The potential and limits of lay post-editing in an online community

Linda Mitchell
CNGL/School of Applied Language and Intercultural Studies
Dublin City University
Dublin, Ireland
linda.mitchell7@mail.dcu.ie

Abstract

This paper aims at exploring the potential of a lay community as post-editors. It focusses on 15 members of an online technology support forum, native speakers of the target language (TL) and some knowledge of the source language (SL) translating content that was machine translated from English into German specific to their own domain. It presents the most predominant errors remaining in the post-edited output and the impact of these on the quality of the post-edited output as measured by domain specialists evaluating adequacy and fluency. This paper further explores examples of these errors and possible solutions to reducing the occurrence of these and maximising the community's potential. The targeted post-editing quality was "'good enough", as determined in the post-editing guidelines. The PE results demonstrate that there is still room for improvement in terms of quality.

1 Introduction

User-Generated Content (UGC) is constantly growing online. With that growth, the demand for translations with fast turnaround times increases, too. Common solutions to meet this demand are MT or a combination of MT and Post-Editing (PE), the correction of automatically translated text. Previous research in the field of PE has predominantly focussed on professional translators, e.g. de Almeida (2013) who investigates correlations between translation experience and post-editing quality, Plitt and Masselot (2010) who compare results of traditional Human Translation to post-editing in an industrial setting using the example of Autodesk, or Guerberof (2009) who compares productivity between translation with Translation Memories (TM) to post-editing of MT content. Participants within these experiments have been found to experience adverse feelings towards post-editing (e.g. de Almeida 2013), which is not correlated to their often excellent performance. There have further been studies with translation students as post-editors, e.g. Koponen (2013), who investigates variation in post-editing preferences or Depraetere (2010), who seeks to establish strategies in post-editing behaviour of translation students. Recently, there have been studies investigating individual lay people or communities as post-editors, such as with subjects who have some knowledge of both the SL and the TL but who are untrained in translation (Aranberri et al. 2014) or domain specialists who are untrained in translation and have no knowledge of the SL (Schwartz 2014).

While the studies presented above have mostly been of hypothetical nature here, we are tapping into a new pool of potential post-editors by focussing on an already existent and real online community. This community is a technology support forum, the Norton[1] Community, a platform facilitating discussion on the Norton products among the users of the products in order to solve problems they may be experiencing with additional guidance from Symantec employees. It is a small community, which sets itself apart from communities discussed in previous translation research, as it is not based on social media, such as Facebook or Twit-

---

[1]Norton is a sub-division of Symantec Corporation.

ter. In order to determine the potential of lay post-editing, we compare the post-editing behaviour of lay post-editors[2] to that of professional translators or any other post-editor types that have been reported on. In the following, such studies are presented to facilitate a comparison with the current research.

2 Related Research

De Almeida (2013) investigates the post-editing behaviour of 18 professional translators with varying degrees of experience in translation and in PE for English → French and English → Brazilian Portuguese. She focusses on essential, preferential changes and newly introduced errors. She finds that her subjects do not only implement essential changes but that they also implement additional preferential changes, e.g. of stylistic or synonymical nature, even though they were instructed not to do so. She uncovers a tendency of professional translators to over-edit the text, which could render the post-editing process less efficient.

Groves and Schmidtke (2009) investigate common edits performed on MT output using the Microsoft Treelet MT engine focussing on EN → DE and EN → FR, employing 3 professional translators with varying degrees of translation experience and productivity. For both language pairs they identify the insertion and deletion of function words as the most common edit, 42% of which consist of determiners (for German), predominantly the insertion of the determiner *die*. They further note changes in punctuation as being among the most common edits in their data and point out the frequent deletion of the pronoun *Sie*, which is typically inserted by their MT system. They hope to resolve these issues with automatic statistical post-editing. Groves and Schmidtke do not report on any errors that remain in the post-edited output.

Depraetere (2010) focusses on post-editing training of translation students looking at the language pair English → French. Hypothetically, these students would not have established automised translation routines yet and would be more open to new translation techniques and would be closer to lay post-editors in behaviour than professional translators. She investigates the post-editing results of 10 such students, and how they intu-

itively approach post-editing with the view to creating appropriate post-editing guidelines that focus on the errors ignored. Depraetere finds that, in contrast to professional translators, there are few stylistic or phrasal ordering changes implemented. More importantly, she finds that the students adhered strictly to the guidelines, which were to "make sure that the source text and the target text were informationally similar and that the target text was grammatically correct" (Depraetere 2010: 3), i.e. they performed minimal edits. Their behaviour involved, for example, accepting literal translations by the MT system that are not equivalent to the source text. She concludes that there are no clear post-editing strategies present on a micro level, i.e. the types of errors corrected.

It can be concluded that professional translators tend to post-edit more systematically, frequently correcting the same errors (Groves and Schmidtke 2009) and that they tend to over-edit MT output by implementing preferential changes to render the writing style closer to their own (de Almeida 2013). Although the manner of measuring errors/changes differs between these studies and the approach taken in this paper, they give an important indication. Furthermore, translation students seem to under-edit the machine translated output and strictly adhere to the guidelines they are presented with (Depraetere 2010).

Moorkens and O'Brien (2015) also focus on differences between novice translators and professionals in the frame a post-editing user interface study. They find that professionals are more efficient but that their working habits and attitudes may prevent them from following the structure of the experiment as intended. They conclude that while novices may be the group that is more easily engaged in research, their results cannot be carried over to professionals.

Čulo et al. (2014) investigate how post-editing affects typical translation strategies for 12 professional translators and 12 translation students (English → German) by comparing translations, monolingually post-edited and bilingually post-edited (with access to the ST) texts. Čulo et al. disprove the claim that bilingual post-editing produces as high a quality as human translation (HT). Based on their examples, they hypothesise that errors and interference effects may be based on the post-editing rules for light PE and on the MT output.

---

[2]We define a lay post-editor as anybody who is not a professional translator or translation student.

PE rules have been addressed on a theoretical level by Rico and Ariano (2014). They seek to establish a framework for developing language dependent PE guidelines (English → Spanish) These were based on an analysis of the MT output and PE patterns that emerged incorporated into a flexible decision tool. While this approach appears to be successful, it is unsuitable for the purpose of the current study. Rico and Ariano do not deal with SMT, nor does their project support German. Their PE guidelines are targeted at professional post-editors or translators and are unsuitable for lay post-editors, as the guidelines require linguistic knowledge to be understood.

The aim of the experiment described in this paper was to uncover the post-editing approach taken by lay post-editors and how it fits into the current body of research. Furthermore, this study aims at identifying the potential of a lay community as post-editors of content that is relevant to their domain. Firstly, we aim to identify the number of errors corrected and the number of errors that remain in the post-edited output of the lay post-editors. This paper subsequently sets out to present strategies to maximise a lay community's potential for post-editing.

3 Experimental Design

The participants for this experiment were recruited online in the German Norton Community by means of private messaging and a publicly posted open call for participation. Fifteen native speakers of German with some knowledge[3] of English, who post-edited bilingually were considered only, with the aim to eliminate outliers based on their language skills. [4] The participants were members of the Norton Community and were familiar with the domain of the Norton products. Each participant post-edited 12 tasks.[5] The texts for the tasks were extracted from the English-speaking Norton community. They were machine translated from English into German using the ACCEPT baseline SMT engine, which is based on Moses, as described in ACCEPT (2012) in the ACCEPT portal.[6] The language pair English → German was

chosen here, as it is particularly challenging for MT engines because of the differences in syntax between the languages.

Figure 1 displays the (German) post-editing interface the lay post-editors used to post-edit the machine translated content.
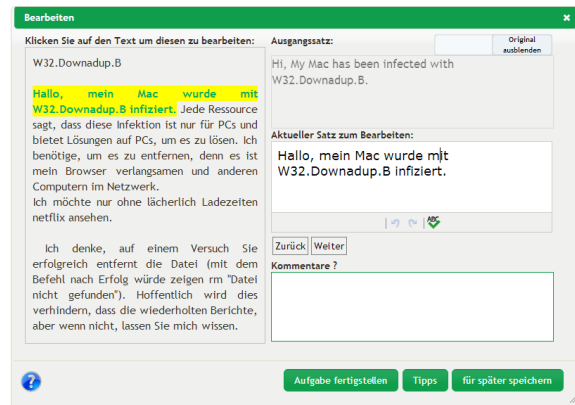

Figure 1: Post-editing Interface (German)

On the left, the machine translated text is displayed with the original version of the current segment displayed on the top right and the current segment to be edited below. Of particular importance here is the middle button on the bottom, "Tipps", which displayed the post-editing guidelines when clicked. These are presented below.
Tips for post-editing:

- Edit the text to make it more fluent and clearer based on your interpretation.

- Try to correct phrasal ordering and spelling, for example, if they make the text hard or impossible to understand.

- Use words, phrases or punctuation as they are, if they are acceptable.

- If you are working with reference to the original text, make sure that no information has been added or deleted.

These guidelines were developed for both monolingual and bilingual post-editing based on TAUS' guidelines to achieve post-editing quality that is "good enough". With these, it was hoped to keep over-editing, identified as being prevalent amongst professional translators (de Almeida 2013), to a minimum. Similar to Depraetere (2010), we sought to establish a baseline of PE behaviour in an online community of lay posteditors.

---

[3] 'Some knowledge' here refers to the categories B1 to C2 as defined in CEDEFOP 2011.

[4] This work is based on the post-editing and post-editor data collected in a study as described in Mitchell (2015).

[5] Each 'task' contained a subject line, a question and the accepted solution to that question.

[6] www.accept-portal.eu

The post-editing results were analysed based on the error categorisation proposed by (de Almeida 2013), which was developed for investigating post-editing behaviour. The error annotation was performed by the author of this paper, a native speaker of German. While it may be argued that one annotator is not sufficient in evaluating the content, it was considered an appropriate solution here, as recent studies have shown that achieving annotator agreement is difficult, due to ambiguity of categories, disagreement on whether a construct is an error or not and disagreement on the spans of errors (e.g. Lommel 2014). Agreement on spans of errors is particularly important for phrasal ordering, an error type that was expected to be of significance in this study, due to the language pair English → German. The categories of preferential and essential changes were dropped, as the participants were assumed to be untrained in translation and would therefore not have experience of these concepts. The main categories used for this experiment were 'Language', 'Accuracy' and 'Format' (based on de Almeida 2013: 95) with the category Language including adjectives, adverbs, capitalisation, conjunctions, determiners, gender, nouns, number, phrasal ordering, prepositions, pronouns, punctuation, spelling and verb tense; the category Accuracy includes extra information, information missing, untranslated information and mistranslation; and Format consists of additional or missing spaces. For the error annotation all machine translated content (322 segments) and 44% of the post-edited content (4 post-edited versions per MT segment), approximately 72 segments per post-editor, was randomly selected and evaluated.

4   Results and Analysis

In order to demonstrate the finding that post-editing by a lay community is feasible, the errors corrected by the participants are presented here. The results can be found in Table 1, displaying the Total number of errors corrected, the number of errors in the Language, Accuracy and Format categories in both absolute numbers and percentages, compared to the errors that had been present in the raw MT output.

It is evident that the lay community was able to correct on average 73% of all errors, with the lowest number being 21% (PE13) and the highest number being 83% (PE8). The average number of errors corrected in the Language category

	Total	%	Lang.	%	Acc.	%	F.	%
PE1	199	66	95	61	100	79	-4	-44
PE2	182	69	88	65	88	74	5	63
PE3	215	81	103	80	101	81	4	80
PE4	254	81	133	81	120	90	1	10
PE5	248	78	116	70	114	88	6	75
PE6	231	87	114	84	107	91	2	40
PE7	164	57	69	44	89	75	-1	-33
PE8	223	83	116	85	94	81	6	67
PE9	215	80	107	80	102	84	-2	-40
PE10	213	77	113	74	94	85	0	0
PE11	222	78	114	75	101	86	-1	-25
PE12	234	81	122	84	105	83	-2	-33
PE13	51	21	67	48	-8	-8	-8	-800
PE14	227	71	131	76	97	75	-9	-225
PE15	193	78	98	77	86	82	1	17
Avg.	205	73	106	72	93	76	0	-57

Table 1: Errors corrected (absolute and in %) for the Total number of errors corrected, errors in the Language, Accuracy and Format category (also absolute and in %)

are 72% with 48% (PE13) as the lowest and 85% (PE8) as the highest number and 76% on average and -8% (PE13) as the lowest, i.e. 8 errors were introduced in the post-edited output, and 91% (PE6) of all errors of the Accuracy category corrected. While none of the post-editors corrected 100% of the errors and there is great variation across the lay post-editors, these results show the potential of lay post-editing, especially with the examples of PE3, PE4, PE6, PE8, PE9, PE12, who correct ≥ 80% of all errors. It should be noted that even in studies with professional translators acting as post-editors, it is often reported that errors remain in the post-edited output. Furthermore, the guidelines used targeted "good enough" post-editing quality, rather than aiming at the best humanly possible quality. Thus, we did not expect a correction rate of 100% of all errors.

In order to interpret this data, the need to investigate the profile of the lay post-editors arises. This was discussed in Mitchell (2015:166-176); Section 7 focusses on the post-editor profile, i.e. language competence, domain competence and psychomotor competence. While the first two were measured by self-reporting, the last was based on key logging data recorded in the ACCEPT portal. We found that there were no correlations between any of these competences and the post-editing quality, represented by both the error annotation and the domain specialist evaluation. Hence, the post-editor background was not deemed to be a helpful variable in the light of this article.

In the following, an overview of the main errors

that remain in the post-edited output and how they affect the quality of the same are presented in order to propose strategies for maximising the potential of a lay community as post-editors. Table 2 displays the average number of errors across all post-editors in the MT output they were editing, as well as in the post-edited content. These are ordered by the most frequent errors present in the MT output. It emerges that errors from the Accuracy category were the most common: mistranslation, information missing and extra information, as well as errors that emerge from the differences in syntax between English and German, i.e. phrasal ordering and verb (tense), the latter of which often manifests itself as a missing part of the verb, determining the correct tense. Additional sub-categories of Language that contained a considerable number of errors in the content annotated were determiner and pronoun.

Category	Errors (MT)	Avg.	Errors (PE)	Avg.
Mistranslation	873	58	162	11
Phrasal ordering	791	53	100	7
Information missing	526	35	161	11
Verb (tense)	271	18	40	3
Extra information	256	17	48	3
Determiner	247	16	50	3
Pronoun	150	10	28	2
Untranslated	146	10	40	3

Table 2: Errors present in MT output and after PE in total and on average across all post-editors

Figure 2 displays the absolute number of errors for the main error category Accuracy for each post-editor, in order to establish how they handled the errors present in the categories extra information, missing information, untranslated and mistranslation individually. The most frequent errors, which post-editors failed to correct/introduced, were either mistranslations or information missing. The number of remaining errors ranged between 3 and 17 for mistranslation and 2 and 13 for the category information missing. Figure 2 also reveals an outlier in the post-editing behaviour, PE13.

While PE13 accounts for nearly 30% of the errors in these two categories, the categories of untranslated and extra information contain numbers of errors that are comparable to those of the other post-editors. The category untranslated information seems to contain the lowest number of errors mostly, followed by the category of extra information, predominantly ≤ 10 errors per post-editor.

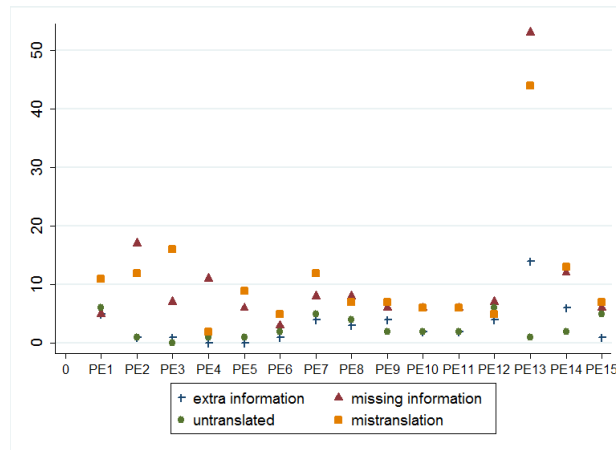Compared to the number of Accuracy errors,

Figure 2: Errors remaining (absolute) in PE output in Accuracy category
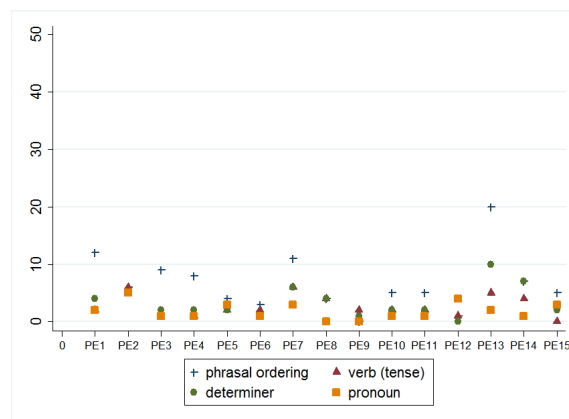

Figure 3: Predominant errors remaining (absolute) in PE output in Language category

as presented in Figure 2, considerably fewer Language errors remained in the MT output, as evident from Figure 3. Phrasal ordering errors, as expected, takes the highest rank of errors remaining in the post-edited output per evaluator. This echoes Depraetere's (2010) finding that post-editors who are not professional translators leave MT output unedited if the sentence is comprehensible and often leave literal translations untouched even though they may not convey the source text meaning accurately. The ratios of types of errors remaining (phrasal ordering, determiner and pronouns) are similar across all post-editors except for PE12 and PE13.

Examples:
The following three examples show minimal editing to no editing. For mistranslation, the error stems from the incorrect conjunction *weil* in

the MT output that was mistakenly retained; for phrasal ordering, the post-editor creates an awkward construction and did not correct the phrasal ordering; for information missing, the machine translated output was left unedited, which results in the conjunction (a function word) missing.

Mistranslation:
ST: ... I bought a new laptop *and*[7] problem was more prevalent.
MT: ... *weil* ich eine neue Laptops und Problem weiter verbreitet wurde.
HT: ... ich kaufte einen neuen Laptop *und* das Problem wurde noch schlimmer.
PE: ... *weil* das Problem durch den Kauf eines neuen Laptop schlimmer wurde.

Phrasal ordering:
ST: Purge *in progress* for 2-1/2 days... help!
MT: Bereinigung *in Fortschritt* für 2-1 / 2 Tage... zu helfen!
HT: Bereinigung seit zweieinhalb Tagen *in Arbeit*...Bitte um Hilfe!
PE: Bereinigung *in Arbeit* seit zweieinhalb Tagen.... Bitte um Hilfe!

Information missing:
ST: First, make sure Teamviewer is not running.
MT: Stellen Sie zuerst sicher TeamViewer wird nicht ausgeführt.
HT: Stellen Sie zuerst sicher, *dass* TeamViewer nicht ausgeführt wird.
PE: Stellen Sie zuerst sicher TeamViewer wird nicht ausgeführt.

The following example focusses on extra information that has been accidentally introduced by retaining the verb from the MT output and inserting it in the wrong position in the sentence. This shows that the placement of verbs not only poses problems for MT engines but also for human lay post-editors.

Extra information:
ST: In almost all cases, the events are caused by legitimate programs or Windows processes...
MT: In fast allen Fällen *verursacht* werden, die Ereignisse von legitimen Programmen oder Windows Prozesse...
HT: In fast allen Fällen werden die Ereignisse von legitimen Programmen oder Windows Prozessen verursacht...
PE: In fast allen Fällen *verursacht* werden diese Ereignisse von legitimen Programmen oder Windows Prozesse *verursacht*...

In addition to retaining errors from the MT, post-editors produce verbal errors. The next example

requires a complex verb construction, which the post-editor failed to produce. While the post-editor does produce a correct sentence compared to the MT output, it does not accurately reflect the meaning of the source text. This may have also been due to insufficient knowledge of the source language.

Verb (tense):
ST: However, NU should not have been deleting this.
MT: Aber NU *sollten nicht gelöscht wurden*.
HT: Aber NU *hätte* diese *nicht löschen sollen*.
PE: NU *sollte* diese Elemente *nicht löschen*.

The following two examples involve the changing of function words, i.e. determiners and pronouns. While the first one shows that the post-editor discards the correct determiner as suggested by the MT system, the second example shows that the post-editor deemed the sentence comprehensible and left it unedited. In addition, the meaning is completely mistranslated, which was not picked up on by the post-editor.

Determiner:
ST: Check if it runs *the* scans or detects any threats.
MT: Überprüfen Sie, ob es führt *die* Scans oder erkennt Bedrohungen.
HT: Überprüfen Sie, ob *die* Scans ausgeführt oder Bedrohungen erkannt werden.
PE: Überprüfen Sie, ob *der* Scans ausgeführt wird oder Bedrohungen erkannt werden.

Pronoun:
ST: All *I* care about is the first C: drive in this list.
MT: *Ich* interessiert, ist das erste Laufwerk C: In dieser Liste enthalten.
HT: *Mich* interessiert nur das erste Laufwerk, C:, in dieser Liste.
PE: *Ich* interessiert, ist das erste Laufwerk C: in dieser Liste enthalten.

The last example shows insufficient knowledge of the source language or the domain. The post-editor left 'Antivirus License Be' unedited, possibly because they assumed it to be the correct term or were unable to translate it.

Untranslated:
ST: Can Norton Antivirus *License Be* Transferred From One Computer To Another?
MT: Kann Norton Antivirus *License Be* übertragen von One Computer So Anderer?
HT: Kann ich die Norton Antivirus *Lizenz* auf einen anderen Computer übertragen?
PE: Kann ich die Norton Antivirus *License Be* auf einen anderen Compter übertragen ?

---

[7]Italics were added in all examples to highlight the translation problem of interest.

In summary, errors may be caused by erroneous source texts, by insufficient knowledge of the source language (or the domain), which leads to mistranslations and retaining errors as introduced by the MT output or introduced by the post-editors. Other times errors may be caused by editing hastily and producing errors that could have been easily avoided. This is not an unexpected situation in a lay post-editing scenario as lay post-editors are assumed to be untrained in translation and proof-reading. Furthermore, post-editors often left segments unchanged that needed editing in regards to syntax, e.g. phrasal ordering and verbs. When it comes to insufficient knowledge of the SL, it would be beneficial for the post-editors to have an option to send the segments in question to another post-editor/professional translator, a solution suggested by Schwartz (2014). Errors stemming from editing too hastily and too little editing could be addressed through revised post-editing guidelines. A factor that may have influenced PE quality negatively is motivation. Revised guidelines could increase the post-editors' knowledge of how to post-edit successfully on a theoretical level. On a practical level, however, they would also need to be motivated to to implement the required changes. We believe that motivation is a complex aspect that has an impact on lay post-editing quality. While motivation was not addressed here, it would be beneficial if it were considered in the future. Hence, we developed the following amended post-editing guidelines focussing on German as a target language.[8] They focus on the most common errors that remain in the post-edited output and do not include a reference to reusing as much of the MT output as possible (cf. Depraetere 2010). The guidelines were written in a clearer and a more concise manner. Extra items have been added for the mistakes that were predominant in the post-edited output: verbs, determiners, pronouns and mistranslations.

Tips for post-editing:

- Correct spelling, grammar and word order errors.

- Pay particular attention to verbs, determiners *(e.g. der, die, das)* and pronouns *(e.g. ich, du, er, mich, dich, sich)*.

- Correct any mistranslated information in the MT output.

- Ensure that no information has been added or deleted from the original text.

With these guidelines, we hope to point lay post-editors in the right direction without confusing them about the goal of their post-editing. Furthermore, they may be a first step in maximising the community's potential and aiming for post-editing quality that is better than "good enough", which is the quality that was targeted for the purpose of this experiment. These revised guidelines would only be a successful solution, however, if they were interpreted correctly by the post-editors and if they had the motivation required to implement them. It remains to be seen whether additional post-editor training in this regard would be helpful and feasible in an online community with volunteer post-editors.

5 Conclusion

It can be concluded from this sample that lay post-editors correct on average around 74% of errors in total. Furthermore, it was found that the most common errors remaining were all categories associated with Accuracy (mistranslation, information missing, extra information and untranslated information) and the following categories associated with Language; phrasal ordering, verb, determiner and pronoun. This fits with findings by Depraetere (2010), in terms of syntactical changes, and with Groves and Schmidtke (2009) in terms of the changing of function words, mainly determiners and pronouns, which occur often in our sample and are not always corrected.

Additionally, the ratios of errors remaining in both the Language and the Accuracy category are quite similar across the post-editors, i.e. they systematically leave some errors uncorrected in the MT output, which corresponds to de Almeida's findings. However, rather than over-editing the text, they adhere to the guidelines and leave literal translations or awkward sentence structures unedited, as described by Depraetere (2010).

Furthermore, the errors remaining in the post-edited output were due to unedited (portions of) segments, insufficient knowledge of English or hurried editing, which could be resolved by passing on 'complicated' segments to more competent editors and by having more 'tuned' post-editing

---

[8]These could be easily adapted to suit other languages by replacing the error categories here with ones specific to these languages.

guidelines, here tailored to German as a target language.

References

ACCEPT Consortium. 2012. *Baseline machine translation systems*. Deliverable 4.1. Available from: `http://www.accept.unige.ch/Products/D_4_1_Baseline_MT_systems.pdf`.

Aranberri, Nora, Gorka Labaka, Arantza Diaz de Ilarraze and Kepa Sarasola 2014. Comparison of postediting productivity between professional translators and lay users. *Proceedings of the Third Workshop on Post-editing Technology and Practice*, Vancouver, Canada. 20–34.

Čulo, Oliver, Silke Gutermuth, Silvia Hansen-Schirra and Jean Nitzke 2014. The influence of post-editing on translation strategies. *IN: Post-editing of Machine Translation* Laura Winther Balling, Michael Carl, Michel Simard, Lucia Specia and Sharon O'Brien (eds.). 200–218.

de Almeida, Giselle 2013. Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages *PhD thesis*, Dublin City University.

Depraetere, Ilse 2010. What counts as useful advice in a university post-editing training context? Report on a case study. *EAMT 2010: Proceedings of the 14th Annual conference of the European Association for Machine Translation,* 27-28 May 2010, Saint-Raphaël, France.

Groves, Declan and Dag Schmidtke 2009. Identification and Analysis of Post-Editing Patterns for MT *MT Summit XII: Proceedings of the twelfth Machine Translation Summit, August 26-30, 2009.* Ottawa, Ontario, Canada. 429–436.

Guerberof, Ana 2012. Productivity and quality in the post-editing of outputs from translation memories and machine translation. *PhD thesis*, Unversitat Rovira.

Koponen, Maarit 2013. This translation is not too bad: an analysis of post-editor choices in a machine translation post-editing task. *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP), 2-6 September.* S. O'Brien, M. Simard and L. Specia (eds.). Nice, France, 1–9.

Linguistic Data Consortium. 2005. Linguistic data annotation specification: assessment of fluency and adequacy in translations. Revision 1.5. Available from: `http://www.ldc.upenn.edu/Catalog/docs/LDC2003T17/TransAssess02.pdf` [Accessed 3 October 2012].

Lommel, Arle, Maja Popovic and Aljoscha Burchardt 2014. Assessing inter-annotator agreement for translation error annotation. *IN: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC),* 26-31 May, Reykjavik, Iceland.

Mitchell, Linda 2015. Community Post-Editing of Machine-Translated User-Generated Content. *PhD thesis*, Dublin City University.

Moorkens, Joss and Sharon O'Brien 2015. Post-editing evaluations: trade-offs between novice and professional participants *IN: Proceedings of the 18th annual conference of the European Association for Machine Translation (EAMT 2015),* 11-13 May, Antalya, Turkey.

Plitt, Mirko and François Masselot 2010. A productivity test of statistical machine translation post-editing in a typical localisation environment *The Prague Bulletin of Mathematical Linguistics*, 93, 7–16.

Rico, Celia and Martín Ariano 2014. Defining Language dependent post-editing guidelines: the case of the language pair English-Spanish *IN: Post-editing of Machine Translation* Laura Winther Balling, Michael Carl, Michel Simard, Lucia Specia and Sharon O'Brien (eds.). 299–322.

Schwartz, Lane 2014. Monolingual post-editing by a domain expert is highly effective for translation triage. *Proceedings of the Third Workshop on Post-editing Technology and Practice*, Vancouver, Canada, 34–44.

TAUS. 2010. MT post-editing guidelines. Available from: `https://www.taus.net/think-tank/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines` [Accessed 24 February 2015].