

What effect does post-editing have on the translation product from an end-user's perspective?

Benjamin Screen, Cardiff University

ABSTRACT

This article details a triangulated eye-tracking experiment carried out at Cardiff University, UK. The experiment sought to compare the quality of final texts, from an end-user's perspective, when different translation modalities (translating and post-editing Machine Translation) were used to translate the same source text. The language pair investigated is English and Welsh. An eye tracker was used to record fixations in a between-groups experimental design as participants read two texts, one post-edited and one translated using the same source text, as well as two subjective Likert-type scales where each participant rated the texts for readability and comprehensibility. Following an analysis of fixation duration, the gaze data of the two groups was found to be statistically identical, and there was no statistically significant difference found between the readability and comprehensibility scores gleaned from subjective Likert-type scales. It is argued following this that post-editing machine translated texts does not necessarily lead to translations of inferior quality in the context of the final end-user, and that this further supports the use of Machine Translation in a professional context.

KEYWORDS

Eye tracking, translation quality, translation end-users, Welsh translation.

1. Introduction: translation in Wales

Language planning and legislation in Wales from the 1960s onwards, first sponsored by the UK Government and then mainly by a devolved Welsh Government after 1999, has proven to be the catalyst for the growth of what is now a relatively small, but extremely important, translation industry in Wales (Kaufmann 2010, 2012). Providing for an official language spoken daily by over 360,000 people (Welsh Government and the Welsh Language Commissioner 2015), translation in Wales was estimated by a 2009 report to be worth £45,000,000 per annum (Prys *et al.* 2009), of which translation between Welsh and English is a large part. However, this was stated in the context of a policy framework based on the Welsh Language Act of 1993 which is gradually being replaced by the provisions of the much stronger Welsh Language (Wales) Measure (Welsh Government 2011). This new 2011 legislation is widely expected to lead to further demand for English to Welsh translation (Richards 2013). Translation in Wales has, however, already grown significantly since the first professional Welsh translation bureau was established in 1966. Then, there were only a handful of translators working between the two languages. Today, there are over 500 (Jones 2015: 91). Such translation is also carried out by a growing number of language companies and their translators, who translate between English and Welsh, provide audio-visual services for S4C (the Welsh language broadcaster) and BBC Wales, and interpreting services from Welsh to

English for a range of clients. These clients include the courts and the education system, to name but two.

As a result of this increase in translation production, technology is recognized by policy actors and language planners in Wales as a development which could not only reduce the amount spent on translation, but could also render translators more efficient and consistent. Evidence of this recognition is provided by the advice note on translation and interpreting provided by the Welsh Language Commissioner (2012), as well as by policy documents from the Welsh Government (2012, 2014, 2017). For example, in relation to translation technology the language regulator advises that “Language technology can help to accelerate and facilitate the work of the translator” (Welsh Language Commissioner 2012: 6). In a 2014 policy document, “The reuse of translations, translation engines and automated translation for post-editing and quality control by humans, so that there can be greater prominence for Welsh” was noted as a priority by the Government (Welsh Government 2014: 11). However, these policy documents do not cite or discuss recent research that has analyzed the effect the use of these technologies may have on the final translated product. Whilst this is not usually the purpose of such documents, this is an extremely important issue in the context of minority languages in bilingual societies, where translation plays a much more important role in ensuring language rights than it does in monolingual societies (González 2005; Núñez 2016). This research regarding the interface between quality from the perspective of the end-user and post-editing will be discussed below, once post-editing has been defined.

2. Quality and machine translation post-editing

Post-editing is a complex linguistic process of editing a raw text that has been automatically produced by an MT system, usually with a minimum of manual labour to ensure the process is more efficient than human translation (Translation Automation Users Society (TAUS) 2010). Earlier definitions were provided by Wagner (1985: 1), namely “Post-editing entails the correction of a pre-translated text rather than translation from scratch”, and by Senz (1998: 289), “The term used for the correction of Machine Translation output by human linguists/editors”. Post-editing can be performed to varying degrees, depending on the target text’s context of use. Allen (2003) distinguishes between rapid post-editing and maximal or full post-editing. This distinction refers to the amount of editing work carried out, whereby light post-editing refers to correcting only the most major errors in language and translation and maximal to correcting the whole text (including style and register), so that it matches the quality of a human translation. TAUS (2016) differentiates between two levels of expected quality (“good enough” and “similar or equal to human translation”). This final distinction will also usually lead to more or less post-editing, depending on the system used. The quality of the system will also inevitably affect the

amount of post-editing. Linguistic characteristics of the input which cause erroneous output (called Negative Translatability Indicators by O'Brien (2006)) and the type of system used (rule-based, statistical, neural or hybrid) will influence the amount and type of issues post-editors have to deal with, as will the text type (Polvsen *et al.* 1998) and language combination.

The most relevant form of post-editing in the context of professional translation is “maximal” or “full” post-editing, whereby the automatic translation is corrected and improved to match the quality achieved by human translation. It is this type of post-editing that has been compared with human translation without machine assistance in a number of studies. This is also the type relevant to the study reported here. Table 1 below lists relevant studies that have analysed potential productivity gains when fully post-editing MT. Studies that have recruited non-professionals (i.e. students) are not included, although MT post-editing has been shown to speed up the translation processes of non-professionals also (Koehn 2009; Castilho *et al.* 2011; Lee and Liao 2011; Garcia 2011; Yamada 2011; Daems *et al.* 2013; Vázquez *et al.* 2013; Läubli *et al.* 2013; Depraetere *et al.* 2014). In the table, rules-based architectures are abbreviated to RBMT and statistical MT systems to SMT. Decreases in time are in standard font, and increases in throughput or words are in italics. If the average difference between the post-editing and “from scratch” translation conditions across participants and languages investigated was not noted, N/A is given. The number of translators providing data for each language, if noted in the original publication, is provided in brackets in the second column. No relevant studies comparing productivity between translating and the post-editing of Neural MT output could be found. All results are different due to the use of a wide range of systems with differing architectures and the large number of language pairs studied. However, there now appears to be a large body of research which shows that post-editing MT can not only reduce cognitive and typing effort compared to manual translation (O'Brien 2006; Carl *et al.* 2015; Koglin 2015; Screen 2016), but also significantly improve the productivity of professionals.

Author(s)	Target Language (English as Source Language)	Average Saving (time/words)	System Architecture
Lange and Bennett (2000)	German	N/A	RBMT
O'Brien (2007)	German (12)	3.96	SMT
Offersgaard <i>et al.</i> (2008)	Danish	67%	SMT
Groves and Schmidtke (2009)	French	14.5%	SMT + RBMT
	Brazilian Portuguese	20%	
	Swedish	8%	
	Danish	26.6%	
	Czech	6.1%	
	Dutch	14.7%	
	Chinese	5.9%	

Author(s)	Target Language (English as Source Language)	Average Saving (time/words)	System Architecture
	German	16%	
Guerberof (2009)	Spanish (9)	25%	SMT
Guerberof (2012)	Spanish (24)	37%	SMT
Flournoy and Duran (2009)	Spanish and Russian	40-45%	SMT
Plitt and Masselot (2010)	French (1)	20-131%	SMT
	Italian (1)		
	German (1)		
	Spanish (1)		
Carl <i>et al.</i> (2011)	Danish	17 secs	SMT
Skadins <i>et al.</i> (2011)	Latvian (1)	181	SMT
Green <i>et al.</i> (2013)	French (1)	N/A	SMT
	Arabic (1)		
	German (1)		
Aranberri <i>et al.</i> (2014)	Basque (Text 1) (12)	28.8%	SMT
	Basque (Text 2) (12)	6.06%	
Elming <i>et al.</i> (2014)	German (5)	25%	SMT
Moran <i>et al.</i> (2014)	Spanish (2)	54%	SMT
	French (2)		
	German (2)		
Silva (2014)	Spanish	86.90%	SMT
Uswak (2014)	German (20)	3.26	RBMT
Zhechev (2014)	French (4)	92.33%	SMT
	Korean (4)	81.93%	
	Italian (4)	63.60%	
	Brazilian Portuguese (4)	63.23%	
	Spanish (4)	59.43%	
	Japanese (4)	59.07%	
	Chinese (4)	58.02%	
	German (4)	44.94%	
Polish (4)	37.13%		
Bowker and Ciro (2015)	Spanish (3)	30%	SMT
Carl <i>et al.</i> (2015)	German	N/A	SMT
Screen (2017b)	Welsh (8)	60	SMT

Table 1. Relevant literature regarding MT and Productivity.

But what effect does this post-editing process have on the final product compared to the process of translation without MT? Before reviewing relevant research about the relationship between quality and post-editing, quality as a concept must be operationalized. In doing this, it must be borne in mind that quality in translation is a multifaceted concept with different meanings to different actors in the translation industry. For an end-user who needs a rough translation of a text in a language he or she doesn't understand, a roughly post-edited machine translation may suffice (e.g. Bowker 2009; Bowker and Ciro 2015). If, however, a professional translation is required, a professional translator is essential whether MT is used or not. Given that translation for the public is the main interest of this article, that is the translation of texts for the public following Welsh

language legislation, Gouadec's 2010 definition of quality is drawn upon here. Gouadec's definition differentiates between intrinsic and extrinsic translation quality. He emphasises three main elements of successful translation, with the first two being intrinsic quality and the final being extrinsic:

1. A target text that is 'faithful' to the original source text, i.e. it contains the same information as the original text¹;
2. A target text that is grammatically correct and appropriate in terms of register and style;
3. A target text that meets the needs of the target audience in that it is acceptable to its end-user in terms of readability and comprehensibility².

Translation quality, then, in the context of professional translation for the public, within a framework of societal bilingualism, may be operationalized as a text that is grammatically and stylistically correct as well as faithful to the original, whilst taking into account the usability of that text in terms of its end user. This means that the text must also be easy to read and comprehend³. It is this final requirement that will be the focus of the investigation into quality in this article. The effect that post-editing has on the intrinsic quality of a text, as compared to manual translation or translation without MT, has already received attention from researchers working in a variety of language pairs. For example, Guerberof (2009, 2012), Plitt and Masselot (2010), Carl *et al.* (2011), García (2011), Skadiņš *et al.* (2011), Daems *et al.* (2013), Läubli *et al.* (2013), Vázquez *et al.* (2013) and Screen (2017a,b) all investigate the relationship between MT post-editing (and editing Translation Memory matches in the case of Guerberof (2009, 2012)), human translation and quality. Guerberof (2009, 2012), Vázquez *et al.* (2013), Läubli *et al.* (2013) and O'Curran (2014) operationalize quality based on the LISA Quality Assessment framework. Similar methodologies based on counting errors according to predefined linguistic categories were used by Plitt and Masselot (2010), Skadiņš *et al.* (2011) and Daems *et al.* (2013). Fiederer and O'Brien (2009) and Screen (2017a,b) used a methodology based on translation accuracy, grammaticality and style on a 1-4 scale. García (2011) used the national translator accreditation framework. Finally, Carl *et al.* (2011) asked reviewers to rate their favourite or preferred translation of the same source text. Skadiņš *et al.* (2011) was the only analysis to find that MT detrimentally affected translation quality in terms of grammaticality and translation accuracy. Fiederer and O'Brien (2009) were the only ones who found that post-edited texts were rated lower for style.

The effect that post-editing has on *extrinsic* quality, however, especially using objective methodologies, has not yet received its due share of academic attention. The relatively few studies that *have* considered the end-user's opinion, rather than solely that of a professional translation

reviewer, will be discussed below as a background to the study carried out here. The reason why this aspect of quality is important, however, is first of all outlined.

2.1. Quality from the end-user's perspective

The studies mentioned above, with the exception of Skadiņš *et al.* (2011) and Fiederer and O'Brien (2009) in relation to style, show that the use of full post-editing in the translation process does not necessarily lead to a final translation of poorer quality compared with a text translated from scratch by human translators. In fact, post-editing may improve quality from a bilingual review perspective (Screen 2017b). The opinion of a qualified reviewer when analysing translation quality, i.e. a translator with considerable experience, is but one side of the coin however. The opinions of those who *use* translations when it comes to comparative quality are arguably equally as important. Few who understand the complexity of translation would deny that only qualified and experienced professionals should *review* translations in a professional context, as noted by Hansen (2009), but that should not mean that the end-users of these translations are ignored. The current paradigm in the analysis of translation quality according to Drugan (2013: 179), despite this, is "active translation agents and passive or unknowable translation recipients", rather than active translation agents (translators and reviewers) and active and known translation recipients (or end-users whose needs and characteristics are understood and respected). In terms of "passive or unknowable translation recipients", research in Wales regarding the use of Welsh language services and public satisfaction with them has allowed Welsh speaking citizens to voice their opinion about the quality of statutory translation, thereby giving researchers an insight into how translation is received in the Welsh speaking community. Work by the Citizens Advice Bureau (2015) and the Welsh Language Commissioner (2015) has shown how little attention has been paid to the type of translation citizens in Wales need and expect. This is neatly summed up in this response to one of the researchers regarding translation (Citizen's Advice Bureau: 53), "[...] and what I see is that when they [service providers] translate material they overcomplicate things, and choose very formal language"⁴. This was also discussed in recent research carried out on behalf of the Welsh Language Commissioner; a significant proportion of those who negatively rated Welsh language services offered by local authorities singled out poor translation, in particular Welsh which was too formal (Beaufort Research 2015: 20). What this shows then is that hitherto the needs of the Welsh speaking community have not been as central as they should be when considering whether or not a translation is fit for purpose, given that two separate pieces of research have shown dissatisfaction towards the quality of statutory translation provision.

The final quality of the translation takes on a further requirement in the case of Welsh, however, and possibly in the case of other bilingual language

communities. The translation must not only be correct, but also clear and written in such a way that all members of the community are able to understand it, members who are arguably on a broad spectrum from 'Welsh dominant' to 'balanced bilingual' to 'English dominant'⁵. Given the importance of good quality translation to the provision of Welsh language services, and the numbers of people who use them, as well as the drive from Government to technologize Welsh translation further, final quality of texts under different translation modalities from the standpoint of the *end-user* is an important and under-researched issue. Even where quality from the perspective of the end-user has been researched for other languages, an analysis of English to Welsh translation is yet to be undertaken. Before describing the current experiment, relevant literature that has analysed the comparative quality of translated and post-edited texts from the perspective of the *end-user* will be reviewed.

Bowker (2009) carried out research to ascertain whether the English and French linguistic communities in Canada would accept post-edited, as opposed to translated, texts. A mere 22.2% favoured the post-edited text rather than the translated one, whereas 37.8% of the English sample in Quebec preferred the post-edited text over the translated one. Bowker and Ehgoetz (2007) performed a similar experiment at Ottawa University, where the participants were required to choose which translation they preferred (the post-edited or the translated text). English to French was the directionality investigated. A large 67.7% of the sample said they would accept a post-edited text, whereas 32.2% said they would not. As well as comparing translated and post-edited texts for accuracy, clarity and style, all participants in Fiederer and O'Brien's 2009 study cited above were asked to choose their most preferred translation also. In terms of favourite or preferred translations, 63% of participants chose the translated texts as opposed to 37% who chose the post-edited texts. Bowker and Ciro (2015) carried out a recipient evaluation of four translations produced using either unedited Google Translate output, a rapid post-edit of this output, a full (maximal) post-edit of it or a human translation. The language directionality investigated was English to Spanish, and the study was predicated on the translation needs of the immigrant Spanish-speaking community in Ottawa. Bowker and Ciro wished to discover whether this community would be prepared to accept translations that were produced using an MT system (unedited or post-edited to various degrees) rather than a human translation, citing the cost of professional human translation in Canada. The authors also asked whether these preferences would change if the participants knew the cost of the production method and how long it took. Bowker and Ciro found that most participants actually preferred the human translation over the raw MT or lightly post-edited versions of all three texts when method, cost and production time were unknown, with fully post-edited texts coming second. However, when this metadata were revealed the picture changes. A majority of participants actually selected the rapidly post-edited texts, with the fully post-edited versions coming second. This

difference is a result of the fact that most respondents either wanted a Spanish translation to process information more quickly, to confirm they understood something or because of limited proficiency. For these purposes, “An elegant text is not required” (Bowker and Ciro 2015: 181). What this shows then is that post-editing, even when stylistic changes are ignored, can in fact provide for the needs of the end-user and provide value for money. A final study to consider is Castilho and O’Brien (2016), who compared the usability of a source text compared to its lightly post-edited and raw machine translated versions, using the definition of usability provided by standard ISO/TR 16982. Using this standard, usability is operationalized as a product that can be used effectively, efficiently and with satisfaction. Fixation duration, fixation count, task time, visit duration (sum of visit length in an area of interest divided by the total number of visits), number of successfully completed goals and results of a Likert-type scale to measure satisfaction were all utilized as dependent variables. No statistically significant differences were found in eye movement data between lightly post-edited and raw machine translated texts, although task time and satisfaction measures showed the group who used the post-edited version to be faster and more efficient at the task they were asked to do. Significant differences were found on all measures between the source text and the post-edited version, whereby users of the source text used this text more effectively, efficiently and with greater satisfaction.

The five studies discussed above have asked what effect post-editing a text rather than translating it has on quality according to the perceptions of the end user. As argued above, with increasing technologisation and the importance of clear and competent translation in Welsh language planning and policy, this question needs to be further investigated. The framework within which this could be done however is an open question. Four of the five studies above have utilized a more subjective methodology whereby participants were asked to choose their preferred option or complete a questionnaire. The current study, however, has chosen to analyse the quality of texts from an end-user’s perspective within the framework of an eye-tracking methodology, by investigating what Gouadec (2010) called the ‘extrinsic’ quality of texts, i.e. their comparative readability and comprehensibility (defined below). As noted above, it is this aspect of translation quality that is yet to receive its due amount of attention, and research in this area has not taken sufficient advantage of more objective, quantitative methodologies, with Castilho and O’Brien (2016) the only study so far. This way of ‘seeing’ translation quality has recently been advocated by Suojanen, Koskinen and Tuominen (2015) with their concept of ‘user-centred translation’. Suojanen *et al.* (2015: 100) advocate eye tracking as an empirical research method that could shed light on how end users react to and cope with different types of translated texts. Whilst eye tracking has been previously used to measure translation *quality* in the context of raw MT Output (Doherty *et al.* 2010), and the effect of controlled language on readability and comprehensibility (Doherty 2012; O’Brien 2010), it has not

yet been used to measure the comparative quality of translated and post-edited texts from an end-user's perspective⁶. The theoretical underpinnings of this methodology will be described in the next section.

3. Methodology

3.1. Hypotheses

Eye tracking is a useful and popular research method, and has been used to date to investigate a number of research questions related to MT and post-editing, including cognitive effort when interacting with MT output (Doherty *et al.* 2010; Carl *et al.* 2015), the role of syntactic variation in translating and post-editing (Bangalore *et al.* 2015), cognitive effort in post-editing as compared to translating metaphors (Koglin 2015), as well as to compare manual translation processes with those of post-editing in terms of gaze behaviour, speed and quality (Carl *et al.* 2011). A recent volume edited by Silvia Hansen-Schirra and Sambor Grucza (2016) also contains two studies where eye tracking has been utilized to investigate post-editing. Nitzke (2016) uses eye tracking to measure cognitive effort when researching during monolingual post-editing and Alves *et al.* (2016) use the methodology within the framework of relevance theory, and collect gaze data to measure comparative cognitive effort between interactive post-editing and standard post-editing.

Analysing the comparative quality of translated and post-edited texts using eye tracking is a new application of this research method. The common underlying assumption of research using eye-tracking data in Translation Studies is the Eye-Mind Assumption (Just and Carpenter 1980), which posits that an object (such as a word) which is fixated on by the eye is currently being processed in working memory, and that the longer this fixation lasts, the more effortful the process is deemed to be. If the process is effortful, we can infer that the text has low readability and comprehensibility. A fixation is defined by Duchowski (2003: 43) as "eye movements which stabilize the retina over a stationary object of interest", and is considered to be a "numerosity measure" by Holmqvist *et al.* (2011), as opposed to a movement, position or latency measure⁷.

Following the research reviewed above in relation to translation quality when texts are post-edited as opposed to translated, as well as comparative quality from the end-user's perspective, the hypothesis in relation to the eye-tracking data is that there should be no difference between the gaze data of the group that read the translated version and the group that read the post-edited version. This should then provide further evidence that the implementation of MT into professional workflows is beneficial not only to the translator in terms of effort, productivity and quality, but also in terms of the actual *users* of post-edited texts. It is this aspect of translation quality, that of the cognitive and subjective reaction of those the text was

intended for rather than solely the translator, that requires further investigation.

Finally, this research question cannot only be measured by recording reading process data using an eye tracker. Triangulation, defined as using multiple research methods to measure the same construct, is useful in empirical research as it can be used to corroborate and support conclusions drawn using other means. As a result, the complexity with which the participants experienced the reading process, objectively recorded above in the first phase, was then probed by reducing this process to the twin concepts of readability and comprehensibility. Participants rated the text they read in their group (either the translated or post-edited text) for readability and comprehensibility on a scale of 1 to 5. Based on the operationalisation of the concepts by Suojanen *et al.* (2015), readability was defined as the ease of *reading* a text, namely how easy or difficult it was for the text reader to process it as a result of grammar, structure, spelling and punctuation, and comprehensibility was defined as the ease with which the participants actually *understood* the message of the text and what the text was about. Although some scholars treat these two concepts as one single conceptual entity, others consider them separately (Doherty 2012: 22). Doherty (2012) is followed here who has analysed readability and comprehensibility apart, and as Suojanen *et al.* (2015: 53) remind us, readability is a feature of the text, while comprehensibility is an interpersonal concept which depends on individual factors such as subject knowledge, motivation to read, intelligence, working memory capacity as well as other factors. The hypothesis that flows from this is that there should be no statistical difference between the Likert-type scale scores for readability and comprehensibility, as a measure of the perception of quality between the translated and post-edited texts.

3.2. Experimental design

A between-groups independent samples design was utilized whereby 6 fluent L1 speakers of Welsh read a text translated by a professional translator and another 6 read a post-edited translation of the same source text. Whilst they read, the participant's gaze behaviour was recorded by a Gazepoint GP3 fixed-position 60Hz remote eye tracker. This remote eye tracker is arguably more ecologically valid than a head-mounted one (O'Brien 2009: 252), as participants read the text on the screen as they normally would. Following the pressing of 'ESC' to end the eye-tracking task, all participants were asked to note their beliefs regarding the readability and comprehensibility of the translation they read. Following successful uses of a 5-point Likert-type scale for measuring perceptions of readability and comprehensibility in translation research by Chang (2011) and Doherty (2012) with no issues reported, this study also used this method in triangulation. A score of 1 on the readability scale meant that the text was deemed unreadable, and a score of 1 on the comprehensibility

scale meant the participant believed that they did not comprehend anything in the text. On the other hand, 5 on the readability scale meant the text was deemed to have perfect readability and 5 on the comprehensibility scale meant that everything was fully comprehended. Both readability and comprehensibility were defined on the print-out of the Likert-type scales (Appendix 1), and participants were invited to read the text again if they so wished.

3.3. Experimental set-up

Prior to the commencement of the experiment, each participant was asked to select a number between 1 and 12. A random number generator was then used to randomly assign the range of numbers into two groups and to randomly order each number within those groups. Each participant was then called according to the list provided. The eye tracker was attached to a 17.5-inch LCD monitor with a non-swivel chair without wheels placed in front of it. This was done to dissuade participants from moving their head and altering their body position, thus affecting the tracker's ability to collect data accurately.

3.4. Participants

All participants (7 female and 5 male) were fluent L1 Welsh speakers enrolled at Cardiff University. Prior to the commencement of the experiment and after signing the research ethics form, all participants were required to read a print-out of a short paragraph in Welsh from the news section on the university website posted over a year previously by a different school. They were then asked to note their beliefs about the readability and comprehensibility of this text, with 1 being the lowest score and 5 the highest. This was done in order to screen participants, as differing levels of ability to read Welsh could have been a confounding variable. The average score for readability across all 12 participants was 4.5, and 4.5 was also the score across all for comprehensibility. As all participants scored highly, none were eliminated from the study. The standard deviation was 0.7 for readability and 0.5 for comprehensibility. There was also a positive correlation between the readability scores of the two groups ($r_s(10) = 1.000, p = .014$) and their comprehensibility scores ($r_s(10) = 1.000, p = .001$), according to Spearman's Rank Order correlation test. Given that participants scored highly and that there was little variation and a high correlation between scores provided by the participants in the two groups, ability to read Welsh was not considered a confounding variable.

3.5. Text selection and description

The source text was about a public-sector organisation's policy on absence from the office. As most translation in Wales is for the public sector, it was deemed important that a text from this sector should be used. The human

translation produced without the aid of any translation technology contained 155 words and 745 characters, whilst the post-edited translation of the same source text contained 144 words and 669 characters. The text was kept short to avoid participant fatigue and to avoid scrolling. As O'Brien (2009: 261) reminds researchers, these are also important factors when considering research validity. O'Brien (2010: 151) also notes that asking readers to comprehend a whole piece of text, sentence by sentence rather than giving them the whole text, can, in fact, hinder proper comprehension. As such, participants could see the whole text and it formed one short paragraph. Google Translate was the MT system chosen to produce the text for the experimental group. No participant knew how the text was produced, i.e. via translating or post-editing. An experienced professional translator translated the source text and another professional translator with similar experience post-edited the raw machine translation. This translator was asked to refrain from making any unnecessary changes, but to ensure the text was grammatically correct, flowed well and preserved all the meaning of the source text. The following readability indices scored the source text as follows: Flesch Reading Ease score (Flesch 1948) - 73.7, Gunning Fog (Gunning 1952) - 9.3, LIX (Björnson 1968) - 38, the SMOG index (McLaughlin 1969) - 6.6. Given these scores, it can be maintained that the text was fairly easy to read and represented a text of average complexity that one would see being used day-to-day⁸. The text was displayed using the Calibri font and at font size 16 with 1.5 spacing between lines. The background was white with black font. All participants confirmed they had not previously seen the text when asked. This was checked as previous knowledge of a text aids subsequent comprehension (Ericson and Kintsch 1995). All were asked to read silently, i.e. the advice of O'Brien (2010: 153) was heeded, as reading aloud is known to affect processing time.

3.6. Dependent variables

The dependent variables under investigation were fixation duration, readability score and comprehensibility score. All but the Likert-type scale scores can be automatically calculated in the spreadsheet provided by the data analysis software attached to the eye tracker (cf. Experimental Set-up). The software used for the analysis was Gazepoint Analysis UX Edition. The Likert-type scale scores were transferred from paper to IBM SPSS once data collection was concluded, thus enabling statistical analysis.

3.7. Data quality

Precautions were taken to ensure the quality of data. Lighting, noise and interruptions are known to affect gaze data (O'Brien 2009: 253). Lighting was kept constant as was the position of the screen, the eye tracker and the table upon which the hardware was placed. The room containing the tracker was quiet and interruptions prevented. Participants were asked to keep as still as possible following successful calibration and it was not

possible to move the chair unless the researcher asked the participant to do so. As well as this, all fixations under 175 milliseconds were manually discarded from data set for each participant, similar to Jensen (2011) who discarded all mean fixations under 175ms. This was done as extremely short fixations are considered unlikely to be related to reading processes (Rayner 1998).

4. Results

4.1. Eye-tracking results

Considering that intrinsic quality appears to be unaffected by full post-editing, and that the end-user also *perceives* the quality of translated and post-edited texts to be identical or just as useful for their own purposes (see Section 2), it was hypothesized that there would be no difference between the gaze data of reading processes when reading professionally translated and post-edited versions of the same source text.

This hypothesis has been confirmed by the data; according to the 'Two One Sided T Test' (TOST) procedure, which tests for equivalence, the gaze from both groups can be considered identical ($p = 0.046$), and a two-sided T Test showed that any difference between the two groups was not significant ($t(10) = .631$, $p = 0.552$). The setting of bounds in the TOST procedure (i.e. the maximum level of difference between two groups that would still render them equivalent) entails a degree of arbitrariness and, since this procedure has not been used frequently in Translation Studies, there is no established way of setting these bounds. However, a lower bound of -50ms and an upper bound of 50ms were used for this procedure, which seemed like a sufficiently narrow difference. The control group mean was 278.5ms and the median 277.5ms, the experimental group mean was 290ms and median 299.5ms. The TOST results are shown in Table 2.

Test	Difference	t	t(Critical Value)	DF	p-value
Upper	-12.167	1.963	1.906	6.725	0.046
Lower	-12.167	-3.225	-1.906	6.725	0.008
				6.725	0.046

Table 2. Results of the TOST procedure with a lower equivalence limit of -50ms and an upper equivalence limit of 50ms.

Full post-editing, when compared to translating, did not alter the reading processes of the participants who took part in this study. In terms of the effect post-editing has on the end-user then, it cannot be said in this case to have detrimentally impacted on the user's experience of reading the translation. Both groups were equally as able to read, and therefore use the translation, regardless of translation modality. This is shown below in Figure

1. The participants’ own subjective perceptions will be discussed in Section 4.2.

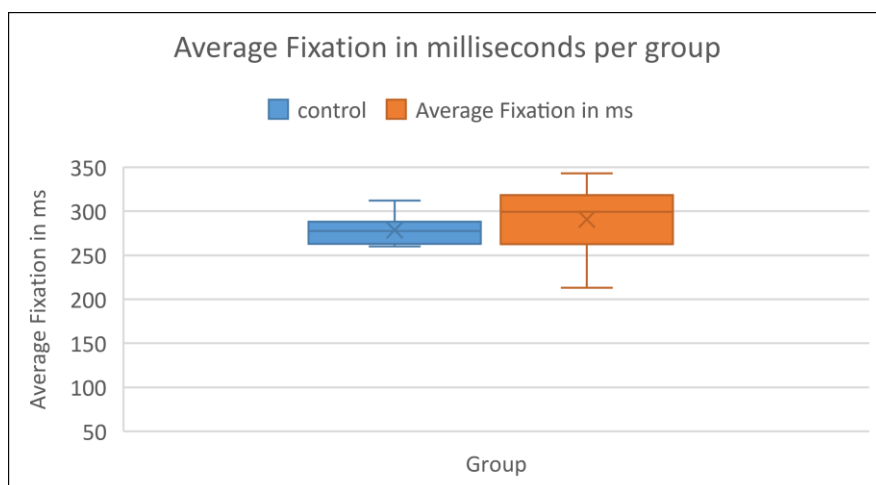


Figure 1: Average Fixation in Milliseconds by Group.

4.2. Likert-type scale results

Summary of Likert-type scale results per participant			
Readability		Comprehensibility	
Control Group	Experimental Group	Control Group	Experimental Group
3	5	5	4
5	5	5	3
4	4	4	3
5	4	5	3
4	5	5	5
4	5	5	5
(Mean= 4)	(Mean= 5)	(Mean= 5)	Mean= 4)
(Median= 4)	(Median= 5)	(Median= 5)	(Median= 3.5)

Table 3. Summary of Likert-type scale results per participant.

For readability, no statistical difference could be found according to a Mann-Whitney test (used as the data was ordinal and from independent samples): $U = 11, z = -1.04083, p = > .05$. The central tendencies of the scores were quite high and also almost identical (control group mean= 4, median= 4; experimental group mean= 4, median= 5), lending support to the conclusion that the scores of the two groups are extremely similar. For comprehensibility, a similar conclusion can be drawn. No statistical difference was found between the datasets according to the same test ($U = 9, z = 1.36109, p = > .05$). Again, the central tendencies of the scores lend support to the conclusion that there is no real difference between the scores given by both groups for comprehensibility (control group mean= 5, median= 5; experimental group mean= 4, median= 3.5). Given these

results, it appears as though both groups perceived the readability and comprehensibility of the texts in the same way. In other words, post-editing did not cause the end-user to perceive the final product as any less readable and understandable than the translated version. A summary of the Likert-type scale results is presented in Table 3.

5. Conclusion

The purpose of this article was to investigate the comparative quality of translated and post-edited texts from the perspective of the end-users, namely the Welsh-speaking community. A brief discussion of the context of translation in Wales and the policies and attitudes towards the place of technology when translating between English and Welsh was first of all provided. Following a review of the literature surrounding quality, which was defined beforehand following Gouadec's 2010 operationalisation, it was argued that research in this area has not yet taken advantage of more objective methodologies. In this regard, eye tracking was identified as a useful methodology. This follows work by Suojanen *et al.* (2015) where the authors suggested that translation quality could be understood in terms of text usability and that eye tracking could be taken advantage of. This study has put that to the test.

It was found that statistically the fixation duration of both groups was identical according to the TOST procedure. This was interpreted as suggesting that the reading processes, and therefore reading experience, of the end-users was not negatively affected by full post-editing as opposed to translation and that both groups were able to read and comprehend the translation *regardless of modality*. In terms of the Likert-type scales and subjective perception of readability and comprehensibility, no differences between the subjective responses given regarding readability and comprehensibility were found.

Following these results, it can be argued that the use of MT followed by full post-editing to translate a standard text from the public sector did *not* negatively affect the reading processes of the participants recruited for this study. In terms of their subjective perceptions regarding readability and comprehensibility, it does not appear that MT post-editing affected this aspect in a *negative* way either, given that there appeared to be no difference between the scores given by both groups. The implications of this for the translation industry is that post-edited texts, given these results, are perceived by end users to be just as readable and comprehensible as translated ones. This adds further justification for the use of MT in professional workflows, as the use of MT post-editing not only speeds up translation and leads to quality texts according to translation reviewers, but it also appears that post-edited texts are received just as well as translated ones.

6. Limitations

It is accepted the sample size was small, and future work will address this issue by recruiting a larger sample. This could have led to non-significant results, but the fact that the mean and median of the fixation data in both groups are actually quite similar does lead one to suspect that the result is sound. The text used was also small, but it should be borne in mind that a smaller text was used so that it would fit the screen without participants having to scroll down. Future work will also incorporate actual measures of comprehension, such as Cloze tests, rather than perceived comprehension. It should be noted however that the effect on the end user was the crux of the present study, rather than solely comprehension, and so the subjective perception of subjects is relevant in this case.

References

- **Allen, Jeffery** (2003). "Post-editing: An Integrated Part of a Translation Software Program." *Language International* 13(2), 26-29.
- **Alves, Fabio, Szpak, Karina, Gonçalves, José, Sekino, Kyoko, Aquino, Marcell, Araújo e Castro, Rodrigo, Koglin, Arlene, de Lima Fonseca, Norma and Bartolomé Mesa-Lao** (2016). "Investigating cognitive effort in post-editing: A relevance-theoretical approach." Silvia Hansen-Schirra and Sambor Grucza (eds) (2016). *Eye-tracking and Applied Linguistics*. Berlin: Language Science Press, 109-143.
- **Aranberri, Nora, Labaka, Gorka, de Ilarraza, Arantza and Kepa Sarasola** (2014). "Comparison of post-editing productivity between professional translators and lay users". Sharon O'Brien, Michel Simard and Lucia Specia (eds) (2014). *Proceedings of the Third Workshop on Post-Editing Technology and Practice*. Association for Machine Translation in the Americas, 20-33. <https://pdfs.semanticscholar.org/af49/45fb16cfb1235847b49699b0f0b98e2bba06.pdf> (consulted 05.12.2018).
- **Bangalore, Srinivas, Behrens, Bergljot, Carl, Michael, Ghankot, Maheshwar, Heilmann, Arndt, Nitzke, Jean, Schaeffer, Moritz and Annegret Sturm** (2015). "The role of syntactic variation in translation and post-editing." *Translation Spaces* 4, 119-143.
- **Beaufort Research** (2015). *Research Report: Local Authority Welsh Language Services*. Cardiff: Beaufort Research.
- **Björnson, Carl Hugo** (1968). *Läsbarhet*. Stockholm: Bokförlaget Liber.
- **Bowker, Lynne** (2009). "Can Machine Translation meet the needs of official language minority communities in Canada? A recipient evaluation." *Linguistica Antverpiensia* 8, 123-155.
- **Bowker, Lynne and Jairo Buitrago Ciro** (2015). "Investigating the usefulness of machine translation for newcomers at the public library." *Translation and Interpreting Studies* 10, 165-186.
- **Bowker, Lynne and Melissa Ehgoetz** (2007). "Exploring user acceptance of machine translation output: A recipient evaluation." Dorothy Kenny and Kyongjoo Ryou (eds)

(2007). *Across boundaries: international perspectives on translation*. Newcastle-upon-Tyne: Cambridge Scholars Publishing, 209-224.

- **Carl, Michael, Dragsted, Barbara, Elming, Jakob, Hardt, Daniel and Arnt Lykke Jakobsen** (2011). "The Process of Post-Editing: A pilot study." Bernadette Sharp, Michael Zock, Michael Carl, Arnt Lykke Jakobsen (eds) (2011). *Proceedings of the 8th International NLPCS Workshop. Special Theme: Human-Machine Interaction in Translation*. Copenhagen: Samfundslitteratur, 131-142. <http://www.l2f.inesc-id.pt/~fmmb/wiki/uploads/Work/dict.ref7.pdf> (consulted 05.12.2018).
- **Carl, Michael, Gutermuth, Silke and Silvia Hansen-Schirra** (2015). "Post-editing Machine Translation: Efficiency, strategies and revision processes in professional translation settings". Aline Ferreira and John Schwieter (eds) (2015). *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 145-174.
- **Castilho, Sheila and Sharon O'Brien** (2016). "Evaluating the impact of light post-editing on usability". Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds) (2016). *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association, 310-316. http://www.lrec-conf.org/proceedings/lrec2016/pdf/539_Paper.pdf (consulted 05.12.2018).
- **Chang, Vincent** (2011). "Translation Directionality and Revised Hierarchical Model: An Eye-tracking study." Sharon O'Brien (ed.) (2011). *Cognitive Explorations of Translation*. London: Continuum, 154-174.
- **Citizens Advice Bureau** (2015). *English by default: Understanding the use and non-use of Welsh language services*. Cardiff: Citizens Advice Bureau.
- **Daems, Joke, Macken, Lieve and Sonia Vandepitte** (2013). "Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE." Sharon O'Brien, Michel Simard, and Lucia Specia (eds) (2013). *Proceedings of the 2nd Workshop on Post-editing Technology and Practice (WPTP-2), Nice, France, September 2*. European Association for Machine Translation, 63-71. <http://www.mt-archive.info/10/MTS-2013-W2-TOC.htm> (consulted 01.12.2018).
- **Castilho, Sheila, Aziz, Wilker and Lucia Specia** (2011). "Assessing the post-Editing effort for automatic and semi-automatic translations of DVD subtitles." Galia Angelova, Kalina Bontcheva, Ruslan Mitkov and Nikolai Nikolov (eds) (2011). *Proceedings of Recent Advances in Natural Language Processing*. Shoumen, 97-103. <http://www.aclweb.org/anthology/R11-1014> (consulted 05.12.2018).
- **Depraetere, Ilse, De Sutter, Nathalie and Arda Tezcan** (2014). "Post-Edited quality, post-editing behaviour and human evaluation: A case study." Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds). (2014). *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing: Newcastle, 78-109.
- **Doherty, Stephen** (2012). *Investigating the Effects of Controlled Language on the Reading and Comprehension of Machine Translated Texts*. PhD Thesis. Dublin City University.

- **Doherty, Stephen, O'Brien, Sharon and Michael Carl** (2010). "Eye-tracking as an MT evaluation technique." *Machine Translation* 24: 1-13.
- **Drieghe, Denis** (2011). "Parafoveal-on-foveal effects on eye movements during reading." Simon P. Liversedge, Iain Gilchrist and Stefan Everling (eds) (2011). *The Oxford Handbook of Eye-movements*. Oxford: Oxford University Press, 839-855.
- **Drugan, Joanna** (2013). *Quality in Professional Translation: Assessment and Improvement*. London/New York: Bloomsbury.
- **Duchowski, Andrew** (2003). *Eye-tracking Methodology – Theory and Practice*. Springer: London.
- **Elming, Jakob, Balling, Laura and Michael Carl** (2014). "Investigating user behaviour in post-editing and translation using the CASMACAT workbench." Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds). (2014). *Post-editing Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, 147-169.
- **Ericson, Anders and Kintsch, Walters** (1995). "Long-term Working Memory." *Psychological Review* 102(2), 211-245
- **Fiederer, Rebecca and Sharon O'Brien** (2009). "Quality and Machine Translation: A realistic objective?" *The Journal of Specialised Translation* 11, 52-74.
- **Flesch, Rudolf** (1948). "A new readability yardstick." *Journal of Applied Psychology* 32, 221-233.
- **Flournoy, Raymond and Christine Duran** (2009). "Machine Translation and document localization at Adobe: From pilot to production." *MT Summit XII The twelfth Machine Translation Summit*. Association for Machine Translation in the Americas. <http://www.mt-archive.info/MTS-2009-Flournoy.pdf> (consulted 05.12.2018).
- **García, Ignacio** (2011). "Translating by post-editing: Is it the way forward?" *Machine Translation* 25(3), 217-237.
- **González, Marta** (2005). "Translation of minority languages in bilingual and multilingual communities". Albert Branchadell and Lovell Margaret West (eds) (2005). *Less Translated Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 105-125.
- **Gouadec, Daniel** (2010). "Quality in translation." Yves Gambier and Luc Van Doorslaer (eds) (2010). *Handbook of Translation Studies, Volume 1*. Amsterdam/Philadelphia: John Benjamins' Publishing Company, 270-275.
- **Green, Spence, Heer, Jeffrey and Christopher D. Manning** (2013). "The efficacy of human post-Editing for language translation." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (ACM)*. Association for Computing Machinery, 439-448. <https://dl.acm.org/citation.cfm?id=2470718> (consulted 21.11.2018).
- **Groves, Declan and Dag Schmidtke** (2009). "Identification and analysis of post-editing patterns for MT." *MT Summit XII – The twelfth Machine Translation Summit International Association for Machine Translation hosted by the Association for Machine Translation in the Americas*. Association for Machine Translation in the Americas, 429-436. <http://www.mt-archive.info/MTS-2009-Groves.pdf> (consulted 21.11.2018).

- **Guerberof, Ana** (2009). "Productivity and quality in the post-editing of outputs from translation memories and Machine Translation." *International Journal of Localization* 7, 11-21.
- — (2012). *Productivity and quality in the post-editing of outputs from Translation Memories and Machine Translation*. PhD Thesis. Universitat Rovira i Virgili.
- **Gunning, Robert** (1952). *The Technique of Clear Writing*. McGraw-Hill: New York.
- **Hansen-Schirra, Silvia and Sambor Grucza** (2016). *Eye-tracking and Applied Linguistics*. Berlin: Language Science Press.
- **Holmqvist, Kenneth, Nystrom, Marcus, Andersson, Richard, Dewhurst, Richard, Jarodzka, Halska and Joost Van Der Weijer** (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.
- **ISO/TR 16982** (1979). *Ergonomics of human-system interaction – Usability methods supporting human centered design*. Geneva, Switzerland: International Organization for Standardization.
- **Jensen, Kristian Tangsgaard Hvelplund** (2011). "Distribution of Attention Between Source Text and Target Text During Translation". Sharon O'Brien (ed.). *Cognitive Explorations of Translation*. Continuum: London, 215-238.
- **Jones, Sylvia Prys** (2015). "Theori ac ymarfer cyfieithu yng Nghymru heddiw". Delyth Prys and Robot Trefor (eds.) (2015). *Ysgrifau a Chanllawiau Cyfieithu*. Caerfyrddin: Coleg Cymraeg Cenedlaethol, 91-105.
- **Just, Marcel Adam and Patricia Carpenter** (1980). "A theory of reading: From eye fixations to comprehension." *Psychological Review* 87, 329-354.
- **Kaufmann, Judith** (2010). "Cyfieithu a Pholisi iaith." *Contemporary Wales* 23, 171-183.
- — (2012). "The Darkened glass of bilingualism? Translation and interpreting in Welsh language planning." *Translation Studies* 5, 327-344.
- **Kennison, Sheila and Charles Clifton** (1995). "Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control." *Journal of Experimental Psychology* 21, 68-81.
- **Koehn, Phillip** (2009). "A process study of computer-aided translation." *Machine Translation* 2, 241-263.
- **Koglin, Arlene** (2015). "An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors." *Translation & Interpreting* 7, 126-14.
- **Lange, Carmen Andres and Scott Winfield Bennett** (2000). "Combining Machine Translation with Translation Memory at Baan". Robert C. Sprung (ed.) (2013). *Translating into Success: Cutting-edge strategies for going multilingual in a global age*. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 203-219.
- **Läubli, Samuel, Fishel, Mark, Massey, Gary, Ehrensberger-Dow, Maureen and Martin Volk** (2013). "Assessing post-editing efficiency in a realistic translation environment." Sharon O'Brien, Michel Simard and Lucia Specia (eds) (2013).

Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice, Nice, September 2, 2013. European Association for Machine Translation, 83-91. <http://www.mt-archive.info/10/MTS-2013-W4-Laubli.pdf> (consulted 05.12.2018).

- **Lee, Jason and Posen Liao** (2011). "A comparative study of human translation and Machine Translation with post-editing." *Compilation and Translation Review* 4, 104-149.
- **McLaughlin, G. Harry** (1969). "SMOG Grading - A New Readability Formula." *Journal of Reading* 12, 639-646.
- **Moran, John, Lewis, David and Christian Saam** (2014). "Analysis of post-editing data: A productivity field test using an instrumented CAT tool." Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds). (2014). *Post-editing of Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, 126-147.
- **Nitzke, Jean** (2016). "Monolingual post-editing: An exploratory study on research behaviour and target text quality." Silvia Hansen-Schirra and Sambor Grucza (eds) (2016). *Eye-tracking and Applied Linguistics*. Berlin: Language Science Press, 83-109.
- **Núñez, Gabriel** (2016). *Translating in linguistically diverse societies: Translation policy in the United Kingdom*. Amsterdam/Philadelphia: John Benjamin's Publishing Company.
- **O'Brien, Sharon** (2006). "Pauses as indicators of cognitive effort in post-editing Machine Translation output." *Across Languages and Cultures* 7, 1-21.
- — (2007). "An empirical investigation of temporal and technical post-editing effort." *Translation and Interpreting Studies* 2, 83-136.
- — (2009). "Eye-tracking in translation process research: methodological challenges and solutions." Inger Mees, Fabio Alves and Susan Göpferich (eds) (2009). *Methodology, technology and innovation in translation process research: a tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur, 251-266.
- — (2010). "Controlled language and readability". Gregory Shreve and Eric Angelone (eds) (2010). *Translation and Cognition*. Amsterdam/Philadelphia: John Benjamins' Publishing Company, 43-165.
- **O'Curran, Elaine** (2014). "Machine Translation and post-editing for user generated content: An LSP perspective." Olga Beregovaya, Mike Dillinger, Jennifer Doyon, Raymond Flournoy, Patti O'Neill-Brown and Chuck Simmons (eds) (2014). *Proceedings of the Eleventh Conference of the Association for Machine Translation in the Americas. Vol2. MT Users Track*. Association for Machine Translation in the Americas, 50-54. <http://www.mt-archive.info/10/AMTA-2014-OCurran.pdf> (consulted 05.12.2018).
- **Offersgaard, Lene; Povlsen, Claus; Almsten, Lisbeth and Bente Maegaard** (2008). "Domain specific MT in use." John Hutchins and Walther v. Hahn (eds) (2008). *Proceedings of the twelfth conference of the European Association for Machine Translation*. Hamburg: HITEC, 150-159. <http://mt-archive.info/EAMT-2008-Offersgaard.pdf> (consulted 05.12.2018).
- **Plitt, Mirko and François Masselot** (2010). "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context." *The Prague Bulletin of Mathematical Linguistics* 93, 7-16

- **Polvsen, Claus; Underwood, Nancy; Music, Bradley and Anne Neville** (1998). "Evaluating text-type suitability for machine translation: A case study of an English-Danish MT system." *Proceedings of the First International Conference on Language Resources and Evaluation*. European Language Resources Association, 27-31. https://www.iissco.unige.ch/en/staff/nancy/lrec_ling98.pdf (consulted 05.12.2018).
- **Prys, Delyth, Prys, Gruffudd and Dewi Bryn Jones** (2009). *Gwell Offer Technoleg Cyfieithu ar gyfer y Diwydiant Cyfieithu yng Nghymru: Arolwg Dadansoddol*. Bangor: Bangor University.
- **Rayner, Keith** (1998). "Eye movements in reading and information processing: 20 years of research." *Psychological Bulletin* 124, 372-422.
- **Richards, Elizabeth** (2013). *Business case: Greater collaborative activity in the delivery and procurement of Welsh translation services*. Cardiff: Welsh Government.
- **Screen, Benjamin** (2016). "What does translation memory do to translation? The effect of translation memory output on specific aspects of the translation process." *Translation and Interpreting* 8(1), 1-18.
- — (2017a). "Machine translation and Welsh: Analysing free statistical machine translation for the professional translation of an under-researched language pair." *The Journal of Specialised Translation* 28(1), 317-344.
- — (2017b). "Productivity and quality when editing Machine Translation and Translation Memory Outputs: An empirical analysis of English to Welsh translation." *Studia Celtica Posnaniensia* 2(1), 119-142.
- **Senez, Dorothy** (1998). "The machine translation help desk and the post-editing service." *Terminologie and Traduction* 1, 289-295.
- **Silva, Roberto** (2014). "Integrating post-editing MT in a professional translation workflow." Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds) (2014). *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, 24-50.
- **Skadiņš, Raivis; Puriņš, Maris; Skadiņa, Inguna and Andrejs Vasiljevs** (2011). "Evaluation of SMT in localization to under-resourced inflected language." Mikel L. Forcada, Heidi Depraetere and Vincent Vandeghinste (eds) (2011). *Proceedings of the 15th International Conference of the European Association for Machine Translation*. European Association for Machine Translation, 35-40. <http://www.mt-archive.info/EAMT-2011-Skadins.pdf> (consulted 05.12.2018).
- **Suojanen, Tytti, Koskinen, Kaisa and Tiina Tuominen** (2015). *User-centered translation*. London: Routledge.
- **TAUS** (2010). *Post-editing in practice*. Amsterdam: Translation Automation Users Society.
- — (2016). *TAUS Post-editing Guidelines*. <https://www.taus.net/think-tank/articles/postedit-articles/taus-post-editing-guidelines> (consulted 04.04.2017).
- **Uswak, Valentina** (2014). "Einsatz der Maschinellen Übersetzung im Übersetzungsprozess in Unternehmen." Jan Mugele (ed.) (2014). *Tagungsband der Nachwuchswissenschaftlerkonferenz* (Magdeburg, 24 April), 70-73.

- **Vázquez, Lucía, Vázquez, Silvia and Pierrette Bouillon** (2013). "Comparing forum data post-editing performance using translation memory and Machine Translation output: A pilot study." Khalil Sima'an, Mikel L. Forcada, Heidi Depraetere and Andy Way (eds) (2013). *Machine Translation Summit XIV 2-6 September 2013, Nice, France*. 249-256. <http://www.mt-archive.info/10/MTS-2013-Morado.pdf> (consulted 05.12.2018).
- **Wagner, Elizabeth** (1985). "Post-editing Systran – A challenge for Commission translators." *Terminologie and Traduction* 3, 1-6.
- **Welsh Government** (2011). *The Welsh Language Measure (Wales) 2011*. Cardiff: Welsh Government.
- – (2012). *A living language: A language for living – Welsh language strategy 2012-2017*. Cardiff: Welsh Government.
- – (2014). *A living language: A language for living – Moving forward. Policy Statement*. Cardiff: Welsh Government.
- – (2017). *Cymraeg 2050: Welsh Language Strategy*. Cardiff: Welsh Government.
- **Welsh Government and Welsh Language Commissioner** (2015). *National Survey for Wales, 2013-2014: Welsh Language Use Survey*. Cardiff: Welsh Government and Welsh Language Commissioner.
- **Welsh Language Commissioner** (2012). *Advice document: Bilingual drafting, translation and interpreting*. Cardiff: Welsh Language Commissioner.
- – (2015). *Annual 5-year Report*. Cardiff: Welsh Language Commissioner.
- **Yamada, Masaru** (2011). "The effect of translation memory databases on productivity." Anthony Pym (ed.). (2011). *Translation research projects 3*. Tarragona: Intercultural Studies Group, 63-73.
- **Zhechev, Ventsislav** (2014). "Analysing the post-editing of Machine Translation at Autodesk." Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds) (2014). *Post-editing of Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, 2-24.

Biography

Ben Screen has recently finished a PhD at the School of Welsh, Cardiff University. His research looked at the relationship between cognitive effort, productivity and final quality, and Machine Translation and Translation Memory output. He is also an accredited translator, and is currently in this role for a public sector organisation. He is also the lead external linguist for a large technology company and is responsible for managing quality and coordinating translation and review teams.



E-mail: screenb@cardiff.ac.uk

Appendix 1. The instrument used in the Likert-type scale data collection phase

Below, you are required to put a cross on a 1-5 scale in order to note your opinion regarding the readability and comprehensibility of the text. Readability means how easy or difficult the text was to read in terms of grammar, structure, spelling and how the text was written. Comprehensibility refers to your opinion regarding how easy or difficult it was for you to understand what the text was describing, or its message. One refers to 'Very low readability/comprehensibility', i.e. you experienced significant problems when reading and had to exert considerable effort to understand the text. Five refers to very high readability/comprehensibility, i.e. you were able to read and understand the text easily without any problems and without any considerable effort.

Readability					Comprehensibility				
1	2	3	4	5	1	2	3	4	5

Appendix 2. The translation read by the Control Group

Mae'n rhaid i chi lenwi ffurflen hunan-ardystio ar eich diwrnod cyntaf yn ôl yn y swyddfa, a hynny er mwyn cofnodi'ch rheswm dros fod yn absennol. Efallai y bydd gofyn i chi weithio rhagor o oriau na'r oriau amodol sydd wedi'u nodi yn eich contract o bryd i'w gilydd. Bydd gofyn i chi weithio rhagor o oriau na'ch oriau amodol, neu weithio ar benwythnosau, Gwyliau Banc neu wyliau Braint. Os oes gennych unrhyw gwestiynau neu bryderon ynglŷn â'r materion uchod, yna cysylltwch â'r Tîm Adnoddau Dynol ar unrhyw adeg. Mae'n rhaid i chi ofyn am ganiatâd eich Rheolwr Llinell ymlaen llaw. Dylech roi cymaint o rybudd â phosibl am eich gwyliau, gan gynnwys eich prif wyliau blynyddol. Pe hoffech gymryd pythefnos o wyliau gyda'i gilydd, mae'n rhaid i chi roi o leiaf fis o rybudd. Unwaith y bydd eich Rheolwr

Llinell wedi caniatâu i chi gymryd gwyliau, llenwch y ffurflen a'i rhoi iddo/iddi i'w llofnodi.

Appendix 3. The post-edited MT output read by the Experimental Group

Mae'n rhaid i chi lenwi ffurflen hunan-ardystio ar eich diwrnod cyntaf yn ôl yn y swyddfa, er mwyn cofnodi natur yr absenoldeb. O bryd i'w gilydd, efallai y bydd gofyn i chi weithio mwy na'ch oriau amodol fel y nodir yn eich contract. Bydd gofyn i chi weithio mwy na'ch oriau amodol, neu ar benwythnosau, gwyliau Banc neu wyliau Braint. Os oes gennych unrhyw gwestiynau neu bryderon ynglŷn â'r uchod, cysylltwch â'r Tîm Adnoddau Dynol ar unrhyw adeg. Mae'n rhaid cael caniatâd eich Rheolwr Llinell o flaen llaw. Rhowch gymaint o rybudd ag y bo modd, gan gynnwys am eich prif wyliau blynyddol. Os ydych yn dymuno cymryd absenoldeb o bythefnos ar yr un pryd, mae'n rhaid i chi roi mis o rybudd o leiaf. Unwaith y byddwch wedi cael caniatâd eich Rheolwr Llinell, llenwch y ffurflen a'i roi iddo / iddi i'w llofnodi.

Notes

¹ Faithfulness, or equivalence, is a problematic concept and is context-dependent. The use of the word here however implies that faithfulness is not a set concept, rather a relationship between source and target that holds at any given time.

² Gouadec (2010, p. 272) also lists functionality and efficiency in this regard (which he doesn't define), however in terms of the quality of texts, how readable and understandable they are is the most important.

³ This explicit link between the usability of a translation as a measure of its quality and how easy it is to read and comprehend is provided by Suojanen *et al.* (2015, p. 49), "However, when we are dealing with the usability of products that are text-based, the user is obviously always also a reader. And it follows that concepts such as readability and comprehensibility are closely related to the usability of texts". Usability can be defined as "The ease of use of a product in a specified context of use [...]" (Suojanen *et al.* 2015, p. 13).

⁴ Author's translation of "[...] a be dw i'n ei weld, ydi pan maen nhw [y darparwyr gwasanaethau] yn cyfieithu deunydd maen nhw'n gorgymhlethu pethau, yn dewis iaith ffurfiol iawn...".

⁵ According to the Welsh Language Use Survey (Welsh Government and Welsh Language Commissioner 2015), there was an increase of 130,000 people since 2006 who said they could speak Welsh but not fluently, and that 13% of the Welsh population are fluent in Welsh (the same figure in fact that uses it daily). Comparing this figure to the 19% of people who said they are able to speak Welsh in the 2011 census, this provides some evidence that Welsh translators are expected to translate for a linguistic community with differing levels of ability.

⁶ The study by Castilho and O'Brien (2016) compared the source text with its lightly post-edited and raw machine translated version, not between the post-edited and translated versions as is done here.

⁷ In terms of *parallel processing* of words in parafoveal view during self-paced reading and spill-over effects, i.e. the processing of words either side of the one actually being fixated on by the fovea centralis in the centre of the visual field, Drieghe (2011) argues in his review that this is still unlikely. That there is "parafoveal preview benefit" however (whereby the word currently fixated upon in foveal vision was previously seen in parafoveal vision, therefore saving processing time (Kennison and Clifton 1995)) is not as controversial. As a result, the operationalisation of the term "fixation" by following the definition of Duchowski (2003) will be used in this article.

⁸ This however does not imply that the text was easy to translate, as O'Brien (2010, p. 144) argues. The purpose of providing these readability index scores was to elucidate more the type of text used, which following the scores can be considered an average text of normal readability.