Baltic J. Modern Computing, Vol. 4 (2016), No. 2, 106-114

# Patterns of Terminological Variation in Post-editing and of Cognate Use in Machine Translation in Contrast to Human Translation

Oliver ČULO, Jean NITZKE

Fachbereich Translations-, Sprach- und Kulturwissenschaft, Universität Mainz
An der Hochschule 2, 76726 Germersheim, Germany

{culo,nitzke}@uni-mainz.de

**Abstract**. Post-Editing and machine translation are often studied from the viewpoint of efficiency (measured e.g. in words processed) or of quality (e.g. human judgement of fluency). Little is known, however, about how post-editing and machine translation change the linguistic profile of the texts produced in contrast to human translations. In this paper, we present a pilot study contrasting lexical profiles of a small collection of texts, focussing on two aspects: variation patterns in terminological translation in post-edited texts, and the translation of cognates in machine translation, both contrasted to purely human translations. The study was conducted for the translation direction English-German.

**Keywords:** Translation, Post-Editing, Terminology, Cognates

## 1. Introduction

Post-editing (PE) is a rather new mode of translation production which is increasingly studied from various angles. A pervasive topic is the raise in productivity which is associated with post-editing, even though to varying degrees depending on such factors as the language pair or the quality of the MT (cf. e.g. Groves and Schmidtke, 2009; O'Brien, 2011) and certainly on the complexity of the setting in which it is used (cf. e.g. Silva, 2014). Recently, there has also been growing interest in how PE differs from human translation (HT), mainly in terms of process-based research (Michael Carl et al., 2011; Elming et al., 2014; Mesa-Lao, 2014), by means of evaluating gaze behaviour and keystroke activities of post-editors.

Little is known, however, about how post-edited texts differ in their linguistic properties from human translations. For machine translated texts, Lapshinova-Koltunski (2013; 2015) provides insights on linguistic properties of texts translated by humans either manually or with the help of CAT tools, or by various machine translation systems. She analyses such features as verbal vs. nominal style and investigates whether typical translation properties such as shining-through (Teich, 2003) or explicitation (Blum-Kulka, 1986) can be observed not only in HT, but also in machine translated texts. Carl and Schaeffer (forthcoming) report that PE products exhibit less lexical variation in translation than HT products, possibly due to the priming of the machine translation output.

In the following, we present a pilot study analysing how texts may differ in terms of their lexical profiles between purely human translations and when machine translation (MT) is added to the process. We focused on two specific kinds of lexemes: terms and cognates. In the first experiment, we contrasted post-editing and human translation along the dimension of term translation within the domain of Languages for Specific Purposes (LSP). In the second experiment we compared translation choices made by humans and machine translation systems in terms of cognates, in order to understand in which other ways MT may change the lexical profiles of texts. Cognates are lexical items that have a similar form and meaning in two languages, like German *System* and English *system*, but are not always the best or preferred translation equivalents.

# 2. Studies

## 2.1. Patterns of terminological variation

The data used for the analysis in this subsection was collected in an experiment in which participants were asked to translate (HT), fully post-edit (FPE) and light-post-edit (LPE) texts from either the technical or the medical domain. The texts were of low formality level and originated from LSP texts freely available on the internet. The three technical texts selected for the experiment are parts taken from a dish washer manual, the three medical texts were taken from package leaflets ranging from a vaccine against measles to human insulin for diabetes patients and medication for treatment of cancer. All texts were about 150 words long. The data collection is a generic collection in the sense that we did not aim at studying one specific phenomenon (e.g. term translation), but it is generally aimed at contrasting the PE and HT processes and products along various dimensions.

The texts were automatically pre-translated by Google Translate for the PE tasks. A permutation scheme was set for the three sessions for each domain so that each text would be translated, fully and lightly post-edited the same amount of times, but by different participants (cf. Table 1). Participants had access to internet search facilities, including online dictionaries and term databases such as IATE[1].

The participant groups consisted of 12 advanced translation students for the technical and 9 for the medical texts, all students at FTSK Germersheim. They had at least two years of training and had passed at least one course on translating in the domain they would translate and post-edit in in the experiment. Some had minor post-editing experience, but given that PE is not part of the regular course offering at the institute, it can be assumed that all participants were better trained in HT than in PE.

The technical texts were thus translated, lightly and fully post-edited each four times, the medical texts each three times. The participants used the Translog-II editor for all three tasks, which logged the participant's key stroke activities, and eye movements were recorded using a Tobii TX 300. The eye-tracking and keylogging data were not used for the present study. The experiment data will be made available through the translation process research database (TPR-DB)[2].

[1] http://iate.europa.eu/
[2] https://sourceforge.net/projects/tprdb/files/

**Table 1.** Permutation scheme for human translation, full, and light post-editing

| Participant | Text 1 | Text 2 | Text 3 |
|---|---|---|---|
| P01 | HT | LPE | FPE |
| P02 | FPE | HT | LPE |
| (...) | | | |

Participants were given instructions on how to post-edit. Part of the instructions for the full PE was to ensure terminological consistency of the post-edited texts. In general, only one term is used to describe one concept in LSP texts. Indeed, in one case where two term naming variants were used synonymously in a package leaflet – the Latin-derived term *varicella* and the equivalent English term *chickenpox*– translators would opt for one variant in human translation: either the more formal Latin-derived variant *Varizellen* or the more colloquial German variant *Windpocken* (cf. Table 5).

Carl and Schaeffer (forthcoming) observe that HT is more varied than PE in terms of lexical variation. A quick look at the number of types used for all texts produced in one translation mode (cf. Table 2) confirms this for our text collection: we counted the types for nouns, verbs adjectives and adverbs for the machine translated texts first, and then for the "bag of words" of all lightly post-edited, all fully post-edited and all human-translated texts. It is not surprising that MT has the lowest number of types, as there was only one machine translated text per source text, where there were multiple translations resp. post-edits. The numbers given in Table 2 also comply with our hypothesis that in light post-editing there would be less lexical variation than in full post-editing, where post-editors are expected to produce a translation of higher quality. For all texts together in HT, however, we get by far the highest number of types, i.e. for multiple translation vs. multiple full post-edits of a text, there seems to be more lexical variation in HT.

As ensuring terminological consistency was not part of the assignment for light PE, we ruled out these data for the analysis presented below.

**Table 2.** Number of lexical types (nouns, verbs, adjectives, adverbs) per translation mode for all texts of that mode combined. Note that for MT, there is only one target text per source text while there are multiple target texts for the other modes

| translation mode | no. of lexical types |
|---|---:|
| MT | 277 |
| LPE | 330 |
| FPE | 384 |
| HT | 488 |

For measuring variation in lexical rendering in translation, Carl and Schaeffer (forthcoming) propose to make use of the Perplexity coefficient. The perplexity coefficient is usually used to measure decision (un)certainty for a translation model. Carl and Schaeffer test whether perplexity and production- and reading-times in translation and post-editing can be correlated. We use the perplexity coefficient as measure of variation in term translation in order to test whether Machine Translation shines through in the final post-editing product.

We manually identified term candidates in the source texts and their counterpart translations; term candidates were verified as term through the use of IATE. We chose terms that appear three times or more often in the source text and checked whether the translations of those varied or whether they were left untranslated; if a term was not translated at one point, we calculated this as variation. There was one borderline case: the phrase *where measles are common* was translated by the German compound noun *Masernvorkommen* 'measles occurrence'. Nominalisation and compounding are typical processes for translations from English to German, but as the word *measles* had been consistently rendered by the noun *Masern* as first part of the German compound, we chose not to count this as variation.

**Table 3.** Mapping of term variant to translation event type for translations of the term *dish washer* in two post-editing sessions

| Session | term translation | term frequ. | event type |
|---|---|---|---|
| P10_FPE | Spülmaschine | 2 | pref.t.trans |
| | Geschirrspülmaschine | 2 | syn.1.trans |
| | Geschirrspüler | 1 | syn.2.trans |
| P21_FPE | Geschirrspülmaschine | 4 | pref.t.trans |
| | Spülmaschine | 2 | syn.1.trans |

Table 3 lists the renderings and their frequencies of translations of the term *dish washer* for two specific post-editing session. One problem becomes apparent when looking at the list of terms used in the translations for one and the same concept in Table 3: As we did not define an a priori list of preferred terms for the tasks, post-editors would vary in their choice of preferred term for concepts which exhibited high terminological variation. We thus needed to map these results onto a scheme which would allow us to measure terminological variation across texts independent of the post-editors choice of preferred term.

We chose to classify term translations into types of "events" consistent with terminology theory (cf., e.g., Arntz, 2014). For each session, we defined a *preferred term* a posteriori, variations of this term were categorized as *synonymous terms*. We thus count two main types of events: translation-by-preferred-term (*pref.t.trans*) and translation-by-synonymous-term (*syn.trans*), where each different synonymous term used counts as a different subtype of event (simply numbered, i.e. *syn.1.trans*, *syn.2.trans*, etc.).[3] For participant P21, defining the preferred term was simple, as one term, namely *Geschirrspülmaschine*, was used more often as a translation of dish washer than the other. The event type syn1.trans covers all translations by means of the synonym *Spülmaschine*. For participant P10, assigning the preferred term status to one of the two synonyms *Spülmaschine* or *Geschirrspülmaschine* is arbitrary, as both appear equally

[3] Indeed, the classification of events could be simpler (e.g. type-1, type-2, …), and opting for one variant as preferred variant is an extra step that is not necessary for computing the perplexity value. This classification is, however, compatible with a theoretical construct well established in Translation Studies and, following the proposals made in (Čulo, 2014), represents, we believe, a mutually understandable concept for both fields (MT and Translation Studies).

frequently; as *Spülmaschine* appeared first in the translation, we chose this as the preferred term. By classifying term variants into these types of events, we were able to aggregate translation probabilities across texts (cf. Table 4). These translation probabilities are then used to calculate the perplexity value for a term over all human translation and post-editing sessions.

**Table 4.** Aggregated translation event type frequency for FPE of participants P10 and P21 for the translation of *dish washer*

| event type | freq. | prob. |
|---|---|---|
| pref.t.trans | 6 | 0.55 |
| syn.1.trans | 4 | 0.36 |
| syn.2.trans | 1 | 0.09 |

Table 5 lists the perplexity values for 10 terms in the MT, the full PE (FPE) and the HT. Our findings reveal levels of variation on the terminological level in the post-edited texts close, but not identical, to those of the machine translation outcomes. MT shows variation for fewer terms than HT, but in cases of variation, usually exhibits stronger variation. In PE, the variation patterns of MT were carried over even though the task description explicitly stated to correct inconsistent terminology.

A Kendall's Tau test comparing FPE and MT values from Table 5 confirms a strong correlation between the variation patterns, which is not true for the comparison between MT and HT. This result indicates a *shining through* (Teich, 2003) of the MT in the post-editing products on the terminological level.

**Table 5.** Perplexity value (MT) and aggregated perplexity value (FPE and HT) per translation of ST term

| ST term (frequency in ST) | MT | FPE | HT |
|---|---|---|---|
| vaccine (3) | 1 | 1 | 1.57 |
| measles (4) | 1 | 1 | 1 |
| varicella (1) / chickenpox (3) | 1.75 | 1.82 | 1 |
| Protaphane (6) | 1 | 1 | 1 |
| anticancer medicine (3) | 1 | 1 | 1.57 |
| dishwasher (text 1: 5) | 2.81 | 2.07 | 1.23 |
| dishwasher (text 2: 5) | 1.96 | 1.48 | 1 |
| rinse aid (4) | 1 | 1 | 1.33 |
| filter (3) | 1 | 1 | 1.42 |
| upper filter assembly (3) | 1.89 | 1 | 1.75 |

## 2.2. Cognate translation

Following the results of the first experiment, we attempted to understand in which other ways the lexical profile of machine translated texts might differ from that human translations. In a pilot study, we compared corpus data (taken from the English-German Translation Corpus of TU Chemnitz[4]) with machine translated data (Google Translate[5])

[4] http://ell.phil.tu-chemnitz.de/search/
[5] https://translate.google.com/

for 13 English-German cognates that occurred more than five times in the corpus data. Cognates "are those translation words that have similar orthographic-phonological forms in the two languages of a bilingual […]; non-cognates are those translations that only share their meaning in the two languages […]" (Costa et al., 2000, p.1285). In the language pair English-German, *system* and *System* are for example cognates, while *government* and *Regierung* are non-cognates. *System* is not always the best translation for *system*, however; depending on the context, translations like *Anlage* (roughly 'installation') or *Verfahren* 'procedure' may be better options. So called *false friends* are different from cognates in that false friends are two words that share the same form, but not the same meaning across two languages, like the English word *actual* (real, existing) vs. the German *aktuell* (current, latest).

**Table 6.** Proportions of cognate translation, human vs. machine translation

| | | HT | | | MT | | |
|---|---|---|---|---|---|---|---|
| Lemma | N | Cognate | Non-cog | Other | Cognate | Non-cog | Other |
| acceptance | 25 | 32 % | 52 % | 16 % | 68 % | 32 % | 0 % |
| **affair** | **7** | **29 %** | **71 %** | **0 %** | **29 %** | **71 %** | **0 %** |
| competence | 25 | 32 % | 56 % | 12 % | 56 % | 44 % | 0 % |
| complexity | 15 | 93 % | 0 % | 7 % | 100 % | 0 % | 0 % |
| **compromise** | **14** | **100 %** | **0 %** | **0 %** | **100 %** | **0 %** | **0 %** |
| fantasy | 9 | 63 % | 37 % | 0 % | 100 % | 0 % | 0 % |
| intelligence | 14 | 64 % | 36 % | 0 % | 100 % | 0 % | 0 % |
| *orientation* | *22* | *55 %* | *23 %* | *23 %* | *41 %* | *59 %* | *0 %* |
| program | 10 | 90 % | 0 % | 10 % | 100 % | 0 % | 0 % |
| reaction | 7 | 71 % | 14 % | 14 % | 86 % | 14 % | 0 % |
| routine | 10 | 80 % | 0 % | 20 % | 100 % | 0 % | 0 % |
| sequence | 31 | 6 % | 84 % | 10 % | 23 % | 77 % | 0 % |
| tendency | 22 | 73 % | 9 % | 18 % | 91 % | 9 % | 0 % |

In a first step, we looked up the cognates in the corpus and counted how often the cognate was translated with the German equivalent cognate and how often with a non-cognate alternative. In the second step, the sentences from the corpus were machine translated and we counted the cognate/non-cognate translations for the data. The results are presented in Table 6. The category *Other* comprises realisations that could not be directly interpreted as cognate or non-cognate, e.g. when the cognate was not translated at all, or when the noun in the source text was verbalised in the target text, as in the following example where the noun *acceptance* is translated by means of the verb *akzeptierten*:[6]

Source: "[…] political stability rested on the acceptance in all classes of the legitimacy […]"

[6] While this may seem counterintuitive at first, there are two reasons for this decision: First, noun and verb, though derived from the same root, may have slightly different meanings, e.g. in the current case the cognate *Akzeptanz* would – intuitively – rather mean ‚rate of acceptance' or ‚degree of willingness to accept' whereas *akzeptieren* can be translated as ‚accept' in terms of ‚not oppose a measure' in the given context. Second, we currently assume that a word class shift needs more cognitive effort (given, e.g. potential shifts in meaning) than simply taking over a cognate form within the same word class and thus do not classify cases involving word shifts as cognate translations.

Target: "[…] beruhte ihre Stabilität darauf, daß alle Klassen die Legitimität […] akzeptierten"
Lit.: […] *rested their stability on that all classes the legitimacy […] accepted*

In nine instances, the cognate was translated more often with the German cognate by the MT system than in the human translation (see non-highlighted elements in Table 6), in two instances the cognate translation appeared equally often (see bold-faced elements in Table 6) and only once did the human translation data make more use of cognate translation than the MT system (see italicised elements in Table 6). A Pearson's Chi-squared test showed significance ($\chi^2$= 10.91, p < .001) for the difference between machine translation and the corpus data considering the selection of cognate and non-cognate translation options. Conclusively, MT output and human translations show different patterns for translating cognates. The category *Other* was not needed for the MT output, highlighting that for the MT system used for translation in this case, word class shifts seem to be rarer at least in cases in which a cognate is available in the same word class in the target language.

## 3. Discussion

In this study, we focused on two aspects of the lexical profile of texts and how these may change when MT is added to the process of translation. We saw in one of the two studies that MT can have an influence on the outcome of PE; even in cases where participants were asked to make certain corrections, these were not made. We do not have an explanation for this, but we can offer one hypothesis: As Mesa-Lao (2014) observes in his pilot study on the differences between the human translation and the post-editing process, there seemed to be no clear initial orientation or final revision phases in most of the post-editing sessions in his study. This pattern might be disadvantageous for a sub-task like ensuring terminological consistency. Further investigation of the typing/correction patterns may reveal more on the role of these phases for terminological consistency.

The study presented here is very limited with respect of the instances investigated. A larger scale study could reveal more on the influence of MT on the lexical (and grammatical) profile of translations (here: from English to German). It should be added that the study on terminological variation presented here was guided by target language norms which hold for German certainly, but not necessarily for other languages or language pairs.

Also, it still needs to be proved whether the cognate profile of post-edited texts will be similar to that of MT. A logical follow-up question to such a finding would, however, be whether in fields, where post-edited texts make up a significant proportion of the texts produced, original language production is influenced by the linguistic profiles of the post-edited texts (e.g., if confirmed for post-edited texts, whether cognates are also used more often in more recent original language production). While the observations made may point into this research direction, the limited scope of the study cannot serve to give any indication about the answer to this question.

Last but not least, it needs to be pointed out that the changes in the lexical profile observed in MT and PE were results of the characteristics of the MT systems used on a specific text type and language pair. Of course, these characteristics may change depending on a variety of factors; from the viewpoint of translation studies, uncovering and understanding these factors is vital, both as potential feedback to MT researchers and in terms of understanding the role of MT in and its influence on translation.

## 4. Conclusions and future work

In further studies, the influence of the machine cognate translation on the post-editing process will be studied. We will focus on the following two questions: Do translators accept the cognate translation of the machine translation system or will they choose another solution in the post-editing process? Will the variation of cognate translation rather direct to the variation in machine translation or the variation of translation from scratch? And ultimately: Are patterns observed in post-edited texts also reproduced in original language? If the latter was the case, then MT might be seen as a driving force of language change in certain areas. We believe that understanding in which way the characteristics of a single text, and through many texts of an LSP as a whole, may change through the use of MT can – depending on which changes (e.g. higher term consistency) may be desirable or undesirable – provide important feedback for MT research and engineering.

## References


Arntz, Klaus-Dirk Reiner; Picht, Heribert; Schmitz. (2014). *Einführung in die Terminologiearbeit*.

Blum-Kulka, Shoshana. (1986). Shifts of Cohesion and Coherence in Translation. In Juliane House and Shoshana Blum-Kulka (eds), *Interlingual and Intercultural Communication*, 17–35. Tübingen: Gunter Narr Verlag.

Carl, Michael, Barbara Dragsted, Jakob Elming, Daniel Hardt, and Arnt Lykke Jakobsen. (2011). The Process of Post-Editing: A Pilot Study. In Bernadette Sharp, Michael Zock, Michael Carl, and Arnt Lykke Jakobsen (eds), *Proceedings of the 8th International NLPSC Workshop. Special Theme: Human-Machine Interaction in Translation*, 131–42. Copenhagen Studies in Language 41. Frederiksberg: Samfundslitteratur.

Carl, M, and M Schaeffer. (forthcoming). Literal Translation and Processes of Post-Editing. In *Translation in Transition: Between Cognition, Computing and Technology*. Amsterdam: John Benjamins.

Costa, Albert, Alfonso Caramazza, and Nuria Sebastian-Galles. (2000). The Cognate Facilitation Effect: Implications for Models of Lexical Access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (5): 1283–96.

Čulo, Oliver. (2014). Approaching Machine Translation from Translation Studies: A Perspective on Commonalities, Potentials, Differences. In *Proceedings of the Seventeenth Annual Conference of the European Association for Machine Translation (EAMT)*, 199–208. Dubrovnik, Croatia.

Elming, Jakob, Laura Winther Balling, and Michael Carl. (2014). Investigating User Behaviour in Post-Editing and Translation Using the CASMACAT Workbench. In Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard, and Lucia Specia (eds), *Post-Editing of Machine Translation*, 147-169. Cambridge Scholars Publishing.

Groves, Declan, and Dags Schmidtke. (2009). Identification and Analysis of Post-Editing Patterns for MT. In *MT Summit XII: Proceedings of the Twelfth Machine Translation Summit*, 429–36. Ottawa, Canada.

Lapshinova-Koltunski, Ekaterina. (2013). VARTRA: A Comparable Corpus for the Analysis of Translation Variation. In *Proceedings of the 6th Workshop on Building and Using Comparable Corpora*, 77–86. Sofia, Bulgaria.

———. (2015). Variation in Translation: Evidence from Corpora. In Claudio Fantinuoli and Federico Zanettin (eds), *New Directions in Corpus-Based Translation Studies*, 93–113. Translation and Multilingual Natural Language Processing 1. Berlin: Language Science Press.

Mesa-Lao, Bartolomé. (2014). Gaze Behaviour on Source Texts: An Exploratory Study Comparing Translation and Post-Editing. In Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard, and Lucia Specia (eds), *Post-Editing of Machine Translation*, 219-245. Cambridge Scholars Publishing.

O'Brien, Sharon. (2011). Towards Predicting Post-Editing Productivity. *Machine Translation* 25 (3): 197–215.

Silva, Roberto. (2014). Integrating Post-Editing MT in a Professional Translation Workflow. In: Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard, and Lucia Specia (eds), *Post-Editing of Machine Translation*, 24–51.

Teich, Elke. (2003). Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts. Vol. 5. *Text, Translation, Computational Processing*. Berlin/New York: Mouton de Gruyter.

Received May 2, 2016, accepted May 4, 2016