Human factors in machine translation and post-editing among institutional translators

Patrick Cadwell *   orcid.org/0000-0002-2371-4378

School of Applied Language and Intercultural Studies / ADAPT Centre

Dublin City University

Dublin 9, Ireland

patrick.cadwell@ adaptcentre.ie

Sheila Castilho   orcid.org/0000-0002-8416-6555

ADAPT Centre

Dublin City University

Dublin 9, Ireland

sheila.castilho@adaptcentre.ie

Sharon O'Brien   orcid.org/0000-0003-4864-5986

School of Applied Language and Intercultural Studies / ADAPT Centre

Dublin City University

Dublin 9, Ireland

sharon.obrien@adaptcentre.ie

Linda Mitchell   orcid.org/0000-0002-1883-9037

Thaler Str. 27

13129 Berlin, Germany

linda.mitchell.dean@gmail.com

Human factors in machine translation and post-editing among institutional translators

In September 2015, the ADAPT Centre for Digital Content Technology carried out a focus group study of 70 translators at the European Commission's Directorate-General for Translation (DGT).[1] The aim was to better understand the factors involved in the translators' adoption and non-adoption of machine translation (MT) during their translation tasks. Our analysis showed that, while broadly positive attitudes to MT could be observed, MT was not consistently adopted for all tasks. We argue that ergonomic factors related to a human translator's needs, abilities, limitations, and overall well-being heavily impacted on participants' decisions to use MT or not in their tasks. We further claim that it is only by taking into account the special institutional circumstances in which the activity of DGT translation is situated that these ergonomic factors can be fully understood and explained.

Keywords: machine translation (MT), post-editing (PE), ergonomics, human factors, Directorate-General for Translation (DGT), focus groups

Introduction

The International Ergonomics Association (IEA) highlights that there are different definitions of 'human factors and ergonomics,' which depend on the discipline (e.g. professional societies, academics, industry). The terms 'human factors' and 'ergonomics' are generally used side-by-side, e.g., the 'Human Factors and Ergonomics Society.' It goes beyond the scope of our article to delve into a discussion about the various definitions of human factors and ergonomics (for more information on the topic see http://www.hfes.org/Web/EducationalResources/HFEdefinitionsmain.html). For our purposes,

we adopt the succinct and clear definition given by the IEA on their website (International Ergonomics Association 2016):

> Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance. Ergonomists contribute to the design and evaluation of tasks, jobs, products, environments and systems in order to make them compatible with the needs, abilities and limitations of people.

This definition includes the core constructs we were interested in for the research reported here: understanding interactions among humans and translation systems, in particular machine translation (MT), optimizing translator well-being, and evaluation of technological design to increase compatibility with human translators' needs, abilities, and limitations.

Research on the use of translation technology

Translators have been using technology for a long time, such that some types of translation can now be conceptualized as human-computer interaction (O'Brien 2012). The effects of digital technology on translation practice have been acute and wide-ranging; new forms of translation activity, such as automatic online translation or crowd-sourced translation, have emerged, and new digital literacies have become required of translators as the focus of translation practice has shifted from the printed page to the computer screen (Cronin 2013). Computers assist translators in a variety of ways—from general computing functions to

specialized translation functions—and can bring many advantages; but translators benefit most when they understand these technologies and are aware of the potential pitfalls of adopting a computerized tool as part of a translation task (Pym, Perekrestenko, and Starink 2006). In particular, contemporary translators are required to master and should be critically aware of the special form of human-computer interaction in translation which takes place through CAT (Computer-Aided Translation) tools. CAT tools generally refer to translation memory, but also to terminology management tools, corpus tools, and specialized localization tools (e.g., Alchemy Catalyst, SDL Passolo). Translators are increasingly also interacting with machine translation (MT), which is traditionally seen as being a separate translation technology relative to CAT tools. These days, however, the most common mode for delivery of MT is within the CAT editing environment, where MT suggestions are added to the typical exact and fuzzy matches within a translation memory. The traditional delineation between CAT tools and MT is, therefore, becoming unstable. This is especially true since the prominent approach to MT, Statistical Machine Translation (or SMT), is enabled by translation memories that have been created through CAT tools. Any highly technologized work environment opens up questions about human factors, i.e., interaction between humans and other elements of a system, as defined above, and the highly technologized translation environment is no exception.

The increased use of MT within CAT environments has raised new concerns and questions among professional translators and translation scholars alike, leading to increased research on the topic. An emerging parallel interest in the cognitive translation process as well as in translation processes in the workplace has contributed to a growing body of research that focuses on human factors, ergonomics, and translation in general (Lavault-Olléon 2011). Somewhat surprisingly, this research has also highlighted discontentment and inefficiencies associated with traditional CAT tools, despite decades of deployment of the

latter. We will briefly review the literature that highlights such discontent before delving into questions about MT specifically and then focusing on research that has viewed translation from an organizational, or situated, embodied cognition perspective.

Researchers such as Lagoudaki (2006, 2008) had already recognized discontent issues with CAT tools a decade ago. One of her conclusions was that the end user's needs were of only subordinate interest to the CAT tool developers of the time. Since then, a number of studies have been carried out that confirm the ongoing dissatisfaction among some translators with the CAT tool environments they have to use. LeBlanc (2013), for example, also found that tool conception and design were causes of dissatisfaction. Workplace observations confirm that some of the tools used by translators slow them down (Ehrensberger-Dow and Massey 2014). A large-scale international survey of ergonomic issues among professional translators validates the assumption that a significant number of translators are irritated by the CAT tools they use (O'Brien, Ehrensberger-Dow, Hasler, and Connolly, submitted.) In that survey, the most commonly mentioned irritant was the complexity of the CAT tool, with issues regarding text segmentation and formatting coming second and third. Respondents also frequently mentioned being irritated by bugs and errors. Yet another survey that specifically focused on features that would support post-editing uncovered much dissatisfaction with the basic CAT tool features (Moorkens and O'Brien, forthcoming). Van den Bergh et al. (2016) conducted surveys and contextual inquiries in five translation companies and produced several recommendations to positively impact on translators' workflows, including improving efficiency, effectiveness, and usability of translation environments, as well as giving translators more control over their text production environments.

It is clearly the case that CAT tools, even today, can lead to translator dissatisfaction, though it must be pointed out that there are also significant advantages to these tools and many translators would not like to return to the era of working without them. More recently,

the translator's work environment has been made even more complex by the addition of machine translation. A common deployment strategy for MT is to integrate the MT service with a translation memory environment, such that MT suggestions are automatically presented to the translator if certain criteria are met (e.g., there is no fuzzy match above a specific fuzzy match value). The rise in usage of MT has led to much research on the task of post-editing. This topic has been covered adequately in previous publications, and so we will not go into detail here. Rather, we will highlight some aspects that serve to demonstrate that the inclusion of MT has led to a need for more in-depth studies on the ergonomics and human factors associated with MT use.

In 2010, a survey of 228 managers by SDL showed that 28% of respondents were using or planning to use MT, and 57% said they would be more likely to adopt MT with post-editing. A survey by the Common Sense Advisory (DePalma et al. 2013) of 1,000 suppliers showed that since 2011 the number of LSPs offering MT with post-editing has grown from 37% to 44%. A report by DePalma and Sargent (2013) on buyers of language services and MT technology featuring 108 respondents who use MT in their companies found that 60% of these companies publish their MT output after some external or internal post-editing, and only 8% publish their MT output immediately. In general, MT output is rarely published without some kind of post-editing.

Several studies on post-editing have demonstrated that post-editing with MT can be faster than translating without MT and that the task does not necessarily have a negative impact on quality (e.g., Guerberof 2009; Plitt and Masselot 2010), or on usability by end users of machine translated text (Doherty and O'Brien 2014; Castilho et al. 2014). However, when translators post-edit, they frequently report that they would be faster without the MT output and would produce better quality. Translators also report that interaction with MT is tiring and tedious, even though, in theory, it means they have to type less if they can reuse

some of the MT suggestions. This points to the possibility that translators are resistant to combining MT with human translation; their perception is that it is a slower and more complex task compared to human translation without MT. A study by Moorkens et al. (2015) probed this issue of perception by asking participants to rate perceived post-editing effort and then compared the effort ratings with actual post-editing effort. Only weak to moderate correlations were found between perception of effort and real effort. Koponen (2012) also investigated this topic and found a tendency to rate longer sentences as requiring more effort when, in fact, they did not. A study by Gaspari et al. (2014) also found an inherent bias in ratings against the task of post-editing. Teixeira (2014) found a lack of correlation between post-editing performance and post-task ratings of performance. It is reasonable to rule out sabotage by study participants here, given that these studies were independent of one another and employed translators in different locations who worked in different language pairs. The studies may point to study participants' inability to properly gauge effort and to a lack of experience with MT. However, it also emphasizes that the current relationship between translators and MT is a complex human-computer interaction (HCI) challenge that warrants further investigation.

In the research on translation technology to date, recurring themes are individual variation (i.e., translators do not behave in a predictable way) and the role that perception plays, but there is also a move towards investigating the interaction with the technology from a socio-technical perspective. It is argued that researchers need to reach beyond the narrow focus of a computer monitor to observe the immediate surroundings and the impact of the work environment, whether that be a freelancer's desk at home or a translator's office in a large organization. In going beyond the level of the computer tool or monitor, it is likely we will gain a greater understanding of the broad range of factors that impact on translators' use of tools. Of note here is the work by Risku (2010) and Risku and Windhager (2013) who

draw on the situated, embodied cognition paradigm to investigate the task of translation. A socio-ergonomic lens is applied by Ehrensberger-Dow (2014) to investigate cognitive challenges for translators in the workplace, while Ehrensberger-Dow and O'Brien (2015) draw on the concept of cognitive friction to compare the experiences of office-based translators and freelance translators along the three dimensions of cognitive, physical, and organizational ergonomics.

While this research demonstrates a move towards broader examination of translator workplace interaction, much of the research has so far been conducted with professional translators, usually in a commercial context. Little attention has been paid to institutional translators (Koskinen 2008), especially in relation to MT. Our motivation for the research study reported here was threefold: (1) we believe that the human factors relating to adoption or non-adoption of MT warrant investigation; (2) we expect that the individual coupled with the organizational aspects will provide a richer understanding; (3) little work has been undertaken to investigate the human factors pertaining to MT in large translating institutions. For these reasons, we undertook a study of human factors and MT at the Directorate General for Translation of the European Commission (DGT).

Context for this study

The DGT is the language department of the European Commission, unique in its size in Europe with its approximately 1,750 translators based in Brussels, Luxembourg, and the EU Member States (Kluvanec 2014). It is divided into 24 language departments with specialised units translating up to 2 million pages per year (European Commission 2012). The DGT covers domains such as legislative and policy documents, reports, web content, papers on legal, economic, scientific and technical matters, and press material. Furthermore, it is the

DGT's policy to draft documents in all 24 official languages and to communicate with any EU citizen in the language of their choice (Kluvanec 2014).

In order to facilitate the translation of the extremely large document volume, the DGT provides a set of tools to its translators, focussing on either administration or translation. This includes Suivi Software to manage translation requests internally, TraDesk to manage the translation jobs for each individual translator and to access earlier translations through archives, a local CAT tool which the translators use to create their translations, and EURAMIS (European Advanced Multilingual Information System), a translation memory storing all linguistic resources that have been collected primarily in the DGT (European Commission 2012). Every translator in the institution uses the EURAMIS translation memory. The essential role it plays is underlined by Koskinen and Ruokonen's (forthcoming) observation that when the translators of the European Commission were given the hypothetical option of writing a 'love letter' or 'break-up letter' to any of their technical tools, they chose to love the DGT's translation memory. Another tool that the DGT provides is MT, which is used primarily by translators as a basis for their work but also by administrative staff for gisting and drafting (Kluvanec 2014).

About MT@EC

From the 1970s to 2010, the DGT developed and used a rule-based MT. The system supported 28 language pairs. Experiments towards an SMT system started in 2009, with the first prototype, called MT@EC, released in 2011 (Kluvanec 2014). The system was tested, evaluated, and improved on a regular basis until the current version (10th generation) to be released at the time of writing.

The MT@EC system is no longer purely statistical but combines the two previous approaches in a hybrid system, consisting of an SMT system based on Moses and additional

rules. Its language model is trained on data from EC translations that have been collected over several years. MT@EC covers 24 languages and 552 language pairs with 62 direct language pairs. In January 2014, the system contained over 200 million segments of training data and 800 million segments across all 24 target languages with an annual growth rate of over 20%. Current focal points for further improvement are morphologic rules, domain adaptation, and pre-processing.

MT@EC is available to all employees of the European Commission, and the use cases include not only DGT translators producing official translations, but also members of other services or of home-state institutions seeking gist translations of documents. MT@EC is, therefore, accessible via a web platform, as a plugin for Outlook, or as an email service for translating email messages (Pilos 2014). The system can also produce translation suggestion options in local CAT tools if there are no TM matches. The translator is able to use this service by importing the MT data into the local CAT tool along with other desired resources.

Feedback on MT@EC

Testing and feedback are essential in order to improve any MT system. The DGT collects comprehensive feedback using BLEU scores and user feedback from their annual user conferences, MT@Work. These conferences were held in 2012 for all translators, in 2013 for all EU staff, and in 2014 for users in member states (Kluvanec 2014). They conducted a maturity check in 2011, for example, with 64 translators from 22 language departments submitting binary quality judgements on the usefulness of MT output. The intention was to identify if and which baseline systems yielded useful results. Subsequently, ten language departments adopted MT in a real-life trial (Eisele 2013).

As part of MT@Work in 2012, a user survey with 763 respondents detailed that 70% (535) had used MT in their work in the preceding six months. In MT@EC 2013, they

discovered that the language patterns varied across analytical, inflected, and agglutinating languages (Kluvanec 2013). The DGT used the TAUS DQF tools to evaluate the fluency and adequacy of segments that had been machine translated from English into ten different languages. On a five-point Likert scale, the average fluency ranged from 2.5 (Lithuanian) to 3.26 (Swedish) and the average adequacy ranged from 3.08 (Lithuanian) to 4.5 (Portuguese) (Bonet 2013).

What emerged from the various feedback sources was that a lack of awareness had been holding back MT@EC, more specifically an understanding of what MT really is and how it works. The DGT approached the matter by introducing MT as a sub-segment retrieval system that should be taken with a 'pinch of salt' and that could not replace humans; they cast it as a system that, in fact, needed human assistance (Kluvanec 2014). Subsequently, the annual MT user conference in 2014 attracted over 300 registrations online in real-time (Rummel 2015). However, while awareness may have increased, the higher usage numbers desired by the DGT have not yet been achieved.

Methods

The aim of the study was to better understand the factors involved in DGT translators' adoption and non-adoption of MT during their translation tasks, and three main research questions were posed:

1. Do participants in a cohort of DGT translators use MT in their work?
2. What reasons do they advance for this adoption or non-adoption of MT?
3. What could account for the reasons put forward by these DGT translators?

Focus group methodology

A focus group methodology was chosen over other forms of discovery to answer these questions because of its strength at interrogating the reasons behind participants' thinking, at differentiating between the shared and variable perspectives that participants hold about these reasons, and at facilitating the interpretation of existing quantitative data about the object of enquiry (Stewart and Shamdasani 2015). Focus groups are also useful for investigating non-uptake and for providing answers to the question 'why not?' (Barbour 2007).

This choice of methodology also imposed limitations on our research. Focus group data are not appropriate for drawing broad generalizations, especially conclusions of a statistical nature (Liamputtong 2011) nor are they amenable to the measurement of phenomena (Barbour 2007). However, neither statistical generalization nor the measurement of the strength of participants' attitudes is being claimed in this paper.

In terms of practical arrangements, two researchers conducted the focus groups. No group included more than six participants at a time. One researcher acted as moderator and guided the discussion through a pre-defined series of topics, while also encouraging the development of spontaneously emerging topics through focused listening. The other researcher acted as co-moderator, kept time, and noted any initial analytical impressions as well as any significant non-verbal information (gestures, group dynamics, incidents, etc.) that were not captured by the two audio devices used to record each discussion. A decision was made to use audio rather than video recording to avoid increasing the discomfort of participants and to circumvent the problems for anonymization that video recording can create. These practical choices about the running of the focus groups were made following a mock focus group that was piloted with colleagues of the researchers to test the discussion protocol and identify the optimum focus group strategy for the research context.

Participants

The recruitment of participants was undertaken by a coordinating business manager in the DGT on behalf of the researchers. He informed the managers of the 24 language departments within the DGT that the focus groups were to be carried out and asked them to request members of their staff to volunteer to participate. All those who volunteered were invited to one of the focus groups. The focus groups were arranged to attempt to gather a broad spectrum of views and to be able to examine similarities and differences between language pairs. These criteria involved ensuring that:

1. At least one participant for each of the 24 target languages was encouraged to participate in the research;
2. Only one participant for the same target language would take part in a focus group at any one time—an exception to this was made for the English target language community whose portfolio comprises translations from more than 23 source languages;
3. Some focus groups would be made up of participants representing target languages which share common features—specifically, certain groups were made up exclusively of participants from the departments of inflected languages (largely Romance family), analytic languages (i.e., Danish, Swedish, Dutch, English, and Bulgarian), highly-inflected languages (largely Slavic, Baltic, Hellenic, Celtic, and Semitic families), agglutinating languages (Finno-Ugric family and, due to extensive composita, German), and English (the procedural source language which accounts for the largest amount of documents in the DGT and which, when used as a target language, accounts for a broad spectrum of all official source languages);

4. The remaining focus groups would be made up of a broad spectrum of participants whose languages did not share similar features, to contrast these participants' experiences from a different perspective.

The researchers emphasized at various points that they wished for translators thought to hold negative opinions of MT to be actively encouraged to participate to reduce bias toward positive views of MT, and they were assured that this was, indeed, the case.

The focus groups took place in the offices of the DGT in Luxembourg and Brussels during a two-week period from September 7th to 18th, 2015. The data from 70 participants across 17 focus groups were gathered. Ethics approval for this project had been granted by Dublin City University's Research Ethics Committee prior to the research and, before each focus group, all participants read a plain language statement and signed an informed consent form.

All 24 languages of the DGT were represented in the cohort. However, participants who spoke an agglutinating language as a mother tongue were somewhat underrepresented in the data, while participants who spoke an analytic language as a mother tongue were somewhat overrepresented in the data, as can be seen from Figure 1.

Mother Tongue of Participants

N.B. Some participants listed more than one mother tongue. Shown here is the mother tongue given which corresponded to their language department.
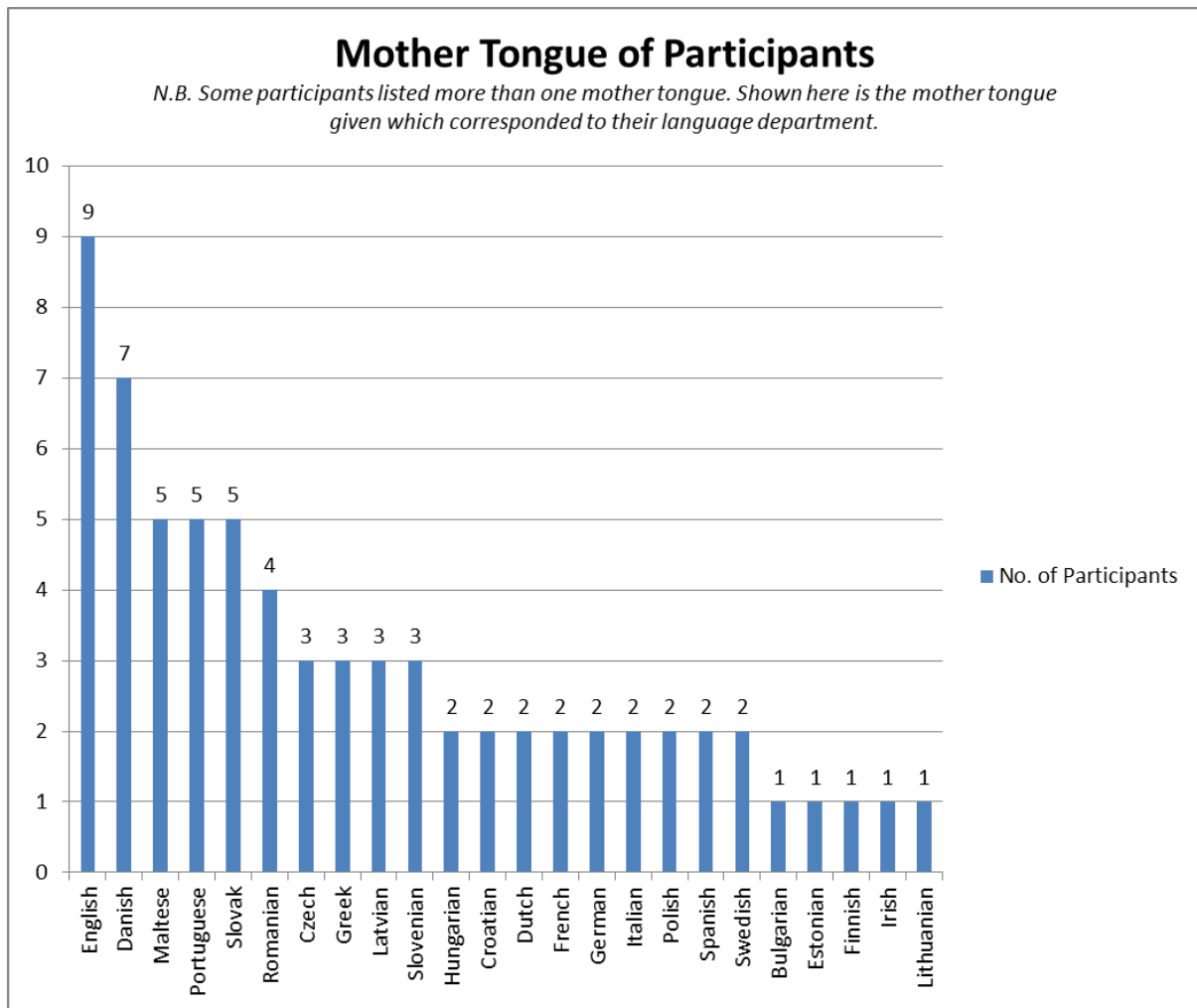
Figure 1. Number of participants in the cohort broken down by mother tongue

Figure 2 illustrates that this cohort of 70 participants split into 17 focus groups was relatively balanced in terms of gender. In addition, participants represented a broad range of ages; the majority of participants were in their 30s, 40s, or 50s at the time of the focus groups, as can be seen from Figure 3.
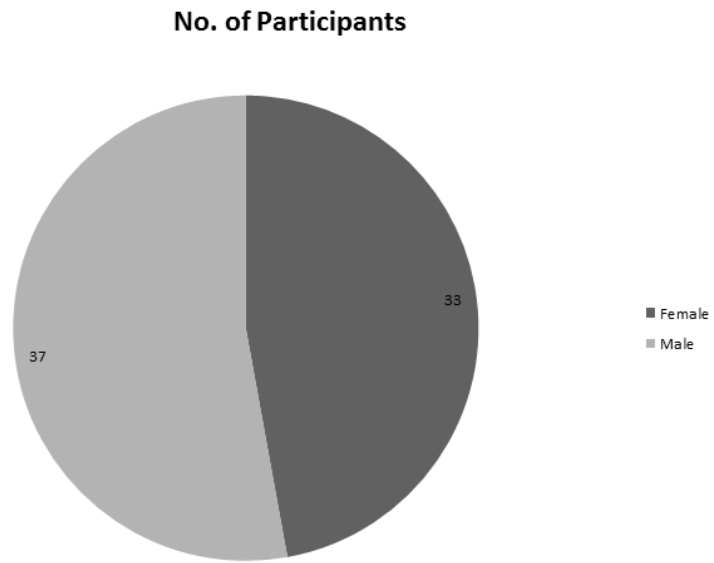
No. of Participants
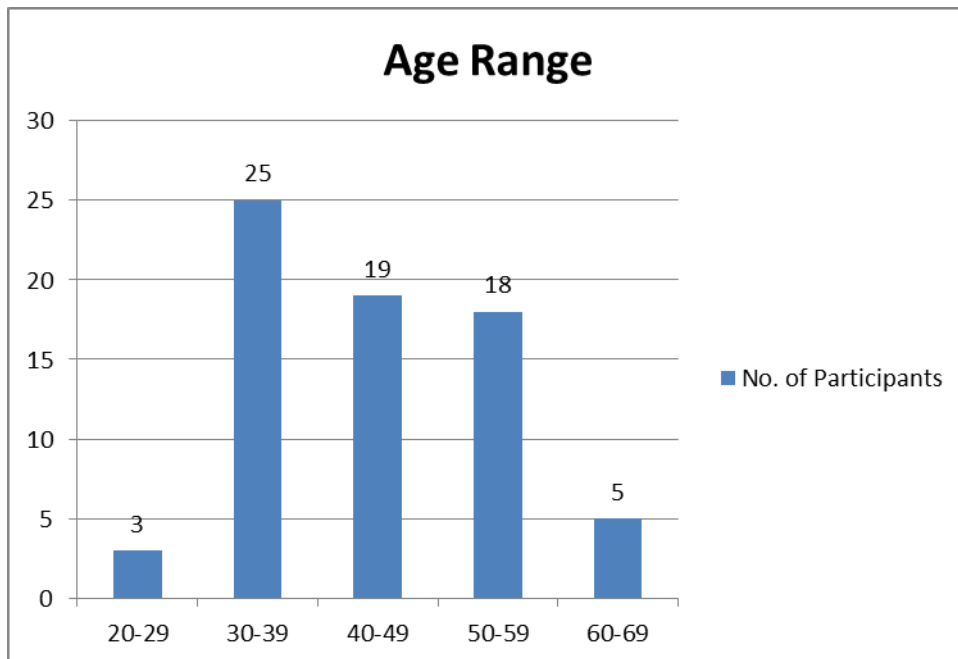

Figure 2. Gender balance of participants

Figure 3. Age range of participants

Furthermore, the cohort was made up, on the whole, of experienced translation professionals; Table 1 shows that the mean number of years participants spent working in the DGT was ten for this cohort and that, on average, participants spent about seven years working as a translator elsewhere.

Table 1. Average years spent working as translators in the DGT and elsewhere

	Years working in DGT	**Years working elsewhere**
Mean	10	7.1
Median	8	6
Mode	2	0
Range	29.5	29

In short, while it is not the goal of this study to claim representativeness of this group of participants and use them to make statistical inferences about the whole population of DGT translators, we feel that the cohort did allow us to access a wide variety of perspectives on our object of enquiry and helped us to understand a broad selection of qualitative experiences.

Coding and analysis

The first step in analyzing the data gathered during the focus groups was to have the audio files transcribed. This task was outsourced to a professional transcription agency, but the completed transcripts were then checked against the audio files by one of the researchers and amended where necessary. It was at this last stage that the contextual, non-verbal information gathered by the co-moderator during the focus groups was added to the transcripts, usually in the form of a parenthesized note.

The unit of analysis ultimately chosen for this research was the group, and the data presented and discussed in the following sections concentrate on patterns and trends observed across groups. Even so, as Morgan (1997, 60) explains, focus groups also allow access to data on individual participants and group interactions, and these data should not be ignored. For this reason, individual contributions to the discussions and intra-group dynamics have also been analysed as part of this research project.

A thematic analytical strategy was operationalised from Braun and Clarke (2006) to examine the focus group data, and coding tasks were facilitated by QSR's NVivo 10 software. Themes were developed over six phases that progressed from participant-led, to interpretive, to abstract analysis. Conducting successive phases of analysis in this way was labour and time intensive, especially with more than 196,000 words of transcript data to read through repeatedly. Nonetheless, these repeated viewings at different times and with different focuses (sometimes participant-led, sometimes interpretive, sometimes abstract) encouraged fresh insights into the data and led to new realizations and valuable questioning of assumptions as the process of analysis progressed. It should also be noted that one third of all data coded was double-checked by another coder to ensure inter-coder reliability.

A total of six themes were developed during this analysis: (1) MT use and development; (2) reasons to use or not use MT; (3) work practices; (4) valued added by human translators; (5) human and machine interfaces and impacts; and (6) training. The first three of these themes combined with the core ergonomic constructs introduced in Section 1 (interactions among humans and systems, optimising translator well-being, and compatibility with needs, abilities, and limitations) allowed us to answer our research questions.

MT use and reasons for MT adoption or non-adoption

In answer to our first research question, we discovered that the participants in our cohort of DGT translators do indeed use MT in their work. A brief survey relating to MT use by participants was taken during each focus group prior to the beginning of the discussion. The survey asked participants to, firstly, indicate the frequency of their use of MT in their translation tasks and to, secondly, indicate their level of agreement on a five-point Likert scale with the following two statements: (1) 'MT is useful in my work' and (2) 'making the most of MT in my work requires skill.'

Figures 4, 5, and 6 show at first glance that MT is perceived by this cohort of DGT translators to be a frequently used and useful tool, the optimization of which requires skill.
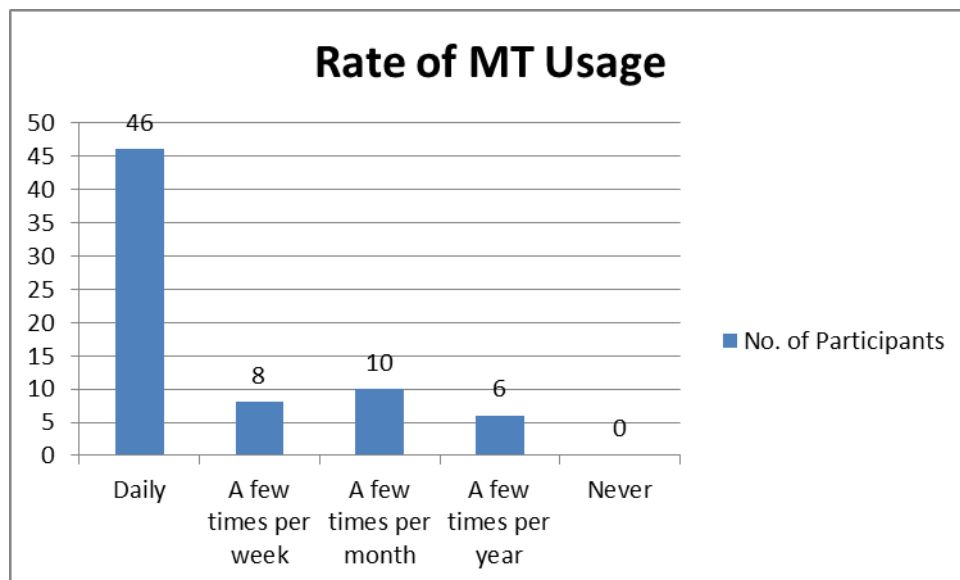

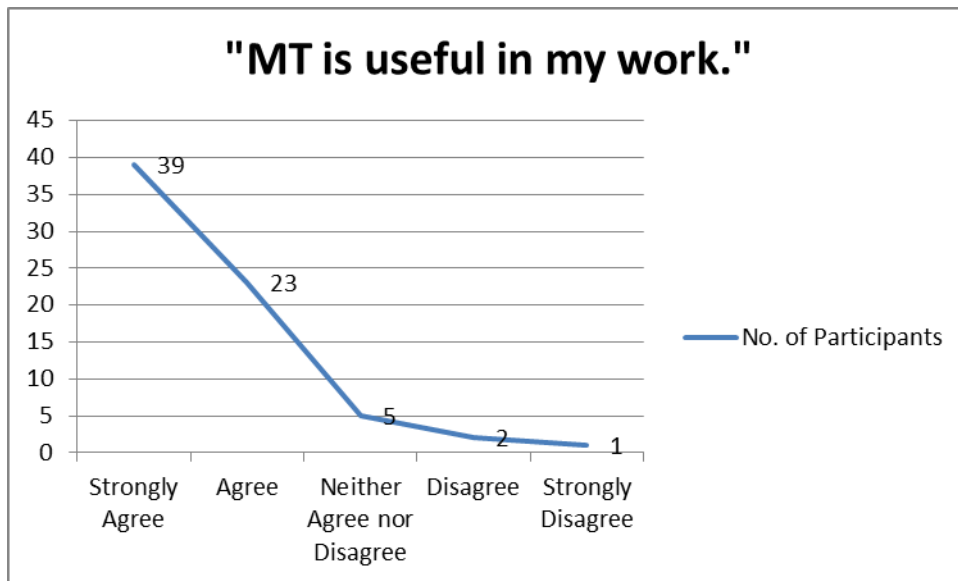Figure 4. Rate of MT usage by participants

Figure 5. Level of agreement with the first MT-related assertion
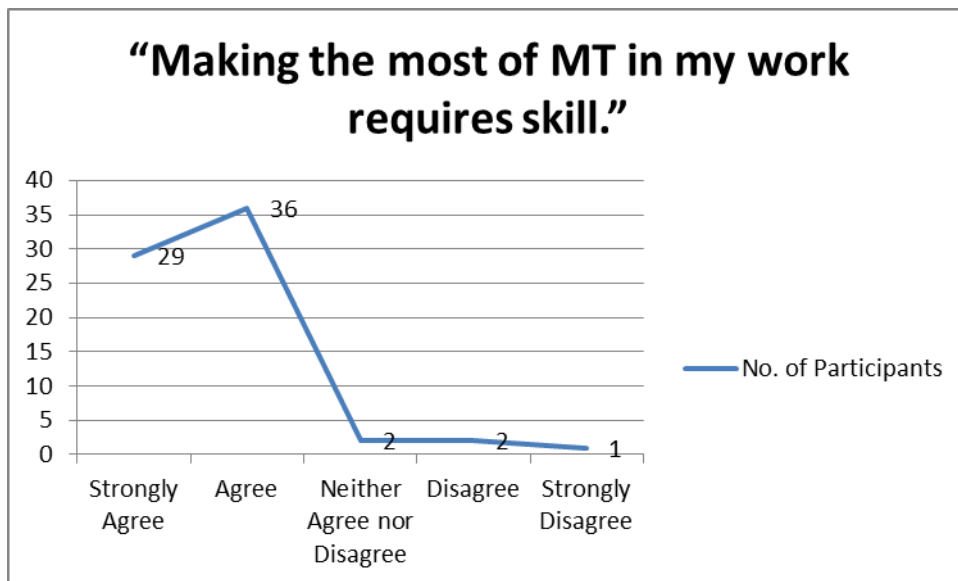

Figure 6. Level of agreement with the second MT-related assertion

In answer to our second research question, we discovered that the participants in our cohort put forward a variety of reasons for their decisions to use or not use MT in their translation tasks. Table 2 lists the reasons advanced by participants for using MT. The wider

the agreement across focus groups on a particular reason, the higher it appears in the table. The first four items in Table 2, in particular, were widely agreed upon by translators regardless of the target language into which they worked. That is to say that, whether working into a Romance, Germanic, Slavic, Semitic, Finno-Ugric, Baltic, Hellenic, or Celtic language, many translators explained that they use MT because: (1) they believe they will experience speed or productivity gains in doing so; (2) they perceive that the quality of the MT output is adequate for their purposes; (3) using MT gives them inspiration or ideas that they would not otherwise have and helps to 'kick-start' the translation process for them; (4) using MT reduces the amount of typing or other data entry they are required to do and reduces their exposure to the threat of repetitive strain injury.

Table 2. Reasons to use MT

1. For speed or productivity gains
2. Because of the perceived good quality of the MT output
3. For inspiration, to kick-start the translation process, or for new ideas
4. To reduce typing or clicking
5. For texts with high retrieval from TMs (e.g., trade, technical, legal)
6. For texts with low retrieval from TMs (e.g., speeches, press releases)
7. Because of a personal interest in technology
8. Because MT output is easily enabled or disabled
9. Out of habit or custom
10. For weaker source languages
11. With the goal of improving the MT system
12. Following direction from a senior or colleague

13. To get a gist understanding
14. Because of MT's positive influence on a translator's abilities
15. To not waste resources
16. For short ST sentences or for short source texts
17. Because a target text will be revised anyway
18. To miss fewer elements of source content
19. For texts which can be easily processed by a machine

Several of the reasons given by DGT participants—for instance, issues of speed or productivity, of perceived quality, or of physical ergonomic benefits—were unsurprising and have been documented in other research settings. Other reasons that were widely agreed upon by participants were more unusual and deserve special mention.

To hear translators explain that they use MT to 'kick-start' their translation process or as a source of inspiration was somewhat unusual, and this was reinforced by the repeated mentions of this idea in the focus group data. Also unusual was the topic of text type as a reason to use MT (as can be seen from the fifth and six items listed in Table 2). There was disagreement over whether to use MT for source texts that tend to have a high retrieval rate from the DGT's proprietary translation memories (such as trade documents and legislative or technical texts) or for source texts that tend to have a low retrieval rate from the DGT's proprietary translation memories (such as speeches, press releases, or more free-form texts). Interestingly, translators working in the Romance, Slavic, and Baltic languages of the European Commission in this cohort tend to use MT for high-retrieval texts but not for low-retrieval texts, while opinion is more divided among translators working in the other languages.

There were also some patterns in the data on reasons for using MT distinctive to particular language groupings. Translators working in the EC's Romance languages in this cohort see a translator's personal interest in technology as being a motivating factor in the adoption or non-adoption of MT. Participants working in the Slavic languages are motivated to use MT by the fact that not doing so would be a waste of the resources available to them. Finally, translators working with English as a target language who took part in the focus groups place particular emphasis on the fact that they use MT when they are working with their weaker source languages.

Table 3 lists the reasons advanced by participants for not using MT. Overall, the agreement was less pronounced among translators for reasons not to use MT in DGT translation tasks than it was for reasons to use MT. However, translators across many languages agreed on three ideas. (Table 3 lists these and other reasons not to use MT; again, the wider the agreement across focus groups on a particular reason, the higher it appears in the list.) Firstly, DGT translators in this cohort do not adopt MT in their translation tasks when they perceive the quality of the MT output to be inadequate for their purposes. This is especially true when the translator has had a bad first experience of using MT. Secondly, participants may be reluctant to use MT for fear that doing so may negatively influence their abilities—making them grow lazy, blocking them into seeing only one possible translation solution, or reducing their opportunities to learn and develop their skills, etc. Thirdly, the DGT translators who took part in these focus groups may avoid using MT because they are conservative and comfortable in their current working style, fear the unknown or fear that MT will eventually make translators obsolete.

Table 3. Reasons not to use MT

1. Because of perceived poor quality of MT output, esp. a bad first experience
2. For texts with low retrieval from TMs (e.g., speeches, press releases)
3. Because of MT's negative influence on a translator's abilities
4. Because of fear (e.g., of the unknown, of being replaced by a machine)
5. Because using MT induces the translator to make particular errors
6. Because of the extra attention required of the translator when using MT
7. Because of a personal lack of interest in technology
8. Because using MT devalues a translator's work
9. For texts with high retrieval from TMs (e.g., trade, technical, legal)
10. Following direction from a senior or colleague
11. Because a translator is not required to use MT in the DGT
12. Because MT cannot be trusted
13. Because using MT diminishes creativity
14. When MT output is not easily accessible
15. For security / confidentiality reasons

Negative perceptions of MT and post-editing have been widely reported in other research settings as reasons not to use MT, as was shown earlier. This also applies to the issue of extra attention that is required when post-editing MT output to correct or ensure the accuracy of word endings and word order, especially in certain language pairs. This latter issue was of particular concern to translators working into the Germanic, Hellenic, Slavic, and Celtic languages in our cohort. In contrast, the findings that MT is not used because it will have a negative influence on a translator's abilities and because MT misleads the translator to make particular errors are relatively unusual and are worthy of more detailed

consideration. It is also worth noting that, while translators in some departments of the DGT are being encouraged to use MT in their tasks (see the twelfth item in Table 2), translators in other departments are being directed by senior members of staff or mentors not to use MT. Mostly, this appears to be because of a desire to encourage novice translators to learn how to translate without the influence of MT at the early stages of their career or because senior colleagues believe using MT is generally counterproductive in that language pair. Nevertheless, the personal preferences of senior members of staff do also feature in participant accounts.

In sum, if we examine Tables 2 and 3, we can see that the majority of reasons to use or not use MT relate to a human translator's needs, abilities, limitations, and especially to the desire to optimize the translator's overall well-being. In this regard, the reasons advanced by this cohort of DGT translators express ergonomic concerns. But why are ergonomic concerns—rather than purely technological, financial, or other concerns—so prevalent in this particular study of MT adoption or non-adoption? To answer this, we must look to the other major construct involved when adopting an ergonomic perspective—interactions among humans and systems. Specifically, we need to understand the interactions of these DGT translators with the other elements of the socio-cultural system in which they operate at the EC.

DGT translation in the EC institutional context
The theme of 'work practices' developed during our analysis of the focus group data helped us to answer our third research question and account for the reasons for MT adoption or non-adoption advanced by our cohort. Data under this theme allowed us to describe certain key features of the work practices of this group of DGT translators, which in turn helped us to understand some of the translators' reasoning.

Features of the work practices of our cohort of translators

According to some participants in the focus groups, the benchmark production for an individual DGT translator is between seven and eight pages a day, but a recurring pattern in the focus groups was discussion of the pressure translators are under to meet tighter deadlines and to achieve higher volumes of production. Participants emphasized, too, that their job is largely to recycle frequently recurring text from a huge database of previous DGT translations and to work with unstable source texts that require frequent updates before a translation can be published. We argue that these features of the translators' work practices are likely to encourage the adoption of MT. At the same time, some contrasting features of their work seem to pull them away from using MT, as for instance in the case of legislative documents.

A recurring pattern across the focus groups was the responsibility translators feel when translating legislation. They see that most of their work involves the production of top quality, legally-binding texts. This obliges the translator to check all texts rigorously and to be suspicious of calls to produce documents which are merely deemed fit for purpose. Indeed, many participants in our research see themselves not just as translators but as part of the quality control of the legislation produced by the EC. For these reasons, the translators we spoke to are reluctant to adopt any technology—not just MT—which might reduce the rigour of their quality checking or the quality of the final target texts associated with their names.

These features of the translators' work, while helping to explain some of the motivations behind using or not using MT, still do not explain why ergonomic concerns relating to a translator's needs, abilities, limitations, and overall well-being are so prominent in our participants' accounts. For this, it is beneficial to focus our attention on the ways in which DGT translators fit into the broader socio-cultural system of the EC.

Interactions between DGT translators and the EC socio-cultural system

From the focus group data coded under the theme 'work practices', we were able to observe that our cohort of DGT translators interacts with the institutions of the EC in a manner which can be characterized by relative job security, recognition, and a freedom not typically enjoyed by professional translators in other settings. We argue that these relatively high levels of job security, recognition, and freedom allow the needs, abilities, limitations, and well being of translators to be matters of central concern when considering MT adoption.

DGT staff translators pass an open recruitment competition in order to be employed as full EU officials. As such, the pay, terms and conditions, and benefits associated with their employment are the same as for any other EU official of the same grade. Most of the participants in our cohort were employed under such a status. The relatively secure and privileged terms of our participants' employment with the EU were underscored by the fact that they repeatedly underlined how the staffing of the DGT is based on political rather than budgetary concerns.

As to recognition, some participants did suggest that other EU officials may be largely uninformed about or uninterested in the DGT, its functions, and its employees. In contrast, though, another pattern in the data related to an increased feeling of recognition by the institutions of the EU for the work of DGT translators. This pattern was characterized by frequent mention of the importance of contact with other EU officials or with national authorities to a DGT translator's work and of the feeling of partnership or mission sharing between translators and other officials in the production of legislation.

Finally, the secure and relatively recognized position of DGT translators is further supplemented by a degree of individual freedom in how to carry out their work. For instance, translators are largely free to adopt or not adopt tools as they see fit in their tasks once their

production targets are achieved, many translators occupy an individual office, and no system is in place at this point in time to track translator edits, log keyboard strokes, or record translation sessions.[2]

Concluding remarks

In conclusion, our study of MT adoption and non-adoption among a cohort of DGT translators showed that a majority of these translators use MT daily in their tasks and perceive it as a useful tool. The participants put forward a variety of reasons for their decisions to use or not use MT in their translation tasks. Many of these reasons, listed in Tables 2 and 3, reflect the definition of ergonomics referenced in the introduction to this paper and relate to the optimization of a translator's well being and to consideration of a translator's needs, abilities, and limitations. For example, many of the reasons mentioned by participants speak to a translator's need for MT output to be of sufficient quality before it can be used, to the positive and negative influences that MT use can have on a translator's abilities, and to the physical and cognitive limitations that MT use by a translator can either impose or remove. Those translators who adopted MT tended to view the technology as meeting their needs and helpful to them to overcome physical or cognitive limitations.

We noted in our review of the literature on translation technology that researchers should go beyond the level of individual translators working at computers when seeking to understand the broad range of factors that impact on a translator's use of tools (Risku 2010; Risku and Windhager 2013). In doing so in this paper, we examined the interactions of DGT translators with the other elements of the socio-cultural system at the EC institutions in which they operate. Our focus group participants highlighted job security, freedom to choose tools, and a feeling of responsibility for the overall mission to produce top quality legislation as reasons

---

[2] Translators who are frequent teleworkers tend to share office space.

for the adoption or non-adoption of MT in their translation tasks. We speculate that the experiences of DGT translators will contrast with the experiences of translators working in commercial settings, based on our previous interactions with the latter. A future goal is, therefore, to compare focus group data gathered from commercial translators with the translators in this institutional setting to test this belief.

References

Barbour, Rosaline. 2007. *Doing Focus Groups*. Thousand Oaks: Sage Publications.

Bonet, Josep. 2013. *Honing Quality in 23 Languages MT @ EC—The Right Foundation*. Unpublished DGT intranet report. Accessed May 25, 2015.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101.

Castilho, Sheila, Sharon O'Brien, Fabio Alves, and Morgan O'Brien. 2014. "Does Post-Editing Increase Usability? A Study with Brazilian Portuguese as Target Language." In *Proceedings of the 17th Conference of the European Association for Machine Translation*, edited by Marko Tadić, Philipp Koehn, Johann Roturier, and Andy Way, 183–190. Dubrovnik: EAMT.

Cronin, Michael. 2013. *Translation in the Digital Age*. Milton Park, Abingdon, Oxon: Routledge.

DePalma, Donald A., and Benjamin B. Sargent. 2013. *Transformative Translation*. Cambridge: Common Sense Advisory.

DePalma, Donald A., Vijayalaxmi Hegde, Hélène Pielmeier, and Robert G. Stewart. 2013. *The Language Services Market: 2013*. Cambridge: Common Sense Advisory.

Doherty, Stephen, and Sharon O'Brien. 2014. "Assessing the Usability of Raw Machine

> Translated Output: A User-Centered Study Using Eye Tracking." *International*
>
> *Journal of Human-Computer Interaction* 30 (1): 40–51.

Ehrensberger-Dow, Maureen. 2014. "Challenges of Translation Process Research at the

> Workplace." *MonTI Monographs in Translation and Interpreting* Special Issue 1:
>
> 355–383. DOI: 10.6035/MonTI.2014.ne1.12.

Ehrensberger-Dow, Maureen, and Gary Massey. 2014. "Translators and Machines: Working

> Together." In *Proceedings of the XXth World Congress of the International*
>
> *Federation of Translators (Vol. I)*, edited by Wolfram Baur, Brigitte Eichner, Sylvia
>
> Kalina, Norma Keßler, Felix Mayer, and Jeanette Ørsted, 199–207. Berlin: BDÜ.

Ehrensberger-Dow, Maureen, and Sharon O'Brien. 2015. "Ergonomics of the Translation

> Workplace: Potential for Cognitive Friction." *Translation Spaces* 4 (1): 98–118. DOI:
>
> 10.1075/ts.4.1.05ehr.

Eisele, Andreas. 2013. "MT@EC: Serving the Multilingual Needs of the European

> Commission." Paper presented at *8th Workshop on Statistical Machine Translation*,
>
> Sofia, August 8.

European Commission. 2012. *Translation Tools and Workflow*. Luxembourg: Publications

> Office of the European Union.

Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves, and Andy Way.

> 2014. "Perception vs Reality: Measuring Machine Translation Post-Editing
>
> Productivity." In *Proceedings of the 11th Conference of the Association for Machine*
>
> *Translation in the Americas: Workshop on Post-Editing Technology and Practice*
>
> *(WPTP3),* edited by Sharon O'Brien, Michel Simard, and Lucia Specia, 60–72.
>
> Vancouver: AMTA.

Guerberof Arenas, Ana. 2009. "Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation." *Localisation Focus* 7 (1): 11–21.

International Ergonomics Association. 2016. *Definition and Domains of Ergonomics.* Accessed June 7, 2016. http://www.iea.cc/whats/index.html.

Koskinen, Kaisa. 2008. *Translating Institutions: An Ethnographic Study of EU Translation*. Manchester: St. Jerome.

Koskinen, Kaisa, and Minna Ruokonen. Forthcoming. "Love Letters or Hate Mail? Translators' Affective Responses to Technology." In *Human Issues in Translation Technology*, edited by Dorothy Kenny. Milton Park, Abingdon, Oxon: Routledge.

Kluvanec, Daniel. 2013. "The Way Forward with MT@EC in 2014 and Beyond." Paper presented at *European Parliament Science and Technology Options Assessment*, Brussels, December 3.

Kluvanec, Daniel. 2014. "Getting the Right Mix: Approaches to Machine Translation in the European Commission." In *Proceedings of the XX*th *World Congress of the International Federation of Translators (Vol. I),* edited by Wolfram Baur, Brigitte Eichner, Sylvia Kalina, Norma Keßler, Felix Mayer, and Jeanette Ørsted, 51–57. Berlin: BDÜ.

Koponen, Maarit. 2012. "Comparing Human Perceptions of Post-Editing Effort with Post-Editing Operations." In *Proceedings of the 7th Workshop on Statistical Machine Translation*, 181–190. New York: Association for Computational Linguistics.

Lagoudaki, Elina. 2006. "Translation Memories Survey 2006: Users' Perceptions around TM Use." Paper presented at *Translating and the Computer 28*, London, November 16–17.

Lagoudaki, Elina. 2008. *Expanding the Possibilities of Translation Memory Systems: From the Translator's Wishlist to the Developer's Design*. PhD Thesis, Imperial College London.

Lavault-Olléon, Élisabeth. 2011. "Une introduction à la problématique « Traduction et Ergonomie » [Introducing Translation and Ergonomics]." *ILCEA Traduction et Ergonomie.* Accessed May 17, 2016. https://ilcea.revues.org/1118.

Le Blanc, Matthieu. 2013. "Translators on Translation Memory (TM): Results of an Ethnographic Study in Three Translation Services and Agencies." *Translation and Interpreting* 5 (2): 1–13. DOI: ti.105202.2013.a01

Liamputtong, Pranee. 2011. *Focus Group Methodology: Principle and Practice.* Thousand Oaks: Sage Publications.

Moorkens, Joss, Sharon O'Brien, Igor A. L. da Silva, Norma B. de Lima Fonseca, and Fabio Alves. 2015. "Correlations of Perceived Post-Editing Effort with Measurements of Actual Effort". *Machine Translation* 29 (3-4): 267–284. DOI: 10.1007/s10590-015-9175-2

Moorkens, Joss, and Sharon O'Brien. Forthcoming. "Assessing User Interface Needs of Post-Editors of Machine Translation." In *Human Issues in Translation Technology*, edited by Dorothy Kenny. Milton Park, Abingdon, Oxon: Routledge.

Morgan, David L. 1997. *Focus Groups as Qualitative Research.* Thousand Oaks: Sage Publications.

O'Brien, Sharon. 2012. "Translation as Human-Computer Interaction." *Translation Spaces* 1: 101–122. DOI: 10.1075/ts.1.05obr

O'Brien, Sharon, Maureen Enhrensberger-Dow, Marcel Hasler, and Megan Connolly. Submitted. "Irritating CAT Tool Features that Matter to Translators" submitted to

Hermes, Journal of Language and Communication, Special Issue on Translation Technology.

Pilos, Spyridon. 2014. "MT@EC – The New Machine Translation of the European Commission." Paper presented at *CEF Information Day*, Luxembourg, January 16.

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. DOI: 10.2478/v10108-010-0010-x.

Pym, Anthony, Alexander Perekrestenko, and Bram Starink. 2006. *Translation Technology and Its Teaching: (With Much Mention of Localization).* Tarragona: Intercultural Studies Group, Universitat Rovira i Virgili.

Risku, Hanna. 2010. "A Cognitive Scientific View on Technical Communication and Translation: Do Embodiment and Situatedness Really Make a Difference?" *Target: International Journal of Translation Studies* 22 (01): 94–111. DOI: 10.1075/target.22.1.06ris.

Risku, Hanna, and Florian Windhager. 2013. "Extended Translation: A Socio-Cognitive Research Agenda." *Target: International Journal of Translation Studies* 25 (1): 33–45. DOI: 10.1075/target.25.1.04ris.

Rummel, Dieter. 2015. *Machine Translation Annual Activity Report 2014*. Unpublished DGT intranet report. Accessed May 25, 2015.

Stewart, D. and Prem N. Shamdasani. 2015. *Focus Groups: Theory and Practice*. Thousand Oaks: Sage Publications.

Teixeira, Carlos S. C. 2014. "Perceived vs. Measured Performance in the Post-Editing of Suggestions from Machine Translation and Translation Memories" In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas:*

Workshop on Post-Editing Technology and Practice (WPTP3), edited by Sharon

O'Brien, Michel Simard, and Lucia Specia, 450–59. Vancouver: AMTA.

Van den Bergh, Jan, Eva Geurts, Donald Degraen, Mieke Haesen, Iulianna van der Lek-

Ciudin, and Karin Coninx. 2016. "Recommendations for Translation Environments to

Improve Translators' Workflows." Paper presented at *Translating and the Computer*

38, London, November 17–18.