User Attitudes to the Post-Editing Interface

Joss Moorkens
CNGL/SALIS
Dublin City University
Ireland
`joss.moorkens@dcu.ie`

Sharon O'Brien
CNGL/SALIS
Dublin City University
Ireland
`sharon.obrien@dcu.ie`

Abstract

At present, the task of post-editing Machine Translation (MT) is commonly carried out via Translation Memory (TM) tools that, while well-suited for editing of TM matches, do not fully support MT post-editing. This paper describes the results of a survey of professional translators and post-editors, in which they chose features and functions that they would like to see in translation and post-editing User Interfaces (UIs). 181 participants provided details of their translation and post-editing experience, along with their current working methods. The survey results suggest that some of the desired features pertain to supporting the translation task in general, even before post-editing is considered. Simplicity and customizablity were emphasized as important features. There was cautious support for the idea of a UI that made use of post-edits to improve the MT output. This research is intended as a first step towards creating specifications for a UI that better supports the task of post-editing.

1   Introduction

The use of machine translation (MT) as part of the localization workflow has mushroomed in recent years, with post-edited MT becoming an increasingly cost-effective solution for specific domains and language pairs. DePalma and Hegde (2010) stated that 42% of language service providers (LSPs) surveyed said that they offered post-edited MT to customers. At present, post-editing tends to be carried out via tools built for editing human-generated translations, such as translation memory (TM) or Translation Environment Tools (TEnT). These environments are fairly well suited to the task for which they were intended. However, it is our opinion that integration with machine translation and support for the post-editing task are not necessarily well catered for in current translation editing interfaces. This lack of support may lead to cognitive friction during the post-editing task and to reluctance among translators to accept post-editing jobs. This paper describes the results of a survey of professional translators and post-editors, in which they chose features and functions that they would like to see in translation and post-editing user interfaces (UIs). The survey is intended as a first step towards creating specifications for UIs that better support the post-editing task. Our starting point is that translators do not require a separate editor for post-editing, but rather that features could be integrated into existing commercial tools in order to better support the task and, ultimately, integration with MT systems.

Research on post-editing has tended to focus largely on rates of productivity. Recent papers have measured translation throughput, cognitive effort, quality (as perceived when compared with human translation), or have attempted to estimate MT quality via comparison of performance with automatic evaluation metrics (AEMs) (e.g. de Almeida and O'Brien, 2010; Specia and Farzindar, 2010, Koponen et al, 2012). This research has involved the use of commercial TM tools such as SDL Trados, proprietary tools such as Crosslang, or purpose-built tools for research that have simple UIs such as Caitra (Koehn, 2009) or PET (Aziz et al., 2012). There has, however, been little focus on the UI itself, or on the functionality required for the job of post-editing.

Vieira and Specia (2011) rated several text-editing tools used for post-editing, using various

criteria, including one of "interface intuitiveness". They acknowledge that this criterion was "highly subjective" as "its judgment was based solely on the experience of a single translator attempting to use the toolkits for the first time" (ibid.). The commercial TM tools rated all "put some effort into assigning intuitive meaning to the interface of the system" by utilizing color codes (ibid.), providing the source and target segments, including concordance search, and including dictionary and other display functions. The tools that they rated highest, however, "show clear evidence of collecting feedback from translators" (ibid.). Their wish list for a post-editing interface includes more sophisticated alignment, accurate confidence scores for MT proposals, and change tracking (included in subsequent versions of SDL Trados Studio), and they conclude that "a number of features deemed desirable for the work of a translator were not satisfactorily found in any of the tools analyzed" (ibid.).

Lagoudaki investigated text editing UIs as part of her TM survey in 2006. She found that, during development, TM users were usually "invited to provide feedback on an almost finished product with limited possibilities for changes" (2006). One translator in Moorkens (2012) said that developers had not understood her feedback as they had not worked as translators and "they don't know the problems you encounter or the things you would like to see". Lagoudaki also had the opinion that industry research is mostly motivated by "technical improvement of the TM system and not how the TM system can best meet the needs of its users" (2008). This runs counter to user-centered design recommendations, whereby a designer defines user profiles, usability requirements, and models before designing the UI (Redmond-Pyle and Moore, 1995).

Lagoudaki also wrote that "systems usability and end-users' demands seem to have been of only subordinate interest" in TM system development (2008). Based on her research, the message from the users of TMs is occasionally conflicting. However, she concludes that an overall message is clear: TM users want simplicity. This does not necessarily mean fewer features; rather they want a streamlined process with compatibility between languages and scripts. They want ease of access, meaning "affordability of the system, not only in terms of purchase cost, but also in terms of upgrade, support and training costs" (2008).

To better understand what features post-editors might require we designed a survey in which the questions focused on five areas in particular. (1) Participants were asked for some biographical details, such as years of professional experience, and about their attitude to technology. (2) They were asked about their current working methods, (3) what they would like to see in their ideal UI, (4) how they would like to see TM matches and MT output presented, and (5) about intelligent functionality that might help combine TM and MT matches. This survey is the first stage in a study that will be followed by interviews and observation, with the aim of creating specifications for a UI dedicated to the task of post-editing. Some interim results from the survey are contained in the following sections.

2 Survey Responses

The survey had 181 participants, of whom 102 completed the survey.[1] 121 participants completed the demographic section. The age range of participants is shown in Figure 1.
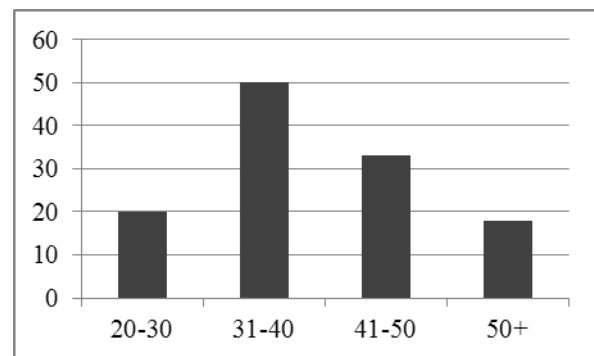

Figure 1: Age range of participants

Most participants reported that they had 8-10 years' experience as a translator, with nine participants claiming over 20 years' experience. Reported post-editing experience was, as may be expected, far shorter. 29 participants said that they had no experience as a post-editor, while most of the others who answered claimed 1-3 years' experience.

40% of participants in this section are free-lancers working independently without an agency, 30% work on a freelance basis with one or more agencies, and 22% are employees of a translation or localization company. Thus, the respondents represent a good spread of work profiles. Over half of the participants (69) said that they like using TM technology, whereas only 23 (19%) said that they like using MT. 79% of participants (95) said that they use TM because it helps their work, and 36% (43) felt the same way about MT. Just over 50% (61) said that MT is still problematic.

Of 114 participants, 100 translate from English, although many listed other source languages too. The target languages are listed in Table 1. The dominance of English as a source language is probably determined by the nature of the respondents and the companies who promoted the survey via their UK or Irish offices, many of whom operate in the IT localization domain. As can be seen, there is a reasonable spread of target languages.

Target Language	No.
Arabic	2
Chinese	16
Czech	4
Danish	1
Dutch	2
English	25
Finnish	3
French	12
German	16
Greek	2
Hindi	1
Hungarian	2
Italian	6
Japanese	4
Korean	1
Malay	1
Norwegian (Bokmål)	1
Polish	1
Portuguese	12
Russian	2
Spanish	8
Swedish	3
Thai	2
Turkish	3
Urdu	2

Table 1. Participants' target languages

2.1 Current Editing Environment

107 participants provided details of the editing environment(s) that they currently use for post-editing. Most participants (76 or 70%) use more than one environment regularly. 74 (69%) use a version of the SDL Trados TM tool. There was little difference in the rate of SDL Trados use between freelancers and company employees; 21 of 28 company employees use SDL Trados. SDL Trados was also listed as the most widely-used tool in Lagoudaki (2006) with a rate of 51% usage among respondents. Lagoudaki also found that company employees are more likely to use multiple tools. In the current survey, 18 of 28 (64%) company employees said that they use multiple tools, with a similar rate of 61% (46 of 75) among freelancers. Interestingly, 44 (41%) use Microsoft Word for post-editing, which suggests that, contrary to what might be deemed best practice, MT and TM are not combined in many instances. Figure 2 shows other tools used for post-editing and the actual number of users among survey participants. Eight participants also listed proprietary tools (Translation Workspace, Helium, and MS LocStudio). Some other tools used by fewer than eight participants were Passolo (6), OmegaT (5), Star Transit (2), TransStudio (1), Alchemy Catalyst (1), and Publisher (1).
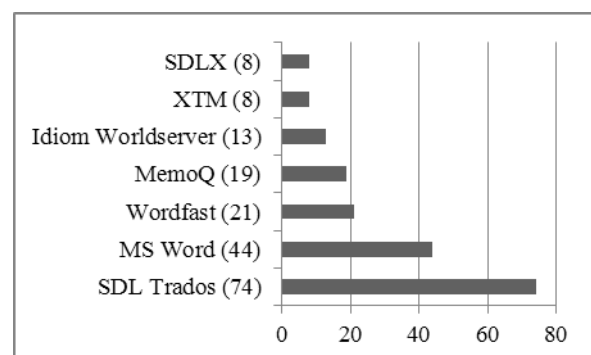

Figure 2. Tools used for post-editing

65% (80) said that they prefer to customize their editor rather than using the default set-up. Of those 80 respondents, 63 (79%) adjust their on-

screen layout, 59 (74%) adjust tag visibility, 57 (71%) adjust font type, and 19 (24%) adjust colors. Of the total number of participants who responded to the question about customization (124), roughly half are unhappy with the default layout, coloring, and display of tags in their editing tool.

2.2 UI Wish List

As might be expected from the results in the previous section, survey participants consider customizability as important. Of 119 respondents, 73 (61%) said that their ideal UI should be customizable, and 70 (59%) said that it should be clean and uncluttered. 63 (53%) always want to see the source segment alongside an MT suggestion, with 37 wishing to see them aligned horizontally and 36 preferring vertical alignment. 72 (58%) always want to see an approved glossary, although 37 (30%) would like to be able to hide or unhide the glossary.

Participants were asked what currently unavailable features they would like to see in a post-editing environment. Several suggested dynamic changes to the MT system in the case of "a recurrent MT error which needs to be fixed many times during the post-edition". While not a trivial demand, this is an important indicator that MT and UI developers need to find efficient solutions for 'on-the-fly' improvements of MT output. Others requested a global find-and-replace function, dictionary plug-ins, reliable concordance features, and QA interoperability (particularly with ApSIC Xbench). What is perhaps most striking about these responses is that only one pertains specifically to post-editing and the other features can be seen as features that are desirable *in general* for supporting the translation task.

Most participants rely heavily on keyboard shortcuts. Of 119, 33% stated that they use keyboard shortcuts often, and 37% use them very often (4% never use them). 82% feel that using keyboard shortcuts improves their productivity. There are some conventions for shortcuts or text selection that users are likely to have become accustomed to. While it may be best to use these conventional shortcuts not to "reinvent too many wheels at once" (Tidwell, 2005), participants in this survey were asked whether certain shortcuts would be useful specifically for the task of MT post-editing, which can demand significant key-

boarding effort. Their responses are shown in Table 2.

Shortcut	No.
Dictionary search	103
One-click rejection of MT suggestion	96
Web-based parallel text lookup	92
Change capitalization	81
Add/delete spaces	67
Apply source punctuation to target	61

Table 2. Keyboard shortcuts requested

Again, the most popular shortcut (dictionary search) and the suggestion of a shortcut for web-based parallel text lookup are not post-editing-specific. The choice by 96 participants (81%) of a keyboard shortcut for a one-click rejection of an MT suggestion is specific to post-editing, but when taken in conjunction with the 50% who had previously said that MT is still problematic, may suggest apprehension about MT quality or usefulness among the participants. Post-editing guidelines often encourage post-editors to use as much of the MT output as possible. At the same time, segments that are completely unusable are still relatively frequent, so this one-click rejection button might actually save time. 68% suggested a keyboard shortcut for changing capitalization, recognizing that letter casing is still problematic in MT output.

Language-specific keyboard shortcut suggestions were less popular, possibly due to the large variety of target languages among participants. The most popular suggested shortcut would change the number of a word (e.g. from singular to plural), but less than half of the participants (59 of 123 or 48%) considered that such a shortcut may be useful. Further responses are shown in Table 3 with proposed shortcuts in one column and the number of respondents who said this would be useful in the other.

Shortcut	No.
Change number (sing./pl.)	59
Adjust word order	58
Change gender	48
Change verb form	46
Add/delete postposition	45
Add/delete preposition	43

Add/delete conjunction	40

Table 3. Language-specific keyboard shortcuts

Participants' comments explained their misgivings relating to these shortcuts. Some were in favor of the shortcuts: "Changing gender and number of words with shortcuts would be something very useful, in my opinion." Many could not understand how they might work in practice. "Finnish is such a complicated language that those kinds of shortcuts probably wouldn't work properly," wrote one. Several thought that manual changes would be easier or less time-consuming than memorizing a large number of shortcuts. "Frankly, oftentimes it takes you less to overwrite/type what you need than learning and applying many shortcuts." Other participants did not consider the shortcuts relevant for their language pairs. "Few of the above suggested features apply to the languages I use." Participants would, however, be in favor of customizable shortcuts. 66% (81) would like to be able to add macros or scripts to adapt the UI functionality, such as adding new keyboard shortcuts. 50% (61) would like to see a guided method to help them create such a macro.

2.3   Match presentation

Most participants (93 or 79%) agreed that they would like to see MT engine confidence scores in the editing environment. Of these 93, 68 (73%) favored the presentation of such scores in the format of percentages, like a fuzzy match score in a TM tool, while 22 (24%) chose a scheme of color coding to denote confidence. Participants were asked what they would like to be shown in the case where an MT match received a higher confidence score than any fuzzy match available from the TM. 106 (90%) said that they would like to see both MT and TM matches. Apprehension about MT quality is again evident as only two participants (less than 2%) said that they would like to see the MT match only, whereas 6 said that they would still like to see the TM match only, ignoring the higher-rated MT match. This apprehension is also evident in responses to the question (shown in Figure 3): below which fuzzy match value would you prefer to see an MT rather than TM fuzzy match?
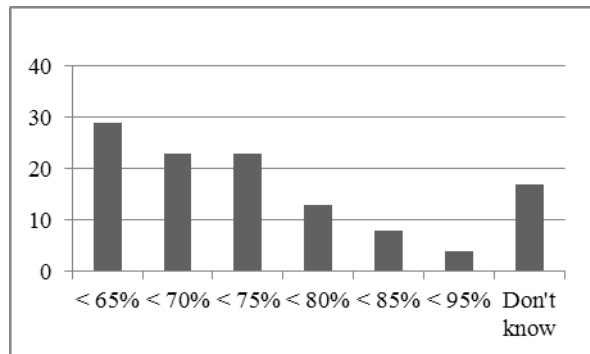

Figure 3. Fuzzy match limits

If the editing environment could combine the best sub-segments from the MT system and the TM to produce an MTM (Machine Translation/Translation Memory) match, 76 participants (64%) said that they felt it would be useful. 85 (72%) would like to see this marked as an 'MTM' match.

Comments for this question showed mixed opinions. Some participants feel positively about a potential MTM match: "Now THAT's a good idea! Somehow dynamically combine the MT output with e.g. MemoQ's longest subsegment concordance." Others are skeptical about the value of any inclusion of MT: "MT is still in baby's shoes, and the quality is horrible, so MT is not very useful in general."

Some participants commented that they would rather have the provenance of MT or TM suggestions kept separate: "It's cognitively difficult enough to distinguish between Fuzzy errors and MT errors." This was shown to be a widespread view when participants were asked whether they would like to see the provenance denoted by color at a sub-segment level. The answer was an overwhelming 'yes' with 104 (88%) in favor. Throughout the survey, it became clear that meta-data showing the origin of match suggestions is important to translators and post-editors. Despite some misgivings about MT, 99 (84%) would like to see 'the best MT suggestion' automatically appear in the UI target window when no TM match is available.

2.4   Intelligent functionality

Participants were asked their opinion of some functions that have been suggested for post-editing of MT, such as interactive machine translation (IMT), whereby human edits are used by an MT system as "additional information to

23

achieve improved suggestions" (Alabau et al., 2012).

72% (83) said that, when working with a client-specific MT system, their edits should be used to improve the MT system. A further 24 (21%) were unsure, with concerns evident in the comments. Some were concerned about issues relating to confidentiality, while others resented further reuse of their translation work. This intellectual property concern was emphasized by one participant to write: "Who would pay a translator for his intellectual work in improving the TM/MTM? Generated content is usually considered property of the agency for which a freelancer works", adding "they'll sell you out any time, any way".

3   Conclusion and future work

Several works have referred to TM tool users' dissatisfaction with their current editing environments (Lagoudaki, 2008; McBride, 2009). In this paper, participants again expressed dissatisfaction with their current editing environment. What was striking here was that many comments pertained to translation editor UIs in general, which seem to still have many short-comings even before post-editing of MT output is considered. Participants emphasized the importance of customizability in their ideal UI. They would like their UI to be clean and uncluttered, and to have plugins for dictionary and Internet search, and for improved concordance search. Most participants currently use a version of SDL Trados for post-editing, and most use multiple tools. 41% use Microsoft Word, suggesting that their current workflow does not combine MT and TM.

Participants would like to see further keyboard shortcuts added to their editing environment for such functions as dictionary search, to remove an MT suggestion, or to change capitalization in their target window. However, they are more circumspect when it comes to complex, language-specific shortcuts that could change word order or gender, due to skepticism over how these functions would work in practice.

Participants have qualified enthusiasm for sub-segment integration of MT and TM, amidst concern over how well this integration might work, and over how to display the provenance of the suggested target text. While some were con-cerned about intellectual property and confidentiality issues, they were largely in favor of dynamic improvements being made to their MT suggestions by incorporating post-edits in an IMT system.

The survey on which this paper is based is still ongoing at the time of writing. Several survey participants have chosen to waive their right to anonymity within the survey in order to participate in follow-on interviews. Based on the survey results and on these interviews, we intend to complete a specifications document for a post-editing UI, and to apply those specifications in building a new prototype UI for test purposes.

4   Acknowledgement

References

Alabau, Vicent, Luis A. Leiva, Daniel Ortiz-Martínez and Francisco Casacuberta. 1997. User Evaluation of Interactive Machine Translation Systems. *Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT)*:20-23.

Aziz, Wilker, Sheila C. M. de Sousa, and Lucia Specia. 2012. PET: a tool for post-editing and assessing machine translation. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 12),* Istanbul, Turkey:20-23.

De Almeida, Giselle and Sharon O'Brien. 2010. Analysing post-editing performance: correlations with years of translation experience. *EAMT 2010: Proceedings of the 14th Annual conference of the European Association for Machine Translation*, 27-28 May 2010, Saint-Raphaël, France. Proceedings ed. Viggo Hansen and François Yvon:8.

DePalma, Donald A., and Vijayalaxmi Hegde. 2010. *The market for MT post-editing*. Commonsense Advisory, Boston, MA.

Koehn, Philipp. 2009. A Web-Based Interactive Computer Aided Translation Tool. *ACL Software demonstration.*

Koponen, Maarit, Wilker Aziz, Luciana Ramos, and Lucia Specia. 2012. Post-editing time as a measure of cognitive effort. *AMTA-2012: Workshop on*

post-editing technology and practice, San Diego, October 28, 2012:10.

Lagoudaki, Elina. 2006. Translation Memories Survey 2006: Users' Perceptions Around TM Use. *Proceedings of ASLIB Translating and the Computer 28*. London, UK. 15-16 November 2006.

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.

McBride, Cheryl. 2009. *Translation Memory Systems: An Analysis of Translators' Attitudes and Opinions*. MA thesis. University of Ottawa.

Moorkens, Joss. 2012. *Measuring Consistency in Translation Memories: A Mixed-Methods Case Study*. PhD thesis. Dublin City University.

Redmond-Pyle, David and Alan Moore. 1995. *Graphical User Interface Design and Evaluation*. Coventry, UK: Prentice Hall.

Specia, Lucia and Atefeh Farzindar. 2010. Estimating machine translation post-editing effort with HTER. *JEC 2010: Second joint EM+/CNGL Workshop - Bringing MT to the user: research on integrating MT in the translation industry*. AMTA 2010, Denver, Colorado, November 4, 2010:33-41.

Tidwell, Jenifer. 2006. *Designing Interfaces.* Sebastopol, CA: O'Reilly.

Vieira, Lucas N. and Lucia Specia. 2011. A Review of Machine Translation Tools from a Post-Editing Perspective. *Proceedings of the Third Joint EM+/CNGL Workshop - Bringing MT to the User: Research Meets Translators (JEC '11)*. Luxembourg, 14 October 2011:33-42.