenues for future work.

2. Related Work

2.1. TD and VR

The idea of using human voice to interact with computers and process texts is as old as the idea of computers themselves. For decades, and in recent years more than ever before, voice input has been widely used in a vast array of domains and applications, from virtual assistants on mobile phones to automated telephone customer services; from professional translation to legal and clinical documentation.

Simply put, VR (also known as voice/speech-to-text or automatic speech recognition) technology recognizes human-voice signals and converts them into digital data. The earliest experiments in VR suggested that voice input was expected to replace other input modes such as the keyboard and the mouse in full natural language communication tasks. However, it was soon discovered that speech often performed better in combination with other input modes such as the keyboard itself, as well as touch, stylus and gesture input on multimodal interfaces (Bolt, 1980; Pausch and Leatherby, 1991; Oviatt 2012).

In translation, there has been a long interest in speaking translations instead of typing them. First, in the 1960s and 1970s professional translators often collaborated with transcriptionists, and dictated their translations either directly to the transcriptionist or into a voice recorder (or dictaphone), before having them transcribed later (a technique often referred to as TD). In the 1990s and 2000s, researchers began to explore VR adaptation for TD purposes. Such developments focused mainly on reducing VR word error rates by combining VR and MT. Hybrid VR/MT systems are presented with the source text and use MT probabilistic models to improve recognition; translators simply dictate their translation from scratch without being presented with the MT output (Brousseau et al., 1995; Désilets et al., 2008; Dymetman et al., 1994; Reddy and Rose, 2010; Rodriguez et al., 2012; Vidal et al., 2006). More recently, further efforts have been made to evaluate the performance of translation students and professionals when using commercial VR systems for straight TD (Dragsted et al., 2009; Dragsted et al., 2011; Mees et al., 2013); to assess and analyze professional translators' needs and opinions about VR technology (Ciobanu, 2014 and 2016; Zapata, 2012), and to explore TD in mobile and multimodal environments (Zapata and Kirkedal, 2015; Zapata, 2016a,b).

2.2. PE and VR

In recent years, the potential of using VR for PE purposes has also been investigated (García-Martínez et al., 2014; Mesa-Lao, 2014; Torres-Hostench et al., 2017). García-Martínez and her collaborators (2014) tested a VR system integrated into a PE environment (both research-level cloud-based systems). They argue that voice input is more interesting than the keyboard alone in a PE environment, not only because some segments may need major changes and therefore could be dictated, but also because, if the post-editor is not a touch typist, the visual attention back and forth between source text, MT text and keyboard adds to the complexity of the PE task.

Mesa-Lao (2014) surveyed student translators, 80% of which (n=15) reported that they would welcome the integration of voice as one of the possible input modes for performing PE tasks. Thus, voice input offers a third dimension to the PE task, making it possible to combine different input modes or to alternate between them according to the difficulty of the task and to the changing conditions of human-computer interaction. Some experiments have also suggested specifically that for certain translators, text types and language combinations, the benefits of VR and PE integration may not be the same (e.g. in terms of efficiency, productivity and cognitive effort) (see Carl et al. 2016a and 2016b).

Tests with VR within a mobile PE app were reported, first by Moorkens et al. (2016), then by Torres-Hostench et al. (2017). Participants were impressed by VR quality and found it useful for long segments. However, they mostly preferred to use the keyboard due to limitations of the software for making minor edits to MT output.

In the following section, we describe our pilot experiment more in detail: our participants' profile and our methodology.

3. Experimental Setup

3.1. Participants' Profile

This experiment included a sample of native (Latin American) Spanish speakers. All four participants are either pursuing or have recently completed a doctoral degree in translation studies. Participants had in common at least a minimum level of acquaintance with the notions of MT, PE and VR. Our sample includes two men and two women between the ages of 26 and 43. Participants reported 3 to 12 years of translation experience, two have training in interpreting, and both of those are regular users of VR (and were therefore familiar with voice commands and other specificities related to dictating with VR). All participants reported to be occasional post-editors.

3.2. Methodology

For this study, we applied a sequential, explanatory mixed-methods design, using the follow-up explanations model, in which the qualitative data is intended to expand upon the quantitative results (Creswell and Plano Clark, 2007:72). We chose this methodology to answer the following two research questions:

1. Can PEVR be as or more productive than comparable approaches, with or without MT and VR?

2. Does the participants' TX suggest that combining MT and VR is feasible for translation projects?

As mentioned in the introduction, four tasks were involved in the quantitative phase of this experiment, namely:

1) Conventional TD;
2) PED;
3) TDVR; and
4) PEVR.

A digital dictaphone was used for tasks 1 and 2. A commercial cloud-based speaker-independent VR system[1] was used on an Android tablet for tasks 3 and 4. (See Zapata and Kirkedal (2015) for a description of the different approaches to VR technology with respect to users (i.e. speaker-dependent, speaker-adapted and speaker-independent systems)).

Source texts were 20-segment sections of newstest 2013 data used in WMT[2] translation tasks. The test sets were analysed using the Wordsmith Wordlist[3] tool to ensure that they were statistically similar, based on measurements for type/text ratio, average sentence length, and average word length. Table 1 shows the statistics of the test set.

Text file	**Type/token ratio (TTR)**	**Mean word length (in characters)**	**Word length std.dev.**	**Sentences**	**Mean (in words)**
Test Set 1	55.12	4.99	2.51	20	18.05
Test Set 2	55.73	4.80	2.63	20	19.65
Test Set 3	54.31	5.00	2.62	22	21.09
Test Set 4	54.20	5.18	2.69	20	17.25

Table 1. Test set statistics for source texts

A commercial-level MT system[4] was used to translate the texts. All texts were printed out separately and presented to the participants in hard copy. Naturally, only in tasks 2 and 4 were participants presented with the segmented source and MT texts. The MT texts for tasks 1 and 3 were used only to calculate HTER scores (Snover et al., 2006); more details are provided in section 4.1.2.

Experiments were run individually (i.e. one participant at a time) over four days. A university study room was booked to perform the experiments.

Tasks were randomized as follows:

[1] Dragon Dictation, integrated in the Swype+Dragon app. See http://www.swype.com/.
[2] http://www.statmt.org/wmt13/
[3] http://lexically.net/wordsmith/
[4] Google Translate. See https://translate.google.com/.

Participant	Order of tasks			
ES1	1	2	3	4
ES2	3	4	1	2
ES3	4	3	2	1
ES4	2	1	4	3

Table 2. Participants and order of tasks

Before performing any of the experimental tasks, participants were briefly instructed how to use the digital dictaphone (for tasks 1 and 2) and the VR system on the tablet (for tasks 3 and 4) (i.e., they were given the opportunity to dictate while testing a few voice commands such as punctuation marks, etc.).

The quantitative experiments consisted of three phases during which task times were measured and some input data were collected:

- Phase I - dictating and post-editing with dictaphone or the VR system on the tablet,
- Phase II - manually transcribing the recordings from tasks 1 and 2 (for TD and PED) on the researcher's laptop; and
- Phase III - revising/editing all four translations on the researcher's laptop.

It is important to highlight that during phase II, participants were instructed not to edit the translation, only transcribe what they heard. The documents in which dictations were performed on the tablet for tasks 3 and 4 in phase I were automatically saved into a cloud-based drive[5] after dictation, and therefore immediately synchronized and available to be edited/revised on the researcher's laptop in phase III.

In phase I, task times were measured using a stopwatch. In both phases II and III, Inputlog (Leijten and Van Waes, 2013) was used. Inputlog is a research-level program designed to log, analyse and visualize writing processes. The program provides data such as total time spent in the document, total time in active writing mode (i.e., of actual keystrokes), total time spent moving/clicking with the mouse, total number of characters typed, total switches between the keyboard and the mouse, etc. Beyond total task times alone, we were interested in collecting this kind of detailed input data, particularly for phase III. We are not reporting data other than task times here given the scope and limitations of this paper; we do consider, however, that input data analysis will be essential in larger-scale experiments.

Thereafter, in the qualitative phase, participants responded to a short online questionnaire, with socio-demographic questions, retrospective questions about the experiment, as well as questions providing insight on the TX with multimodal/mobile VR-enabled TD and PE applications (more details to be provided in section 4.4).

In the following section, some of the data collected is presented and analysed.

4. Results and Analysis

4.1. Task Times Measures (Quantitative Phase)

In order to investigate the effects on productivity of integrating PE with VR and TD in the quantitative phase of this research, we have conducted analysis of the task times as follows:

[5] Dropbox. See https://www.dropbox.com.

1. Comparing tasks of the same nature with and without VR, that is, a) TD vs. TDVR (see 4.1), and b) PED vs. PEVR (see 4.2)
2. Comparing translation vs. PE within phases, that is: a) TD vs PED (4.3) and b) TDVR vs. PEVR (4.4).

We consider:

a) Translation and/or PE time (phase I + phase II), that is, the time participants needed to translate and/or post-edit, as well as the transcription time (for TD and PED);
b) Revision duration (phase III), that is, the total time participants needed to review/edit their translation/post-editing;
c) Total task time (phase I + phase II+ phase III), that is, the total time the participants needed to perform each task.

TD *versus* TDVR

When comparing both TD tasks (Table 3), i.e. the one performed with a dictaphone (TD) and the one performed with a VR program (TDVR), we can see that the total translation time is always shorter when participants use VR. A reminder to the reader that the total translation time in the dictaphone task includes the time participants need to transcribe their translations (phase II).

Regarding revision duration, however, tasks performed with VR seem to take longer to be completed. We speculate that this is because during the revision time, participants do not only review their translation but also must correct errors produced by the VR program.

Participants	**Task**	**Translation Time**			**Revision Time**	**Total Task Time**
		Translation time	Transcription time	Total		
ES1	TD	537	716	1253	402	1655
	TDVR	796	n/a	**796**	656	**1452**
ES2	TD	688	1197	1885	405	**2290**
	TDVR	1330	n/a	**1330**	1191	2521
ES3	TD	846	1116	1962	227	2189
	TDVR	377	n/a	**377**	722	**1099**
ES4	TD	700	1432	2132	454	2586
	TDVR	460	n/a	**460**	1046	**1506**

Table 3. TD vs TDVR (in seconds)

Overall, when considering all phases, total task time seems to be lower for TDVR, apart from participant ES2, who shows lower time when performing TD.

PED *versus* PEVR

Results for both PE tasks (PED and PEVR) were also compared (table 4). We notice that the PE time (total) is lower for all participants in the VR condition. As for revision, the time is higher in PEVR, which we assume is for the same reason described in above: that participants also need to correct errors produced by the VR application. However, when considering all phases, participants were still faster post-editing with VR than with the dictaphone.

To compare how much PE was performed for each task, we have calculated the translation edit rate (HTER) (Snover et al. 2016). The HTER score is a measure that compares the raw MT output and the post-edited version, and goes from 0 to 1, where the higher number, the more modifications were made in the raw MT output. We can see in table 4 that most of the participants have an average score of 0.2 – which indicates that little post-editing was performed. However, participant ES3 displays more post-editing performed for the PED task (0.52).

Participants	**Task**	**PE Time**			**Revision Time**	**Total Task Time**	**HTER**
		PE time	Transcription time	Total			
ES1	PED	633	692	1325	238	1563	0.24
	PEVR	623	n/a	**623**	776	**1399**	0.23
ES2	PED	822	604	1426	537	1963	0.24
	PEVR	910	n/a	**910**	606	**1516**	0.17
ES3	PED	612	1366	1978	270	2248	**0.52**
	PEVR	344	n/a	**344**	475	**819**	0.25
ES4	PED	396	1725	2121	654	2775	0.26
	PEVR	1176	n/a	**1176**	1007	**2183**	0.14

Table 4. PED vs PEVR (times are in seconds)

TD *versus* PED

As mentioned above, we also decided to consider the differences between translation and PE when both were performed in the same manner; that is TD and PED; and TDVR and PEVR.

Table 5 compares the results for TD and PED. When looking at the results for translation and PE translation time (total task time; last column), we notice that the results are mixed: while participants ES1 and ES2 were faster with TD, the other two participants (ES3 and ES4) were faster with PED. Interestingly, the transcription time is inversely higher, that is, participants ES1 and ES2 had higher transcription time for the TD tasks, whereas ES3 and ES4 had higher transcription time in PED. Now, when considering the total translation/PE time, we can see that the results are very close, the more visible differences lying for ES1 and ES2, where the former is faster with TD and the latter with PED.

In sum, when looking at the different time measures across phases, we notice no trend in the results. This indicates that, in general, there were not many differences between TD and PED.

Participants	Task	Translation/PE Time			Revision Time	Total Task Time
		Translation/ PE time	Transcription time	Total		
ES1	TD	**537**	716	**1253**	402	1655
	PED	633	**692**	1325	**238**	**1563**
ES2	TD	**688**	1197	1885	**405**	2290
	PED	822	**604**	**1426**	537	**1963**
ES3	TD	846	**1116**	**1962**	**227**	**2189**
	PED	**612**	1366	**1978**	270	2248
ES4	TD	700	**1432**	**2132**	**454**	**2586**
	PED	**396**	1725	**2121**	654	2775

Table 5. TD vs PED (in seconds)

Table 6 compares the results for TDVR and PEVR. We can see that total task times are lower for the first three participants when post-editing with VR than translating from scratch. Only participant ES4 was faster in the translation task. Interestingly, participant ES4 displayed close times for revision, whereas participant ES1 showed lower times to revise the translation. In sum, only participant ES4 showed higher times when post-editing than when translating from scratch, which suggests that PE with the help of VR could generally lead to higher productivity.

Participants	Task	Translation/PE Time	Revision Time	Total Task Time
ES1	TDVR	796	**656**	1452
	PEVR	**623**	776	**1399**
ES2	TDVR	1330	1191	2521
	PEVR	**910**	**606**	**1516**
ES3	TDVR	377	722	1099
	PEVR	**344**	**475**	**819**
ES4	TDVR	**460**	**1046**	**1506**
	PEVR	1176	**1007**	2183

Table 6. TDVR vs PEVR (in seconds)

4.2. TX Analysis (Qualitative Phase)

In the follow-on, qualitative phase of this experiment, participants responded to an online questionnaire with sociodemographic questions (see *Participant's profile* in section 3.1 above) and retrospective questions about the experiment, as well as questions providing insight on the TX with multimodal/mobile VR-enabled TD and PE applications. The notion of TX is inspired from the notion of user experience (UX) – extensively investigated in the field of human-computer interaction – and is defined as "a translator's perceptions of and responses to the use or anticipated use of a product, system or service" (Zapata, 2016a).

In this section, we report on the results of our questionnaire.

Subjectively Experienced Productivity

The questionnaire included an item to ask participants to indicate which one of the four translation tasks they *felt* made them most productive, and which one made them least productive. Three participants believed that TDVR made them most productive when in fact they had performed the PEVR task faster. Two participants felt that they were slowest in the PED condition. This perception of slower pace when MT has been introduced, contradicting quantitative measurements that recorded increased speed, has been seen elsewhere by Plitt and Masselot (2010) and Gaspari et al. (2014). When compared to their actual productivity times, we note that apart from ES1 regarding TD (where he/she is least productive), the other participants perceive it differently from the actual numbers. Table 7 below shows the perceived productivity against the actual productivity, where l/L = least, m/M = most, lower-case letters are for the perceived productivity and capital letters for the actual productivity.

Participant	TD	PED	TDVR	PEVR
ES1	l/L		m	M
ES2		l	m/L	M
ES3		l	m/L	M
ES4	m	L	l/M	

Table 7. Subjectively experienced productivity against actual productivity

Subjectively Perceived Quality

The questionnaire also included an item to ask participants to indicate which one of the four translation tasks they *felt* would result in the best quality, and which one would result in the worst quality (that is, quality of the final target text). Table 8 shows that two of the four participants were confident enough in the PEVR process, that they expected the output texts from that process to be of high quality.

Participant	TD	PED	TDVR	PEVR
ES1	worst			best
ES2			worst	best
ES3		worst	best	
ES4	best			worst

Table 8. Subjectively perceived quality

Challenges for VR-enabled TD and PE

A further question asked participants to elaborate on what they thought are the challenges of VR, on the one hand, and of MT, on the other hand, to provide translators with a useful VR-enabled TD and PE tool.

Participants found VR to be reasonably accurate, but with room for improvement, particularly regarding “proper names and figures”. Participants preferred translation without the ‘constraint’ of MT as they considered the suggestions artificial. Participant ES2 wrote that “the Spanish translation sounded more like a transliteration of a technical text in English, and this is not translation as far as I understand”. The added cognitive load when MT is added to source and target texts may be initially off-putting for translators, and may add to the perception of decreased speed when MT is introduced to the workflow. They recognized that VR and MT could aid productivity, but would prefer to add MT electively. Participant ES1 wrote that “a translator or post-editor should have the option to translate from scratch by default, and request the help from the machine only when needed”. Participant ES2 agreed: “For quality purposes, I prefer the [VR] translation from scratch or post-editing from [translation memories] where you have more leeway.” In the opinion of participant ES4, “MT makes work faster but not necessarily better. It somehow guides the work towards the paradigmatic level. I think the overall cohesion of the document is affected.”

Advantages and Disadvantages of Mobile *versus* PC-based TD and PE

Finally, participants were asked to elaborate on the perceived advantages and disadvantages of using a mobile TD and PE tool (i.e., on a mobile device such as a smartphone or a tablet) *versus* a laptop- or PC-based tool. Several mentioned the flexibility of a mobile device, and participant ES2 suggested that “it may help translators to develop interpreting strategies; such as segmentation, quick thinking, anticipation, short-term memory, etc.” Two participants mentioned the difficulties of working in a noisy environment and of speaking translations in a public place. Participant ES3 felt that, although PEVR felt fast to him/her, it was difficult to edit retrospectively. He/she added that if there was “a way to make it more seamless between the keyboard and the mic, a balance so to say, then that'd be amazing.”

5. Conclusion and Future Work

We have reported a pilot experiment on the use of a cloud-based voice recognition (VR) application for translation dictation (TD) and post-editing (PE), using both quantitative and qualitative methods.

In answer to our first research question, based on this small-scale pilot experiment, PE with VR can be as or more productive than comparable approaches, with or without machine translation (MT) and VR. When looking at quantitative data alone, our results showed that, in general, PE with the aid of a VR system was the most efficient method, being the fastest for three of the participants. Interestingly, PE in dictation mode (PED) was the slowest for two participants, followed by TD and TD with VR (TDVR). In the quantitative data, however, we observe that most participants perceived productivity to be higher in the TDVR condition, and expressed a preference to translate/dictate from scratch and have PE added as an option.

One of the issues we identified in our experiment is high revision/editing times in the VR tasks; transcriptions by the VR system were far from flawless, leading to higher revision/editing times. VR applications may produce errors due to translators’ lack of familiarity with TD and insufficient training in how to speak to a VR system, especially for properly adding punctuation using the appropriate commands. Trainers and researchers in translation have explicitly affirmed that training in sight translation, TD, and VR will be essential to succeed with (mobile) voice-enabled tools and devices (Mees et al. 2013; Zapata and Quirion, 2016). We noted also that some foreign-language words (e.g. Russian names) in the source texts caused a few misrecognitions in Spanish VR. Moreover, we noticed that some participants would often wait until the software had transcribed a sentence or chunk of a sentence onto the word processor page to continue speaking, which tends to confuse the system (as opposed to when the dictation is continuous). Lastly, if the user pauses for several seconds, the VR system "stops listening" and disconnects, which also causes both the system and the user to lose the flow of the dictation.

Another point to highlight is that the participants' typing skills may considerably affect translation times. If our time task measures excluded the transcription time in TD and PED, the whole productivity picture would change. Considering this and the issues described in the previous paragraph, the ideal scenario would be one in which translators do not need to transcribe their dictation, either in TD or PE. Instead, they would have a VR system with human-like transcription capabilities, keeping dictation, transcription, and editing/revision times (as well as recognition errors) to a minimum.

In answer to our second research question, participants' TX suggests that combining MT and VR is indeed feasible for translation projects, with some caveats. When asked about their experience with the tasks, our participants seem to have preferred translation without the 'constraint' of MT as they considered the suggestions artificial, though the quantitative results show that the PE task was more efficient than that of translation from scratch. The results of this small-scale experiment suggest that PE with VR (PEVR) may be a usable way to add MT to a translation workflow, and is worth testing at a larger scale.

For future work, we intend to carry out experiments with more participants and language pairs. Further experimentation will include input logging, as well as eye-tracking technologies to collect empirical data on cognitive effort when using VR for TD and PE. We also seek to evaluate the impact of training translators in TD and VR over a period of time before performing TDVR and PEVR tasks. Also, we will include objective measures of quality (with the participation of expert evaluators) to compare it with the participants' perceived quality of the target texts. Another avenue for future work is to investigate a collaborative scenario in which translators/post-editors collaborate with transcriptionists and/or revisers who would take part in the different phases of the experiment. This list of ideas for future work is of course non-exhaustive; the possibilities seem endless.

The unprecedented robustness of VR technology and its availability on mobile devices via the cloud opens a world of possibilities for human-aided MT and human translation environments. By keeping human translators at the core of research, with strong consideration of their perceptions and preferences for new technologies and applications, we can advance towards finding the right balance in translator-computer interaction (O'Brien, 2012), towards establishing what it is that the machine can do better than humans, and what it is that humans can do better than the machine.

Acknowledgement

We would like to thank our anonymous participants for their time and involvement in this pilot experiment. This work was supported by the ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and co-funded under the European Regional Development Fund.

References

Bolt, R. A. (1980). "Put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the SIGGRAPH'80*, pages 262–270. ACM Press.

Brousseau, J., Drouin, C., Foster, G., Isabelle, P., Kuhn, R., Normandin, Y., & Plamondon, P. (1995). French speech recognition in an automatic dictation system for translators: The TransTalk project. In *Proceedings of Eurospeech'95*, pages 193-196, Madrid, Spain.

Carl, M., Aizawa, A., & Yamada, M. (2016a). English-to-Japanese Translation vs. Dictation vs. Post-editing: Comparing Translation Modes in a Multilingual Setting. In *The LREC 2016 Proceedings: Tenth International Conference on Language Resources and Evaluation*, pages 4024–4031, Portorož, Slovenia.

Carl, M., Lacruz, I., Yamada, M., & Aizawa, A. (2016b). Comparing spoken and written translation with post-editing in the ENJA15 English to Japanese Translation Corpus. In *The 22nd Annual Meeting of the Association for Natural Language Processing (NLP2016)*, Sendai, Japan.

Chartered Institute of Linguists, European Commission Representation in the UK, and the Institute of Translation and Interpreting. (2017). *UK Translator Survey: Final Report*. Technical Report. Chartered Institute of Linguists (CIOL), London, UK.

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.

Désilets, A., Stojanovic, M., Lapointe, J.-F., Rose, R., and Reddy, A. (2008). Evaluating Productivity Gains of Hybrid ASR-MT Systems for Translation Dictation. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 158-165, Waikiki, USA.

Dragsted, B., Hansen, I. G., and Selsøe Sørensen, H. (2009). Experts Exposed. *Copenhagen Studies in Language*, *38*: 293–317.

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.

Dymetman, M., Brousseau, J., Foster, G., Isabelle, P., Normandin, Y., and Plamondon, P. (1994). Towards an Automatic Dictation System for Translators: the TransTalk Project. In *Fourth European Conference on Speech Communication and Technology*, page 4, Yokohama, Japan.

Garcia-Martinez, 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.

Gaspari, F., Toral, A., Kumar Naskar, S., Groves, D., Way, A. (2014). Perception vs Reality: Measuring Machine Translation Post-Editing Productivity. In *Proceedings of AMTA 2014 Workshop on Post-editing Technology and Practice*, pages 60-72, Vancouver, Canada.

Leijten, M., and Van Waes, L. (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication*, *30*(3): 358–392.

Lommel, A. R and DePalma, D. A. (2016). *Europe's Leading Role in Machine Translation: How Europe Is Driving the Shift to MT.* Technical Report. Common Sense Advisory, Boston, USA.

Mees, I. M., Dragsted, B., Hansen, I. G., and Jakobsen, A. L. (2013). Sound effects in translation. *Target, 25*(1): 140–154.

Mesa-Lao, B. (2014). 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

Moorkens, J., and O'Brien, S. (2017). Assessing user interface needs of post-editors of machine translation. In *Human Issues in Translation Technology: The IATIS Yearbook*, pages 109-130. Taylor & Francis.

Moorkens, J., O'Brien, S., and Vreeke, J. (2016). Developing and testing Kanjingo: a mobile app for post-editing. *Tradumàtica, 14*: 58-65.

O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces*, 1(1): 101–122.

Oviatt, S. (2012). Multimodal Interfaces. In J. A. Jacko (Ed.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (3rd ed., pages 415-429). Lawrence Erlbaum Associates.

Pausch, R., and Leatherby, J. H. (1991). An Empirical Study: Adding Voice Input to a Graphical Editor. *Journal of the American Voice Input/Output Society 9*(2): 55-66.

Plitt, M., Masselot, F. (2010). A productivity test of statistical machine translation post-editing in a typical localization context. *Prague Bulletin of Mathematical Linguistics* 93: 7-16.

Reddy, A., and Rose, R. C. (2010). Integration of Statistical Models for Dictation of Document Translations in a Machine Aided Human Translation Task. *IEEE Transactions on Audio, Speech and Language Processing*, *18*(8): 1-11.

Rodriguez, L., Reddy, A., and Rose, R. (2012). Efficient Integration of Translation and Speech Models in Dictation Based Machine Aided Human Translation. In *Proceedings of the IEEE 2012 International Conference on Acoustics, Speech, and Signal Processing*, 2: 4949-4952.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of Association for Machine Translation in the Americas*, pages 223-231, Cambridge, USA.

Teixeira, C. S. C. (2014). Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. In *Proceedings of AMTA 2014 Workshop on Post-editing Technology and Practice,* pages 45-59, Vancouver, Canada.

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., and 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. (2012). *Traduction dictée interactive : intégrer la reconnaissance vocale à l'enseignement et à la pratique de la traduction professionnelle*. M.A. thesis. University of Ottawa.

Zapata, J. (2016a). Translating On the Go? Investigating the Potential of Multimodal Mobile Devices for Interactive Translation Dictation. *Tradumatica*, 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., and Kirkedal, A. S. (2015). Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows. In *Proceedings of the 20th Nordic Conference of Computational Linguistics*, pages 201-210, Vilnius, Lithuania.

Zapata, J., and Quirion, J. (2016). La traduction dictée interactive et sa nécessaire intégration à la formation des traducteurs. *Babel*, 62(4): 531-551.