

Article

# Speech Synthesis in the Translation Revision Process: Evidence from Error Analysis, Questionnaire, and Eye-Tracking

Dragoş Ciobanu <sup>1,\*</sup> , Valentina Ragni <sup>2</sup> and Alina Secară <sup>1</sup>

<sup>1</sup> Centre for Translation Studies, University of Leeds, Leeds, LS2 9JT, UK; a.secara@leeds.ac.uk

<sup>2</sup> School of Modern Languages, University of Bristol, Bristol, BS8 1TE, UK; v.ragni@bristol.ac.uk

\* Correspondence: d.i.ciobanu@leeds.ac.uk

Received: 1 July 2019; Accepted: 5 November 2019; Published: 11 November 2019



**Abstract:** *Translation revision* is a relevant topic for translator training and research. Recent technological developments justify increased focus on embedding speech technologies—speech synthesis (text-to-speech) and speech recognition (speech-to-text)—into revision workflows. Despite some integration of speech recognition into computer-assisted translation (CAT)/translation environment tools (TEnT)/Revision tools, to date we are unaware of any CAT/TEnT/Revision tool that includes speech synthesis. This paper addresses this issue by presenting initial results of a case study with 11 participants exploring if and how the presence of sound, specifically in the source text (ST), affects revisers’ revision quality, preference and viewing behaviour. Our findings suggest an improvement in revision quality, especially regarding Accuracy errors, when sound was present. The majority of participants preferred listening to the ST while revising, but their self-reported gains on concentration and productivity were not conclusive. For viewing behaviour, a subset of eye-tracking data shows that participants focused more on the target text (TT) than the source regardless of the revising condition, though with differences in fixation counts, dwell time and mean fixation duration (MDF). Orientation and finalisation phases were also identified. Finally, speech synthesis appears to increase perceived alertness, and may prompt revisers to consult external resources more frequently.

**Keywords:** translation revision; speech synthesis; computer-assisted translation (CAT); eye-tracking; error detection; error analysis

## 1. Introduction: Literature Review

*Translation revision* is “an emerging topic in the translation industry, in translator training and in translation research” [1] due to its strategic role in the private and public sector quality assurance processes—including international organisations (IOs). This is a direct consequence of the significant increase in translation outsourcing [2]. A decade ago the term *revision* was rather more ambiguous, appearing in “*self-revision*”, “*other-revision*”, “*unilingual revision*” and “*post-editing (i.e., revision)*” [3]. In the meantime, standardisation efforts such as ISO 17100 [4] (and BS 15038 before it) have sought to separate these tasks terminologically, too, assigning them the terms *check*, *revision*, *review*, *proofread* and *post-edit*, respectively.

Moreover, following the effort to disambiguate these related, but essentially separate tasks, as with any trainable activity, it is important to clarify several dimensions: *what* each task consists of, *who* should be performing it, *when* it happens in the translation or localisation workflow, *how* it is performed and *which success criteria* there are. This article aims to contribute to the *how* dimension of *revision*, but we will briefly introduce the other dimensions, too, in order to set the context.

In ISO 17100, the *what* dimension of *revision* is defined as “bilingual examination of *target language content against source language content* for its suitability for the agreed purpose” [4] (p. 2). The standard also specifies that this source-target comparison should include several aspects, which the translator should have addressed, from “compliance with specific domain and client terminology” through to ensuring that the “target audience and purpose of the target language content” have been considered [4] (p. 10). Thus, *revision* is distinct from *review* (“monolingual examination of *target language content* for its suitability for the agreed purpose”), *post-edit* (“edit and correct *machine translation output*”), *check* (“examination of *target language content* carried out by the translator”) and *proofread* (“examine the revised *target language content* and applying *corrections* before printing”) [4] (pp. 2–3).

In terms of who can be a reviser, the ISO 17100 standard specifies that it should be “a person other than the translator”, who has the same competencies and qualifications as the ones listed by the standard for the translator, together with “translation and/or revision experience in the domain under consideration” [4] (p. 10). Having another (ideally, more knowledgeable) pair of eyes comparing the target text (TT) against the source text (ST) makes intuitive sense and has been shown to be effective even in non-standard settings such as crowdsourcing, where integrating one or occasionally two stages of revision ensured that the final quality of the translated product was within accepted professional standards [5]. However, in certain IOs, the progression from the role of the Translator to the much-coveted one of Reviser also includes the role of Self-revising translator, which assumes that translators with more experience can always effectively identify their own errors. This is a risky assumption to make, nevertheless, given that all the studies mentioned in [3], the error analysis performed in [5], and our own experiment, have highlighted that even experienced linguists miss a number of translation errors during revision.

Although our experiment was not specifically designed to study self-revising translators, but revisers as individuals different from an initial translator, the finding that the revisers in our study did not correct all of the errors present in the initial translation (see Section 3.1) highlights the need for new methods to improve this situation. Should a third person not be available—as is most often the case due to budget and time constraints—we hypothesised that alternative attention-raising technology such as automatic speech synthesis could be integrated into current CAT/TEtT/Revision tools to enhance the effectiveness of the traditional silent revision process.

As for the *when* dimension of *revision*, the Translation Workflow proposed in ISO 17100 suggests that Revision be performed after the stages of Translation and Check and before the optional, client-negotiable stages of Review and Proofreading. Therefore, revision is not an optional component of an effective translation workflow that aims to produce a translation of publishable quality. However, the widespread practice of spot-checking (selective revision) can undermine the claimed effectiveness of the compulsory stage of revision.

Our case study aims to contribute to the *how* dimension of *revision*. Given the recent development of speech technologies such as speech synthesis (text-to-speech) and speech recognition (speech-to-text), revisers have very slowly started to integrate these tools into their workflows, as well as CAT/TEtT set-ups [6,7]. Today, CAT tools are certainly widespread, with only 1% of companies and 13% of language professionals recently surveyed not using them [8]. Moreover, speech recognition (dictation) now ranks third for companies and professionals in the category of ‘other technologies’ used in the translation process, “but it is clearly more popular with the individuals than with the companies” [8] (p. 19).

We believe that the investigation of a speech-enabled CAT/TEtT is timely, especially given that, although speech technologies are still not ubiquitous, professional translators already use “reading aloud” to various degrees in the self-revision process: in a 2017 survey of professional translators, 20% of the 55 participants stated that “they regularly read the translated text aloud, 44% never read aloud the translation they are revising, and the remaining 36% read the translation aloud only occasionally or when the sound and rhythm effects are particularly hard to recreate” [9] (p. 15). Our study

aims to complement such self-reported evidence by combining eye-tracking, error annotation and questionnaire data to investigate the impact of hearing the source text on the revision process.

To date, research suggests that monolingual revision involving only the target language is ineffective and even dangerous—which should also send a serious warning message to supporters of post-editing machine translation (PEMT) carried out without knowledge of the source language. For example, Brunette et al. [10] showed that bilingual revision consistently achieves better results and outperforms monolingual revision in terms of accuracy, readability, appropriateness and linguistic coding. They also reported the use of monolingual revision to be “an irrational practice, even less helpful than no revision” (ibid.).

At the same time, linguists appear to mix and match approaches depending on personal preferences rather than scientific research: an empirical study of revision practices in the Danish translation profession—with particular focus on translation companies—revealed that “the preferred procedure seems to be that the target text is first checked on its own and that a comparison with the source text is only carried out where it is deemed necessary or relevant” [11] (p. 114). The researchers also found that “some revisers prefer to do it the other way round, i.e., starting with a full comparative revision followed by a unilingual revision” [11] (p. 114). Moreover, a recent professional translator survey [9] indicated that, of the 55 respondents, 40% of them start revising by reading the ST beforehand (only 8% read the ST in full) and 52% check the ST against TT line by line. These practices are present in even higher proportion in trainee translators, with an eye-tracking study of 36 participants indicating that 55.6% of students read the ST in the initial planning stages (which will be called the *orientation phase* in the current article), 88.9% check the ST against TT in the revision per se, and none of the participants focus solely on the TT at the expense of the ST [12]. Lack of time seems to be the main factor preventing professional translators from applying these practices to the whole text, while trainees are influenced by their translation proficiency, as well as text type, length and complexity [12].

In an attempt to describe an “ideal” revision process, a three-step activity is proposed by some researchers [10], consisting of an initial reading of the ST, followed by a comparative reading of ST and TT (referred to as “bilingual revision”), and finally a correction and re-reading of the TT. Conversely, others advocate for “unilingual revision,” that is, the reviser’s reading of the target text alone, going back to the source text only when the reviser detects a problem and subsequently makes a change [3]. This unilingual re-reading “may well produce a translation that is not quite as close in meaning to the source as a comparative re-reading will produce. On the other hand, it will often read better because the reviser has been attending more to the flow and logic of the translation” [3] (p. 116). Another benefit of “reading the draft translation without looking at the source text,” can be that revisers “have the unique opportunity to avoid coming under the spell of source-language structures.” [13] (p. 13).

Leaving aside the drive to improve target language fluency, currently the hype around monolingual target language revision feeds on the hype around the occasionally surprising quality of neural machine translation (NMT) and the constantly-publicised need for ever-faster translations. However, Robert and Van Waes [14] found that a bilingual revision does not take significantly more time than the ineffective (especially from the point of view of ensuring the accuracy of the translation) monolingual approach in their study that investigated the correlation between *quality* and *revision method* (monolingual; bilingual; bilingual followed by monolingual; and monolingual followed by bilingual) among 16 professional translators. Moreover, although some professional linguists still seem to prefer combining a monolingual revision stage with bilingual revision, Robert and Van Waes [14] found that to be unnecessary. Following the analysis of key-logging, revisions made, and think-aloud protocols, they concluded that contrary to their expectations, for full revision “there is no significant difference between the bilingual revision procedure and the two procedures involving a bilingual revision together with a monolingual revision” (a comparative re-reading followed by a monolingual re-reading, and a monolingual re-reading followed by a comparative re-reading). This led them to state that “re-reading a second time does not seem to be worth the effort”. This is a very

important suggestion, especially given the growing productivity expectations of the language services industry [15].

As more translators are translating and revising using CAT and speech recognition and/or synthesis tools together, the challenge becomes identifying the optimum integration and combination of such tools when carrying out specific tasks such as revision. A piece of technology that is proving helpful in this sense is eye-tracking, a well-established method in psychology that is being increasingly exploited in Translation Studies. More recently, eye-tracking has been used in Translation Process Research (TPR) to investigate viewing habits of different categories of language users (translators, revisers, post-editors, subtitlers, even language learners). In our experiment, eye-tracking is used, first of all, to describe participant behaviour when revising with and without sound. It is also used to investigate whether there is any relationship between the eye measures considered, the quality of the revised output, and the participants' perception of the speech technology used.

Eye-tracking studies can take into account different types of eye movements—e.g., *fixations* and *saccades*. A fixation can be defined as “an instance of gaze that remains for longer than a predefined duration on the same point on the screen” [16] (p. 35). A saccade, on the other hand, is a “rapid jerk-like movement ( . . . ) necessary to direct the gaze to a new location” during which, however, “no meaningful new visual information is gathered” [17] (p. 2). In this paper we will only report on fixation data. Among these data, some of the most commonly used measures are fixation counts, fixation durations and mean fixation durations (MFD). Generally speaking, fixation counts refer to the number of fixations made on a particular item; fixation durations represent the amount of time spent processing that item (usually calculated in milliseconds—ms); and mean fixation duration (MFD) is a secondary measure calculated by dividing fixation durations on one item by fixation counts. When all fixation durations on a particular item are added together, dwell time is obtained.

Following current interpretations of eye movements in reading research, we work on the assumption that fixation counts and dwell time provide a measure of overall attention and total cognitive effort dedicated to the ST and TT, while MFD provides a measure of depth of processing, which includes ease of access to word meanings and word integration into the sentence being read [16–19]. On average, a longer fixation duration is “often associated with a deeper and more effortful cognitive processing” [20] (p. 381), which is relevant to the present investigation as the aim is to establish whether speech synthesis hinders or facilitates the revision task.

Our case study investigated the impact of adding sound produced by speech synthesis tools to what has been traditionally a silent, text-only process—namely, revision in a CAT tool. We acknowledge working with a small number of revisers—with all the limitations that this approach entails—addressed specifically in Section 5.

## 2. Materials and Methods

### 2.1. Research Questions

The study investigated three core research questions using a variety of research methods and types of analysis (see Table 1):

- (RQ1) Which of the two conditions (with source sound [SS] versus without source sound [NoSS]) is conducive to spotting and correcting more translation errors in the revision process?
- (RQ2) What are the participants' views on the integration of text-to-speech in the revision process?
- (RQ3) Is viewing behaviour while revising affected by the presence of source sound [SS]?

**Table 1.** Research questions, methodology used to investigate them, and type of analysis used.

Research Question	Method	Type of Analysis
RQ1	Error counts	Error correction analysis (based on the TAUS DQF-MQM error typology)
RQ2	Questionnaire	Qualitative and quantitative analysis of participant responses
RQ3	Eye-tracking	Eye movement analysis—fixation counts, dwell time, Mean Fixation Duration (MFD), weighted average MFD; External Resources (ER) examined

## 2.2. Participants

Thus far, our study has involved five professional translators (based in Leeds, UK, with a minimum of one year of professional experience and an average age of 36.4 years) and six trainees (whose average age was 25). The trainees were postgraduate students enrolled on the MA programmes offered by the Centre for Translation Studies at the University of Leeds. All participants were native English speakers translating and revising from French, and they were remunerated for their contribution. They were recruited by e-mailing the local student and professional linguists' groups.

## 2.3. Experimental Texts

The participants revised the English translation of a text written originally in French: a financial report (Text 1–T1; 298 words in the French original; 265 words in the English translation; 11 segments in the CAT tool memoQ). This text was provided to the research team by one of the researchers' industry collaborators—Sandberg Translation Partners (STP)—with the specification that the English translation had failed the STP's freelancer recruitment test for French into English translators. Further errors were inserted into the translation by the research team to test the revisers' attention to detail, and the structure of certain literal English translations was reversed in order to investigate whether viewing patterns change when revising them in a with-sound condition—for example, the French original "Le management réitère sa confiance d'une croissance d'au moins 10% en 2013, étant donné les opportunités dans les pays émergents, le positionnement de ses marques Ray-Ban et Oakley, et la nouvelle licence Giorgio Armani." had been initially translated as "The management reiterates its trust in growth of at least 10% in 2013 given the opportunities in emerging countries, the positioning of its brands Ray-Ban et Oakley and the new Giorgio Armani license". However, in our experiment, revisers worked on the structurally-modified alternative "Given the opportunities in emerging countries, the positioning of its brands Rayban and Oaxley and the new Georgio Armani license, the management reiterates its trust in growth of at least 10% in 2013". While the change in structure was not regarded as an error to be corrected because the meaning is preserved in both alternatives, the *Fluency* (i.e., in this example spelling and grammar) and *Style* (i.e., awkward style) errors present in the translation were expected to be corrected by the revisers.

Overall, the revisers had a total of 33 errors to correct in T1 (10 Accuracy, 17 Fluency and six Style) [21].

STP also provided an alternative T1 English translation as an example of a translation that had passed the company's recruitment test; the experiment participants' revision output was manually scored against this 'gold standard' text by two assessors working collaboratively.

Following best practices in experiment design, all participants first became familiar with the experimental environment by revising a 100-word English practice translation of a 111-word French original text on computer encoding practices (Text 0—T0; 10 segments in the CAT tool memoQ).

The short length of both T0 and T1 allowed them to be viewed by participants without the need for scrolling.

## 2.4. Experimental Design and Procedure

We randomly split our 11 participants into two groups—G1\_1 and G1\_2—while still ensuring a balanced representation of trainees and professionals. Given that professional revisers rarely use speech synthesis habitually in their work [6,7], all participants started by revising T0 in 10 min using

source sound (SS). The *source sound* was generated using a French synthetic computer voice. This practice text was followed by the revision of T1 in a maximum of 25 min, as follows: G1\_1 revised T1 in silence (*no source sound* condition—NoSS); and G1\_2 could request to listen to *source* text sound (*source sound* condition—SS) while revising T1 (see Table 2). Overall, G1\_2 participants requested to hear the T1 source sound in 98% of the cases (only one segment was not requested by one participant). T0 was just a practice text (although participants were not aware of this aspect at the time); therefore, the errors corrected during this initial revision were not logged in the experiment results.

**Table 2.** Distribution of experimental groups and tasks.

Name of Group	Professionals	Trainees	T0 (Training Text)	T1 (Financial Report)
G1_1 ( <i>n</i> = 6)	3	3	Source sound (SS)	No source sound (NoSS)
G1_2 ( <i>n</i> = 5)	2	3	Source sound (SS)	Source sound (SS)

Participants had their eye movements tracked while revising to publishable level the two texts T0 and T1 in memoQ within the already-mentioned time limits that matched a professional expectation of approximately 5000 words/day [15]. Data were collected using an EyeLink 1000 Plus eye tracker unit (SR Research) with a 35 mm lens and a sampling rate of 500 Hz. For a fixation-based study like ours, such a frequency value can be deemed more than adequate, given that a threshold of 60 Hz is often considered sufficient for this type of study, and 500 Hz hardware is commonly used in saccade-based studies [16].

To mirror real working conditions, participants were allowed to use online external resources and switch freely between memoQ and Google Chrome throughout the revision. After finishing the T1 revision task, participants were also asked to record their impressions in an online questionnaire whose results, alongside the eye-tracking data and the error analysis results, provide a fuller perspective on the process of revision.

The experiment obtained ethical approval (reference LTS LCS-081) from the University of Leeds Arts, Humanities and Cultures Faculty Research Ethics Committee, and all participants read a summary of the study procedure and signed a consent form before taking part.

### 2.5. Hardware and Software Set-Up

The French ST and the corresponding English TT to be revised were presented in the CAT tool memoQ in the Tahoma typeface, size 14. We chose memoQ (version 8.5) for our study because we designed our experiment to incorporate source sound triggered by voice commands uttered by the reviser, and at the time of the design memoQ was the most speech-friendly CAT tool available on the market, with the highest compatibility level with Dragon NaturallySpeaking (DNS, version 15 Professional), the most popular dictation tool for professional linguists [6].

However, at the time of the experiment (December, 2018–March, 2019), our experimental hardware running concurrently memoQ 8.5, DNS 15 Professional, Google Chrome and the eye-tracking software EyeLink 1000 Plus (SR Research) with its built-in Screen Recorder software resulted in a lag of approximately 8 seconds between users uttering a “Read Sentence” command in DNS and hearing the ‘spoken’ output of that command.

During a series of successively-tuned pilot studies with volunteers, it became clear that 8 seconds was too long for revisers to wait and, since no alternative software combination met the research criteria, the research team created separate audio files for all individual source text segments using the Microsoft Word built-in Read Aloud French synthetic voice. During the experiment, revisers uttered the command “Read source 1/2/3/*n*” to listen to that particular recording played without delay by a research team member.

In order to investigate our third research question regarding the impact on viewing behaviour of incorporating sound into the revision process, we used eye-tracking technology. Initially, for ecological validity, we set out to use a mobile eye-tracking approach, which would allow participants to move

their heads freely during the tasks. However, during the pilot phase, a series of calibration and tracking issues arose, which impacted the quality and reliability of the recorded data. The chosen alternative option of using a tower mount with a fixed head-and chin-rest ensured a much higher eye-tracking accuracy and eliminated tracking issues caused by the participant moving out of the head-box space (the trackable area when working with remote eye-trackers).

The experiment was run remotely on a display laptop (with an Intel i7 processor, 8 GB of RAM, and running Windows 10) that was connected to an external keyboard, mouse, and 22-inch LCD monitor screen to improve the ergonomic set-up for the participants. The display laptop was also connected to the Host PC, where the calibration process was run by the experimenter. Data were recorded using Screen Recorder, a dedicated piece of software provided by SR Research, which interfaces with the eye-tracker hardware, allowing researchers to simultaneously record eye movements, ambient sound and screen activity, including keyboard typing and mouse clicks. Screen Recorder produces not only the eye-tracking data file (.edf) but also a separate video (.mpeg) and audio (.wav) file for each participant recording.

### 2.6. Data Preparation for Analysis

Our experiment produced three types of data: revision error analysis data; participant questionnaire data; and participant eye-tracking data.

First, the revised English text T1 was exported as MS Word files out of memoQ for subsequent error analysis. Following industry practices, the error annotation was performed using the TAUS harmonised DQF-MQM error typology [21], which is well-known in the localisation industry. Two independent assessors worked collaboratively to identify how many of the initial T1 Accuracy, Fluency and Style errors had been corrected by each participant.

Secondly, the post-eye-tracking questionnaire gathered the following information:

- Demographic information on the participants' gender, age, translation, interpreting and revision experience, as well as previous exposure to CAT and speech technologies;
- Participants' perceptions regarding revision quality, productivity and concentration when using speech synthesis in the experiment;
- Participants' preferences regarding future use of speech technologies in the process of translation and revision.

The questionnaire consisted of close-ended questions, following a three-level Likert scale approach, except for the final question, which asked participants to comment on their overall evaluation of speech technologies in the process of revision. The questionnaire was designed in Google Forms and was filled in by participants electronically in the lab. The results were downloaded from Google Forms in MS Excel format for analysis.

Thirdly, the software Data Viewer (SR Research) was used to visualise the eye-tracking and screen activity data, and to produce the fixation data files. In the Data Viewer replay function, the experimenters manually identified which time intervals (called 'interest periods' IPs) of each participant's video recording were spent revising in the CAT tool (labelled 'memoQ intervals'), and which time intervals were spent consulting external resources (labelled 'web searches'). The beginning and end of each of these IPs were marked in Data Viewer, so that fixation data can be extracted on these two types of IPs, either individually (a particular memoQ interval or web search) or cumulatively (e.g., aggregation of all or some memoQ intervals). In this study, we extracted data cumulatively, which means that eye movement measures on each memoQ interval were collated in one dataset by Data Viewer, which was then exported for analysis. Importantly, the number and length of the memoQ intervals differs between participants working on the same text because each participant had the freedom to make as many or as few web searches as they deemed necessary. Data were extracted for each memoQ interval on the two areas of interest (AOIs) relevant to the study, namely the source text (ST) and the target text (TT). To ensure our eye-tracking data were genuinely collected only on ST

and TT, the AOIs were drawn *only* around the texts themselves; thus, any looks anywhere else on the computer screen (e.g., at the time, taskbar, segment numbers, or fuzzy match percentages in memoQ) were not counted, preventing artificial inflation of fixation measures.

As already briefly mentioned in Section 1, fixation count, dwell time and MFD data were collected and analysed in this study. Specifically, we defined *fixation counts* as the total number of fixations falling within an AOI. We defined *dwell time* as the sum of the durations of all fixations falling in an AOI. In this study, therefore, dwell time refers to the amount of time spent processing the ST or the TT *while being capable of processing visual information*, i.e., excluding saccades. MFD in each memoQ interval was obtained by dividing dwell time on an AOI by fixation counts on that AOI. Alongside these values, for two participants' full-length revision session screen recordings, we present a *weighted average MFD* for both ST and TT in order to acknowledge the relative importance of the separate memoQ interval fixation counts. These two participants' experiment ids are *s08* and *s09*. When extracting data for analysis, fixation counts and dwell times are primary measures calculated by the eye-tracking software, whereas MFD and weighted average MFD are secondary measures calculated from the primary ones.

Finally, given the different word counts between source text and target texts, we have normalised our participants' fixation counts per 1000 words.

### 3. Results

For the error analysis (RQ1) and the questionnaire analysis (RQ2), this paper will present data from 11 participants belonging to experimental groups G1\_1 and G1\_2 (see Table 2). For the eye movement analysis (RQ3), we will present a comparison of eye-tracking data of two participants—*s08* and *s09*—belonging to G1\_1 and G1\_2, respectively. Time and resource constraints have prevented us from analysing the eye-tracking data of all 11 participants, but we trust that, for a case study, this reduced amount of data still offers interesting insights.

#### 3.1. Error Data (RQ1)

In this section we will specifically address RQ1, namely: Which of the two conditions (with source sound [SS] versus without source sound [NoSS]) is conducive to spotting and correcting more translation errors in the revision process? The error annotation task identified 10 Accuracy errors, 17 Fluency errors, and six Style errors in the French ST to be corrected in the T1 English translation (in total 33 errors). As a reminder, G1\_1 worked in the NoSS condition, while G1\_2 worked in the SS condition. Table 3 presents the percentage of errors that participants corrected in each category, as well as the average error correction scores for each group (G\_1 and G\_2).

**Table 3.** Percentage of corrected errors grouped by DQF-MQM category for T1.

Test Group	T1 Condition	T1_Accuracy	T1_Fluency	T1_Style	T1_Total_Errors_Corrected	Participant ID
G1_1	NoSS	0.40	0.76	0.67	0.64	s01
G1_1	NoSS	0.30	0.24	0.50	0.30	s07
G1_1	NoSS	0.40	0.47	0.50	0.45	s08
G1_1	NoSS	0.60	0.71	0.83	0.70	s11
G1_1	NoSS	0.30	0.41	0.33	0.36	s13
G1_1	NoSS	0.20	0.47	0.83	0.45	s14
G1_1 group average		0.37	0.51	0.61	0.48	
G1_2	SS	0.50	0.76	0.67	0.67	p01
G1_2	SS	0.70	0.47	0.67	0.58	s04
G1_2	SS	0.80	0.59	0.67	0.67	s09
G1_2	SS	0.60	0.18	0.83	0.42	s10
G1_2	SS	0.70	0.59	0.67	0.64	s12
G1_2 group average		0.66	0.52	0.70	0.60	

### 3.2. Questionnaire Data (RQ2)

In this section we will specifically address RQ2, namely: What are the participants' views on the integration of text-to-speech in the revision process?

Out of 11 participants, five reported to have no interpreting experience, and four more stated that they had less than one year's interpreting experience. We believe that this was an important question to ask because of the audio element which we introduced in our experiment. Traditional translation and revision training and professional practices are mostly silent, while interpreters are much more likely to have performed tasks that involve sound. We were interested in capturing whether previous interpreting experience would correlate with a particular attitude towards using sound in the revision process.

Regarding the participants' translation, revision and interpreting experience, as well as exposure to CAT tools, Table 4 summarises their answers. Moreover, Table 5 details whether participants habitually used CAT tools for translation and revision.

**Table 4.** Experiment participants' self-reported experience with *translation, revision, CAT tools* and *interpreting*.

Experience	No Experience	Less than 1 Year	1–3 Years	3–5 Years	Over 5 Years
Translation	0	3	3	1	4
Revision	0	4	2	1	4
Working with CAT tools	4	4	1	0	2
Interpreting	5	4	1	0	1

**Table 5.** Experiment participants' self-reported habitual environment for *translation* and *revision*.

Environment	In a CAT Tool	Outside CAT Tools
Translation	6	5
Revision	5	6

The perceived impact on concentration, productivity and quality when revision included SS was rather mixed and is summarised in Table 6 (+ represents a positive impact, – a negative impact, and *neutral* no impact):

**Table 6.** Perceived impact on concentration, productivity and quality of revision when SS is present.

With SS	Concentration	Productivity	Quality
+	3	4	3
–	4	4	0
Neutral	4	3	8

Regarding the participants' preference for revising *with* or *without* sound, at the end of the experiment seven participants chose the *with* option. In addition, when asked whether the experiment prompted them to investigate further the possibility of integrating speech technologies into their revision process, five said *yes*, four *maybe* and only two *no*.

When asked whether hearing the source sound led them to be more alert, five participants reported enhanced alertness on the source segments, three on the target, two felt no change, and one reported lowered levels of alertness on both source and target due to a perceived difficulty in concentrating, especially for longer segments.

As mentioned above, the last question was open-ended, to enable participants to express their thoughts regarding employing speech technologies in the translation and revision process in general. The most relevant insights from these answers will be presented in the discussion below (Section 4).

### 3.3. Eye Movement Data (RQ3)

This section presents the data relevant to the last research question (RQ3): *Is viewing behaviour while revising affected by the presence of source sound [SS]?* This analysis reports data from a subset of two participants: one from G1\_1 (s08, NoSS condition) and one from G1\_2 (s09, SS condition). Both participants were female, professional, over 30, with similar experience translating and revising, and also similarly familiar with CAT tools. Table 7 illustrates a variety of eye movement measures, namely fixation counts, dwell time and weighted average MFD, as well as the number of T1 memoQ intervals. For an explanation of all these measures, please see Section 2.6. Given the different word counts of the ST and TT for T1 (see Section 2.3), the total fixation counts have been normalised per 1000 words.

As one can see from Table 7, both participants concentrated more on the TT during the revision task, as all measures, including normalised TFC and weighted average MFD, display higher values in T1 TT. The eye-tracking data collected further shows that, as opposed to s08, s09 had an orientation phase of 3.7 min, during which s09 carried out a parallel reading of all source and target segments in silence. Despite the difference in the number of memoQ intervals made by s08 ( $n = 14$ ) and those made by s09 ( $n = 20$ ), the average memoQ interval duration was similar: 13.4 s for s08 ST versus 13.9 s for s09 ST, and 31.6 s for s08 TT versus 29.9 s for s09 TT.

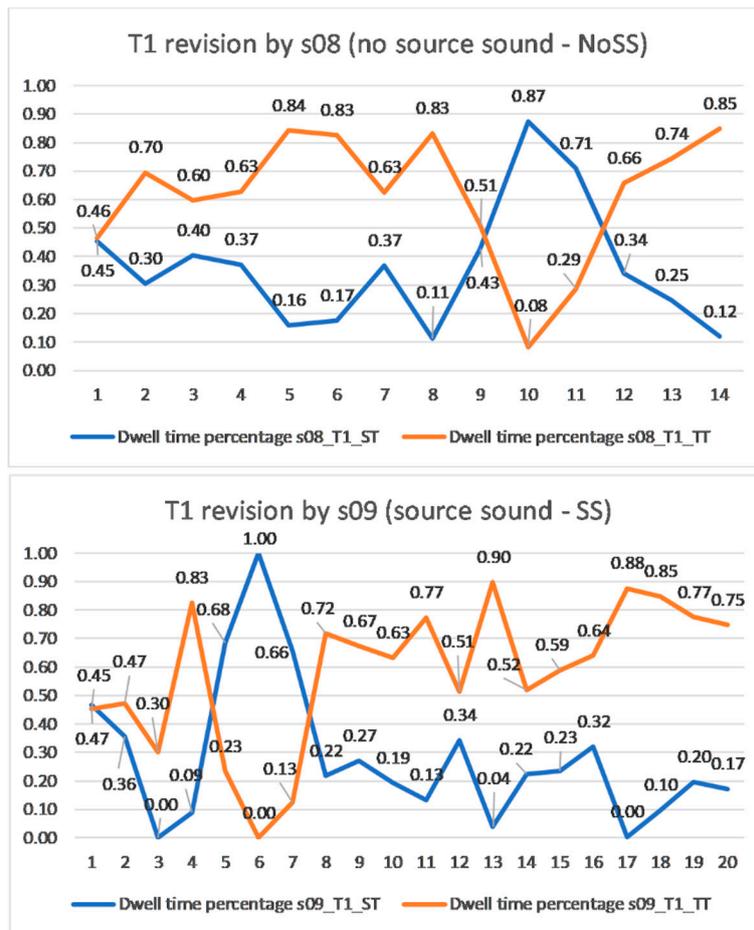
As already mentioned, a *web search* in our current study represents a distinct interval between two memoQ intervals, in which a participant consulted one or several external resources (ER)—e.g., WordReference, Reverso, Linguee and Google Chrome. s08 performed 13 such web searches, against s09's 19.

In total, eye-tracking screen recordings show that s08 spent approximately 16 min on the T1 revision: 10.5 min in memoQ and 5.5 min on web searches (66% and 34% of the total revision task, respectively), while s09 spent the full 25 min allowed for the T1 revision: 14 min in memoQ and 11 min on web searches (56% and 44% of the total revision task, respectively).

Moreover, Figure 1 displays the percentage of time spent looking at ST and TT across memoQ intervals for s08 and s09. In these graphs, the X axis marks the individual memoQ intervals (i.e., time periods spent in the CAT tool between web searches) made by each participant (14 for s08, 20 for s09), while the Y axis indicates the fixation dwell time expressed as a percentage of the total dwell time per memoQ interval. The data in the NoSS graph shows that, during the revision task, s08 spent more time on the TT, with the exception of intervals 10 and 11. The data in the SS graph shows that, when revising with sound, s09 also spent more time on the TT, with the exception of intervals 1, 5, 6 and 7.

**Table 7.** T1 fixation data on ST and TT for s08 and s09 within memoQ intervals.

	No of T1 MemoQ Intervals		Total Fixation Counts (TFC)		TFC Normalised by 1000 Words		Dwell Time		Weighted Average MFD (Mean Fixation Duration)	
	s08	s09	s08	s09	s08	s09	s08	s09	s08	s09
<b>T1—Source Text (ST)</b>	14	20	807	1336	2708	4483	188,141 ms (3.14 min)	263,383 ms (4.39 min)	233 ms	197 ms
<b>T1—Target Text (TT)</b>	14	20	1781	2735	6720	10,320	442,532 ms (7.38 min)	567,327 ms (9.46 min)	248 ms	207 ms



**Figure 1.** s08 and s09’s dwell time on T1 ST and TT expressed as a percentage of the memoQ intervals.

Figure 1 is useful for investigating the proportional distribution of time spent by s08 and s09 looking at the ST and TT in each memoQ interval. In addition, Figure 2 shows the number of fixations (fixation counts) made by s08 and s09 within these AOIs. In these graphs, the X axis represents again the same memoQ intervals linked to each participant, while the Y axis represents fixation counts. The two graphs in Figure 2 show different patterns of fixations in the two conditions. When working without sound (NoSS, top graph), data show that more fixations were consistently made on the TT in most memoQ intervals (except for interval 9, 10 and 11). When working with source sound (SS, bottom graph), data show a more varied behaviour, though still within the trend of looking more at the TT.

Figure 3 below shows MFD values by memoQ interval in the two conditions. MFD considers dwell time and fixation count data together, providing a measure of depth of processing. Therefore, considering MFD alongside dwell time and fixation counts helps provide a more precise picture of cognitive processing on an AOI (in our case, ST and TT). The top panel reveals that s08, working without sound, processes the TT more deeply in eight memoQ intervals (2, 4–9, 12) and the ST in six memoQ intervals (1, 3, 10, 11, 13, 14). The bottom panel shows that s09, working with sound, on average processes the TT more deeply in 15 memoQ intervals and the ST in five memoQ intervals (5, 6, 7, 15, 16).

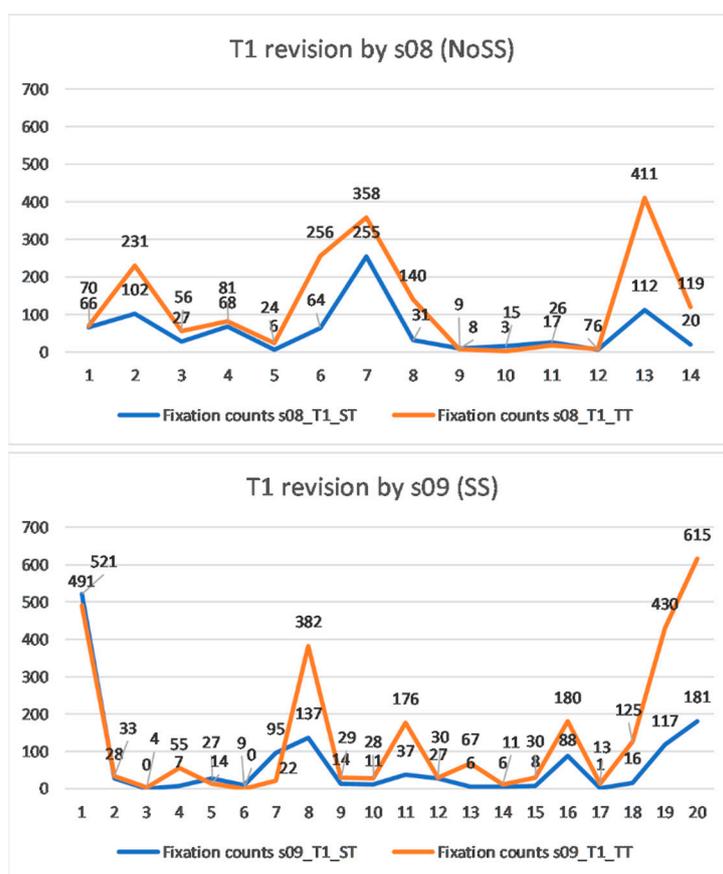


Figure 2. s08 and s09’s fixation counts for T1 ST and TT during memoQ intervals.

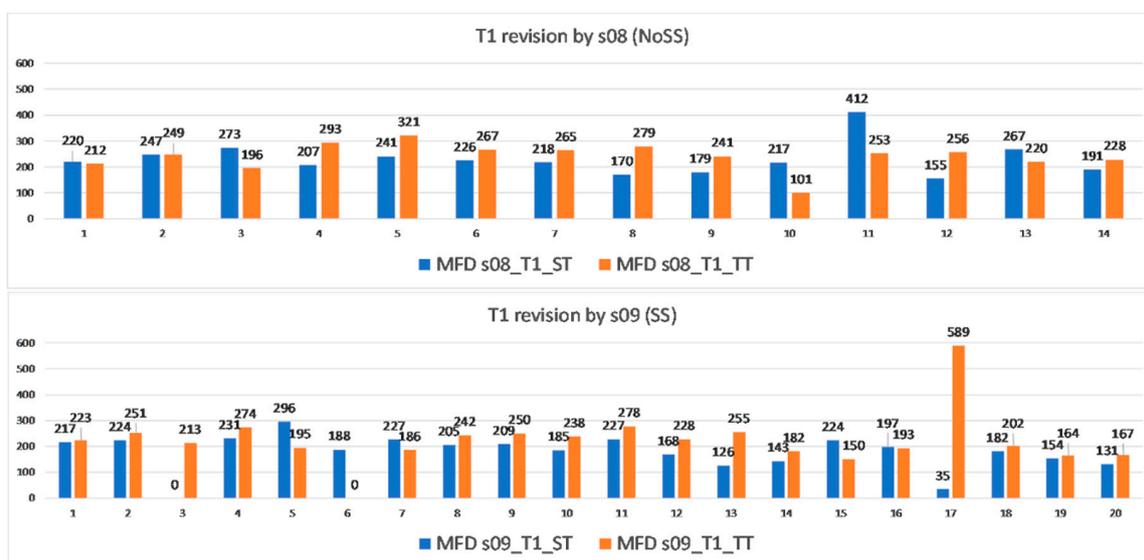


Figure 3. s08 and s09 MFD for T1 ST and TT during memoQ intervals.

#### 4. Discussion

In answer to our RQ1, the error analysis performed suggests that the presence of source sound was conducive to better revision quality overall. All the TAUS DQF-MQM categories used—i.e., Accuracy, Fluency and Style—registered improvements for G1\_2 (Table 3). The largest improvement was for Accuracy, where G1\_2 achieved an average score of 66% compared to 37% for G1\_1. Style was the category which benefitted next—G1\_2 scored 70% compared to 61% for G1\_1, while the scores for

Fluency are very close: G1\_2 achieved 52% compared to 51% for G1\_1. Overall, G1\_2 corrected a total of 60% of all errors, while G1\_1 only managed 48%.

The open-ended part of our questionnaire helped explain this better performance from G1\_2, who had been revising with source sound: “[source sound] helped me to understand the flow of the speech better”; “[it] alerted me to particular kinds of errors (such as spelling errors) that I may have missed from reading the source text”; “[I spotted] different kinds of errors”; “having the source text read to me ensured I was more focused on it”; “speech technology made me focus back on the ST”; “it allowed me to check that all of the info (and the same logic) was to be found in the target text”; “I noticed errors in the translation as a result of hearing the source, it was useful to have these ‘jump out’ at me in this way”; “it was very useful for picking up subtle content errors”; “checking for typos, proof-listening, if you like”; “I concentrated better in this case than if I had only been reading the text”.

We find these improvements in revision quality encouraging given that T1 had an increased level of difficulty based on the majority of readability scores produced by StyleWriter 4, a popular tool used by content editors to assess the difficulty and appropriateness of English content for different target audiences (for details on how StyleWriter 4 calculates the scores, see Table A1 in Appendix B).

One of the known features of challenging texts is segment length—in T1, ST segments ranged between 3 and 51 words (average size 27.09), and TT segments ranged between 2 and 43 (average size 24.09). The issue of segment length was also raised by our participants: “I found my concentration drifting with longer segments” and “[revising with sound] would take some getting used to in longer chunks”. Therefore, we suggest that the presence of sound during the revision process will have different impacts depending on the length and complexity of the source and target segments—with the emerging evidence presented here pointing to a higher likelihood that users will find it more comfortable to adapt to this new working mode on shorter segments.

Secondly, in connection with RQ2, as our participants had no consistent experience with speech technologies, their views on their impact on the translation and revision processes only reflect their limited exposure to these technologies during our experiment. Overall, a majority of seven participants said that they would prefer hearing the source sound when revising, while four said that they would prefer not having source sound (SS) during revision. From their comments, it appears that using synthetic sound effectively takes practice. After attempting to carry out visual bilingual text processing concomitantly with aural source processing, some of our participants stated: “I simply couldn’t read the source and target and listen to the source at the same time”, as well as “I am not used to having to think about hearing the source, so it was an additional thing to do and slightly disrupted my usual way of approaching a revision”, and “occasionally I would get distracted while trying to edit the target segment and listen to the source segment at the same time”. At times, though, this disruption was welcomed: “it was slowing me down as I stopped writing to listen; [ . . . it] alerted me to particular kinds of errors that I may have missed”.

We believe that, with specialised training, linguists can learn to progress from “shorter segments, especially segments with figures, [where] I found the dictation quite useful” to longer and more complex segments, especially as most of our participants saw clear benefits in using speech. Nine out of 11 participants expressed an interest in continuing to explore using speech technologies—“it would help to get used to doing actual translation in this way”—and four even thought that they were more productive when using it. Such progress is certainly possible, given that even 1 hour in the lab was considered useful: “I quickly adapted”.

Our questionnaire shows that, when listening to source sound, three translators reported to have been more alert to the corresponding target segment, five to the source segment being read, and the remaining three reported no change. Such alertness was beneficial for identifying more subtle errors: “it was very useful for picking up subtle content errors and I feel that in time, with practice it would be very helpful in improving my productivity”; “[it] allowed me to focus on which words or expressions were difficult for me, which sped up the process”. All those who reported enhanced alertness to the target had interpreting training. On the other hand, other participants reported that source sound

was “difficult to process”, and that their listening skills were not always sufficiently developed to enable them to cope with the multimodal processing of information. As all of these skills are part of interpreting training, we believe that they should be included in future translation courses exploring the integration of speech technologies in translation and revision workflows.

Moreover, such training needs to be customised because all four participants who, at the end of the experiment, expressed a preference for revision without sound, in fact had interpreting training and/or experience. An enhanced focus on the very challenging task of bilingual text processing concomitantly with aural input processing will be needed for revisers to perform comfortably in this environment.

Having said that, even without customised training and during a very short experiment such as ours, none of the participants perceived any adverse effect on quality posed by the presence of source sound—see Table 6. All those who reported SS to have had a positive impact on the quality of their revision also reported SS to have led them to focus more on the source segments. We believe that this increased focus also contributed to Accuracy scores being much higher for the group which heard the source sound than the group who revised T1 in silence.

Except for the Quality category in Table 6, the reported data on the other two indicators—Concentration and Productivity—is less conclusive. Similarly, we found no correlations between the participants’ profiles as summarised in Tables 4 and 5 and their attitudes and preferences towards using source sound in the revision process.

Finally, the difference of the artificial voice from a human one—“not sound[ing] like a natural human speaker”—was also mentioned by participants: “the robotic voice was difficult to listen to”. As speech technologies continue to evolve and are gradually integrated into CAT/TEnT environments, we hope that this inconvenience will be addressed.

Thirdly, in response to our initial RQ3, we would like to reiterate that the data reported here on s08 and s09 does not allow us to make any generalisations—we chose to report on these two participants because their professional profiles were most similar, but there are no data available indicating they are also representative of their respective groups in terms of reading behaviour. Therefore, our discussion below only highlights interesting comparative elements.

First of all, the fixation counts presented in Table 7 show that, as expected, both participants focused more on the target text than the source text. s08 made fewer fixations on ST and TT than s09, yet we cannot state whether this was due to the impact of introducing source sound in s09’s condition (G1\_2). However, it is interesting to observe, when comparing the normalised fixation counts per 1000 words, that, despite the rather different values between s08 and s09 (2708 versus 4483 for ST and 6720 versus 10,320 for TT), the ratio of TT fixations to ST fixations is quite similar: 2.48 for s08 and 2.30 for s09.

Focusing more closely on s08 and s09, Figures 1 and 2 show that s08 had a finalisation phase, while s09 had both an orientation and a finalisation phase. The orientation phase we observed for s09 is similar to the stage discussed in [22] for translation tasks, which encompasses activities carried out before typing begins, and, in our case, before starting the revision proper with source sound (SS). The presence of the orientation phase for s09 is in line with Scocchera’s [9] and Huang’s [12] findings, which show that 40% of professional revisers and 55.6% of student translators prefer to read the ST before revising the TT.

Moreover, our eye-tracking data showed the presence of a finalisation phase, in which both s08 and s09 had a larger number of fixations on the TT while checking once more the TT against the ST segment by segment. The data captured for both participants also confirm the results of previous studies where fewer referrals to the ST are made in the finalisation phase [12,23,24]. The questionnaire results suggest that source sound is perceived by some as actively supporting the finalisation phase: “[speech synthesis was] useful as a kind of final check, as it allowed me to go through the ST more quickly than if I had only been reading the text”.

The weighted average MFD values which we calculated in order to mitigate the relative importance of the separate memoQ interval fixation counts were consistently higher for s08 than s09 (233 ms versus

197 ms for ST, and 248 ms versus 207 ms for TT). By comparing only two participants, however, it is not possible to establish whether this is due to the presence of source sound. As expected, from these values, as well as the ones presented in Figure 3, both participants appear to concentrate more on the target text. Further analysis at group level will enable us to see whether the presence of source sound generally leads to shorter weighted average MDF values or whether s09 is an exception.

Eye-tracking data also showed that s09 had more memoQ intervals than s08, which means that s09 stopped more frequently to perform web searches in websites such as Reverso, Linguee, or in Google. Daems and colleagues [25] investigated the use of external resources (ER) by trainee translators for translation (T) and post-editing (PE) tasks. Their results showed that, overall, more time was spent by trainees on ER for T than PE, but that the type of ER used and task (T or PE) were not significantly inter-dependent. Moreover, spending a longer time on ER correlated with better quality when translating, but not when post-editing. Once we have analysed all the eye-tracking data from all our participants, it will be interesting to see whether G1\_2 as a whole (the group who worked with source sound—SS) made more web searches than G1\_1 (the group who worked without sound—NoSS), whether there were any differences in the nature of these web searches and whether spending longer time on ER correlates with revision quality.

## 5. Limitations

Our small case study presents a number of limitations inherent to our methodological approach, which make the findings not generalisable. First of all, we were limited in the total number of participants we could recruit ( $n = 11$ ) due to the experiment set-up. However, we believe the number we reported on to be not only sufficient for an exploratory study, but also comparable to similar studies.

Secondly, we collected data from both professional and trainee translators. However, carrying out separate sub-analyses was not warranted, as the results would lack reliability due to the size of the sub-groups. Therefore, we analysed the data collectively, but made sure that the groups would be split almost equally (see Table 2), which means that the data we report on is balanced.

Thirdly, we only looked at one text, which naturally confines the applicability of the results to the genre analysed—in our case finance. Financial texts present a set of specific challenges, thus it has to be recognised that the results may differ for other genres.

Finally, as previously mentioned, the eye-tracking analysis reports data for a subset of two participants, further limiting the generalisability of the findings. However, the two participants were matched quite closely (see Section 3.3) and analysing a smaller sample allowed us to present in-depth data (see Table 1, Figures 1 and 2) from which more fine-grained comparisons can be made.

## 6. Conclusions

This article addressed the applicability and relevance of integrating speech synthesis in the revision process, in particular using source segment sound, and presented detailed information on our experiment design and set-up, as well as the revision error analysis and participant questionnaire results. It was very interesting to notice that the presence of source sound led to the correction of more translation errors, especially Accuracy errors.

Our experiment shows that linguists are open to such speech-enabled revision environments, produce higher-quality output using them, and need enhanced training specifically for revision with sound which combines interpreting and translation training methods.

Although in this article we only presented eye movement data for a limited number of participants, in future articles we plan to expand the eye-tracking data analysis across participant groups and report on responses received from further participants exposed to target segment sound. Further research is needed in order to determine the impact of source sound, as well as target sound on texts belonging to different genres, with different segment lengths and difficulty. We hope that CAT/TEnT/Revision tools will soon integrate both source and target sound, in order to increase the ecological validity of such future research.

**Author Contributions:** All three named authors contributed equally to the study design, data analyses and manuscript drafting.

**Acknowledgments:** This research was supported by the University of Leeds LCS Strategic Research Development Fund—SRDF 2018/2019 3. The authors are also grateful to Sandberg Translation Partners, and especially to Anu Carnegie-Brown and Raisa McNab, for providing the original source and target texts used in this experiment, as well as for their constant enthusiasm and hard-work in supporting collaborations between the language services industry and the academia.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A

Below is a copy of the questionnaire, the results of which were analysed when answering RQ2 above.

### “Using speech technologies in the computer-assisted translation process—Questionnaire”

Many thanks for taking part in the Centre for Translation Studies study investigating the influence of speech synthesis on the revision process. Your answers to the following questions are extremely valuable in order to gain a better understanding of your experience with speech technologies during the revision process. Thank you in advance for taking the time to answer our questions. This questionnaire should take no longer than 15 min.

1. Please enter your assigned experiment ID \*
2. Are you \* (Mark only one oval) Female/Male/Prefer not to say
3. Age \*
4. Are you \* (Mark only one oval) A professional translator/A student
5. Have you received interpreting training (more than a module or short CPD) or have professional interpreting experience? \* (Mark only one oval) less than 1 year/1–3 years/3–5 years/Over 5 years/No interpreting training or experience/Other:
6. How long have you been translating? \* (Mark only one oval) less than 1 year/1–3 years/3–5 years/Over 5 years
7. How long have you been revising translations? \* Mark only one oval. less than 1 year/1–3 years/3–5 years/Over 5 years/No revision experience
8. How long have you been using CAT tools? \* Mark only one oval. less than 1 year/1–3 years/3–5 years/Over 5 years/No CAT tools experience
9. Where do you carry out the majority of your translation work? \* Mark only one oval. In a CAT tool/Outside a CAT tool (for example, in a word processor such as Microsoft Word)/Other:
10. Where do you carry out the majority of your revision work? \* Mark only one oval. In a CAT tool/Outside a CAT tool (for example, in a word processor such as Microsoft Word)/Other:
11. How often do you dictate your translations through an automatic speech recognition system such as Dragon Naturally Speaking? \* Mark only one oval. Never/Occasionally/Often
12. How often do you revise translations using a text-to-speech technology that reads out the SOURCE segment for you? \* Mark only one oval. Never/Occasionally/Often
13. How often do you revise translations using a text-to-speech technology that reads out the TARGET segment for you? \* Mark only one oval. Never/Occasionally/Often
14. After experiencing the two conditions of this study, which one do you prefer? \* Mark only one oval. Revision without hearing the source/target segment/Revision when it is possible to hear the source/target segment
15. Did you notice any changes in your approach to revision as a result of listening to the source/target segment? \* Mark only one oval. I felt I concentrated more on the source segment/I felt I concentrated more on the target segment/I felt no change/Other:

16. Did you notice any changes in your productivity as a result of listening to the source/target segment? \* Mark only one oval. I felt more productive/I felt less productive/I felt no change/Other:
17. Did you notice any changes in your concentration as a result of listening to the source/target segment? \* Mark only one oval. I felt less focused on the revision task/I felt more focused on the revision task/I felt no change/Other:
18. What do you think was the impact of hearing the source/target segments on the quality of your revision? \* Mark only one oval. My revisions were of better quality/My revisions were of lower quality/No change in quality at all/Other:
19. Has this experiment prompted you to investigate the possibility of integrating speech technologies into your revision process? \* Mark only one oval. Yes/No/Maybe
20. We would really appreciate it if you could take a couple of minutes to tell us your thoughts about using speech technologies in the translation and revision processes. If possible, please comment on specific instances in this experiment when you found the speech to be of help, and instances when it was a hindrance. Your comments will enable us to better understand how/if speech technologies could/should be used in the translation process. Thank you!"

## Appendix B

StyleWriter 4 generates automatic readability and text difficulty scores for the T1 original English translation.

**Table A1.** StyleWriter 4 scores for the original T1 English translation to be revised.

StyleWriter 4 Categories	T1—to Be Revised
Bog	75 (Poor)
Pep	6 (Poor)
Passive	10 (Excellent)
Style	76 (Poor)
Sentence	26 (Poor)
Jargon	4.2% (Bad)
Glue	41% (Fair)
Grade	16.2 (Difficult)

The following explanations are offered by StyleWriter:

*Bog Index:* The Bog Index is a measure using a weighted difficulty score for more than 200,000 graded words, sentence length and style issues. It gives credit for many writing features (Pep) that make documents easier to read. A Bog Index below 20 shows a clear and readable style. To see how *StyleWriter* calculates its Bog Index, see Bog Index Explained. The Bog Index incorporates *StyleWriter's* Style Index.

*Pep Index:* Pep is writing that make the reader's job easier and more enjoyable. It could be lively verbs, interesting nouns, people's names or conversational style using contractions, personal pronouns, questions and short sentences. *StyleWriter* subtracts the Pep Index from Bog to calculate the Bog Index (see Bog Index Explained).

*Passive Index:* The Passive Index shows whether you have used too many passive verbs and helps you decide how many you need to change into active verbs. Too many passive verbs make writing tedious and difficult to read.

*Style Index:* The Style Index measures all the plain English problems in your text, including a weighted score for long sentences. It then converts this measure into an index. The best writing consistently scores below 20, equivalent to two style faults for every 100 words.

*Average Sentence Length:* The Average Sentence Length is an important statistic in writing style. Easy reading uses a sentence average below 20, sometimes lower depending on your audience and

your writing task. If you have a high sentence average, look for ways to cut out unnecessary words and phrases or split the long sentences into shorter sentences.

*Jargon Percentage:* This statistic shows the percentage of jargon words in the document. Jargon is a language of specialised terms used by a group of people or profession. It is common shorthand among experts and can be useful within these groups if used sensibly. However, if writers use jargon outside these groups, readers are unlikely to understand the message fully.

*Glue Words:* Glue words are the 200 or so most common words in the English language (excluding personal pronouns). They are necessary to link nouns, verbs, adverbs and adjectives in any sentence. Most writers use too many glue words and almost every document could benefit from running an editorial pen through unnecessary glue words.

*Stanton-Wright Reading Grade:* The Stanton-Wright Reading Grade is the most accurate formula yet developed to predict the reading grade level of a document. It combines average sentence length with *StyleWriter's* weighted readability score for more than 200,000 graded words, instead of simply counting syllables or word length as in other readability measures."

## References

1. Ipsen, A.H.; Dam, H.V. Translation revision: Correlating revision procedure and error detection. *Hermes J. Lang. Commun. Bus.* **2016**, *55*, 143–156. [CrossRef]
2. Prioux, R.; Rochard, M. Économie de la révision dans une organisation internationale: Le cas de l'OCDE'. *JoSTrans* **2007**, *8*, 21–41.
3. Mossop, B. Empirical studies of revision: What we know and need to know. *JoSTrans* **2007**, *8*, 5–20.
4. International Organization for Standardization. ISO 17100:2015—Translation Services—Requirements for Translation Services. 2015. Available online: <https://www.iso.org/standard/59149.html> (accessed on 20 March 2019).
5. Yang, J. Translation quality assessment: Exploring the importance of engaging "revision" in translation training (unpublished). In Proceedings of the didTRAD—PACTE: Fourth International Conference on Research into the Didactics of Translation, Barcelona, Spain, 20–22 June 2018.
6. Ciobanu, D. Of dragons and speech recognition wizards and apprentices. *Tradumàtica Technol. Traducció* **2014**, *12*, 524–538. [CrossRef]
7. Ciobanu, D. Automatic speech recognition in the professional translation process. *Transl. Spaces* **2016**, *5*, 124–144. [CrossRef]
8. ELIA; EMT; EUATC; FIT Europe; GALA; LIND. Language Industry Survey—Expectations and Concerns of the European Language Industry. 2018. Available online: [https://ec.europa.eu/info/sites/info/files/2017\\_language\\_industry\\_survey\\_report\\_en.pdf](https://ec.europa.eu/info/sites/info/files/2017_language_industry_survey_report_en.pdf) (accessed on 10 March 2019).
9. Scocchera, G. Translation revision as rereading: Different aspects of the translator's and reviser's approach to the revision process. *Mémoires Du Livre* **2017**, *9*, 1–21.
10. Brunette, L.; Gagnon, C.; Hine, J. The GREVIS Project: Revise or court calamity. *Across Lang. Cult.* **2005**, *6*, 29–45.
11. Wølch Rasmussen, K.; Schjoldager, A. Revising translations. A survey of revision policies in Danish translation companies. A survey of revision policies in Danish translation companies. *JoSTrans* **2011**, *15*, 87–120.
12. Huang, J. Working styles of student translators in self-revision, other-revision and post-editing. In *Eye Tracking and Multidisciplinary Studies on Translation*; Walker, C., Federici, F.M., Eds.; John Benjamins: London, UK, 2018; pp. 145–184.
13. Künzli, A. Teaching and learning translation revision: Some suggestions based on evidence from a think-aloud protocol study. In *Current Trends in Translation Teaching and Learning*; Garant, M., Ed.; Helsinki University: Helsinki, Finland, 2006; pp. 9–24.
14. Robert, I.S.; Van Waes, L. Selecting a translation revision procedure: Do common sense and statistics agree? *Perspectives* **2014**, *22*, 304–320. [CrossRef]
15. Matis, N. *Comment Gérer vos Projets de Traduction*; EdiPro: Liège, Belgium, 2010; ISBN 978-287-496-052-9.

16. Teixeira, C.S.C.; O'Brien, S. Overcoming methodological challenges of eye tracking in the translation workplace. In *Eye Tracking and Multidisciplinary Studies on Translation*; Walker, C., Federici, F.M., Eds.; John Benjamins: London, UK, 2018; pp. 33–54.
17. Cop, U.; Drieghe, D.; Duyck, W. Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE* **2015**, *10*, e0134008. [[CrossRef](#)] [[PubMed](#)]
18. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **1998**, *124*, 372–422. [[CrossRef](#)] [[PubMed](#)]
19. Rayner, K. Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **2009**, *62*, 1457–1506. [[CrossRef](#)] [[PubMed](#)]
20. Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; Oxford University Press: Oxford, UK, 2011; ISBN 978-019-873-859-6.
21. TAUS. Harmonized DQF-MQM Error Typology. Available online: <https://www.taus.net/evaluate/qt21-project#harmonized-error-typology> (accessed on 10 March 2019).
22. Jakobsen, A.L. Translation drafting by professional translators and by translation students. *Cph. Stud. Lang.* **2002**, *27*, 191–204.
23. Shih, C. Revision from translators' point of view: An interview study. *Target* **2006**, *18*, 295–312. [[CrossRef](#)]
24. Shih, C. Translators' 'end-revision' processing patterns and maxims: A think-aloud protocol study. *Arab World Engl. J. Awej* **2013**, *4*, 31–44.
25. Daems, J.; Carl, M.; Vandepitte, S.; Hartsuiker, R.; Macken, L. The effectiveness of consulting external resources during translation and postediting of general text types. In *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*, 1st ed.; Carl, M., Bangalore, S., Schaeffer, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 111–133. ISBN 978-3-319-20357-7.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).