FACULTY OF ARTS AND PHILOSOPHY

AN EXPLORATORY STUDY OF PROFESSIONAL POST-EDITS BY ENGLISH-DUTCH DGT TRANSLATORS

Word count: 16,334

Luca Desmet
Student number: 01606648

Supervisor: Prof. Dr. Sonia Vandepitte
Co-supervisor: Prof. Dr. Lieve Macken

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master in Translation

Academic year: 2020 - 2021

GHENT UNIVERSITY

Abstract

The ever-increasing quality and adoption of neural machine translation (NMT) has spiked the need for and interest in post-editing – the act of correcting machine translation output by a human translator. While the combination of the two is generally believed to speed up translation processes, the study into its impact on and adoption in high-quality translation environments such as the European Commission Directorate-General of Translation (DGT) has only just started (Rossi and Chevrot, 2019; Vardaro, Schaeffer and Hansen-Schirra, 2019; Arnejšek and Unk, 2020; Macken, Prou and Tezcan, 2020; Vandevoorde, Weintraub and Arabadjieva, 2021), and has not yet covered the DGT Dutch Language Department. This paper therefore presents an analysis of post-edits carried out by highly professional English-Dutch translators at the DGT Dutch Language Department. Post-edits by nine translators were manually annotated and categorised by means of a purpose-built typology, according to which several characteristics of post-edits have been classified as compared to both the source text and the NMT output. The post-edits were annotated on three different levels: a translation quality, translation norm, and text-linguistic level, with the aim of assessing their correctness and necessity, and determining the types of post-edits most frequently implemented. Though preliminary, the results of this analysis can be of importance in establishing specific high-quality post-editing guidelines, and in making suggestions for further improvement of NMT systems.

(220 words)

Verklaring i.v.m. auteursrecht

Acknowledgements

This Master's dissertation could not have been put into practice if it weren't for the following people, whom I would very much like to express my gratitude to.

I would first of all like to thank my supervisor Prof. Dr. Sonia Vandepitte and my co-supervisor Prof. Dr. Lieve Macken for their guidance throughout the entire process of this study, for providing keen insights, contributions, and feedback, for their encouraging and kind words, and for helping me out whenever I had a question or felt like I was at a roadblock. I sincerely wish to thank them for their commitment to making this dissertation a success, for all the work they have put into it, and for keeping me motivated and on track. I would like to thank Dr. Arda Tezcan as well, for converting the provided SDLXLIFF files to workable Excel files.

I would also very much like to thank the ten translators at the Dutch Language Department of the European Commission Directorate-General for Translation who agreed to share their translations for the present research project. Without their data, this study would not have been possible in the first place, for which their contribution is greatly appreciated. I sincerely hope that I have done their involvement justice and have provided them with interesting new insights into their very own translation and post-editing work.

And lastly, I want to express my gratitude to my family, my friends, and my partner (in crime) for their continuous support and patience, and for believing in me even when I did not quite believe in myself.

Table of Contents

1 Introduction 1

2 Related Research 3

2.1 Machine Translation 3

2.1.1 The Road to Neural Machine Translation 3

2.1.2 Post-Editing of SMT and NMT 4

2.2 Post-Editing Guidelines 8

2.3 Translation Quality 11

2.3.1 Machine Translation Quality 13

2.3.2 Quality of the Post-Edited Product 14

2.4 Machine Translation at the European Commission DGT 17

2.5 Professional Translators 19

2.6 Research Questions 20

3 Method 23

3.1 Data 23

3.2 Sample Selection and Post-Edit Extraction 25

3.3 Typology 28

3.4 Annotation 31

4 Results 37

4.1 General Results 37

4.1.1 Translation Quality Parameter 37

4.1.2 Translation Norm Parameter 38

4.1.3 Text-Linguistic Characteristics 40

4.2 Results per Translator 49

5 Discussion 55

6 Conclusions 61

Bibliography 65

Appendix 75

Appendix A Post-Edit Typology 75

Appendix B Decision Tree for the Text-Linguistic Annotation 93

Appendix C Post-Edit Analysis 101

Appendix D List of Undesirable Changes 103

List of Abbreviations

CAT (tool)	Computer-Assisted Translation (tool)
DGT	Directorate-General for Translation
LSP	Language Service Provider
MT	Machine Translation
MTPE	Machine Translation Post-Editing
MTQA	Machine Translation Quality Assessment
NMT	Neural Machine Translation
NMTPE	Neural Machine Translation Post-Editing
PBSMT	Phrase-Based Statistical Machine Translation
PE	Post-Editing
RBMT	Rule-Based Machine Translation
SMT	Statistical Machine Translation
SMTPE	Statistical Machine Translation Post-Editing
TM	Translation Memory
TQA	Translation Quality Assessment

List of Tables

Table 1. Schematic overview of the final data set on segment level 24

Table 2. Schematic overview of the final data set on word level 25

Table 3. Numbers of post-edits extracted per translator and in total 27

Table 4. Schematic overview and annotation codes of the Translation Quality Parameter 29

Table 5. Schematic overview and annotation codes of the Translation Norm Parameter 29

Table 6. Schematic overview and annotation codes of the Text-Linguistic Characteristics 31

Table 7. Distribution of the Text-Linguistic categories per parameter 44

List of Figures

Figure 1. Fragment of the sample in the Excel annotation file (taken from Translator 5) 27

Figure 2. Fragment of the annotation in the Excel annotation file (taken from Translator 6) 35

Figure 3. Distribution of the Translation Quality Parameter categories 37

Figure 4. Distribution of the Translation Norm Parameter categories 38

Figure 5. Translation Quality Parameter and Translation Norm Parameter in relation to each other. 39

Figure 6. Distribution of the Text-Linguistic parameters 41

Figure 7. Distribution of the Text-Linguistic parameters in relation to the Translation Quality Parameter 42

Figure 8. Text-Linguistic category frequency 45

Figure 9. Text-Linguistic categories occurring at least 10 times, in relation to the Translation Quality Parameter 47

Figure 10. Distribution of the Translation Quality Parameter per translator 49

Figure 11. Distribution of the Translation Norm Parameter per translator 51

Figure 12. Distribution of the Text-Linguistic parameters per translator 52

1 INTRODUCTION

Translation has been paramount to globalisation – being central to global exchange of information (Bielsa, 2005) – and in recent years, globalisation and digitalisation in turn have led the translation industry to substantially evolve. Translation is more omnipresent than ever, and as is the case in many other industries, the working pace in its industry is ever-increasing: more pages are to be translated within shorter time frames, while upholding similar or even higher levels of quality. This has been enabled through digitalisation of the translation process, with a major player also having been machine translation (MT). MT, and in particular neural machine translation (NMT), has seen substantial improvement and as a result also significant uptake in the industry (Cho et al., 2014; Bentivogli, Bisazza, Cettolo, & Federico, 2016). Not only is NMT nowadays adopted in publicly accessible machine translation engines such as Google Translate[1] and DeepL Translator[2], and is it used for automatic fast yet lower-quality translation of e.g. TripAdvisor reviews (Cenni, 2019) and Facebook posts (van Belle, 2020), it is also increasingly becoming integrated into professional translation workflows (e.g. Macken, Prou, & Tezcan, 2020). This implies that within their computer-assisted translation (CAT) environment, professional translators can rely on MT as a translation aid when no translation memory (TM) match is available for a segment. The MT output is then corrected by the human translator, which is referred to as machine translation post-editing (MTPE) or simply post-editing (PE).

As MTPE is generally believed to increase translation speed and reduce translation effort (e.g. Läubli, Fishel, Massey, Ehrensberger-Dow, & Volk, 2013), interest in PE has in translation studies been on the rise. Research has laid focus on, for instance, student PE and how PE can be integrated into translator training, but on MT and PE error analysis as well (e.g. Daems, Macken, & Vandepitte, 2014), given that PE time can significantly be reduced through awareness of the types of errors to be expected in MT output (Martínez, 2003). Even NMT is not yet perfect and appears to often contain specific error types (Castilho et al., 2017), for which PE has proven to be of crucial importance, especially in very high-quality translation environments such as the European Commission Directorate-General for Translation (DGT) (e.g. Arnejšek & Unk, 2020). The DGT NMT system eTranslation, operative since November 2017, is integrated into their CAT tool to provide the aid of MT alongside the extensive central

[1] https://translate.google.com/
[2] https://www.deepl.com/translator

translation memory EURAMIS (European Advanced Multilingual Information System) and the terminology database IATE (Inter-Active Terminology for Europe).

The adoption of eTranslation NMT and its impact on the DGT translation workflow specifically, have been investigated at a number of different language departments (Lesznyák, 2019; Arnejšek & Unk, 2020; Macken et al., 2020; Stefaniak, 2020), not including, however, the DGT Dutch Language Department. In addition, highly professional post-edits have in general not often been analysed for the English-Dutch language pair. These considerations raise questions such as: *To what extent do English-Dutch DGT translators make use of MT? Which types of changes are recurring NMT post-edits in high-quality English-Dutch translations?* And *to what extent are these post-edits essential to reach the extremely high DGT quality standards?* An analysis of DGT post-edits may thus provide an indirect view on what is considered highly qualitative translation in the industry, and may moreover yield insights of use for further improvement of NMT systems. It would be of interest and relevance to investigate what English-Dutch DGT translators dedicate their post-editing time to, as it can additionally lay bare if and how the post-editing process in this qualitative translation workflow could still be optimised. Lastly, it was not until recently that the importance of post-editing guidelines was acknowledged by industry and academia, for which these remain relatively scarce, and as they also vary along both MT systems used and target text functions, very high quality (DGT) MTPE will likely require a tailored set of PE guidelines that is yet to be established.

The present dissertation therefore provides an exploratory study of post-edits carried out by highly professional English-Dutch DGT translators, seeking to lay bare recurrent types of English-Dutch NMT post-edits, and their correctness and necessity. The remainder of this paper is structured as follows: Chapter 2 provides an overview of related literature on machine translation, post-editing, and MT in the context of the DGT specifically, after which the four research questions drawn up for the present study are listed. Chapter 3 discusses the data that was provided by the ten DGT translators having participated in the study, and the selection of the sample for analysis, as well as the typology that was devised for the annotation, and the method that was adopted for it. The results of the analysis are then presented and elaborated on in Chapter 4, after which they are discussed in the light of the research questions and related research in Chapter 5. Finally, Chapter 6 draws a number of conclusions, and touches on suggestions for future research in the field of professional post-editing.

2 RELATED RESEARCH

2.1 Machine Translation

2.1.1 The Road to Neural Machine Translation

As a result of the rapid improvements in machine translation (MT) in recent years, the translation industry has become increasingly accepting of and curious towards it. MT has come a long way since the first rule-based machine translation (RBMT) systems, which have now made way for statistical machine translation (SMT) and, more recently, neural machine translation (NMT). Relying on artificial neural networks trained on large bilingual corpora, NMT (Cho et al., 2014) generally proves to outperform both RBMT and SMT for a wide range of different language pairs (Bentivogli, Bisazza, Cettolo, & Federico, 2016; Junczys-Dowmunt, Dwojak, & Hoang, 2016; Wu et al., 2016; Koehn & Knowles, 2017; Toral & Sánchez-Cartagena, 2017; Daems & Macken, 2019; Stasimioti, Sosoni, Kermanidis, & Mouratidis, 2020), with more fluent (acceptable) target texts appearing to be the main aspect of improvement (Kelleher, 2016; Bentivogli et al., 2016; Castilho et al., 2017; Toral & Sánchez-Cartagena, 2017; Klubička, Toral, & Sánchez-Cartagena, 2017). Wu et al. (2016), for example, found that Google's Neural Machine Translation system (GNMT) reduced translation errors by 60% in comparison to the phrase-based machine translation (PBMT) system used prior, and Bentivogli et al. (2016) showed that NMT impressively outperformed phrase-based statistical machine translation (PBSMT) in the placement of verbs for the English-German language pair, while additionally producing fewer lexical and morphological errors, and substantially fewer word order errors (Bentivogli et al., 2016). Results of Koponen, Salmi and Nikulin (2019), however, indicated that while NMT also generated fewer word order errors than both RBMT and SMT, it did generate more lexical errors.

Mixed results have thus been reported as well when comparing SMT and NMT. Although automatic evaluations in their study did show NMT outperforming SMT, Castilho et al. (2017, p. 109) still warned "not to oversell" NMT, largely based on the mixed human evaluations obtained. Accuracy errors (infidelities to the source text) such as mistranslations, additions and omissions, moreover, appear to be recurrent problematic error types in NMT output (Castilho et al., 2017; Vardaro, Schaeffer, & Hansen-Schirra, 2019), and challenges for NMT systems have proven to include texts with a rich morphology or long sentences (Bentivogli et al., 2016; Koehn and Knowles, 2017; Toral & Sánchez-Cartagena, 2017). Koehn and Knowles (2017)

also pointed out that, when exposed to only limited training data or confronted with significantly different conditions than the training conditions, NMT systems seemed to fall short (Koehn & Knowles, 2017). A 2018 study even found that for the English-Irish language pair, the NMT system generated lower quality translations than a tailored SMT system (Dowling, Lynn, Poncelas, & Way, 2018), which was indeed ascribed to the scarcity of available data on this language pair, and to the rich morphology and long sentences that are intrinsic to the Irish language. Correspondingly, results from Skadiņa and Pinnis (2017) for the English-Latvian language pair showed that although the SMT output in their study was not as fluent as the NMT output, the SMT system did outperform the NMT system in terms of accuracy errors. The NMT system could not learn enough from the small amount of data, while the SMT system had more comprehensively acquired specific terminology and phrasing. A final prominent disadvantage of NMT systems to which amongst others Bentivogli et al. (2016), Koehn and Knowles (2017), and Vardaro et al. (2019) have drawn attention, is their much less transparent and decipherable nature. Vardaro et al. (2019) found that 60% of the NMT errors in their study were register errors, suggesting that NMT errors may indeed more often be hidden, and Yamada (2019, p. 87) mentioned NMT having "more advanced, human-like translation abilities" than SMT. In this respect, Koehn and Knowles (2017, p. 1) noted that there is "a clear need to develop better analytics for NMT".

Overall, it can be argued that while automatic translation has improved substantially with the introduction of NMT, particular areas do appear to remain problematic in NMT systems. They sometimes generate more mistranslations than SMT systems, and moreover need larger amounts of training data, which are unavailable for some language pairs. They also have the disadvantage that the errors generated are often more difficult to detect and comprehend, as a result of the underlying deep neural network structure. NMT does, however, prove to produce more fluent translations than SMT – often containing substantially fewer lexical, morphological, and word order errors – for which it is still generally regarded as the state of the art in MT.

2.1.2 Post-Editing of SMT and NMT

The adoption of MT in the translation industry gave rise to the need for machine translation post-editing (MTPE), hereafter referred to as post-editing (PE), to produce larger volumes of translations at the same level of quality, within a shorter time frame. The activity of PE has

been included in the ISO 18587:2017 standard, in which it is defined as "edit and correct machine translation output" (ISO 18587, 2017). In the literature as well, PE has been defined in a number of ways, including by Allen (2003, p. 297), as "the task of editing, modifying and/or correcting pre-translated text that has been processed by a MT system from a source language into a target language", and by TAUS (2010), as "the process of improving a machine-generated translation with a minimum of manual labour." The latter (TAUS, 2010) lays focus on the "minimum of manual labour" involved in PE, of which Massardo (2018) argued that it is a key component of the definition, and clearly marks the distinction with translation revision. PE is often compared to revision, but differs from it in that changes are made to MT output instead of human translated text. In addition, the focus would in MTPE be on grammatical, terminological, and mechanical (spelling and punctuation) errors and other issues in MT output, while a revisor is to carry out a more in-depth contrastive analysis of the source and target text, in which style and grammar should be adjusted as appropriate as well (Massardo, 2018). Massardo (2018) moreover noted that the skill sets required for the both differ: profound knowledge of source and target languages as well as of translation, is essential for revision – preferably accompanied by knowledge on the domain of the text – whereas a post-editor predominantly needs to master the target language, and possess domain-specific knowledge (Massardo, 2018).

However, with NMT producing increasingly high quality and even human-like output (Yamada, 2019), and being increasingly adopted for higher-quality translation, the line between post-editing and revision may currently more often be blurred. A 2020 study by Daems and Macken, for example, investigated the degree to which PE is in practice approached differently than revision by professional translators, through having them revise/post-edit human and neural machine translated texts without knowing which of the two it was they were truly doing. Their results surprisingly indicated that the NMT output was adjusted more often when the translators thought they were revising human translation (HT), and that HT texts were of higher quality after assumingly being post-edited. The question is thus whether the distinction between the two should and will be maintained (Daems & Macken, 2020). After all, large translation systems and translation memories (TMs) have over the last few years made of translation an integrated workflow, in which TM matches and automatic MT suggestions are frequently alternated (e.g. Macken, Prou, & Tezcan, 2020). Seldom, or only in specific contexts, are texts fully post-edited; rather is PE increasingly becoming part of translation workflows. What is more, the 2020 European Language Industry Survey revealed that 78% of the language service

providers (LSPs) taking part in the survey planned on adopting MT or PE or increasing their use in the future. The overall strongest technology trend appeared to be MT as well, and the most popular new service MTPE.

This is for good reason, as research seems to agree that MTPE results in higher translation productivity than translating from scratch (Plitt & Masselot, 2010; de Almeida & O'Brien, 2010; Läubli, Fishel, Massey, Ehrensberger-Dow, & Volk, 2013; Guerberof, 2014; Depraetere, De Sutter, & Tezcan, 2014; Daems, Vandepitte, Hartsuiker, & Macken, 2017, Toral, Wieling, & Way, 2018; Koponen, Sulubacak, Vitikainen, & Tiedemann, 2020). Considerable variation, however, has been observed depending on the language pair and the translator as well (Koponen et al., 2020), and the productivity gain has proven to additionally vary along the type of MT used. A 2018 study by Toral, Wieling and Way showed that post-editing of both PBMT and NMT lead to productivity gains, but the increase when using NMT was twice the increase obtained through post-editing of PBMT (Toral et al., 2018). While a lower number of keystrokes was also reported for both SMT and NMTPE in comparison to translating from scratch, the reduction with NMT was over double the one obtained with SMT (Toral et al., 2018). Interestingly, as opposed to processing fuzzy matches (TM matches overlapping for 80-90%), Guerberof (2009) reported higher productivity when post-editing SMT, whereas Sánchez-Gijón, Moorkens and Way (2019) found NMTPE not resulting in higher productivity than TM segment editing. Though the NMT post-edits were fewer in number, the translators appeared to spend more time on them (Sánchez-Gijón, Moorkens, & Way, 2019).

In terms of translation quality, several studies report an equal or even higher level of quality of post-edited MT output compared to from-scratch translation. Depraetere et al. (2014), for example, concluded that SMT quality does not negatively impact the quality of its post-edited final translation, and Daems, Macken and Vandepitte (2013) even found the post-edited SMT in their study to be of higher quality than the human translated text. Similarly, Guerberof (2009) found post-edited SMT to be of higher quality than edited fuzzy TM segments. The quality of the MT system used evidently seems to influence the quality of the final translation, for which NMT has been observed to further improve the quality of final post-edited translations. Yamada (2019) found that the final product of NMT output post-edited by student translators was of higher quality – containing fewer errors – than the post-edited SMT output obtained in the similar 2014 study (Yamada, 2014). The post-edited NMT output of the 2019 study with these students did, however, not live up to professional quality standards, which suggests that

NMTPE may pose a greater challenge to translation students because NMT systems generate more advanced and human-like output than SMT systems (Yamada, 2019). NMTPE was found to result in translations of similar quality as from-scratch translations by Daems et al. (2017) and Jia, Carl and Wang (2019), amongst others, and it appears that the largest differences in both PE quality and productivity are found between individual translators (Koehn & Germann, 2014), which is why stress is often laid on the importance of integrating PE in translator training (de Almeida & O'Brien, 2010; de Almeida, 2013; Daems et al., 2017; Jia, Carl, & Wang, 2019; Yamada, 2019).

Similar to PE productivity and quality comparisons, Bentivogli et al. (2016) found NMTPE to be an improvement on SMTPE when it comes to overall PE effort as well – meaning that NMTPE would require less PE effort than SMTPE. PE effort can be explained through its division into three types by Krings (2001): temporal, technical and cognitive effort. Temporal effort is the time needed to post-edit MT output for a certain translation task until the desired level of quality is reached; cognitive effort comprises the mental processes involved in the identification of MT errors during PE; and technical effort is represented by the actual edit operations following those mental processes, i.e. the insertions, deletions and rearrangements implemented in the MT output. Temporal effort can readily be measured by timing a post-editor throughout a translation task, with possible additional keystroke logging as a means for obtaining the most precise results (e.g. to take into account any breaks or disruptions). Technical effort is measured through automatic metrics, such as HTER (Human-mediated Translation Edit Rate, Snover, Dorr, Schwartz, Micciulla, & Makhoul, 2006), which measures the number of edit operations required to convert the MT output into a targeted reference translation. It should be noted, however, that HTER is in essence an assessment of the end product and does not necessarily correlate with the cognitive load of a PE task (Daems et al., 2017; Stefaniak, 2020), for which it should always be considered in relation to temporal and cognitive effort. For process-based measuring of technical effort, keystroke logging has often been used (Krings, 2001; O'Brien, 2005; Koponen, Salmi, & Nikulin, 2019). Contrary to the former two types of PE effort, cognitive effort cannot be measured directly. Krings (2001) used think-aloud protocols (TAPs) as indicators for this type of effort, whereas extensive work by Viera (2016) included subjective ratings, seconds per word, eye fixations per word and average fixation duration, pause-to-word ratio, pause ratio and average pause ratio – all of which proved to be interrelated with each other. Others, too, have relied on eye fixation (Jakobsen & Jensen, 2008; Doherty & O'Brien, 2009; Daems, Vandepitte, Hartsuiker, & Macken, 2015) or pause recording

(O'Brien, 2006; Toral, Wieling, & Way, 2018; Jia, Carl, & Wang, 2019) for measuring of cognitive effort.

Jia, Carl and Wang (2019) found that while PE was in their study not significantly faster than from-scratch translation for general texts, i.e. that temporal effort was not reduced, PE still required significantly less cognitive effort. In line with this, de Almeida (2013, p. 199) stressed that "PE effort and PE performance involve a high level of complexity that cannot be explained only by analysing temporal values." Daems et al. (2015) additionally suggested that PE effort can be predicted by the quality of the MT output, as they found MT error types to be predictors of various post-editing effort indicators. Translation duration (temporal effort), for example, was most impacted by coherence issues (Daems et al., 2015). Interestingly as well, Sánchez-Gijón, Moorkens and Way (2019) found that perceived NMTPE effort was in line with actual post-editing performance/productivity.

2.2 Post-Editing Guidelines

To some frustration of both translators and LSPs, PE guidelines were long lacking: amongst others de Almeida and O'Brien (2010), DePalma (2013), de Almeida (2013), and Hu and Cadwell (2016), have referred to the absence of general PE guidelines. As both the need for and interest in PE have been on the rise in recent years, however, so has the development of PE guidelines. In this respect, Hu and Cadwell (2016) did highlight that PE guidelines will always vary along the individual needs of an LSP and the function of the target text, which is why LSPs often draw up their own internal guidelines and do not wish or feel the need publish them. They are tailored and therefore considered only internally applicable (Hu & Cadwell, 2016).

Concluding their comparative analysis of five sets of guidelines available, Hu and Cadwell (2016) agreed with DePalma (2013) and Densmer (2014) that clients and LSPs should discuss and establish the required level of quality and the expectations of the final translation before any PE job. This is in line with an important one of the more general sets of guidelines freely accessible: the PE guidelines by the Translation Automation User Society (TAUS, 2010; 2016) – a global language data network providing LSPs with data as well as knowledge. The TAUS PE guidelines were drawn up in partnership with the Centre for Global Intelligent Content (CNGL) to provide a general baseline for companies to start from (Hu & Cadwell, 2016). As specified in these as well, the desired level of quality of the end product should be established

before a costumer or LSP can determine which specific guidelines are in place for the particular translation task. To attain this pre-determined expected level of quality, either light post-editing or full post-editing (Allen, 2003) may be needed. Light post-editing consists of correcting raw MT output merely in terms of accuracy, and yields end products that are destined for internal use ("fit for purpose"), whereas full post-editing focusses on both accuracy and fluency (appropriateness of the target language) problems, and is the method adopted when the target text needs to be of high quality or is intended for publication ("high-quality human translation and revision") (TAUS, 2016). Evidently, both types of post-editing confront the post-editor with different difficulties and require a different post-editing approach. It should be highlighted, however, that TAUS (2016) distinguishes between guidelines for the two *levels of expected quality* – fit for purpose or high-quality – rather than for light and full post-editing in themselves, because it may be the case that good quality raw MT output needs no more than light post-editing to attain a (relatively) high-quality end product (TAUS, 2016).

De Almeida and O'Brien (2010) implemented in their pilot project the GALE Post-Editing Guidelines (Post Editing Guidelines for GALE Machine Translation Evaluation, 2007). They devised an error typology for classification of post-edits, which was later fine-tuned and adopted in a post-editing study by de Almeida (2013) as well. The GALE Post-Editing Guidelines (Post Editing Guidelines for GALE Machine Translation Evaluation, 2007) are therefore worth mentioning as well. They were initially developed for the Chinese-English and Arabic-English language pairs, but have proven to allow for flexibility (de Almeida & O'Brien, 2010; de Almeida, 2013). Flanagan and Christensen (2014), on the other hand, established their own set of guidelines, specifically for translator trainees, based on the TAUS (2010) guidelines. The results of the PE task in their study revealed that those by TAUS (2010) were difficult for trainees to interpret in terms of style, which led them to spend too much time on unnecessary stylistic changes (Flanagan & Christensen, 2014). These unnecessary or "preferential" changes appear to be an often-observed phenomenon in PE, found with professional translators as well (de Almeida, 2013; Flanagan & Christensen, 2014; Koponen & Salmi, 2017; Koponen, Salmi, & Nikulin, 2019). They were described by de Almeida (2013, p. 100) as follows: "a change is considered preferential if the sentence from the raw MT output would still be grammatically correct, intelligible and accurate in relation to the source text, even if the change in question was not implemented." Most of the available guidelines stipulate that translators be wary of over-editing, i.e. making any preferential change at all, which especially holds true for light post-editing, where style, grammar, and syntax should be ignored, and the final translation

should merely be comprehensible and correct (Hu & Cadwell, 2016). In this respect, de Almeida and O'Brien (2010) attribute three essential skills to a successful post-editor: the ability to (1) identify issues in the raw MT output that need to be addressed, referred to as "Essential Changes", and fix them appropriately, (2) carry out the post-editing task with reasonable speed, i.e. meet the expectations of daily productivity for this type of activity (approximately 5,000 words post-edited per day, on average), and (3) adhere to the PE guidelines, so as to minimise the number of "Preferential Changes", which are normally outside the scope of PE (de Almeida & O'Brien, 2010, p. 2).

Refraining from making such preferential changes, however, has proven to pose difficulties to translators. It is common knowledge that translators often have strong stylistic preferences (Flanagan & Christensen, 2014), or preferences for specific phrasing and words (Koponen, 2013). The results of the pilot project by de Almeida and O'Brien (2010) even indicated that the more experienced, the more inclined a translator is to implement preferential changes. Koponen and Salmi (2017) found that for the English-Finnish language pair, a significant number (34%) of the post-edits analysed in their study were, though mostly correct, unnecessary. Similarly, Koponen, Salmi and Nikulin (2019) report high numbers of preferential changes in NMT, SMT, and especially RBMT output. A 2019 Master's dissertation analysing post-editing behaviour and attitudes of translators at the KBC language department in Leuven also found that overall, most of the post-edits carried out were preferential lexical changes, and to a lesser degree also "optional" grammar and other linguistic changes, and consistency changes (Man, 2019, p. 60). Only 15 to 20% of the post-edits were found to be true corrected MT errors. In English publicity texts, however, over half of all post-edits were deemed necessary when consistency and merging (for improved readability) were regarded as essential changes as well, whereas English financial texts were in that case more often altered out of preference (Man, 2019). In terms of overall adoption of MT, great variation was observed between both all English-Dutch and all French-Dutch post-editors. It was also remarked that post-editors appeared to "repeatedly correct the same machine mistranslation in the same text" (Man, 2019, p. 61).

A final note should be made in view of the present study analysing post-edited texts from professional translators at the European Commission Directorate-General for Translation (DGT). In their study at the DGT German Language Department, Vardaro et al. (2019, p. 8) made the important note that "they [translation experts from the DGT German Language

Department] are not trained to be highly efficient post-editors, where the best quality is not always necessary and stylistic changes are usually not supposed to be inserted, but rather used to putting a lot of thought and time into their correlations to create a translation that is as perfect as possible, which leads to possible over-editing." Another recent study by Vandevoorde, Weintraub and Arabadjieva (2021) investigated the extent to which MT retrievals (MTRs) meet the extremely high quality standards at the Translation Service of the Council of the European Union, annotating Human Translation Actions (post-edits) in English-Dutch translations for type, necessity and impact on the translation quality. They found no less than 91% of the changes substantially impacting the quality of segments, with 43 to 68% of all changes appearing to be essential (Vandevoorde, Weintraub, & Arabadjieva, 2021). In terms of the level on which the actions were carried out, 85% were on a lexico-semantic level, 28% of which were corrections, and 23% synonyms. The researchers moreover noted that to achieve the elevated Council quality standards, even optional changes are not superfluous, as all of them combined result in the high-quality end product that is ultimately to be obtained (Vandevoorde et al., 2021). These two studies on high-quality post-editing once again highlight the importance of the text purpose and expected level of quality being taken into account both when drawing up guidelines for a PE task, as LSP or client, and when post-editing, as translator. MT at the DGT specifically is further elaborated on in Section 2.4, after first addressing translation quality.

2.3 Translation Quality

The notion of quality is a delicate matter in the field of translation. Research agrees that translation quality does not have one single meaning, but is dependent on variables such as context, target audience and function of the target text in the target culture (Van Slype, 1979; Koponen, 2010; Drugan, 2013; House, 2014; Castilho, Doherty, Gaspari, & Moorkens, 2018; van Egdom & Pluymaekers, 2019). This lack of one "single objective way to measure quality" (Drugan, 2013, p. 35) has, however, led to much disagreement and debate about how translation quality should then be assessed. Van Egdom & Pluymaekers (2019), for instance, urged translators not to meekly submit their translations to abstract quality standards. In terms of the evaluation of translation quality, amongst others Castilho et al. (2018) have highlighted that views and approaches differ substantially between industry, research, and training, and even in the public domain (Castilho et al., 2018; Valdez & Vandepitte, 2020).

In attempts to increase objectivity or even obtain generalisation in human translation evaluation, various models for Translation Quality Assessment (TQA) have been drawn up in research as well as in the industry. Each approaching quality from a (slightly) different perspective, these TQA models most often comprise some kind of translation error typology, or fine-tune/merge already-existing typologies (Daems, Macken, & Vandepitte, 2013). As Castilho et al. (2018) put it: in error analysis "errors found in (samples of) the translated text are counted, classified and weighted according to their severity by a senior translator or reviewer." Many of these typologies distinguish between adequacy and fluency errors, with adequacy (or accuracy) referring to "how much of the meaning expressed in the gold-standard translation or the source is also expressed in the target translation" (Linguistic Data Consortium, in TAUS, 2017), and fluency indicating the extent to which the translation is "one that is well-formed grammatically, contains correct spellings, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker" (Linguistic Data Consortium, in TAUS, 2017). Another widely used error type division distinguishes between adequacy and acceptability (Toury, 1995), with acceptability being more or less equivalent to fluency in that it reflects the degree to which the translation complies with the norms and expectations of the target text and its readers (Daems et al., 2013; Castilho et al., 2018). These types of error analyses often require the data segments to be at least at sentence level (TAUS, 2017), which may also constitute a disadvantage considering coherence throughout the text as a whole may more easily be overlooked when individual sentences are analysed (Daems et al., 2013). With the increasing adoption of MT more recently, automatic evaluation methods and PE effort measuring have been added to the mix of TQA-models (Nießen, Och, Leusch, & Ney, 2000; Papineni, Roukos, Ward, & Zhu, 2002; Snover, Dorr, Schwartz, Micciulla, & Makhoul, 2006; Lavie & Agarwal, 2007; Castilho et al., 2018). Castilho et al. (2018) noted, in that respect, that the increasing harmonisation of HT and MT may lead to the TQA approaches for these different types of translation eventually converging.

In her extensive overview of TQA approaches and developments, House (2014) argued that TQA ideally consists of a separate ideational and interpersonal component: a linguistic analysis based on research is necessary, but must be followed by value judgements, since "analysing without judging is pointless" (House, 2014, p. 262). Correspondingly, van Egdom and Pluymaekers (2019, p. 158) concluded that while the extent of post-editing has an influence on the quality of a final translation, "the additional effort associated with higher degrees of post-editing does not necessarily lead to more positive judgments about text quality." An extensive

overview of TQA methods would be outside the scope of the present dissertation, but as both MT and PE quality will be assessed and compared, it is relevant to discuss a number of TQA models and metrics in greater detail.

2.3.1 Machine Translation Quality

The importance of MTQA is generally acknowledged in translation studies and industry; LSPs using MT are to deliver target translations of a certain level of quality, and research is aimed at continuously studying and improving MT. MTQA is either carried out by means of automatic or human evaluation, or by measuring PE effort. It is predominantly error-based (Castilho et al., 2018), meaning that MT output is compared to reference or post-edited translations. Widely used automatic metrics include Word Error Rate (WER, Nießen et al., 2000), in which the number of edits – inserted, deleted, and substituted words – between the MT output and a reference translation are counted, and Bilingual Evaluation Understudy (BLEU, Papineni et al., 2002), which calculates numbers of n-grams co-occurring in MT output and its human reference translation. Though BLEU is extensively referenced and has shown to sometimes correlate well with human QA (Agarwal & Lavie, 2008; Farrús, Ruiz Costa-Jussà, Popovic, & Henriquez, 2012), it is most often deemed inadequate for measuring quality due to its many limitations and lack of consistent correlation with human judgements (Callison-Burch, Osborne, & Koehn, 2006; Snover et al., 2006; TAUS, 2017; Way, 2018). Additionally taking into account matches that are simple morphological variants or synonyms, METEOR (Banerjee & Lavie, 2005) was devised as an improvement on BLEU, and does indeed seem to correlate better with human judgements. Translation Error Rate (TER, Snover et al., 2006) and Human Translation Rate (HTER, Snover et al., 2006), both derived from WER, are defined as "the minimum number of edits needed to change a hypothesis so that it exactly matches one of the references, normalized by the average length of the references", with the human variant HTER involving a targeted reference solely created for measuring HTER (Snover et al., 2006, pp. 225-226). A final mention should be made of F-measure (van Rijsbergen, 1979), which is the harmonic mean of precision – measuring the quality of a MT system – and recall – measuring its quantity (completeness). Automatic metrics provide faster and cheaper TQA than human evaluation, but have been proven to not always adequately represent the quality of a translation (Callison-Burch, Osborne, & Koehn, 2006; Snover et al., 2006; TAUS, 2017; Way, 2018).

Human evaluation of MT, on the other hand, has the intrinsic disadvantages of being more subjective and more labour-intensive than automatic evaluation methods (Snover et al., 2006; de Almeida, 2013; Daems et al., 2014). Nevertheless, a 2013 study by Doherty, Gaspari, Groves and van Genabith found it the preferred method over both automatic and internally developed methods by about 70% of the nearly 500 translation and localisation buyers and vendors filling out the survey. Human MTQA has in research frequently been carried out using error analyses (Vilar, Xu, Luis Fernando, & Ney 2006; Koponen, 2010, Daems, Macken, & Vandepitte, 2014; Lommel, Uszkoreit, & Burchardt, 2014; Costa, Ling, Luís, Correia, & Coheur, 2015; Tezcan, Hoste, & Macken, 2017b; Tezcan, Hoste, & Macken, 2020), which are often based on commonly used error typologies in the localisation industry such as SAE-J2450 (2001), LISA (2011) and EN 15038 (2006). Two error typologies, the TAUS Dynamic Quality Framework (DQF, O'Brien, Choudhury, Van der Meer, & Aranberri Monasterio, 2011) and the Multidimensional Quality Metrics (MQM, 2014), were developed with the aim of providing a standardised yet flexible TQA model, and have received attention as well in both the localisation industry and translation studies (O'Brien, 2012; Melby, Fields, & Housley, 2014; Lommel, 2018; Yamada, 2019; Vardaro, Schaeffer, & Hansen-Schirra, 2019). It should be noted, though, that the latter includes principles from the LISA QA Model (LISA, 2011) and is applicable to both MT output and HT (Castilho et al., 2018), whereas this is not the case for the DQF. A merger of the two, called the DQF-MQM Error Typology, was proposed in 2014 and "has seen significant uptake in industry, research, and academia" in recent years, further eliminating the much-debated subjectivity and inconsistency in TQA (Lommel, 2018, p. 109). As manually classifying errors in accordance with a typology has proven to be a very time-consuming and often difficult task (de Almeida, 2013; Daems et al., 2014), automatic error classification tools have been devised as well, including Hjerson (Popović, 2011), Addicter (Zeman, Fishel, Berka, & Bojar, 2011) and Hjerson+ (Popović, Arcan, Avramidis, Burchardt, & Lommel, 2015).

2.3.2 Quality of the Post-Edited Product

A number of studies devising and adopting error typologies to assess MT quality have evaluated the quality of post-edited MT – often by translation students – and compared it to MT output quality, or compared students' to professional translators' work. Alongside automatic metrics, also error typologies have been used for this: Koponen and Salmi (2017), for example, relied on HTER and only a very limited post-edit categorisation (consisting of no more than six

categories) to investigate the correctness and necessity of translation student post-edits. Daems, Macken and Vandepitte (2014) investigated the quality of MT output and post-edited end products by translation students as well, using their own two-step TQA approach devised and tested the year prior (Daems, Macken, & Vandepitte, 2013). This two-step 2013 error typology is based on existing typologies (SAE-J2450, 2001; LISA, 2011; EN 15038, 2006) and distinguishes between adequacy and acceptability errors (Toury, 1995). In the adequacy step, the (only) main category is *meaning shift*; the acceptability error types are divided into five main categories: *grammar and syntax*, *lexicon*, *spelling and typos*, *style and register*, and *coherence* (Daems, Macken, & Vandepitte, 2013, p. 65). In both the first test experiment by Daems et al. (2013) and the follow-up 2014 experiment (Daems et al., 2014), error weights from 0 to 4 were assigned to each of the categories, depending on the text type and the impact the error would have on the readability and comprehensibility of the text. A '*meaning shift – terminology*' error would receive a high error weight in technical texts, for example (Daems et al., 2013). The findings of the 2014 experiment indicated that the main error types in Google Translate[3] SMT of newspaper articles, are grammatical errors such as word order errors and missing words, followed by wrong collocations and word sense disambiguation errors. As wrong collocations and word sense errors were still found in the post-edited target texts – alongside three types of spelling errors – the researchers concluded that most of the PE errors made by translation students appeared to be caused by the MT errors (Daems et al., 2014). Interestingly, using both the MQM framework and Hjerson for NMT and NMTPE error annotation, Vardaro et al. (2019) also observed a priming effect of MT output on highly professional translators post-editing at the German Language Department of the DGT. The most common error types in both the NMT output and the post-edited texts in their analysis, included stylistic/register, lexical, and function word errors, and mistranslations.

The two-step TQA approach by Daems et al. (2013) was also adopted in the product analysis of a 2017 study by Daems, Vandepitte, Hartsuiker and Macken, which consisted of a process and product comparison of both post-editing to human translation of newspaper articles (general texts), and students to professional translators. All texts in this study were annotated following the fine-grained TQA typology (Daems et al., 2013) and using the BRAT rapid annotation tool (Stenetorp, Pyysalo, Topic, Ananiadou, & Aizawa, 2012), resulting in no statistically significant difference between translation product quality of human translation and post-editing.

[3] https://translate.google.com

Students and professional translators also appeared to attain similar levels of quality, which the researchers ascribed to the fact that the translation students' training had mostly involved general texts. They therefore hypothesised to see greater differences between students and professional translators for specialised texts (Daems, Vandepitte, Hartsuiker, & Macken, 2017). An additional comparison of the error types most common for all four variables showed that students struggled most with meaning shifts (adequacy errors), although not as much when post-editing as opposed to translating from scratch. Interestingly, professionals made many spelling mistakes and made more of them when translating from scratch than when post-editing, whereas students made fewer spelling mistakes overall, but included more when post-editing than when translating from scratch. Another remarkable finding was that in the post-edited texts by student translators, the most common error type was '*logical problem*', whereas this was not at all observed by Daems et al. (2014).

Studies on the quality of post-edited end products have thus also laid focus on the relationship between post-editing performance and translation experience. De Almeida and O'Brien (2010) showed that more experienced translators were faster post-editors, but also made more preferential changes. Similarly, Depraetere (2010) noted a "striking difference in the mindset between translation trainees and professionals" (p. 6), as students did not rewrite passages of which the meaning was clear. In contrast, de Almeida (2013) did not find any clear correlations between post-editing performance and translation experience or post-editing experience. As stated above, a customised translation error typology was established in her extensive study, combining and adjusting error categories from the Gale PE Guidelines and the LISA QA Model and aimed at describing post-editor changes rather than evaluating the quality of the translation products. The typology did therefore not include the levels of severity adopted in the LISA QA Model. Instead, she opted for four major categories: "Essential changes", "Preferential changes", "Essential changes not implemented" and "Introduced errors", all of which encompassed a same set of subcategories. For the classification of overlapping MT errors, a strategy was adopted that had been used prior by Krings (2001, p. 266) as well: every assignment of a category was counted as "a discrete error". In line with de Almeida's results (2013), Guerberof (2014) also found that professional experience did not have a significant impact on processing speed in the post-editing of fuzzy match segments nor MT; the more experienced translators in the study performed similar to the novice translators. Notwithstanding, translators with more experience appeared to make significantly fewer mistakes than those with less experience, obtaining higher-quality end products. The MT output

did, however, seem to have had a levelling effect on the difference between the numbers of errors.

2.4 Machine Translation at the European Commission DGT

As is the case for the whole of the translation industry, in recent years the number of pages to be translated at the European Commission Directorate-General for Translation (DGT) has gone up tremendously whereas resources are increasingly limited and quality standards remain equally elevated (Rossi & Chevrot, 2019; Macken, Prou, & Tezcan, 2020; Stefaniak, 2020). MT is therefore increasingly being relied on (Macken et al., 2020). The DGT is the largest translation service in Europe and "[a]rguably the largest translation agency in the world" (Koskinen, 2008, p. 69) and MT has been used since the 1980s, when the RBMT system based on Systran was in practice. In 2013 this first system made way for a new SMT system called MT@EC that was based on Moses, and MT@EC has in turn gradually been replaced by the new NMT-based system eTranslation since November 2017 (Macken et al., 2020). MT is at the DGT integrated as a tool within their CAT environment based on SDL Trados Studio, in the form of translation memory exchange (TMX) files, and can be used alongside the large EURAMIS central translation memory, implying that the translators can make use of MT suggestions when no TM match above the 75% threshold is available, but still have the option to translate from scratch as well (Macken et al., 2020; Stefaniak, 2020). MT suggestions can moreover be directly inserted into the segment, or used in autosuggest mode (Stefaniak, 2020). Internal statistics revealed that for an average of 70% of their translation projects, translators at the DGT enable MT, which indicates that most DGT translators either systematically or sporadically make use of MT (Macken et al., 2020). Interestingly, Rossi and Chevrot (2019) also found that the degree to which DGT translators perceive MT as useful and actually use it, is influenced by their perceptions of MT.

In studies at the DGT, the extraordinarily high standards for Commission translations – often involving legal or political implications – are regularly highlighted (Svoboda, 2017; Rossi & Chevrot, 2019; Vardaro, Schaeffer, & Hansen-Schirra, 2019; Arnejšek & Unk, 2020; Macken et al., 2020; Stefaniak, 2020). MT has, due to these elevated quality standards, proven to be considered "just a tool" (Arnejšek & Unk, 2020, p. 9), and even perfect MT suggestions need verification by a human post-editor – requiring a minimum of post-editing time (Macken et al., 2020). In accordance with the ISO 17100:2015 standard, moreover, post-edited texts are to be

revised as would any other translation product (International Organization for Standardization, 2015; Vardaro et al., 2019). As also noted by Rossi and Chevrot (2019) and Macken et al. (2020), the embedding of MT in the highly advanced CAT workflow may limit its impact on productivity and therefore the extent to which it is perceived as useful and is actually used by DGT translators. The EURAMIS translation memory is extensive, and suggested MT segments are easily dismissible (Rossi & Chevrot, 2019). Several analyses of translators' MT perceptions and adoption have been conducted at different DGT Language Departments and have aimed at assessing this impact of MT on the DGT translation workflow. Rossi and Chevrot (2019), for example, collected ethnographic data and conducted semi-directed interviews, from which they concluded that technology was overall largely accepted at the DGT. Of the 89 translators from 15 different language departments (of the total of 24 departments) having taken part in the study, moreover, just over a third mentioned using MT suggestions mainly to save time or as a typing aid; the remaining part appeared to use MT for terminology, as a source of inspiration, or for specific contexts only. Some negative answers were given as well, and although good overall knowledge of MT was observed among their subjects, they did also find a significant correlation between the lack of knowledge on MT and the perception of MT as a threat – through the 11% of the participants that did not have any MT knowledge.

Lesznyák (2019) observed more mixed opinions on NMT at the DGT's Hungarian Language Department – not included in Rossi and Chevrot (2019). Though many of the Hungarian translators perceived NMT as useful overall, it was not implemented on a daily basis by many, claiming its quality varied along the segment and therefore considering it not always reliable. Among the benefits of MT, the translators did report reductions in typing effort and psychological benefits such as not having to start translating from a blank segment (Lesznyák, 2019). Similar to Lesznyák (2019), Stefaniak (2020, pp. 6-7) reported on the risk of reduced quality with usage of NMT for the English-Polish language pair at the Polish Language Department, concluding that "the productivity gain when NMT is used to complement TM matches is still modest" for this language pair, and that NMT was at this department generally not perceived as very useful. Macken, Prou and Tezcan (2020) analysed the impact of English-French SMT and English-Finnish NMT on the translation workflow at the French and the Finnish Language Departments respectively, and found PE leading to translation speed gains, but relatively limited ones in comparison to results of studies having analysed different text types in different experimental conditions. They attributed their findings to the fact that they investigated the highly professional translators working in their normal working conditions,

which had not been done before, and that it is difficult in general to compare productivity gains. As is often the case in comparable studies (e.g. Stefaniak, 2020), considerable variation was also observed between individual translators (Macken et al., 2020), and reduction of typing effort and psychological benefits were again listed among the main reasons for using MT (Macken et al., 2020). Having analysed the reported errors of eTranslation NMT at the Slovene Language Department, Arnejšek and Unk (2020) concluded that "NMT output cannot be used as is" (p. 9) and argued that a higher awareness and a better understanding of NMT error types through training and education are essential. Their results authenticated the recurring error categories in NMT output, such as inconsistencies and semantic and lexical errors, with additionally punctuation appearing to be problematic. Despite the low quality of the NMT output as compared to human translation, the Slovene translators did prove to both highly appreciate and adopt NMT, deeming it an improvement on the SMT system used prior.

2.5 Professional Translators

As the aim of this study is to investigate which alterations are made to NMT output by professional DGT translators in order to obtain high-quality translations – thus lay bare professional post-editing behaviour – through an analysis of the types of post-edits carried out, their frequencies, and their necessity, a last brief note should be made on what is considered a "professional" in the translation industry. This notion of what makes for a "professional", has been approached from various angles in the literature, a common one of which is years of translator experience. Eszenyi (2016), however, defined a professional in terms of the modern translation industry, and used as the base for his definition the six translator competences drawn up by the EMT (European Masters in Translation) Expert Group of the European Commission DGT (Gambier, 2009):

1. translator service provision competence,
2. language competence,
3. intercultural competence,
4. information mining competence,
5. technological competence/mastery of tools,
6. thematic competence.

Building on these six competences, Eszenyi (2016) described a modern professional translator as a linguist eager to continuously expand their knowledge and an expert mastering beyond-average linguistic knowledge and belonging to a professional community, but at the same time an entrepreneur with adequate business skills and a technician managing various programmes and online tools such as databases and dictionaries. Professional translators' competences are therefore versatile and dynamic; professionals are willing to evolve along with both innovations in the profession and changes in the world (Eszenyi, 2016).

A professional translator, however, does not necessarily make for a good post-editor. As pointed out by de Almeida & O'Brien (2010), translator experience does not always positively correlate to post-editing performance, and post-editing would require a different set of skills than translating from scratch. In this respect, de Almeida (2013, p. 105) made a note on what can be considered a good post-editor: "a good post-editor would meet the requirements of the specific PE task at hand, making changes and corrections only according to the guidelines provided, and delivering a final text with the required level of quality in the time-frame specified."

2.6 Research Questions

A conclusion that can be drawn from the literature review presented in this chapter, is that while much has been investigated regarding MT, PE (effort) and evaluation metrics for both MT and its post-edited end product, and several studies have analysed post-editor behaviour and attitudes in the setting of (a department within) the European Commission Directorate-General for Translation (DGT), adoption and perceptions of MT within the DGT Dutch Language Department have not yet been thoroughly studied. In 2017, the department was still listed among those in which MT – SMT at the time – was least adopted (Kluvanec, 2017, slide 20), and though this was observed with NMT as well in 2019 (Rossi & Chevrot, 2019), the increasing integration of NMT in translation workflows and the overall improvement of NMT may have brought about a shift in its adoption at this department as well. Other than the recent study by Vandevoorde, Weintraub and Arabadjieva (2021) at the Translation Service of the Council of the European Union, few studies have also in general analysed the types of post-edits most frequently occurring in very high-quality English-Dutch translations, and their necessity, for which it would especially be of interest and relevance to do so at the DGT Dutch Language Department. Awareness of the types of MT errors may moreover increase PE productivity and quality (Martínez, 2003), and may be of importance for possible establishment

of DGT-tailored high-quality English-Dutch PE guidelines, in turn increasing PE productivity and quality at this department as well.

The present dissertation is therefore aimed at analysing the types of post-edits most frequently occurring in high-quality English-Dutch DGT translations, as well as their correctness and necessity. For this, taking into consideration the above-mentioned literature and concepts, the following four research questions have been drawn up:

RQ1 To what extent do English-Dutch translators at the Dutch Language Department of the European Commission Directorate-General for Translation make use of MT?

RQ2 How many post-edits carried out by English-Dutch DGT translators can be considered essential, preferential, and undesirable changes?

RQ3 Which types of changes are most often carried out by English-Dutch DGT translators when post-editing English to Dutch MT?

RQ4 How many post-edits carried out by English-Dutch DGT translators are changes related to the meaning of the source text (Toury's norm of adequacy), and how many are in function of the acceptability of the target text in the target culture and its suitability for its function (Toury's norm of acceptability)?

3 METHOD

This chapter covers the method that was adopted for the collection and analysis of the data used for the present study. First, the data provided and the sample selection process are discussed, after which the typology developed for the analysis is elaborated on, to conclude with the method for the actual annotation and analysis of the post-edits.

3.1 Data

A total of 69 SDXLIFF files containing English source and Dutch target texts along with metadata, was provided by the DGT Dutch Language Department, in two sets of 3 and 66 files translated by 1 and 9 translators respectively. As these 10 translators remained anonymous, they will hereafter be referred to as Translators 1 to 10. The files included formal (legislative) texts such as contracts, reports, regulations, directives, policies and plans, communications of the Commission, etc., though translations by Translator 10 included informal texts such as blog posts as well. Both sets of SDXLIFF files were converted to Excel files containing the aligned and segmented source and target texts along with the metadata for each of the segments. As mentioned earlier, the DGT translation workflow allows translators to use the extensive central EURAMIS translation memory, and supplement the translation process with eTranslation NMT as they see fit. In which way a segment was translated, is indicated in the metadata by means of different codes. Segments for which either a TM match or MT is used, are split up into *tmedited* and *tmunchanged*, and *mtedited* and *mtunchanged*, respectively. Other than those four options, the segment types occurring in the data sets were *autopropedited*, *autopropunchanged*, *copysource*, *copysourceedited*, *emptytarget*, *fromscratch* and *unclassified*, all self-explanatory. An important note to make here is that the raw MT output itself was not included in the SDXLIFF files provided by the DGT, but was manually re-generated using the eTranslation EU Formal Language MT system – of which the output was verified through comparison with *mtunchanged* segments – on 28 January 2021 (first set) and 18 March 2021 (second set). The MT output was then inserted into the Excel files in a new column in between the source and target segment columns, for each of the machine translated segments.

The first set of three translations (aligned source and target texts) were all texts translated by 1 translator, and was made up of 2,096 segments in total, of which 73 were MT output segments: 63 edited MT segments and 10 unchanged MT segments. The second set of 66 translations by 9 translators, received later in the study, consisted of 14,884 segments total, of

which 3,307 were MT output segments: 2,996 edited MT segments and 311 unchanged MT segments. After a test analysis of the first data set (as further discussed in Section 3.3), the two sets were merged into one Excel file, bringing the total number of segments in the 69 translations to 16,980, of which 3,380 MT segments: 3,059 edited and 321 unchanged. This means that of all translations provided by the DGT Dutch Language Department, MT was used for one fifth (20%) of the segments. In addition to this, 91% of these machine translated segments were post-edited; very few MT translations were left unchanged. Table 1 provides a schematic overview of the final data set on segment level, including (from left to right)

#TRANSL	the number of translations;
TOT SEG	the total number of segments;
#TRANSL+MT	the number of translations in which MT was used (included because translator 5 did not implement MT in every translation);
TOT MT SEG	the total number of machine translated segments (both edited and unchanged);
%MT/TOT SEG	the total number of machine translated segments (both edited and unchanged) in relation to the total number of segments;
TOT MT ED SEG	the total number of post-edited MT segments;
%MT ED/TOT MT SEG	the total number of post-edited MT segments in relation to the total number of machine translated segments;
%MT ED/TOT SEG	the total number of post-edited MT segments in relation to the total number of segments.

	#TRANSL	TOT SEG	#TRANSL +MT	TOT MT SEG	%MT/ TOT SEG	TOT MT ED SEG	%MT ED/TOT MT SEG	%MT ED/TOT SEG
Translator 1	3	2096	3	73	3%	63	86%	3%
Translator 2	8	1598	8	852	53%	785	92%	49%
Translator 3	4	298	4	77	26%	71	92%	24%
Translator 4	10	2349	10	602	26%	528	88%	22%
Translator 5	10	892	2	162	18%	157	97%	18%
Translator 6	10	491	10	190	39%	174	92%	35%
Translator 7	7	2935	7	356	12%	319	90%	11%
Translator 8	1	4484	1	253	6%	232	92%	5%
Translator 9	6	1023	6	679	66%	611	90%	60%
Translator 10	10	814	10	136	17%	119	88%	15%
Total	**69**	**16980**	**61**	**3380**	**20%**	**3059**	**91%**	**18%**

Table 1. Schematic overview of the final data set on segment level

As, however, both (perfect) TM matches and unchanged MT segments are very regularly short segments (such as titles, list items, links, etc.), an analysis of the words was carried out for the full data set as well, to obtain a more accurate view. For this, the numbers of *ST tokens* of each of the (MT) segments, shown in the metadata within the Excel file, were cumulated. Of the total of 258,384 words, 75,659 were machine translated – amounting to 29% – of which in turn 71,803 were edited MT words – constituting 95% of all machine translated words and 28% of the total word count. These figures on word level are, indeed, higher than those obtained at segment level, and as they were for the individual translators anywhere between 1 to even 26% higher than on segment level as well, a second schematic overview has been included in Table 2. The columns overlap with those of Table 1 on segment level above; only the two columns with the numbers of translations (in which MT was used) have been excluded from the table.

	TOT WORDS	TOT MT WORDS	%MT/ TOT WORDS	TOT MT ED WORDS	%MT ED/ TOT MT WORDS	%MT ED/ TOT WORDS
Translator 1	34363	2049	6%	1919	94%	6%
Translator 2	24480	19083	78%	18269	96%	75%
Translator 3	5922	2588	44%	2499	97%	42%
Translator 4	41217	14076	34%	13042	93%	32%
Translator 5	17992	3717	21%	3597	97%	20%
Translator 6	7596	4457	59%	4255	95%	56%
Translator 7	22174	6447	29%	6138	95%	28%
Translator 8	78620	5877	7%	5688	97%	7%
Translator 9	19726	15122	77%	14260	94%	72%
Translator 10	6294	2243	36%	2136	95%	34%
Total	**258384**	**75659**	**29%**	**71803**	**95%**	**28%**

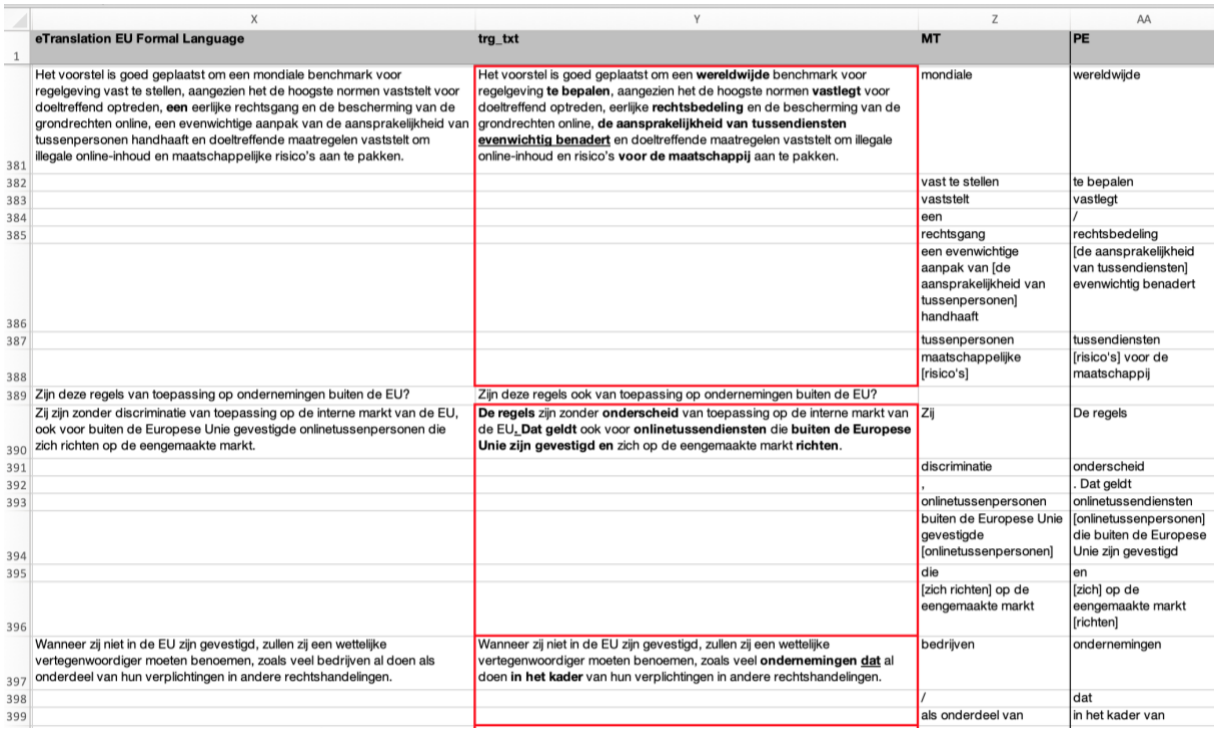
Table 2. Schematic overview of the final data set on word level

3.2 Sample Selection and Post-Edit Extraction

In a separate Excel sheet for each of the translators, the segments were first sorted by translation, next by segment type (alphabetically), and lastly by segment number (consecutively), to obtain a clear view of exactly how many segments within each text of each translator were post-edited. As the first set of three translations by Translator 1 was provided first, a test annotation of the 63 post-edited MT segments in these 3 translations was carried out, to provide a starting point

for the development of the post-edit typology. For this test analysis, *all* post-edits in the target texts were indicated, extracted, and annotated. For the remainder of the translators, however, a selection of post-edited segments needed to be made due to the size of the final data set in comparison to the limited scope of this study and time available. A choice was made to use MT segments from that text of each post-editor in which the most *mtedited* segments occurred, and select the first 10 of those segments that were of sufficient length for analysis of the post-edits carried out in them. No exact character threshold was established; perceptibly short segments were dismissed. Of the segments selected, the target text cells were marked with a red frame so as to be clearly visible in the sheet. For the definite annotation of Translator 1 (after the typology had been devised and the other translators had been annotated), this same sample selection procedure was followed. Translator 10 was excluded from the analysis, as the text type (blog posts) and register of the only texts with sufficient *mtedited* segments differed from that of the other translators' texts – more informal as opposed to the other, very formal texts.

All differences between the MT output and the post-edited target text – i.e. all post-edits – in the 90 segments selected for annotation were manually indicated by placing the words or punctuation marks in bold in the target text cells, or in the corresponding MT output cell in case of a deletion. Post-edits occurring just after other separate ones in a sentence, and punctuation marks that were not sufficiently clearly marked in bold alone, were additionally underlined to make them more visible or to clarify the distinction between different post-edits in one sentence, with the objective of easing the later extraction and annotation processes. All post-edits were then extracted from the segments and placed in two columns next to the target text column: one with the original MT translations (**MT**), and one with the post-edits (**PE**). Text that could aid in understanding a post-edit – without having to consult the full MT and target sentences – but that was not actually part of the post-edit, was included between square brackets (e.g. in case of a change within the sentence order). Deletions were indicated by a slash (/) in the PE column, additions by a slash in the MT column. Figure 1 provides a screenshot (taken from Translator 5) of a fragment of the Excel annotation file, for clarification of the above.



	X	Y	Z	AA
1	**eTranslation EU Formal Language**	**trg_txt**	**MT**	**PE**
381	Het voorstel is goed geplaatst om een mondiale benchmark voor regelgeving vast te stellen, aangezien het de hoogste normen vaststelt voor doeltreffend optreden, **een** eerlijke rechtsgang en de bescherming van de grondrechten online, een evenwichtige aanpak van de aansprakelijkheid van tussenpersonen handhaaft en doeltreffende maatregelen vaststelt om illegale online-inhoud en maatschappelijke risico's aan te pakken.	Het voorstel is goed geplaatst om een **wereldwijde** benchmark voor regelgeving **te bepalen**, aangezien het de hoogste normen **vastlegt** voor doeltreffend optreden, eerlijke **rechtsbedeling** en de bescherming van de grondrechten online, **de aansprakelijkheid van tussendiensten evenwichtig benadert** en doeltreffende maatregelen vaststelt om illegale online-inhoud en risico's **voor de maatschappij** aan te pakken.	mondiale	wereldwijde
382			vast te stellen	te bepalen
383			vaststelt	vastlegt
384			een	/
385			rechtsgang	rechtsbedeling
386			een evenwichtige aanpak van [de aansprakelijkheid van tussenpersonen] handhaaft	[de aansprakelijkheid van tussendiensten] evenwichtig benadert
387			tussenpersonen	tussendiensten
388			maatschappelijke [risico's]	[risico's] voor de maatschappij
389	Zijn deze regels van toepassing op ondernemingen buiten de EU?	Zijn deze regels ook van toepassing op ondernemingen buiten de EU?		
390	Zij zijn zonder discriminatie van toepassing op de interne markt van de EU, ook voor buiten de Europese Unie gevestigde onlinetussenpersonen die zich richten op de eengemaakte markt.	**De regels** zijn zonder **onderscheid** van toepassing op de interne markt van de EU**. Dat geldt** ook voor **onlinetussendiensten** die **buiten de Europese Unie zijn gevestigd en** zich op de eengemaakte markt **richten**.	Zij	De regels
391			discriminatie	onderscheid
392			,	. Dat geldt
393			onlinetussenpersonen	onlinetussendiensten
394			buiten de Europese Unie gevestigde [onlinetussenpersonen]	[onlinetussenpersonen] die buiten de Europese Unie zijn gevestigd
395			die	en
396			[zich richten] op de eengemaakte markt	[zich] op de eengemaakte markt [richten]
397	Wanneer zij niet in de EU zijn gevestigd, zullen zij een wettelijke vertegenwoordiger moeten benoemen, zoals veel bedrijven al doen als onderdeel van hun verplichtingen in andere rechtshandelingen.	Wanneer zij niet in de EU zijn gevestigd, zullen zij een wettelijke vertegenwoordiger moeten benoemen, zoals veel **ondernemingen dat** al doen **in het kader** van hun verplichtingen in andere rechtshandelingen.	bedrijven	ondernemingen
398			/	dat
399			als onderdeel van	in het kader van

Figure 1. Fragment of the sample in the Excel annotation file (taken from Translator 5)

From the 90-segment sample of the 9 translators, a total of 493 post-edits were extracted, constituting an average of 55 post-edits per translator, spread over their 10 segments. Table 3 provides an overview of the numbers of post-edits extracted and annotated for each of the translators.

Translator 1	58
Translator 2	41
Translator 3	92
Translator 4	37
Translator 5	61
Translator 6	53
Translator 7	57
Translator 8	51
Translator 9	43
Total	**493**
Average	**55**

Table 3. Numbers of post-edits extracted per translator and in total

3.3 Typology

The test annotation of the first data set by Translator 1, was largely carried out using a combination of categories from the two-step translation quality assessment by Daems, Macken & Vandepitte (2013; 2014; Daems, 2016) and the post-edit typology devised by de Almeida (2013). These two typologies were opted for because they were thought to be most suitable for the research objectives of the present study, and because they encompass categories from TQA models widely used in the localisation and translation industries, such as the LISA TQA model (2011) and the GALE post-editing guidelines (2007), which in turn had also already been tailored to researching PE corrections and PE and MT errors (Daems et al., 2013; de Almeida, 2013; Daems et al., 2014). Both typologies, however, lay focus on PE (and MT) *errors*, whereas this is not the case for the present study: rather than PE errors, professional PE *corrections and alterations* are examined. In addition to that, all of the above-mentioned studies were conducted on SMT, whereas the current MT system at the DGT Dutch Language Department, eTranslation, relies on NMT. NMT systems have already proven to make distinct errors from SMT errors, in turn leading to different necessary PE corrections and alterations (e.g. Bentivogli, Bisazza, Cettolo, & Federico, 2016). For these two reasons, the present study could be considered sufficiently different from Daems et al. (2013; 2014; Daems, 2016) and de Almeida (2013) in terms of research material and objectives, for which a new typology *based on* the two typologies was devised, after the test annotation. The typology was continuously re-evaluated and adjusted throughout the entire annotation process of the second data set (Translators 2-9), and was then used for the definite annotation of the segments selected from Translator 1.

A choice was made to categorise post-edits from three different (and independent) perspectives, and distinguish between six different parameters. Firstly, a Translation Quality Parameter based on de Almeida (2013) was used to annotate a post-edit for its correctness and necessity. A distinction was made between *style guide changes*, *consistency & intertextuality changes*, *MT errors*, *preferential changes*, and *undesirable changes*. It should be highlighted that style guide changes are a specific type of essential change, and consistency & intertextuality changes are in turn a type of style guide change (and therefore essential as well), insofar as a style guide is provided for a translation, or internal quality guidelines are applicable. For high-quality translations this is very likely to be the case, and at the DGT, language-specific style guides (such as for Dutch) are indeed numerous and extensive, prescribing especially many (internal)

punctuation and capitalisation standards (e.g. Svoboda, 2017). Terminology as well, should not be deviated from (e.g. documents often have legislative implications). To ease the annotation process in the Excel file, codes were assigned to all categories in the typology. Table 4 shows a schematic overview of the Translation Quality Parameter with the annotation codes for each of the categories. The definitions of each of the categories from all parameters can be found in Appendix A, along with examples taken from the post-edited DGT texts. Only the text-linguistic categories *comparative/superlative of an adjective/adverb*, *merged sentences* and *other stylistic change* did not occur in any of the segments analysed, for which no examples of them are included in the typology. They were, however, not eliminated from the typology, as they may still occur in other post-edited segments.

Translation Quality Parameter				
Style guide changes	Consistency & intertextuality changes	MT errors	Preferential changes	Undesirable changes
SG	CON	MT ERROR	PREF	UNDES

Table 4. Schematic overview and annotation codes of the Translation Quality Parameter

In a second step, post-edits were annotated in terms of the relation between the target text and the source text (Toury's norm of *adequacy*), and the target text and the target language (Toury's norm of *acceptability*): the Translation Norm Parameter, based on the two-step TQA by Daems et al. (2013). As a *combination* of the two norms is possible as well, this third category was added. Table 5 shows the three Translation Norm Parameter categories with the annotation codes for each of them.

Translation Norm Parameter		
Adequacy	Acceptability	Adequacy + acceptability
AD	ACC	AD+ACC

Table 5. Schematic overview and annotation codes of the Translation Norm Parameter

The final four parameters: the Semantics Parameter, Syntax & Morphology Parameter, Style & Register Parameter and Spelling & Punctuation Parameter, were grouped under the umbrella category of Text-Linguistic Characteristics, as they all serve to classify a post-edit from a (text-)linguistic perspective. The categories in these four parameters were based on categories

from both de Almeida (2013) and Daems et al. (2013) as well, but their focus was shifted from MT and PE errors to PE alterations. Not all types of categories from the two typologies were incorporated in the present typology, and entirely new categories for often-occurring phenomena in the sample were added as appropriate (e.g. *nominalisation*). Table 6 provides a schematic overview of the Text-Linguistic categories with the annotation codes for each of them.

Semantics Parameter	Addition	Addition of source text information missing from the MT output	ADD ST INFO
		Explicitation of ST and MT output	EXPLIC
		Hyponymy	HYPONYMY
		Other addition	OTH ADD
	Deletion	Deletion of ST and MT information	DEL LEX INFO
		Deletion of extra information in MT output	DEL EXTRA INFO
		Hyperonymy	HYPERONYMY
	Other meaning shift	Action	AC
		Time	T
		Place	P
		Agent	AG
		Modality	MOD
		Other meaning change	OTH MEAN CH
Syntax & Morphology Parameter	Order	Internal phrase order	PHRASE ORDER
		Internal sentence order	SENT ORDER
	Agreement	Article-noun agreement	ART-NOUN
		Noun-adjective agreement	NOUN-ADJ
		Subject-verb agreement	SUBJ-VERB
		Reference	REF
	Structural change	Nominalisation	NOM
		Reduction of nominalisation	RED OF NOM
		Active → passive	ACT-PAS
		Passive → active	PAS-ACT
		Other structural change	OTH STRUCT CH
	Morphological change	Comparative/superlative of an adjective/adverb	COMP/SUPERL
		Singular/plural noun	SING/PLUR
		Named entity	NAMED ENT
		Other morphological change	OTH MORPH CH

Style & Register Parameter	Synonymy	SYNONYMY
	Preposition	PREP
	Lexical change	LEX CH
	Register	REG
	Untranslated text in MT	UNTRANS IN MT
	Split sentence	SPLIT SENT
	Merged sentences	MERG SENT
	Repetition	REP
	Reduction of repetition	RED OF REP
	Coherence	COH
	Conjunction	CONJ
	Other stylistic change	OTH STYL CH
Spelling & Punctuation Parameter	Capitalisation	CAP
	Compound	COMP
	Punctuation	PUNCT
	Numeral	NUMERAL
	Other spelling change	OTH SPEL CH

Table 6. Schematic overview and annotation codes of the Text-Linguistic Characteristics

3.4 Annotation

As stated above, after the extraction of all post-edits, each of them was analysed and categorised in accordance with the typology drawn up yet throughout the annotation process still subject to change. The annotation was carried out in the central Excel file containing all texts of all translators, as this way, an overview of both the context and the terminology and notations used throughout the (source) text / in (the) other texts (of the translator), could more easily be maintained. The post-edits were manually annotated by one researcher only, but the full annotations of two translators (Translator 2 and Translator 3) were checked by the supervisor and the co-supervisor of the present dissertation, as a control.

For the six columns next to the PE column, the relevant annotation codes were entered in Data Validation lists for each of the parameters, after which categories were assigned to the post-edits. Emphasis should again be laid on the stand-aloneness of the different parameters. The final column was used for comments on categorisation choices. For the annotation of all parameters, but particularly the Translation Quality Parameter and the Translation Norm

Parameter, the definitions in Appendix A were departed from. Unlike the four parameters under the Text-Linguistic Characteristics, for these two a category was assigned to every one of the post-edits. This is because a post-edit always pertains to one of the categories under the first two parameters (necessity/correctness and translation norm), but very rarely consists in a semantic, syntactical/morphological, style/register, and spelling/punctuation change all at the same time. The category *style guide change* was only assigned to post-editor corrections that are clearly specified in the internal Dutch Style Guide[4]. To establish whether a post-edit was a *consistency & intertextuality* change, words and phrases were looked up in IATE[5] and in TM segments throughout the text in question and the other texts in the Excel file. A post-edit was only assigned the *adequacy + acceptability* category when the MT output clearly did not convey the same meaning as the source text, while simultaneously causing a problem in the target language, whether it be logical, grammatical, in terms of spelling, etc.

For the annotation of the Text-Linguistic Characteristics, a decision tree based on the category definitions was drawn up, as the categories of those four parameters are more extensive and complicated. This decision tree is included in Appendix B. To compare translation units, validate hyponyms, hypernyms, and synonyms, and look up grammatical rules, parts of speech, capitalisation conventions and named entities, the following external resources were used: DeepL Translator[6], Google Translate[7], Van Dale's Great Dictionary of the Dutch Language and English-Dutch and Dutch-English Dictionaries[8], Cambridge Dictionary[9], Longman Dictionary of Contemporary English[10], and Google[11]. Certain post-edits were assigned a category from more than one parameter within the Text-Linguistic Characteristics (e.g. a word that was both moved in the sentence and substituted for a synonym). If, however, a syntactical change other than *internal phrase order* or *internal sentence order* automatically led to a change in the sentence structure – the order of phrases within the sentence – as well, then the *internal sentence order* category was not additionally attributed to that post-edit. Combinations of five of the six *other meaning shift* categories (excluding the final *other meaning change* category) were

[4] Publicly accessible via https://ec.europa.eu/info/sites/default/files/styleguide_dutch_dgt_nl.pdf
[5] https://iate.europa.eu/home
[6] https://www.deepl.com/translator
[7] https://translate.google.com/?sl=en&tl=nl&op=translate
[8] Consulted online via https://pakket95.vandale.nl/, through https://athena.ugent.be/
[9] Consulted online via https://dictionary.cambridge.org/dictionary/english/
[10] Consulted online via https://www.ldoceonline.com
[11] https://www.google.com/

possible (cf. Appendix B), as illustrated by the example below, in which an *action* and a *time* change were simultaneously implemented in one post-edit by Translator 9.

T9	**Action + Time**
(ST)	Though Europe's response **has demonstrated** strengths
(MT)	Hoewel de respons van Europa sterke punten **heeft aangetoond**
(PE)	Hoewel de respons van Europa ook sterke punten **kende**

The same goes for all types of Spelling & Punctuation changes: combinations of up to two of the five categories may occur in post-edits as well (cf. Appendix B). None were found in the sample, however, for which an example cannot be provided.

A choice was made to differentiate between synonymy and other lexical changes. As only different words between the MT output sentence and the post-edited sentence (and not the full different meaning units) were marked as post-edits and placed in the columns (cf. Figure 1), a single word in the PE column was sometimes not a synonym of the MT word, yet the post-edit had not brought about a meaning shift in the target sentence. Though the distinction is often minor, it was made from a linguistic point of view, to provide another option when the words extracted from the segment were not (entirely) synonymous. Below is an example of a synonymy annotation and a lexical change annotation of Translator 2 and Translator 3 respectively, to clarify the distinction.

T2	**Synonymy**
(ST)	however
(MT)	echter
(PE)	evenwel

T3	**Lexical change**	Back translation
(ST)	individuals **playing** sports on them	
(MT)	personen die sporten op hen **uitoefenen**	persons **exercising** sports on them
(PE)	mensen die daarop een sport **beoefenen**	people **practising** a sport on them

Substitutions of relative clauses or post-nominal phrases for adjectives and vice versa were classed as *other structural changes*. Below are two examples, by Translator 4 and Translator 2.

The definitions of all categories in the typology along with more examples can, again, be consulted in Appendix A.

T4	**Other structural change**	Back translation
(ST)	**interested** services providers	
(MT)	**geïnteresseerde** dienstverleners	**interested** service providers
(PE)	dienstverleners **die belangstelling hebben**	service providers **who are interested**

T2	**Other structural change**	Back translation
(ST)	**high diversity** landscape features	
(MT)	landschapselementen **met een grote diversiteit**	landscape elements **with great diversity**
(PE)	**diversiteitsrijke** landschapselementen	**diversity-rich** landscape elements

For clarification of the annotation process, Figure 2 provides a screenshot of (a fragment of) the Excel annotation file (taken from Translator 6); the Excel file containing all sample segments with the post-edits analysed, can be found in Appendix C (digital). For brevity considerations, this appendix does not contain the full translations.

	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	**MT**	**PE**	**TQP**	**TNP**	**SEMANTICS PARAMETER**	**SYNTAX & MORPHOLOGY PARAMETER**	**STYLE & REGISTER PARAMETER**	**SPELLING & PUNCTUATION PARAMETER**	**COMMENTS**
	Het	het	SG	ACC				CAP	
308	/	,	PREF	ACC				PUNCT	
309	de werking van [de media-activiteiten]	de media-activiteiten	PREF	AD	DEL LEX INFO				
310	Deel	deel	SG	ACC				CAP	
311	Vertegenwoordiging	vertegenwoordiging	CON	ACC				CAP	
312	/	dat	MT ERROR	ACC		OTH STRUCT CH			
313	1	één	SG	ACC				NUMERAL	
314	extra	bijkomende	PREF	ACC			SYNONYMY		
315	personen	persoon	MT ERROR	ACC		SING/PLUR			
316	ten aanzien van	ten overstaan van	UNDES	ACC			SYNONYMY		Van Dale: ten overstaan van = niet algemeen voor ten aanzien van
317	dienst	Dienst	CON	ACC				CAP	
318	Trends	trends	SG	ACC				CAP	
319									
320	te analyseren	analyseren	MT ERROR	ACC		OTH STRUCT CH			
321	toezicht te houden op	volgen	MT ERROR	AD	AC				
322	/	hierover	MT ERROR	AD	EXPLIC				
323	uit te brengen	uitbrengen	MT ERROR	ACC		OTH STRUCT CH			
324	/	aan	MT ERROR	AD	ADD ST INFO				
325	/	deze	MT ERROR	ACC	ADD ST INFO				
326	te informeren	informeren	MT ERROR	ACC		OTH STRUCT CH			
327	te waarschuwen	waarschuwen	MT ERROR	ACC		OTH STRUCT CH			
	voor zover van toepassing	indien nodig	PREF	ACC			SYNONYMY		
328	Toezicht	toezicht	SG	ACC				CAP	
329	Toezicht houden op en toezicht houden	toezicht [en controle] uitoefenen	PREF	ACC			LEX CH		
330	toezicht houden	controle uitoefenen	MT ERROR	AD	AC				
331	eerbiediging	inachtneming	PREF	ACC			SYNONYMY		
332	aanzienlijke	belangrijke	PREF	ACC			SYNONYMY		
333	[ten minste 2 jaar voltijdse beroepservaring	na het behalen van het in punt 3.3.1 bedoelde	PREF	ACC		SENT ORDER			

Figure 2. Fragment of the annotation in the Excel annotation file (taken from Translator 6)

4 RESULTS

This chapter presents the results of the post-edit analysis. The general results for all post-edits of all 9 translators are discussed first, after which a comparison is made between the individual translators. A number of annotation codes have been elaborated, for clarification.

4.1 General Results

4.1.1 Translation Quality Parameter

Figure 3 shows the distribution of the five Translation Quality Parameter categories among the total of 493 post-edits.

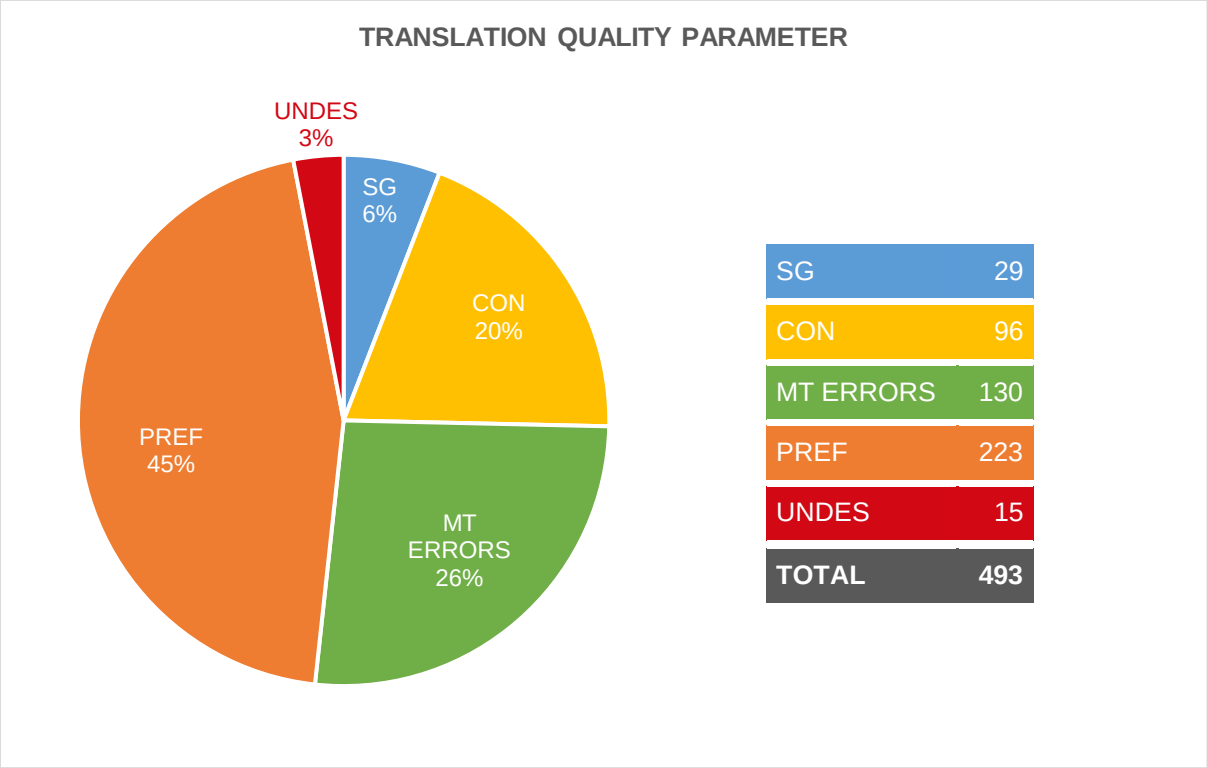


Figure 3. Distribution of the Translation Quality Parameter categories

Figure 3 indicates that 223 of all post-edits were preferential in nature (PREF), by which preferential changes constitute almost half (45%) of the total number. A quarter of all changes (26%) were MT errors (130), and combined with the 6% style guide changes (SG – 29) and 20% consistency changes (CON – 96), the two other specific types of essential changes in high-quality translations, the total number of essential changes (ALL ESS) amounts to 255, comprising just over half of all post-edits (52%). About as many essential changes as

preferential changes were thus carried out, the essential post-edits, however, prevailing. Only 15 (3%) of all post-edits could be considered undesirable (UNDES), most of which were (slight) meaning shifts in comparison to the ST and the MT output (e.g. deletions of ST information), minor grammatical errors, or the occasional adoption of inconsistent terminology (for which the MT system had generated the word or phrase generally used). A list of all undesirable changes can be found in Appendix C.

4.1.2 Translation Norm Parameter

The distribution of the three Translation Norm Parameter categories among the 493 post-edits, is depicted by Figure 4.

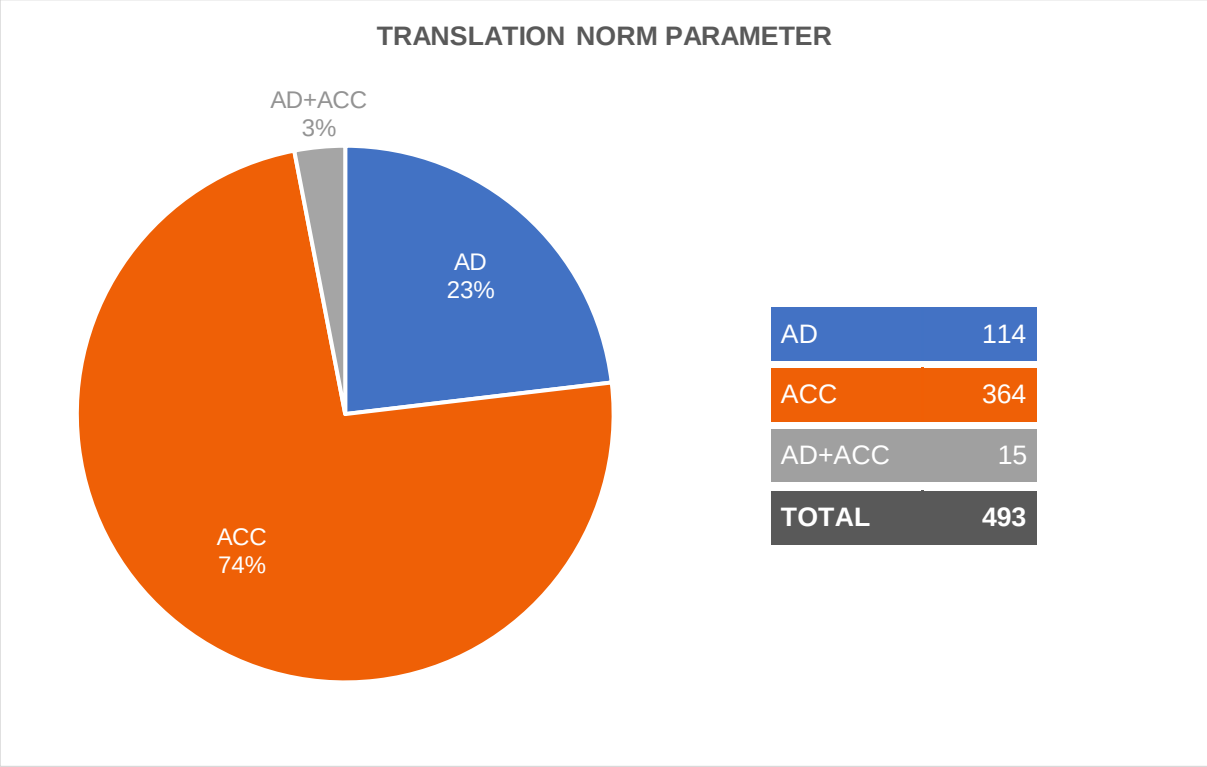
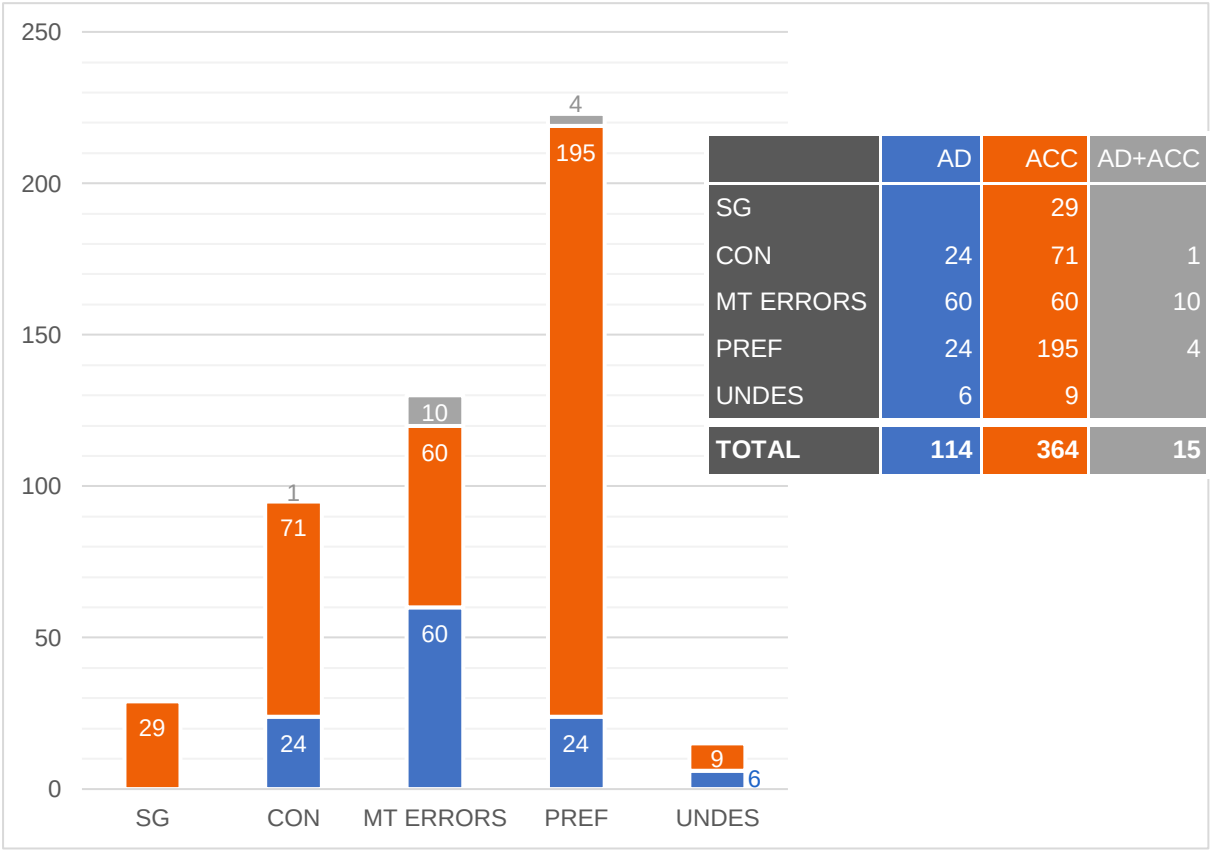


Figure 4. Distribution of the Translation Norm Parameter categories

As can be clearly observed in Figure 4, 75% (364) of all post-edits were related to acceptability (the relation between target text and target culture/function – ACC), and 23% (114) were changes in terms of adequacy (the relation between source and target text – AD). These will hereafter also be referred to as simply acceptability and adequacy post-edits/changes respectively. Nearly three times as many acceptability post-edits were thus carried out compared to adequacy post-edits, which suggests that the MT output was less frequently

incorrect as compared to the ST, than it was deemed inadequate for the target text's function, in terms of language use. This additionally indicates that target text norms are fairly strong among translators. Only few post-edits (15 – 3%) were considered simultaneously bringing about an adequacy and an acceptability change (AD+ACC) in the target text.

Figure 5 illustrates the Translation Quality Parameter and the Translation Norm Parameter in relation to each other.



	AD	ACC	AD+ACC	TOTAL
SG		29		29
CON	24	71	1	96
MT ERRORS	60	60	10	130
PREF	24	195	4	223
UNDES	6	9		15
TOTAL	114	364	15	493

Figure 5. Translation Quality Parameter and Translation Norm Parameter in relation to each other

A complex situation can be deducted from Figure 5: the two prevailing categories in the parameters discussed above (ALL ESS – ACC) do not seem to overlap. Though it is true that of all essential changes (SG + CON + MT ERRORS), most were changes in terms of acceptability (160 of 255 or 63%), over half of all *acceptability* changes were preferential (195 of 364 or 54%), and in turn, most preferential changes were acceptability changes as well (195 of 223 or 87%). The preponderance of the acceptability category amongst the essential changes is at the very least interesting, but can partly be explained by the 29 style guide changes all being acceptability changes, as the DGT Dutch Style Guide specifies Dutch written language

usage, and the majority of the essential acceptability changes (71) being carried out for consistency considerations – consistency to be achieved amongst *target texts*. What stands out, however, is the exactly equal number of corrected MT errors being adequacy as acceptability post-edits. This may be due to the MT output system having often translated English text into Dutch too literally, preserving English sentence constructions that are not accepted in Dutch. A closer look at the Text-Linguistic Characteristics of these acceptability MT errors revealed that most often, syntax or morphology needed to be adjusted: agreement problems were solved (article-noun disagreement of named entities, incorrect referring expressions), sentence order was altered or other structural changes were implemented, and named entities were adjusted. Stylistic changes were less often essential acceptability post-edits; only a number of lexical items were incorrectly translated by the MT system.

Most adequacy post-edits, on the other hand, were essential (84 of 114 or 74%): 24 consistency changes brought about a meaning shift, and 60 MT errors were infidelities to the source text. The 24 preferential adequacy post-edits mostly comprised hypernyms, hyponyms, explicitations, or deletions of lexical information – all when implemented without (heavily) affecting the meaning conveyed by the source text sentence. Interestingly, most AD+ACC changes were MT errors as well (10 of 15 or 67%). Most of the undesirable changes were acceptability changes, revealing that slightly more language errors were made by the post-editors, than interpretation mistakes. The difference (with the adequacy category) is, however, very small.

4.1.3 Text-Linguistic Characteristics

The results for the Text-Linguistic Characteristics lay bare which types of changes were most frequently made as regarded from a linguistic perspective. First an overview of the four main types (parameters) of text-linguistic post-edits is provided, after which the most frequent categories under them are discussed. It should be reiterated that it was possible for a single post-edit to fall under a category from more than one of these four parameters, but not for it to be assigned more than one category within each parameter (cf. Figure 2; only one annotation per parameter column for each post-edit). This is the reason why the total number of text-linguistic *annotations* does not amount to 493 (the total number of post-edits), but to 524, spread across the 493 post-edits.

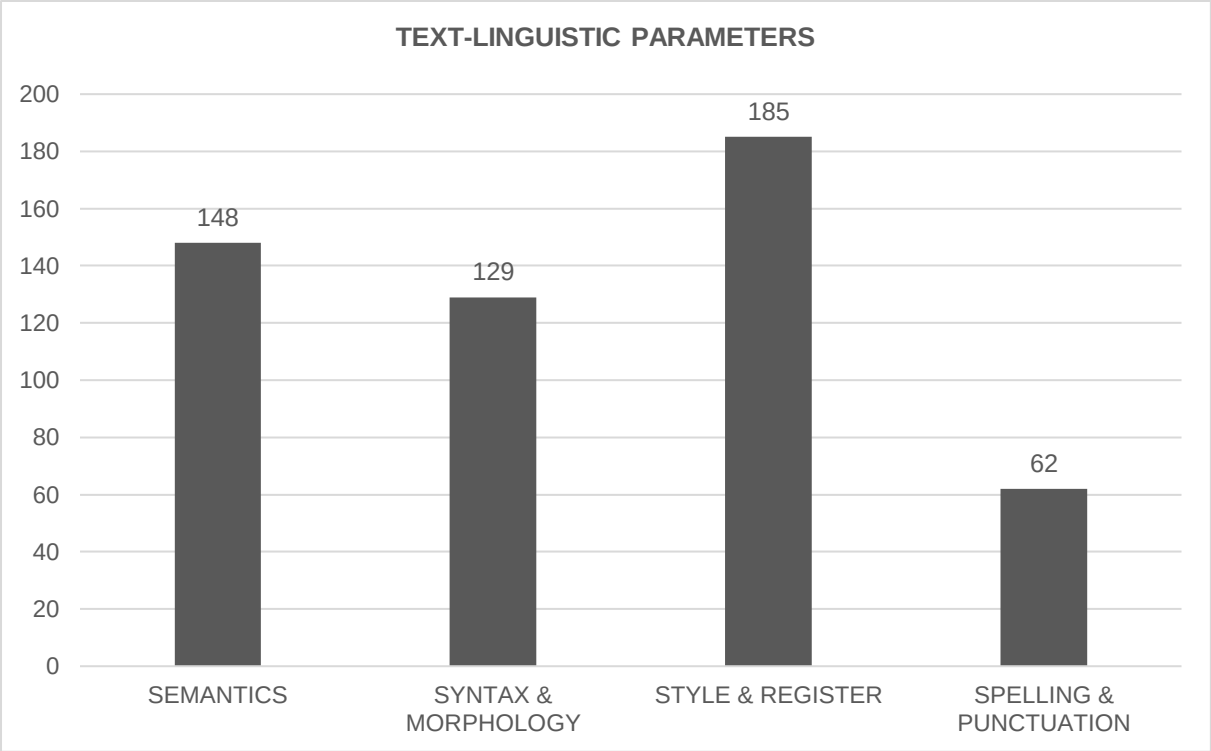


Figure 6. Distribution of the Text-Linguistic parameters

Figure 6 shows the total number of annotations for each of the four main types. Style or register changes were most implemented (185 post-edits), followed by semantic changes (148) and alterations related to syntax or morphology (129). 62 spelling and punctuation changes were made, making this the main post-edit type least frequently occurring. In Figure 7, the relation between the four parameters and the Translation Quality Parameter is illustrated, i.e. the number of times a type of change was a style guide change, consistency change, MT error, preferential change, and undesirable change. The lengths of the bars represent the relative values (in comparison to the total number of times the parameter was annotated); the numbers in the bars represent the absolute values (the numbers of annotations).

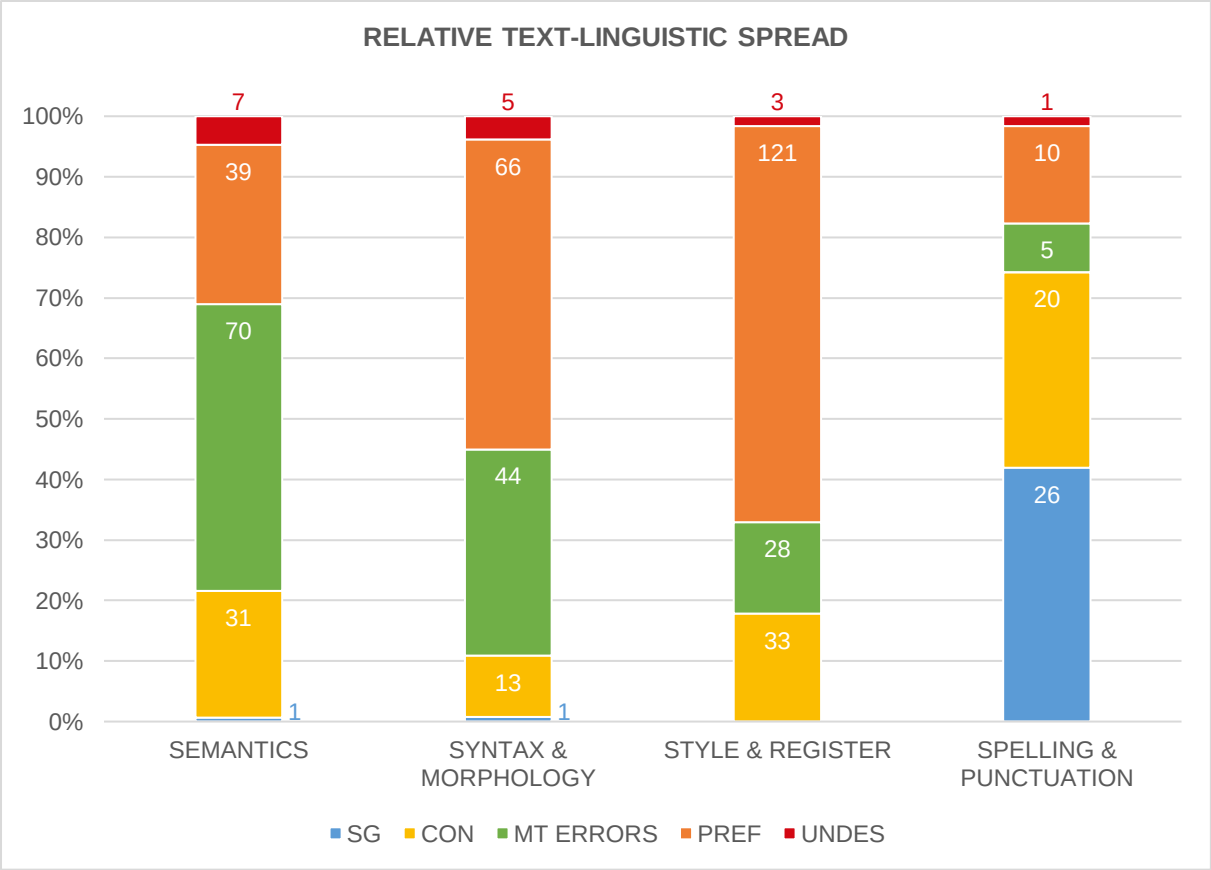


Figure 7. Distribution of the Text-Linguistic parameters in relation to the Translation Quality Parameter

A number of findings can be deducted from Figure 7. Of all semantic post-edits, first of all, just under half were MT errors, pushing the total necessary share (ALL ESS) of semantics-related post-edits to nearly 70%. This type of change was also most often undesirable, which is at the same time both surprising, as translators are expected not to (severely) alter the meaning conveyed by a source text they are given, and unsurprising, as meaning deviations from the source text will more quickly be deemed undesirable than some other types of changes. A quarter (39) were slight meaning changes out of preference (hypernyms, explicitations, etc.). What immediately stands out as well, is that 65% of all style & register changes were preferential. When considering that 176 of the 185 style and register changes were acceptability post-edits and 61% of preferential acceptability post-edits were related to style and register, this could already be expected from the Translation Norm Parameter findings; most preferential changes were acceptability-related. A total of 61 (33%) of the stylistic post-edits, however, were essential, with 28 of them even being MT errors. Most were related to lexical choices and (reductions of) repetitions, and 3 untranslated English words or phrases were found in the MT output as well. Below are two examples of these stylistic MT errors: a lexical change, and untranslated text.

	T3 **Lexical change**
(ST)	individuals **playing** sports on them
(MT)	personen die sporten op hen **uitoefenen**
(PE)	mensen die daarop een sport **beoefenen**

	T7 **Untranslated text in MT**
(ST)	through capacities of robustness, adaptability and **transformability**
(MT)	door middel van robuustheid, aanpassingsvermogen en **transformability**
(PE)	door middel van robuustheid, aanpassingsvermogen en **veranderbaarheid**

Other than those types of errors, prepositions were sometimes incorrect, and synonyms needed to be used (for consistency). Interestingly as well, alteration of syntax or morphology was more often preferential or even undesirable than it was necessary for the sentence to be grammatically correct.

At first sight, it may seem surprising that not one style and register change was classified as a style guide post-edit. The reason, however, is that a consistency change was a more *specific type of style guide change* (as specified in Appendix A), and that stylistic changes (such as synonyms) were very often implemented to maintain consistent terminology, for example, throughout the text and other Commission texts. The 28 stylistic MT error corrections were all more text-dependent stylistic problems rather than falling under the specific rules listed in the Dutch Style Guide. As, however, the guide does elaborately cover punctuation, capitalisation, numerals, and spelling (including compounds), it comes as no surprise that three quarters of all spelling and punctuation post-edits, were style guide and consistency changes. Spelling and punctuation also included the least undesirable changes (1 post-edit; 2% of all spelling post-edits), and only 16% of spelling or punctuation changes were preferential. This can be explained by the many rules established in the Dutch language for spelling and punctuation in general.

The distribution of the Text-Linguistic categories within each of the four main types is shown in Table 7; the categories are sorted from most to least frequently occurring.

SEMANTICS	
TOTAL	**148**
HYPONYMY	22
ADD ST INFO	22
AGENT	21
EXPLIC	17
HYPERONYMY	14
ACTION	13
DEL LEX INFO	9
ACTION + AGENT	6
MODALITY	5
TIME	5
OTH MEAN CH	5
PLACE + AGENT	5
OTH ADD	3
ACTION + TIME	1

SYNTAX & MORPHOLOGY	
TOTAL	**129**
SENT ORDER	40
OTH STRUCT CH	29
SING/PLUR	13
ACT-PAS	10
REFERENCE	9
ART-NOUN	7
OTH MORPH CH	5
NAMED ENT	5
NOM	4
RED OF NOM	4
NOUN-ADJ	1
SUBJ-VERB	1
PAS-ACT	1

STYLE & REGISTER	
TOTAL	**185**
SYNONYMY	98
LEX CH	39
PREPOSITION	10
REPETITION	9
CONJUNCTION	7
COHERENCE	6
RED OF REP	6
REGISTER	5
UNTRANS IN MT	4
SPLIT SENT	1

SPELLING & PUNCTUATION	
TOTAL	**62**
PUNCT	30
CAP	23
COMP	4
OTH SPEL CH	3
NUMERAL	2

Table 7. Distribution of the Text-Linguistic categories per parameter

Figure 8 then provides a full overview of all Text-Linguistic categories, again sorted from most to least frequently occurring. A grey colour coding was added to maintain a view of which of the four parameters a category belongs to, without creating confusion with the (colour coding of the) parameter categories discussed in Sections 4.1.1 and 4.1.2.

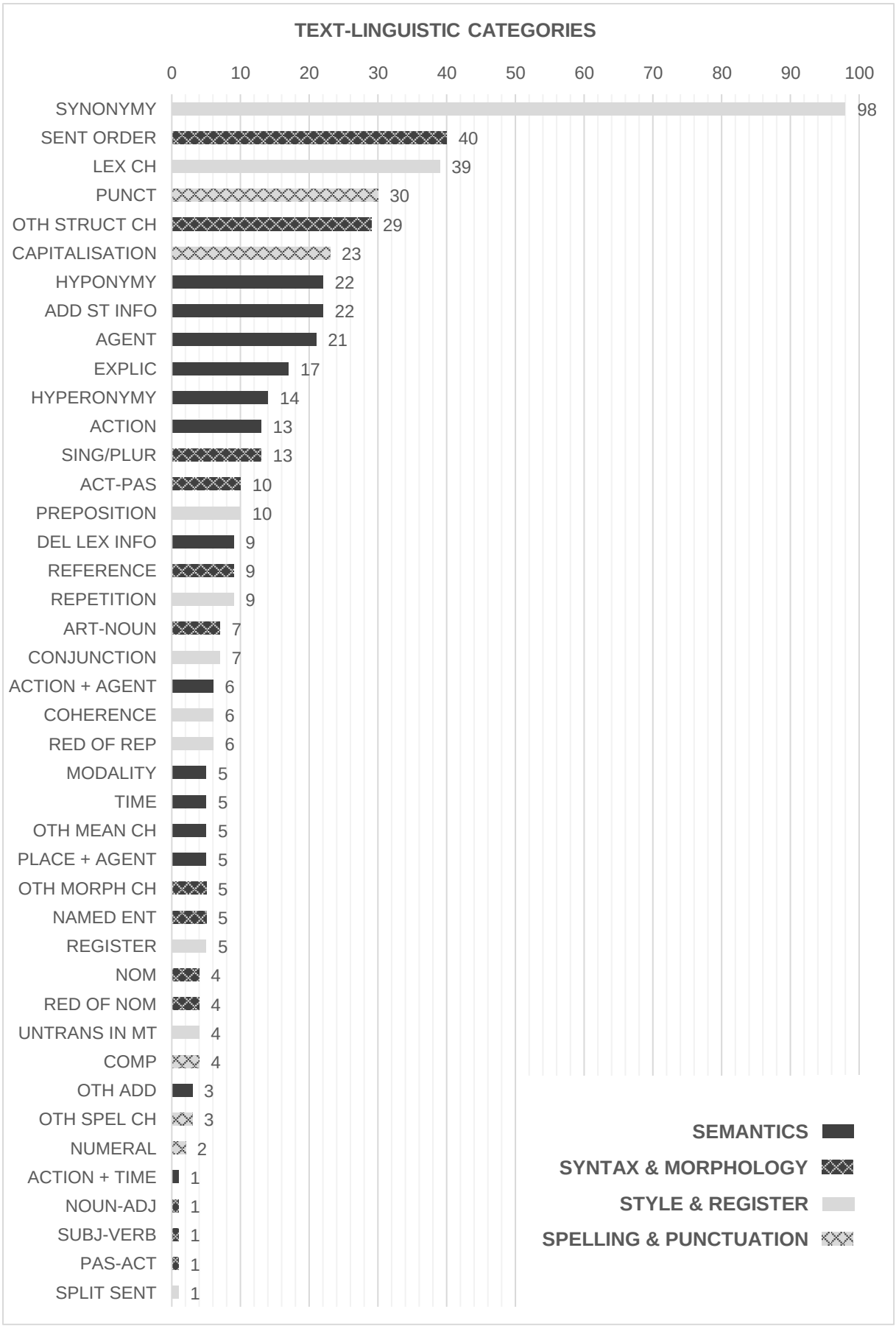


Figure 8. Text-Linguistic category frequency

The type of post-edit that was most implemented, is by far synonymy: 98 times a synonym was substituted for a word, which is over double the next most frequent post-edit type, internal sentence order. The category lexical change was, as mentioned in Chapter 3, assigned to substitutions that were not synonyms of the extracted MT output words/phrases, but as these substitutions did not lead to a meaning change in the sentence, the category could be more or less grouped together with synonymy. A full constituent was moved within a sentence 40 times, whereas a transposition within a single phrase (e.g. geworden is → is geworden) did not occur in the data at all. Though noteworthy, this may be due to the limitedness of the sample analysed, however. Other than word order, structural changes were often implemented, and article-noun agreement was adjusted more frequently than both noun-adjective and subject-verb agreement. Active constructions were substituted by passive ones more often than the other way around, which is remarkable, as the Dutch Style Guide specifies that active sentences are preferred over passive ones. This finding can, however, be put down to the extremely formal language used in the texts. Although spelling and punctuation was the main parameter least occurring throughout the sample (cf. Figure 6), relatively speaking, punctuation and capitalisation could be believed problematic for the NMT system: 30 punctuation marks were altered, and capitalisation was changed 23 times. Very few compound nouns, on the other hand, were misspelled, and only two numerals – the numbers one and two – were written in numbers by the MT system and substituted for words by the post-editors, thereby adhering to the Dutch Style Guide.

Only after 7 categories related to formative aspects, a series of semantics categories is listed. MT output words were substituted for hyponyms of them 22 times, which was mostly either because the MT word was not sufficiently specific, or because the hyponym was the consistently used term or expression throughout the translation(s) and possibly IATE. Hyponyms were used preferentially as well – when the MT output would have sufficed without post-editor intervention. Substitutions for hypernyms, on the other hand, were carried out as well, though less frequently than for hyponyms. What stands out, is that ST information apparently very often needed to be added (ADD ST INFO) or was made explicit (EXPLIC) by the post-editors. To a lesser extent they felt the need to or were forced to delete information from the MT output. In terms of meaning shifts other than additions or deletions of information, subjects and objects carrying out or undergoing actions in sentences (AGENT) and the actions happening themselves (ACTION), were most frequently modified – including a number of combinations of the two as well. Clause modality and the time and place of an action were not altered many times, and also only very rarely, something was added by a post-editor without it

having been stated in the ST (OTH ADD). Interestingly, only one sentence was split into two sentences (SPLIT SENT) by a post-editor, while the English language is known to more often tolerate long sentences than the Dutch. No sentences were merged.

Figure 9 provides a more detailed overview of the 15 Text-Linguistic Characteristic categories occurring 10 or more times throughout the sample: in relation to the Translation Quality Parameter.

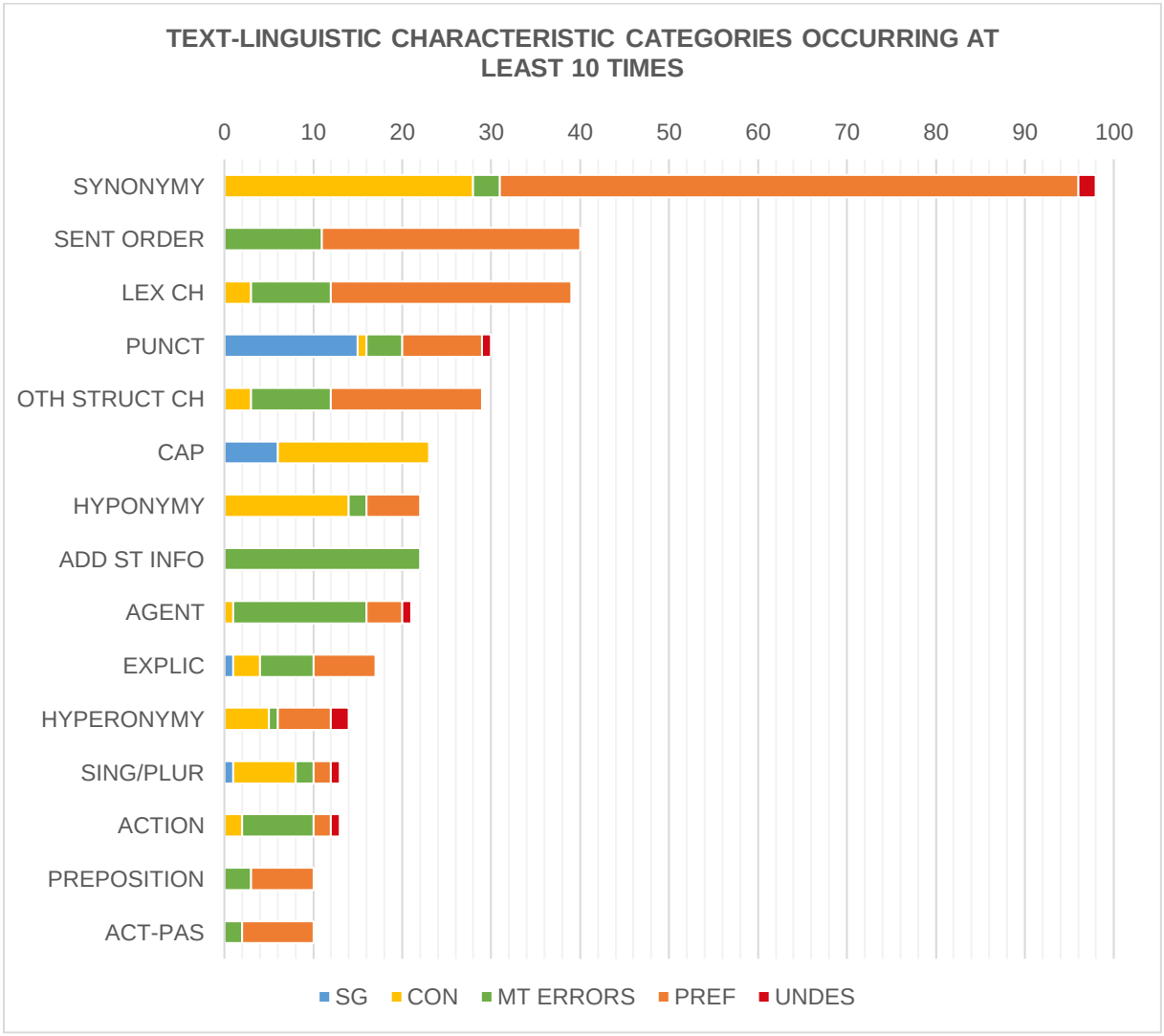


Figure 9. Text-Linguistic categories occurring at least 10 times, in relation to the Translation Quality Parameter

The top three types of post-edits: synonymy, internal sentence order, and lexical change, were all implemented out of preference more often than out of necessity. Punctuation and capitalisation, on the other hand, were more often (or only) essential, which can be ascribed to the elaborate Dutch Style Guide to be adhered to, or to consistency needing to be achieved. Below is an example of such a style guide punctuation change (by Translator 8).

T8	Punctuation
(ST)	Articles 29 **(3)**, **(4)** and **(5)** of Directive (EU) 2018/2001 of the European Parliament and of the Council
(MT)	de artikelen 29 **(3)**, **(4)** en **(5)** van Richtlijn (EU) 2018/2001 van het Europees Parlement en de Raad
(PE)	artikel, leden **3**, **4** en **5**, van Richtlijn (EU) 2018/2001 van het Europees Parlement en de Raad

Post-edits that were additions of ST information, were always MT errors and therefore essential changes as well, but explicitation of ST information was nearly as often the translator's preference as it was essential (though still mostly essential). Alterations of both agents and actions appeared to be mostly necessary corrections of MT errors, whereas over half of all hyponyms were used to achieve consistent terminology, as was the case for the following example (by Translator 3):

T3	Hyponymy
(ST)	granules
(MT)	korrels
(PE)	granulaat

Interestingly, changes from the singular form of a noun to the plural form or vice versa, were the only type of morphological change that was more often essential than preferential, which could often be ascribed to terminology needing to remain consistent. No syntactical change was more frequently essentially implemented than preferentially.

The changes that were mostly preferential, were predominantly related to either style and register, or syntax. The stylistic post-edits included synonymy and other lexical changes not causing meaning shifts, as mentioned above, but preposition changes as well. In terms of syntax post-edits aside from internal sentence order changes, other structural changes such as switches from relative clauses to adjectives were, likewise, more frequently preferential than essential. Active constructions, too, were mostly preferentially converted to passive constructions. The one mostly preferentially implemented post-edit type (from the 15 included in the graph) related to change of meaning, was hyperonymy. As hypernyms sometimes lead to loss of information present in the source text as well, this type was also one of those most often deemed undesirable. The other frequently undesirable change, though to a much smaller proportion than hyperonymy, was synonymy, usually when causing inconsistent terminology to be introduced.

4.2 Results per Translator

The results per translator are presented in function of (1) the Translation Quality Parameter, (2) the Translation Norm Parameter, and (3) the Text-Linguistic Characteristics. Before elaborating on them, however, emphasis should again be laid on the limitedness of the segments analysed for each translator. Figure 10 shows the distribution of the Translation Quality Parameter for each of the nine translators.

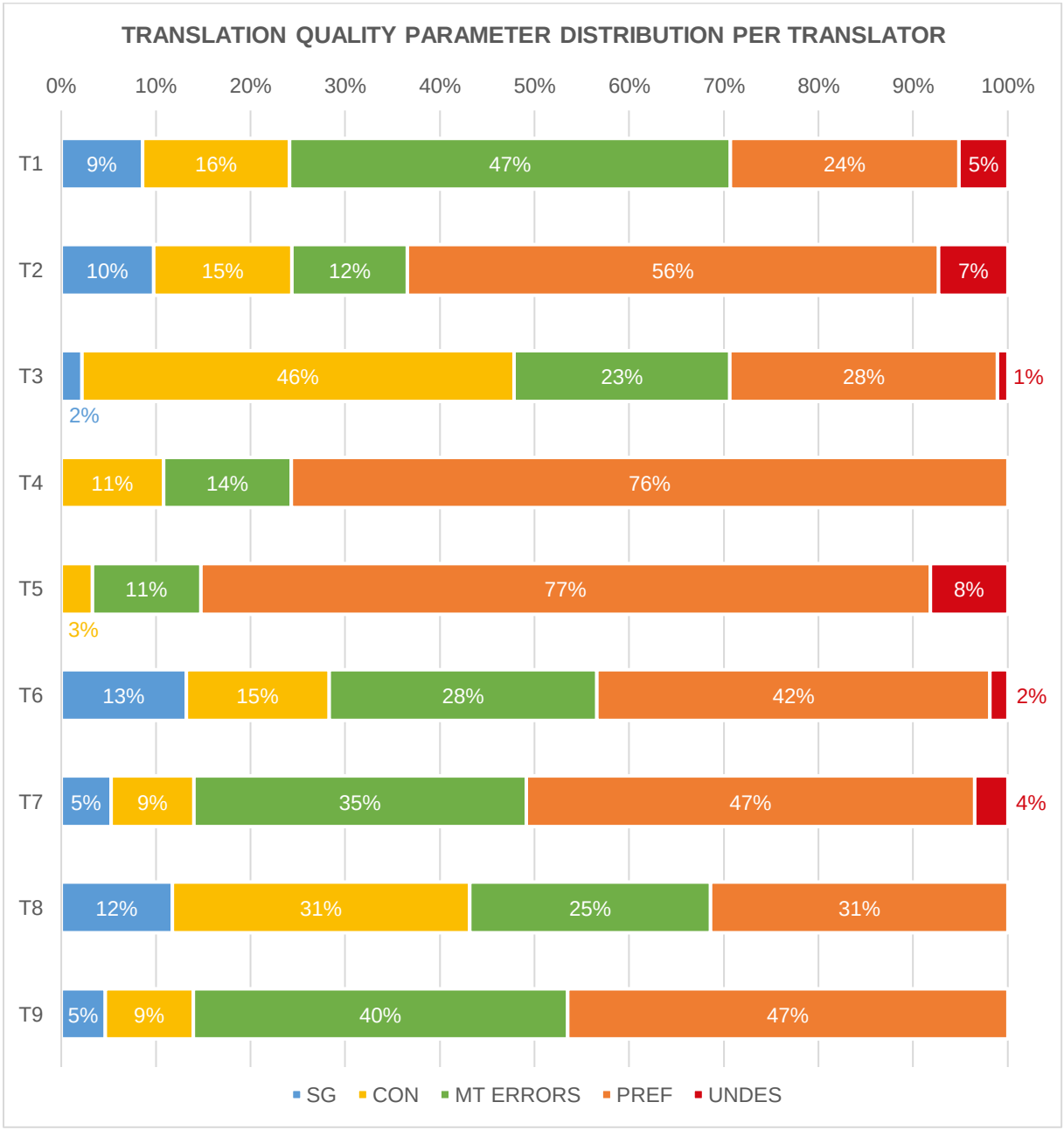


Figure 10. Distribution of the Translation Quality Parameter per translator

In terms of the correctness and necessity of their post-edits, the differences between the translators are striking. The shares of preferential changes were largest for Translator 4 and Translator 5 (76% and 77% respectively), with the latter appearing to have also carried out the least essential changes (14%), and even the most undesirable changes (8%). Translator 1 and Translator 3, on the other hand, implemented far more essential changes (ALL ESS): just over 70% of their post-edits could be considered essential. Between them as well, however, differences in terms of the types of essential post-edits can be observed: 46% of the changes by Translator 3 were consistency-related, while Translator 1 corrected about that many MT errors, and made fewer consistency changes (16%). Five of the nine translators carried out more essential changes (ALL ESS) than preferential ones: Translators 1, 3, 6, 8 and 9. Interestingly, Translator 5 and Translator 2 carried out relatively many undesirable post-edits while also having large shares of preferential changes, which may indicate that their attitude of altering much out of preference may have led to them making more undesirable changes as well. This association does, however, not hold true for Translator 1 nor Translator 4. Besides Translator 4, Translator 8 and Translator 9 appeared to not have implemented any undesirable change at all either.

In terms of style guide changes, Translator 4 and Translator 5 did not carry out any – which may also partially account for their lower shares of essential changes – whereas Translator 6 and Translator 8 did need to alter relatively many items conflicting with the Dutch Style Guide (very often punctuation and capitalisation): 13% and 12% respectively. Translator 8 appeared to additionally be confronted with several consistency problems (31%).

Figure 11 illustrates the distribution of the Translation Norm Parameter per translator.

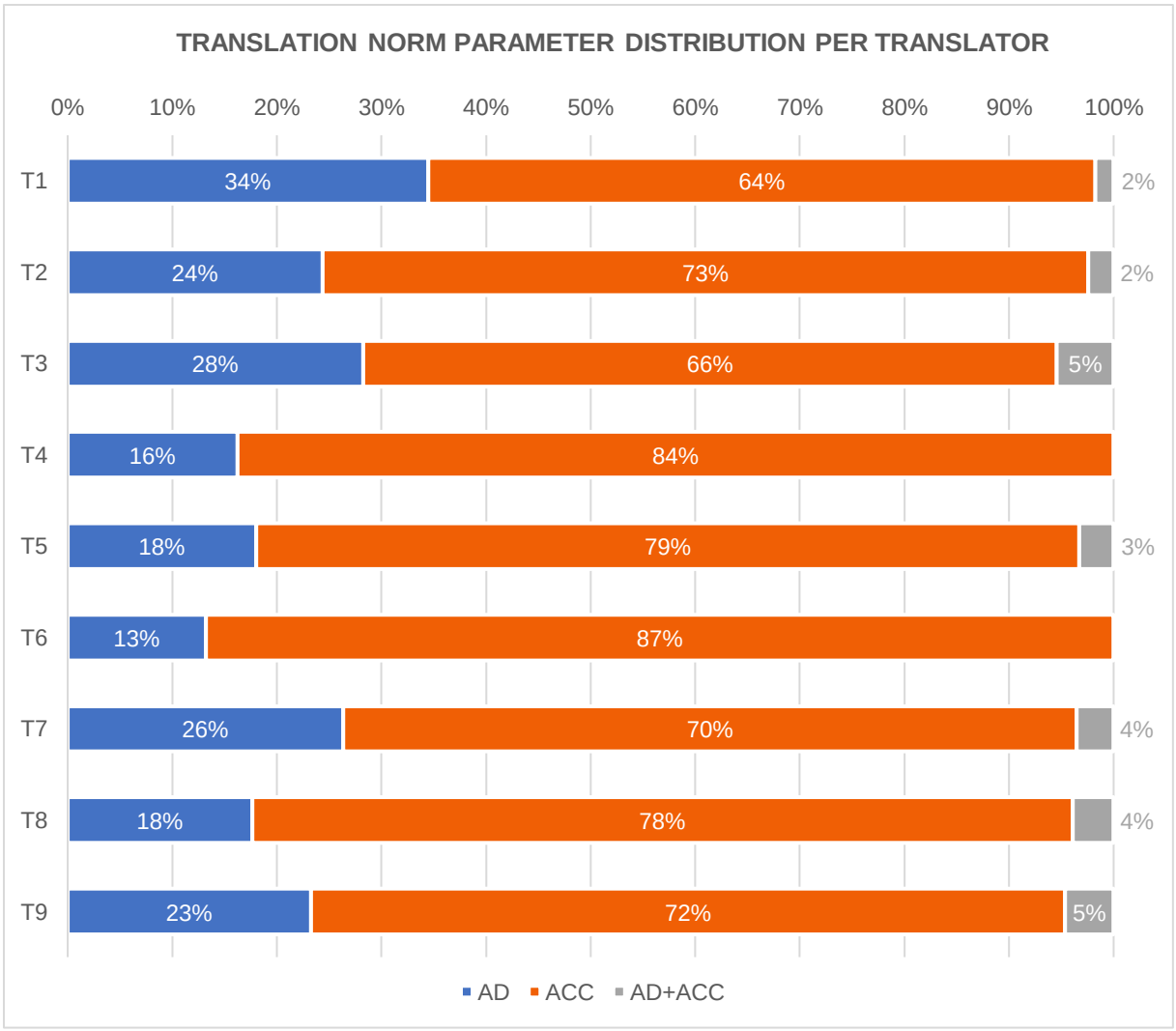


Figure 11. Distribution of the Translation Norm Parameter per translator

Much less variation between the translators can be immediately observed when it comes to the translation norm level. Well over half of all post-edits from all translators were changes in terms of acceptability, with all of them ranging in between 64 to even 87%. A number of interesting links can, however, be established with the translation quality distribution (Figure 10). Translator 1 implemented the most adequacy-related changes, while also being the translator with the largest share of corrected MT errors out of all translators. This suggests that many of these corrected MT errors were adequacy-related problems, as no style guide changes were logically related to adequacy and overall only few consistency changes were (cf. Figure 5). Interestingly, Translator 6 made the fewest adequacy changes and implemented the most acceptability-related post-edits, though not being the translator with the least ALL ESS changes. Quite the contrary: over 55% were essential, which again suggests that making acceptability-related adjustments to the MT output is often not optional.

Translator 4 and Translator 5, who had the largest shares of preferential changes, appeared to make the second most acceptability changes (84% and 79% respectively) and the second least post-edits in terms of adequacy (16% and 18% respectively). Translator 3 implemented the most consistency-related post-edits (46%), while having the second largest share of adequacy post-edits (28%), which suggests that these consistency changes (such as terms) sometimes also implied slight (yet adequate) meaning shifts.

Lastly, the distribution of the Text-Linguistic main categories for all translators can be seen in Figure 12.

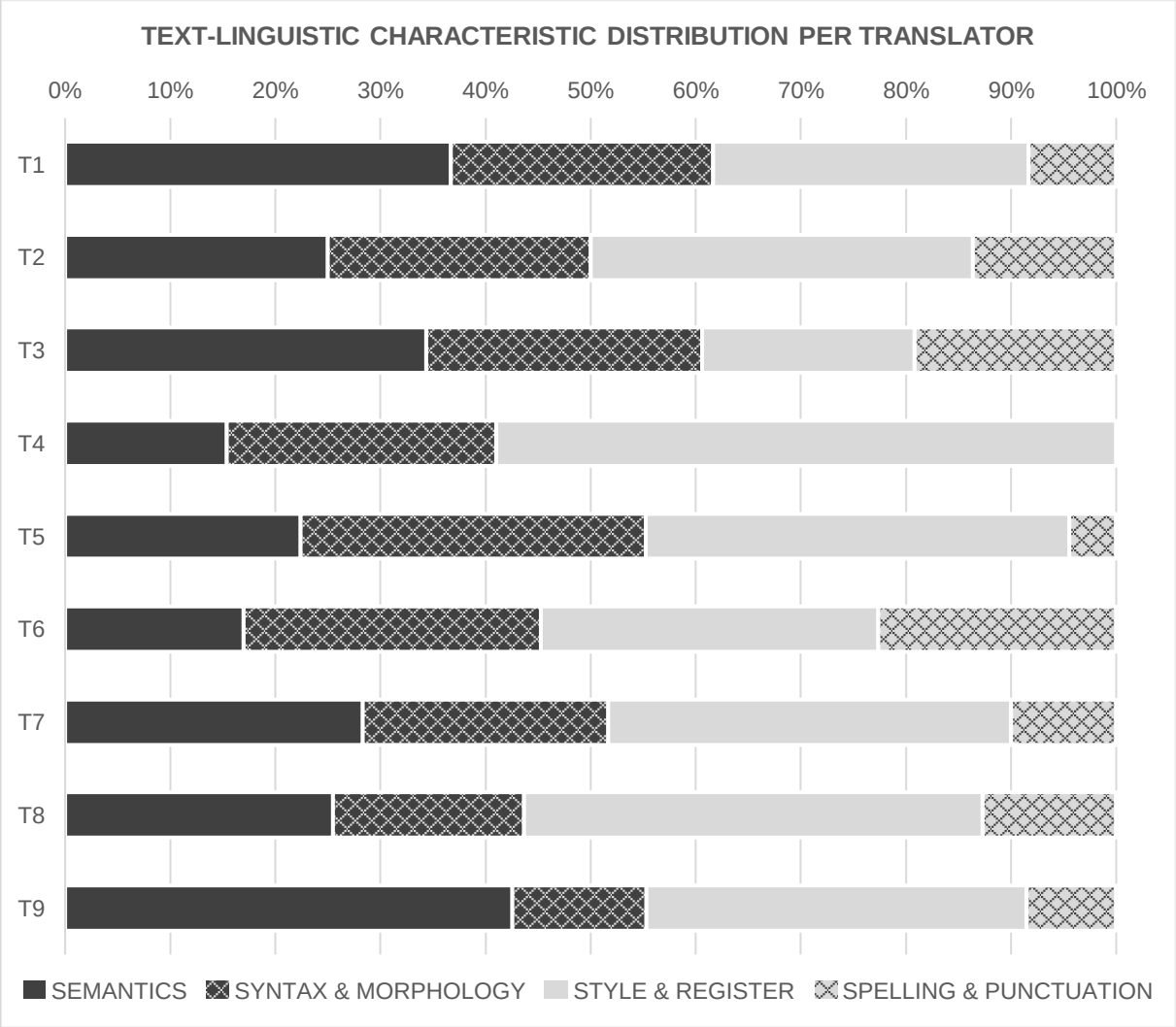


Figure 12. Distribution of the Text-Linguistic parameters per translator

From a text-linguistic point of view, greater variation can again be observed between the translators.

Translator 9 changed most in terms of semantics, while being situated exactly in the middle of the shares of adequacy-related post-edits (cf. Figure 11: 23%), and implemented least syntax and morphology changes out of all translators. The highest relative share of syntax and morphology changes is Translator 5's, who appeared to have also made many preferential changes. When considering this translator's 22 structural changes, only 3 of them were, in fact, essential changes; all other ones were preferential (and one undesirable). Translator 4 most frequently carried out stylistic changes, and as style and register post-edits are related to target text acceptability, this seems to overlap with the fact that the second highest share of acceptability changes belonged to this translator as well.

No spelling or punctuation change was, on the other hand, carried out by Translator 4, which appears to be in line with the absence of style guide changes (cf. Figure 10), and is likely to also partially account for the elevated share of preferential changes. This may hold true for Translator 5 as well: no style guide changes were found either, and correspondingly only a very limited share of post-edits appeared to be related to spelling or punctuation. The few spelling and punctuation changes made by this translator, included a preferential substitution of commas for dashes, for example. Translator 6 carried out the most spelling and punctuation changes out of all translators, which is in line with the highest share of style guide changes also pertaining to this translator (13%). Interestingly, however, relatively many post-edits by Translator 3 appeared to be related to spelling and punctuation as well, whereas only 2% of post-edits by this translator were style guide changes. This alleged discrepancy can be explained through the many capitalisation issues in this translator's text being consistency-related.

5 DISCUSSION

This section reviews the results discussed in Chapter 4 in the light of the four research questions listed in Section 2.6 and related research elaborated on in Chapter 2.

RQ1 To what extent do English-Dutch translators at the Dutch Language Department of the European Commission Directorate-General for Translation (DGT) make use of MT?

The analysis of all 69 English-Dutch SDLXLIFF files provided by the 10 DGT translators (cf. Table 1; Table 2) showed that machine translated segments, both edited and unedited, made up 20% of the total number of segments, and machine translated words 29% of the total word count, spread over the 61 texts in which MT was used. In only 8 texts no MT was used at all – all by Translator 5, who did receive a total of 72% (perfect) TM matches. In terms of MT adoption, considerable variation could be observed between the translators, as was the case in Man (2019) and Koponen, Sulubacak, Vitikainen, and Tiedemann (2020) as well. Translator 9 and Translator 2 made the most use of MT, for 66% and 53% of their segments respectively, whereas MT was least implemented by Translator 1 (3%) and Translator 8 (6%). These differences could be explained by the extent to which each translator was given TM matches for their segments, which is in line with Rossi and Chevrot (2019) and Macken, Prou and Tezcan (2020) noting that MT adoption may be limited as a result of the embedding of MT in the highly advanced DGT CAT workflow. Translator 1, for example, could rely on 86% TM matches (86% of which in turn were unchanged TM matches), while Translator 9 was only given TM matches 21% of the time. Both of them, however, practically did not translate from scratch at all (1% and 0,3% respectively), which could hint at MT suggestions being preferred given the psychological benefit of not having to start translating from a blank segment, as found by e.g. Lesznyák (2019) and Macken et al. (2020). Translator 7, in contrast, who also had one of the lowest percentages of MT adoption (12%), was the only one who translated over 10% of segments from scratch (21%). Translator 6 used MT for nearly half of the segments (39%), but the remainder of the translators were more within the same range: MT usage varied between 12% and 26% of their segments. Interestingly, when considering word count, the percentages for all translators were anywhere between 1 to even 26% higher than on segment level, and translator variation was larger as well, which indicates that (perfect) TM matches and unchanged MT segments are often short segments.

The latter of the two appears to additionally be confirmed by the finding that only 9% of all MT segments – equating to 5% of all MT words – were left unchanged, meaning that nearly all MT suggestions were post-edited (91% of MT segments and 95% of MT word count). Translator variation was in this respect also much more limited. These figures are likely to be due to the extremely high translation quality standards at the DGT as well, as emphasised by amongst others Vardaro, Schaeffer and Hansen-Schirra (2019), Arnejšek and Unk (2020), and Macken et al. (2020).

RQ2 How many post-edits carried out by English-Dutch DGT translators can be considered essential, preferential, and undesirable changes?

Of all 493 post-edits analysed, over half (52%) were essential, of which one half were MT errors, and the other half consistency and style guide issues. These figures are in line with Man (2019), who also found over half of all post-edits in English-Dutch publicity texts at the KBC Language Department being necessary when in addition to MT errors, consistency, merging and other edits were considered essential. They are moreover very much in line with the results of Vandevoorde, Weintraub and Arabadjieva (2021) as well, who found 43 to 68% of the HTAs (human translation actions) in English-Dutch NMTPE in their study being essential and 24% of them lexico-semantic corrections (comparable to MT errors). In this respect, Arnejšek and Unk's view (2020) that MTPE is essential to achieve high-quality (DGT) translations is also reflected by the present results. In addition, they indicated that nearly half of all post-edits were of preferential nature, which is in line with findings by de Almeida and O'Brien (2010), Koponen and Salmi (2017), and Koponen, Salmi and Nikulin (2019), having observed many preferential post-edits, and which reinforces the idea noted by Vandevoorde et al. (2021) that preferential ("optional") changes are not superfluous in these types of very high-quality translations, as all of them combined result in the elevated level of quality that is ultimately to be achieved.

As was the case for amongst others Man (2020), much variation could be observed in terms of the ratios of preferential to essential post-edits between the different translators, which suggests that some translators may be more inclined than others to make alterations merely out of preference. The very few undesirable changes overall found, furthermore, once again demonstrate both the DGT translation quality standards, and the professionalism of DGT

translators. The fact that some were found, however, is a reminder that on the one hand even highly professional DGT translators are only human, and that on the other revision remains a paramount stage of the translation process (particularly for high-quality translations).

RQ3 Which types of changes are most often carried out by English-Dutch DGT translators when post-editing English to Dutch MT?

The three post-edit types most frequently carried out included two stylistic types of change – synonymy and other lexical changes (these in themselves being closely related) – and one syntactic type – internal sentence order. That many stylistic post-edits were carried out, is in line with Koponen et al. (2019), who found many word substitutions in NMTPE, and Vandevoorde et al. (2021), having found a relatively large share of synonyms. It may moreover corroborate amongst others Flanagan and Christensen (2014) and Koponen (2013), who not only claim that target text norms are fairly strong among (professional) translators, but that translators have strong stylistic preferences as well. Though these frequent stylistic changes were mostly implemented preferentially, synonyms often also needed to be used to achieve consistency and were therefore essential, which can be linked to Vandevoorde et al. (2021), who found a number of changes related to what they referred to as jargon. In terms of lexical choices, some errors were made by the MT system as well. The two most recurrent MT errors, however, were changes in agents (subjects and objects carrying out or undergoing the action in a sentence), and deletions of ST information, which was often re-introduced by the post-editors. These findings appear to be consistent with Toral and Sánchez-Cartagena (2017), Castilho et al. (2017), Vardaro et al. (2019), and Arnejšek and Unk (2020), having found terminology/lexical errors, omissions, mistranslations, and stylistic/register errors to be the most prominent error categories in NMT output. They moreover could be assumed to support the ideas of Koehn and Knowles (2017) – that NMT systems in themselves are (insufficiently) interpretable – and of Yamada (2019) – that NMT has human-like translation abilities, the errors therefore likely being more challenging to detect and correct.

Besides by Bentivogli Bisazza, Cettolo and Federico (2016), who had already mentioned a problematic area for NMT systems being reordering of particular linguistic constituents that require a deep semantic understanding of text, the present finding that some MT errors were indeed made in terms of the placement of constituents, has not often been mentioned in the

literature. This, combined with the finding that internal sentence order was moreover also very frequently preferentially implemented, could make of this post-edit type a NMT challenge requiring further investigation and improvement. What stood out as well, is that although spelling changes overall represented the post-edit type implemented least frequently, punctuation and capitalisation were fairly regularly altered by the post-editors, as the MT output did often not comply with the specific rules in the Dutch Style Guide, or inconsistent capitalisation had been generated throughout the text. Punctuation appearing particularly problematic for eTranslation NMT, had already been remarked by Arnejšek and Unk (2020) as well. Equally noteworthy is the observation that active constructions were substituted by passive ones 10 times throughout all segments analysed, and passives were substituted for actives only once, whereas the style guide stipulates that active sentences are preferred over passives. Similarly, nominalisation was introduced by the post-editors as often as it was reduced (4 times), though according to the guide, verbs should be preferred over nouns, and nominalisation avoided. These two findings may, however, be ascribed to the very formal nature of the contracts, reports, regulations, directives, policies, communications of the Commission, etc. – often also comprising very long sentences – and the style guide being applicable to less formal text/translation types as well.

RQ4 How many post-edits carried out by English-Dutch DGT translators are changes related to the meaning of the source text (Toury's norm of adequacy), and how many are in function of the acceptability of the target text in the target culture and its suitability for its function (Toury's norm of acceptability)?

Overall, many more post-edits were related to acceptability than to adequacy, and this was the case for all translators. This finding again suggests that target text norms are strong among translators, though the extremely high quality standards for DGT translations may account for it as well. Remarkably, the essential changes, too, were more frequently related to acceptability than to adequacy. The main reasons for this were that all style guide post-edits were acceptability changes, and that many inconsistencies had been generated by the MT system – inconsistent terminology, for example, throughout the *target* text and already-existing Commission translations. Perhaps even more noteworthy, is that exactly half of all MT errors were acceptability issues as well. The essential changes needed for target text acceptability thus varied along the type of essential change: style guide changes mostly covered punctuation

issues; for consistency, synonymy and capitalisation were most frequently adjusted; and in terms of MT errors, the acceptability post-edits were various syntactical and stylistic types of changes. Referring expressions and named entities were corrected, and lexical choices, sentence order and other structural changes appeared to sometimes be problematic for the MT system as well. In addition, incorrect articles were found, though those were mostly linked to named entities. These problems with a number of lexical choices, and named entities and their referring pronouns and abbreviations, can again be linked to Arnejšek and Unk (2020), who found eTranslation NMT having problems with proper nouns and abbreviations, as well as complex structures and text fragments.

Though adequacy post-edits were most often corrected MT errors such as additions of ST information, a small share of adequacy-related changes appeared to be preferential. These included hypernyms and hyponyms used although the ST information had not heavily been affected by the MT system, and explicitations and deletions of lexical MT information that were not entirely necessary for the ST information to be clear, while not entirely unwelcome either. Lastly, also only few post-edits simultaneously brought about an adequacy and an acceptability change in the target text, but the mere occurrence of this category, combined with the finding that most of these changes were corrected MT errors, does hint at NMT systems producing human-like output, as was noted by Yamada (2019). This means that NMT errors may sometimes be complex and challenging for post-editors (Yamada, 2019), and may therefore often require the simultaneous application of different types of post-editor solutions.

6 CONCLUSIONS

This pilot study aimed at investigating the types and necessity of English-Dutch post-edits carried out by highly professional translators at the Dutch Language Department of the European Commission Directorate-General for Translation (DGT). Ninety machine-translated segments post-edited by nine different translators were compared to the original NMT output, yielding a total of 493 post-edits for analysis. A typology based on the TQA typologies by Daems, Macken and Vandepitte (2013; 2014; 2016) and de Almeida (2013) was devised, by means of which each of the post-edits was annotated at three different (and independent) levels: a translation quality level, a translation norm level, and a text-linguistic level. From the results of this threefold post-edit analysis, a number of conclusions can be drawn.

Firstly, though MT appears to be adopted by English-Dutch DGT translators, the extent to which considerably depends on the readiness of TM matches, indicating an increasingly integrated translation workflow as often referred to in research and observed in the industry. MT proves to be a useful and advanced tool, but post-editing of its output remains of vital importance for high-quality DGT translation. Many of the same errors were remarkably also repeated by the MT system – therefore needing to be repeatedly corrected by the post-editors – as was found by Man (2019) as well. This means that the NMT system does not learn during the translation process, which could be a point of further improvement of (tailored) NMT systems: the NMT system learning throughout the translation process of merely one text. In addition, a possibility to somehow integrate terminology lists in NMT systems themselves – through which the MT system could automatically and consistently generate correct terms – could be a substantial improvement of NMT systems (i.e. partially going back to rule-based MT, within NMT, in a sense). Similarly, the integration of fixed English to Dutch DGT punctuation conversions for specific items listed in the DGT Dutch Style Guide (e.g. brackets in English – no brackets in Dutch), could save translators post-editing time.

Secondly, while lexical changes frequently need to be carried out to achieve consistency, the use of synonyms or even hyponyms very often appears to be preferential as well, and may therefore be something that post-editors could save time on. The same goes for sentence reordering and some other structural changes, especially since an attitude of making many preferential changes could possibly lead to an increased introduction of undesirable post-edits. DGT translators may also benefit from the sensitisation to keep an eye out specifically for

punctuation errors and omissions of ST information, since both appear to be recurrent issues in eTranslation NMT output. Though preferential (stylistic) changes are perhaps sometimes unnecessarily implemented, it could be argued that all these combined do result in the highest possible translation quality that is ultimately to be achieved for Commission translations (cf. Vandevoorde, Weintraub, & Arabadjieva, 2021).

A number of limitations of the present study should, however, be emphasised. Given that manual post-edit annotation is a labour-intensive and difficult process, this dissertation serves as a pilot study and could include only a very limited number of post-edited segments. These preliminary findings may thus not apply to all English-Dutch DGT translators or even all translations of the translators included in the study, for which they cannot be generalised. The limited scope and time of this dissertation also resulted in the annotation being carried out by only one person, meaning that no inter-annotator agreement could be relied on. As a control means, however, the annotation of two translators was checked by the supervisor and the co-supervisor of the dissertation. Despite great effort to objectify the annotation as much as possible (through the extensive use of external resources such as dictionaries), manual annotation will also always to a certain degree remain subjective. Many post-edits could be considered difficult to annotate, and reviewing the correctness and nature of post-edits can sometimes be difficult when done without the knowledge of the DGT translation that the translators themselves do have. Lastly, the typology devised was based on the formal texts included in the data set such as regulations, contracts, policies, etc. and though it appeared to serve well for these formal text types analysed, it, still, may not be watertight, and will likely need (slight) alteration for the examination of texts of other (more informal) text types. This became apparent when annotating post-edits from Translator 10, who had submitted more informal texts (blog posts) than the other translators.

Nevertheless, the preliminary findings present interesting insights into post-editing behaviour of highly professional translators, and open a number of options for further research. Future research could first of all include more segments in the analysis, and possibly investigate the adoption of MT by different translators, and their types of changes, in relation to their opinions on MTPE. It would be of particular interest as well to hear from the translators themselves why they (preferentially) altered a certain item while post-editing, possibly revealing new relations and common threads. Another interesting investigation could include a post-edit analysis of less formal translations by professional (DGT) translators. Though clear tendencies could already

be observed in the present study, it may not yet in itself be sufficient for establishment of true, tailored English-Dutch DGT post-editing guidelines. More data would first be needed, to confirm or disprove the present findings, and these post-editing guidelines would moreover need to be determined in close consultation with the DGT Dutch Language Department itself.

BIBLIOGRAPHY

Agarwal, A., & Lavie, A. (2008, June). Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output. *Proceedings of the Third Workshop on Statistical Machine Translation*, 115-118.

Allen, J. (2003). Post-editing. *Benjamins Translation Library, 35*, 297-318.

Arnejšek, M., & Unk, A. (2020). Multidimensional assessment of the eTranslation output for English–Slovene. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 383-392. European Association for Machine Translation. Retrieved from https://www.aclweb.org/anthology/2020.eamt-1.41

Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* (pp. 65-72).

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: a case study. *arXiv preprint arXiv:1608.04631*.

Bielsa, E. (2005). Globalisation and translation: A theoretical approach. *Language and Intercultural Communication, 5*(2), 131-144.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006, April). Re-evaluating the role of BLEU in machine translation research. *11th Conference of the European Chapter of the Association for Computational Linguistics*.

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. *Translation Quality Assessment* (pp. 9-38). Springer, Cham.

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art?. *The Prague Bulletin of Mathematical Linguistics,* (108).

Cenni, I. (2019). Multilingualism 2.0: language policies and the use of online translation tools on global platforms. *Argentinian Journal of Applied Linguistics 7*(1), 79-92.

Cho, K., van Merriënboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder–decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724-1734, Doha, Qatar, October. Association for Computational Linguistics.

Costa, Â., Ling, W., Luís, T., Correia, R., & Coheur, L. (2015). A linguistically motivated taxonomy for Machine Translation error analysis. *Machine Translation, 29*(2), 127-161.

Daems, J. (2016). *A translation robot for each translator?: a comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude* (Doctoral dissertation, Ghent University).

Daems, J., & Macken, L. (2019). Interactive adaptive SMT versus interactive adaptive NMT: a user experience evaluation. *Machine Translation, 33*(1), 117-134.

Daems, J., & Macken, L. (2020). POST-EDITING HUMAN TRANSLATIONS AND REVISING MACHINE TRANSLATIONS. *Translation Revision and Post-editing: Industry Practices and Cognitive Processes*, 50.

Daems, J., Macken, L., & Vandepitte, S. (2013, September). Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+ PE. *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice* (Vol. 2, pp. 63-71).

Daems, J., Macken, L., & Vandepitte, S. (2014, May). On the origin of errors: A fine-grained analysis of MT and PE errors and their relationship. In *LREC* (pp. 62-66).

Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2015). The impact of machine translation error types on post-editing effort indicators. *4th Workshop on Post-Editing Technology and Practice (WPTP4)* (pp. 31-45). Association for Machine Translation in the Americas.

Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2017). Translation methods and experience: A comparative analysis of human translation and post-editing with students and professional translators. *Meta: Journal des traducteurs/Meta: Translators' Journal, 62*(2), 245-270.

de Almeida, G., & O'Brien, S. (2010, May). Analysing post-editing performance: correlations with years of translation experience. *Proceedings of the 14th annual conference of the European association for machine translation* (pp. 27-28).

de Almeida, G. (2013). *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages* (Doctoral dissertation, Dublin City University).

Densmer, L. (2014). Light and Full MT Post-Editing Explained.

DePalma, D. (2013). Post-editing in practice. *Translation and Localization.*

Depraetere, I. (2010, May). What counts as useful advice in a university post-editing training context? Report on a case study. *Annual conference of the European Association for Machine Translation*.

Depraetere, I., De Sutter, N., & Tezcan, A. (2014). CHAPTER FOUR POST-EDITED QUALITY, POST-EDITING BEHAVIOUR AND HUMAN EVALUATION: A CASE STUDY1. *Post-editing of machine translation: processes and applications*, 78.

Doherty, S., Gaspari, F., Groves, D., & van Genabith, J. (2013). Mapping the industry I: Findings on translation technologies and quality assessment. GALA.

Doherty, S., & O'Brien, S. (2009). Can MT output be evaluated through eye tracking. *Proceedings of MT Summit XII*, 214-221.

Dowling, M., Lynn, T., Poncelas, A., & Way, A. (2018). SMT versus NMT: Preliminary comparisons for Irish.

Drugan, J. (2013). *Quality in professional translation: assessment and improvement* (Vol. 9). A&C Black.

EN 15038. (2006). *Translation services - Service requirements*

Eszenyi, R. (2016). What Makes a Professional Translator? The Profile of the Modern Translator. *The Modern Translator and Interpreter*, 17-28.

European Union of Associations of Translation Companies (EUATC). (2020). 2020 European Language Industry survey. Retrieved from https://euatc.org/industry-surveys/2020-european-language-industry-survey-full-slide-set/

Farrús, M., Ruiz Costa-Jussà, M., Popovic, M., & Henriquez, C. A. (2012). Study and correlation analysis of linguistic, perceptual and automatic machine translation evaluations. *Journal of the Association for Information Science and Technology, 63*(1), 174-184.

Flanagan, M., & Christensen, T. P. (2014). Testing post-editing guidelines: how translation trainees interpret them and how to tailor them for translator training purposes. *The Interpreter and Translator Trainer, 8*(2), 257-275.

Gambier, Y. (EMT Expert Group). (2009). *Competences for professional translators, experts in multilingual and multimedia communication.* http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf

Guerberof Arenas, A. (2009). Productivity and quality in MT post-editing.

Guerberof Arenas, A. (2013). What do professional translators think about post-editing. *JoSTrans The journal of specialised translation, 19*, 75-95.

Guerberof Arenas, A. (2014). The role of professional experience in post-editing from a quality and productivity perspective.

House, J. (2014). Translation quality assessment: Past and present. *Translation: A multidisciplinary approach*, 241-264. Palgrave Macmillan, London.

Hu, K., & Cadwell, P. (2016). A comparative study of post-editing guidelines. *Baltic Journal of Modern Computing, 4*(2), 346-353.

International Organization for Standardization (ISO) 17100:2015. (2015). *Translation services-requirements for translation services*. Technical Committee ISO/TC37, 2015. Retrieved from https://edu.mynbn.be/nbnframework/index.php/pdfMeta/RO/538433?l=E

International Organization for Standardization (ISO) 18587:2017. (2017). *International Standard. Translation services — Post-editing of machine translation output — Requirements*. Retrieved from https://edu.mynbn.be/nbnframework/index.php/pdfMeta/RO/583010?l=E

Jakobsen, A., & Jensen, K. (2008). Eye Movement Behaviour Across Four Different Types of Reading Task. *Copenhagen Studies in Language, 36*, 103-124.

Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation, 31*, 60-86.

Junczys-Dowmunt, M., Dwojak, T., & Hoang, H. (2016). Is neural machine translation ready for deployment? A case study on 30 translation directions. *arXiv preprint arXiv:1610.01108.*

Kelleher, J. D. (2016). Fundamentals of machine learning for neural machine translation.

Koehn, P., & Germann, U. (2014, April). The impact of machine translation quality on human post-editing. *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*, 38-46.

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. *arXiv preprint arXiv:1706.03872.*

Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2017). Fine-grained human evaluation of neural versus phrase-based machine translation. *arXiv preprint arXiv:1706.04389*.

Kluvanec, D. (2017). From machine translation at the European Commission to European language resource coordination (ELRC). Paper presented at *EULITA: #TranslatingEurope Workshop* (Vienna, 30 March 2017). http://eulita.eu/wp/772-2/.

Koponen, M. (2010, January). Assessing machine translation quality with error analysis. *Electronic proceeding of the KaTu symposium on translation and interpreting studies*.

Koponen, M. (2013, September). This translation is not too bad: An analysis of post-editor choices in a machine translation post-editing task. *Workshop Proceeding: Workshop on Post-editing Technology and Practice (WPTP-2)* (pp. 1-9).

Koponen, M., & Salmi, L. (2017). Post-editing quality: Analysing the correctness and necessity of post-editor corrections. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 16*, 137–148.

Koponen, M., Salmi, L., & Nikulin, M. (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation, 33*(1), 61-90.

Koponen, M., Sulubacak, U., Vitikainen, K., & Tiedemann, J. (2020, November). MT for subtitling: User evaluation of post-editing productivity. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 115-124.

Koskinen, K. (2008). *Translating institutions: An ethnographic study of EU translation*. Manchester: St. Jerome.

Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes* (Vol. 5). Kent State University Press.

Läubli, S., Fishel, M., Massey, G., Ehrensberger-Dow, M., Volk, M. (2013). Assessing post-editing efficiency in a realistic translation environment. *MT Summit XIV Workshop on Post-editing Technology and Practice*, 83-91.

Lavie, A., & Agarwal, A. (2007, June). METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. *Proceedings of the second workshop on statistical machine translation* (pp. 228-231).

Lesznyák, Á. (2019). Hungarian translators' perceptions of neural machine translation in the European Commission. *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks* (pp. 16-22).

Localization Industry Standards Association. (2011). LISA QA Model 3.1.

Lommel, A. (2018). Metrics for translation quality assessment: a case for standardising error typologies. *Translation Quality Assessment* (pp. 109-127). Springer, Cham.

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*, (12), 455-463.

Macken, L., Prou, D., & Tezcan, A. (2020, June). Quantifying the Effect of Machine Translation in a High-Quality Human Translation Production Process. *Informatics, 7*(12), 19pp. Multidisciplinary Digital Publishing Institute.

Man, C. T. (2019). Post-Editing at Work: A Study into a Real-Life Post-Editing Environment at the KBC Language Department.

Martínez, L. G. (2003). *Human Translation versus Machine Translation and Full Post-editing of Raw Machine Translation Output.* Univ..

Massardo, I. (2018, 23 December). *What Is Post-Editing? Here Are 3 Approaches*. Wordbee Translation Management and Localization Tools. Retrieved on 20 May 2021 from https://www.wordbee.com/blog/localization-industry/what-is-post-editing-here-are-3-approaches/.

Melby, A. K., Fields, P. J., & Housley, J. (2014). CHAPTER TWELVE ASSESSMENT OF POST-EDITING VIA STRUCTURED TRANSLATION SPECIFICATIONS. *Post-editing of Machine Translation: Processes and Applications*, 274.

Moorkens, J., & O'Brien, S. (2015, May). Post-editing evaluations: Trade-offs between novice and professional participants. *Proceedings of the 18th Annual Conference of the European Association for Machine Translation* (pp. 75-81).

Nießen, S., Och, F. J., Leusch, G., & Ney, H. (2000, May). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. *LREC*.

NIST & LDC. (2007). Post editing guidelines for GALE machine translation evaluation. Guidelines developed by the National Institute of Standards and Technology (NIST), and the Linguistic Data Consortium (LDC).

O'Brien, S. (2005). Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine translation, 19*(1), 37-58.

O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures, 7*(1), 1-21.

O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces, 1*(1), 101-122.

O'Brien, S., Choudhury, R., van der Meer, J., Aranberri Monasterio, N. (2011). Dynamic Quality Evaluation Framework. Retrieved from https://www.taus.net/component/rsfiles/download-file/files?path=Reports%2FFree+Reports%2Ftausdynamicquality.pdf.

Okpor, M. D. (2014). Machine translation approaches: issues and challenges. *International Journal of Computer Science Issues (IJCSI), 11*(5), 159.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).

Plitt, M., & Masselot, F. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics 93*(1), 7–16.

Popović, M. (2011). Hjerson: An open source tool for automatic error classification of machine translation output. *The Prague Bulletin of Mathematical Linguistics, 96*(1), 59-67.

Popović, M., Arcan, M., Avramidis, E., Burchardt, A., & Lommel, A. (2015). Poor man's lemmatisation for automatic error classification. *The 18th annual conference of the European Association for Machine Translation (EAMT 2015)*, 105–112.

Rossi, C., & Chevrot, J. P. (2019). Uses and perceptions of machine translation at the European Commission. *The Journal of specialised translation* (*JoSTrans*).

SAE-J2540. (2001). *Quality Metric for Language Translation*.: Society of Automotive Engineers.

Sánchez-Gijón, P., Moorkens, J., & Way, A. (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation, 33*(1), 31-59.

Skadiņa, I., & Pinnis, M. (2017, November). NMT or SMT: case study of a narrow-domain English-Latvian post-editing project. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 373-383).

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006, August). A study of translation edit rate with targeted human annotation. *Proceedings of association for machine translation in the Americas* (Vol. 200, No. 6).

Stasimioti, M., Sosoni, V., Kermanidis, K. L., & Mouratidis, D. (2020, November). Machine Translation Quality: A comparative evaluation of SMT, NMT and tailored-NMT outputs. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 441-450).

Stefaniak, K. (2020, November). Evaluating the usefulness of neural machine translation for the Polish translators in the European Commission. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 263-269.

Stenetorp, P., Pyysalo, S., Topic, G., Ananiadou, S., & Aizawa, A. (2012). Normalisation with the BRAT rapid annotation tool. *SMBM 2012*, 87.

Svoboda, T. (2017). Translation manuals and style guides as quality assurance indicators: The case of the European Commission's Directorate-General for Translation. *Quality aspects in institutional translation, 8*, 75.

Translation Automation User Society (TAUS). (2010, March). *Post-editing in Practice Report*.

Translation Automation User Society (TAUS). (2010, 2016). Machine Translation Post-editing Guidelines produced in partnership with CNGL (Centre for Next Generation Localisation). Retrieved on 29 January 2021 from https://info.taus.net/mt-post-editing-guidelines.

Translation Automation User Society (TAUS). (2017). Quality Evaluation using Adequacy and Fluency Approaches. Retrieved on 29 January 2021from https://cdn2.hubspot.net/hubfs/2734675/Reports,%20ebooks/Adequacy-Fluency%20Best%20Practice%20Guidelines.pdf.

Tezcan, A., Hoste, V., & Macken, L. (2017b). Scate taxonomy and corpus of machine translation errors. In G. C. Pastor & I. Durán-Muñoz (Eds.), *Trends in e-tools and resources for translators and interpreters* (Vol. 45, pp. 219–244). Brill | Rodopi.

Tezcan, A., Hoste, V., & Macken, L. (2020). Estimating Word-Level Quality of Statistical Machine Translation Output Using Monolingual Information Alone. *Natural Language Engineering 1*(1), 1-31. Cambridge University Press.

Toral, A., & Sánchez-Cartagena, V. M. (2017). A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. *arXiv preprint arXiv:1701.02901*.

Toral, A., Wieling, M., & Way, A. (2018). Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities*, *5*, 9.

Toury, G. (1995). The Nature and Role of Norms in Translation. In G. Toury (Ed.), *Descriptive Translation Studies and Beyond* (pp. 53-69). Amsterdam-Philadelphia: John-Benjamins.

Valdez, S., & Vandepitte, S. (2020). Exploring a two-way street: revisers' and translators' attitudes and expectations about each other in biomedical translation. *Translation revision and/or post-editing: industry practices and cognitive processes* (pp. 148-164). Routledge.

van Belle, M. C. (2020). Evaluation of Facebook's NMT with monolingual post-editing: An exploratory investigation of error types and comprehension.

van Egdom, G. M. W., & Pluymaekers, M. (2019). Why go the extra mile? How different degrees of post-editing affect perceptions of texts, senders and products among end users. *Journal of specialised translation*, *31*, 158-176.

van Rijsbergen, C. (1979). *Information Retrieval*. London: Butterworths.

Van Slype, G. (1979). *Critical Methods for Evaluating the Quality of Machine Translation. Prepared for the European Commission Directorate-General Scientific and Technical Information and Information Management*, Report BR-19142. Bureau Marcel van Dijk.

Vandevoorde, L., Weintraub, R. & Arabadjieva, M. (2021). Sustained quality in Council translations: assessing the importance of human translation actions. Presentation at JIAMCATT 2021.

Vardaro, J., Schaeffer, M., & Hansen-Schirra, S. (2019, September). Translation quality and error recognition in professional neural machine translation post-editing. *Informatics* (Vol. 6, No. 3, p. 41). Multidisciplinary Digital Publishing Institute.

Vieira, L. N. (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation, 30*(1-2), 41-62.

Vilar, D., Xu, J., Luis Fernando, D. H., & Ney, H. (2006, May). Error Analysis of Statistical Machine Translation Output. *LREC*, 697-702.

Way, A. (2018). Quality expectations of machine translation. *Translation quality assessment* (pp. 159-178). Springer, Cham.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Yamada, M. (2014). Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings. *Machine Translation, 29*(1), 49-67.

Yamada, M. (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation, 31*, 87-106.

Zeman, D., Fishel, M., Berka, J., & Bojar, O. (2011). Addicter: what is wrong with my translations?. *The Prague Bulletin of Mathematical Linguistics*, *96*(1), 79-88.

APPENDIX

APPENDIX A POST-EDIT TYPOLOGY

1 Post-Edit Typology

1.1 Translation Quality Parameter

Definition Categorisation of a post-edit from the perspective of correctness and necessity.

Style guide changes

Definition A post-edit is considered a style guide change if the change has been made in order for the target text to adhere to (the) internal (Dutch) style guide(s).

e.g.: (ST) the Farm to Fork Strategy
(MT) „van boer tot bord"-strategie
(PE) "van boer tot bord"-strategie

Consistency & intertextuality changes

Definition A specific type of style guide change: the elimination of inconsistent use of terms or notations throughout the MT output, in order to achieve conformity between the post-edited segments and TM matches, other segments in the text, similar other texts (of a similar text type), and terminology lists (i.e. IATE).

e.g.: (ST) Similarly, while the agricultural factor income per worker is well above the EU average, its evolution is following a slightly negative trend because expenses are growing faster than revenue.
(MT) Ook al ligt het landbouwfactorinkomen per werknemer ruim boven het EU-gemiddelde, dan volgt de ontwikkeling ervan een licht negatieve trend omdat de uitgaven sneller stijgen dan de inkomsten.
(PE) Ook bij het agrarisch factorinkomen per werkende, dat nochtans ruim boven het EU-gemiddelde ligt, is er een licht negatieve trend omdat de uitgaven sneller stijgen dan de inkomsten.

MT errors

Definition A post-edit is considered an MT error if without the change, the sentence (or part of it) becomes either:
a) grammatically incorrect (i.e. it would have breached a grammatical rule specified in accepted grammar books), or
b) grammatically correct, but rendering a blatantly different meaning in comparison to the source text (i.e. it would not have contained all the information present in the source text, or it would have contained extra information that was not present in the source text).

e.g.: (ST) In its view, the Commission admitted that state presence is rather limited in citrus fruits industry, but then side-tracked this issue

(MT) Volgens haar heeft de Commissie erkend dat de aanwezigheid van de staat in de citrusvruchtenindustrie vrij beperkt is, maar heeft zij deze kwestie gevolgd

(PE) Volgens de CCC heeft de Commissie erkend dat de overheidsdeelneming in de citrusvruchtenindustrie vrij beperkt is, maar heeft zij deze kwestie terzijde geschoven

Preferential changes

Definition A post-edit is considered a preferential change if even without the change, the target sentence from the MT output is already grammatically correct, intelligible, and adequate in relation to the source text.

e.g.: (ST) However, the gap between agricultural and non-agricultural income has been growing since 2012.

(MT) De kloof tussen het landbouwinkomen en het niet-landbouwinkomen neemt echter toe sinds 2012.

(PE) Sinds 2012 neemt de kloof tussen het landbouwinkomen en het niet-landbouwinkomen evenwel toe.

Undesirable changes

Definition A post-edit is considered an undesirable change if the MT output is both adequate and acceptable and if the target sentence becomes grammatically incorrect, does not follow the internal Dutch style guide and internal quality guidelines, or is inadequate in relation to the source text, as a result of the post-edit.

e.g.: (ST) In addition, specific rules will be introduced for very large platforms given their systemic impact in facilitating public debate, economic transactions and the dissemination of information, opinions and ideas.

(MT) Daarnaast zullen specifieke regels worden ingevoerd voor zeer grote platforms, gezien hun systemische impact op het bevorderen van het publieke debat, economische transacties en de verspreiding van informatie, meningen en ideeën.

(PE) Voor zeer grote platformen zullen er daarnaast specifieke regels worden ingevoerd, gezien hun wezenlijke impact op het openbaar debat, economische transacties en de verspreiding van informatie, meningen en ideeën.

1.2 Translation Norm Parameter

Definition Categorisation of a post-edit from the perspective of the relation between the target text and the source text (Toury's norm of adequacy), and the target text and the target language (Toury's norm of acceptability).

Adequacy

Definition A post-edit is considered a change in terms of adequacy if the MT output information is different from the ST information and the post-edited sentence

restores or was implemented with the intention to restore the information from the ST.

e.g.: (ST) professional and amateur outfield players and goalkeepers
(MT) professionele en amateursporten en goalkekers
(PE) professionele en amateurbuitenveldspelers en -doelverdedigers

Acceptability

Definition A post-edit is considered a change in terms of acceptability if it was implemented with the intention of having the target sentence adhere to the norms and rules of the target language and culture and be suitable for its function in the target culture.

e.g.: (ST) However, the gap between agricultural and non-agricultural income has been growing since 2012.
(MT) De kloof tussen het landbouwinkomen en het niet-landbouwinkomen neemt echter toe sinds 2012.
(PE) Sinds 2012 neemt de kloof tussen het landbouwinkomen en het niet-landbouwinkomen evenwel toe.

Adequacy + acceptability

Definition A post-edit applying to both the adequacy and the acceptability norms.

e.g.: (ST) Given the importance of steel as input and in view of the fact that all inputs
(MT) Gezien het belang van staal als grondstof en gezien het feit dat alle inputs
(PE) Gezien het belang van staal als basisproduct en gezien het feit dat alle basisproducten

1.3 Text-Linguistic Characteristics

Definition Categorisation of a post-edit from a (text-)linguistic perspective.

1.3.1 Semantics Parameter

Definition Categorisation of a post-edit bringing about a difference in meaning as opposed to the MT output and/or the source text.

Addition

Addition of source text information missing from the MT output

Definition Information from the source text that was not present in the MT output – for which a meaning shift had been caused in the MT output – has been added.

e.g.: (ST) and that movements of (1) millions of people in numerous cities demonstrates the existence of mobility of people (2) in China
(MT) en dat (1) miljoenen mensen in tal van steden het bestaan van mobiliteit (2) in China aantonen
(PE) en dat de verplaatsingen van (1) miljoenen mensen in tal van steden het bestaan van mobiliteit van personen (2) in China aantonen.

Explicitation of ST and MT output

Definition Information that **could be derived from, i.e. was implicit in both the source text and the MT output**, and **did not necessarily need to be expressed** for the reader to understand the information presented in the text, has been made explicit. The MT output sentence may have been a too literal translation for it to be acceptable in the target language, therefore requiring explicitation.

e.g.:
(ST) To perform the risk characterisation of the granules or mulches containing the eight PAHs, the concentration limits of the individual PAHs in entry 28 of Annex XVII to Regulation (EC) No 1907/2006 cannot simply be summed up
(MT) Om de risicokarakterisering van de korrels of mulches die de acht PAK's bevatten, te verrichten, kunnen de concentratiegrenzen van de afzonderlijke PAK's in vermelding 28 van bijlage XVII bij Verordening (EG) nr. 1907/2006 niet eenvoudig worden opgeteld.
(PE) Om de risicokarakterisering van het granulaat dat of de mulch die de acht pak's bevat, te verrichten, kunnen de concentratiegrenzen van de afzonderlijke pak's in vermelding 28 van bijlage XVII bij Verordening (EG) nr. 1907/2006 niet eenvoudigweg bij elkaar worden opgeteld.

Hyponymy

Definition A MT output word or phrase has been substituted for a **hyponym** of that word/phrase, through which information has been added in the form of specification.

e.g.:
(ST) Non-perennial crops
(MT) Niet-blijvende gewassen
(PE) Eenjarige gewassen

Other addition

Definition A meaningful element that **could not be derived from the source text** and was not expressed in the MT output, has been added.

e.g.:
(ST) In accordance with Article 50, the Union negotiated with the United Kingdom the Withdrawal Agreement.
(MT) Overeenkomstig artikel 50 heeft de Unie met het Verenigd Koninkrijk onderhandeld over het terugtrekkingsakkoord.
(PE) Overeenkomstig dit artikel 50 heeft de Unie met het Verenigd Koninkrijk na onderhandelingen het terugtrekkingsakkoord gesloten.

Deletion

Deletion of ST and MT information

Definition A meaningful element of the source text or a relation that was present in both the source text and the MT output, has been deleted.

e.g.:
(ST) Assist the Head of Representation in defining the strategy and policies of the team, and in programming and organising the functioning of the media activities in the Representation;
(MT) Het hoofd van de vertegenwoordiging bijstaan bij het bepalen van de strategie en het beleid van het team en bij het programmeren en organiseren van de werking van de media-activiteiten in de vertegenwoordiging;

(PE) het hoofd van de vertegenwoordiging bijstaan bij het bepalen van de strategie en het beleid van het team, en bij het programmeren en organiseren van de media-activiteiten in de vertegenwoordiging;

Deletion of extra information in MT output

Definition Information from the MT output that was not expressed in the source text, has been deleted.

e.g.:
(ST) Moreover, the CCC submitted that the Appellate Body ruling in dispute DS473 set the rules how to determine normal values and that rulings of the Appellate Body must be adhered to by the EU.
(MT) Bovendien heeft het CCC aangevoerd dat in de uitspraak van de Beroepsinstantie in geschil DS473 de regels zijn vastgesteld voor de vaststelling van de normale waarden en dat de uitspraken van de beroepsinstantie door de EU moeten worden nageleefd.
(PE) Bovendien voerde de CCC aan dat in de uitspraak van de Beroepsinstantie in geschil DS473 de regels voor de vaststelling van normale waarden zijn vastgelegd en dat uitspraken van de beroepsinstantie door de EU moeten worden nageleefd.

Hyperonymy

Definition A MT output word or phrase has been substituted for a **hypernym** (a superordinate term or expression) of that word/phrase.

e.g.:
(ST) In its evaluation, the Agency recommended lowering the concentration limit of the eight PAHs in granules used in synthetic turf pitches
(MT) In zijn beoordeling heeft het Agentschap aanbevolen de concentratiegrens voor de acht pak's in granulaat voor gebruik in kunstgrasvelden te verlagen
(PE) In zijn evaluatie heeft het ECHA aanbevolen de concentratiegrens voor de acht PAK's in korrels die in kunstgrasvelden worden gebruikt, te verlagen

Other meaning shift

Action

Definition The type of action happening in the MT output sentence has been changed.

e.g.:
(ST) In its view, the Commission admitted that state presence is rather limited in citrus fruits industry, but then side-tracked this issue
(MT) Volgens haar heeft de Commissie erkend dat de aanwezigheid van de staat in de citrusvruchtenindustrie vrij beperkt is, maar heeft zij deze kwestie gevolgd
(PE) Volgens de CCC heeft de Commissie erkend dat de overheidsdeelneming in de citrusvruchtenindustrie vrij beperkt is, maar heeft zij deze kwestie terzijde geschoven

Time

Definition The post-edit has brought about any sort of shift in time in comparison to the MT output.

e.g.:
(ST) the CCC submitted
(MT) heeft het CCC aangevoerd

(PE) voerde de CCC aan

Place

Definition The post-edit has caused the action to take place in another place than it did in the MT output sentence.

e.g.:
(ST) Therefore, the concentrations of the eight PAHs in granules used as infill material in synthetic turf pitches
(MT) Daarom moeten de concentraties van de acht PAK's in korrels die als infiltratiemateriaal worden gebruikt in kunstgrasvelden
(PE) Daarom moeten de concentraties van de acht pak's in granulaat voor gebruik als instrooimateriaal voor kunstgrasvelden

Agent

Definition The agent executing the action in the MT output sentence has been changed by the post-editor.

e.g.:
(ST) By the deadline for submission of applications, and in addition to the diploma and when applicable to the one year experience required above, candidates must have at least 2 years' full time professional experience gained after obtaining the diploma required under Point 3.3.1 directly related to the duties as described under Point 1 above.
(MT) Uiterlijk op de uiterste datum voor de indiening van sollicitaties moeten kandidaten, naast het diploma en, indien van toepassing, de hierboven vereiste werkervaring van één jaar, ten minste 2 jaar voltijdse beroepservaring hebben opgedaan na het behalen van het in punt 3.3.1 bedoelde diploma dat rechtstreeks verband houdt met de in punt 1 beschreven taken.
(PE) Uiterlijk op de uiterste datum voor de indiening van sollicitaties moeten kandidaten, naast het diploma en, indien van toepassing, de hierboven vereiste werkervaring van één jaar, na het behalen van het in punt 3.3.1 bedoelde diploma, ten minste twee jaar voltijdse werkervaring die rechtstreeks verband houdt met de in punt 1 beschreven taken hebben opgedaan.

Modality

Definition The modality of a clause from the MT output has been altered.

e.g.:
(ST) Adding the United Kingdom to the to the list of countries included in the EU001, should not negatively affect the security of the Union or international security.
(MT) De toevoeging van het Verenigd Koninkrijk aan de lijst van landen in de EU001 mag geen negatieve gevolgen hebben voor de veiligheid van de Unie of de internationale veiligheid.
(PE) De toevoeging van het Verenigd Koninkrijk aan de lijst van landen waarvoor uitvoervergunning EU001 geldig is, zal naar verwachting geen negatieve gevolgen hebben voor de veiligheid van de Unie of de internationale veiligheid.

e.g.:
(ST) At the same time, online intermediaries will also benefit from the legal clarity of the liability exemptions and from a single set of rules when providing their services in the EU.

(MT) Tegelijkertijd zullen onlinetussenpersonen ook profiteren van de juridische duidelijkheid van de vrijstellingen van aansprakelijkheid en van één enkele reeks regels bij het verlenen van hun diensten in de EU.

(PE) Aan de andere kant zullen onlinetussendiensten, wanneer zij hun diensten in de EU verlenen, kunnen profiteren van de juridische duidelijkheid over de ontheffingen van aansprakelijkheid en van een gemeenschappelijke reeks regels.

Other meaning change

Definition Any other change regarding the relation between the target text, the source text and the MT output.

e.g.:

(ST) In this respect, the Commission noted that the productivity of this industry is influenced by the quality and the quantity of the available fresh fruit meaning that in a year when the harvest is good productivity goes up and when the harvest is bad productivity goes down.

(MT) In dit verband merkte de Commissie op dat de productiviteit van deze sector wordt beïnvloed door de kwaliteit en de kwantiteit van het beschikbare verse fruit, wat betekent dat in een jaar waarin de oogst een goede productiviteit is en de oogst slecht is, de productiviteit daalt.

(PE) In dit verband merkte de Commissie op dat de productiviteit van deze bedrijfstak wordt beïnvloed door de kwaliteit en de kwantiteit van het beschikbare verse fruit, wat betekent dat in een jaar met een goede oogst de productiviteit stijgt, terwijl deze daalt wanneer de oogst slecht is.

1.3.2 Syntax & Morphology Parameter

Definition Categorisation of a syntactical or morphological change made to the MT output by the post-editor.

Order

Internal phrase order

Definition The order of the words within a single phrase (e.g. a verb phrase) in the MT output has been altered.

e.g.:

(ST) In reply to these comments, the Commission pointed out that its above analysis concerning distortions according to Article 2(6a)(b), sixth indent, is based on objective evidence which was placed on the file and on which the CCC was given an opportunity to comment.

(MT) In antwoord op deze opmerkingen heeft de Commissie erop gewezen dat haar bovenstaande analyse met betrekking tot verstoringen overeenkomstig artikel 2, lid 6 bis, onder b), zesde streepje, is gebaseerd op objectieve gegevens die in het dossier zijn opgenomen en waarover het CCC opmerkingen heeft kunnen maken.

(PE) In antwoord op deze opmerkingen wees de Commissie erop dat haar bovenstaande analyse met betrekking tot verstoringen overeenkomstig artikel 2, lid 6 bis, onder b), zesde streepje, gebaseerd is op objectieve gegevens die in het dossier zijn opgenomen en waarover de CCC opmerkingen heeft kunnen maken.

Internal sentence order

Definition A constituent from the MT output sentence has been moved in the post-edited sentence.

e.g.: (ST) the factors of production used in the manufacturing of canned mandarins
(MT) de productiefactoren die worden gebruikt bij de vervaardiging van mandarijnen in blik
(PE) de productiefactoren die bij de vervaardiging van mandarijnenconserven worden gebruikt

Agreement

Article-noun agreement

Definition An article from the MT output has been substituted, with the objective of having it agree with the relevant noun.

e.g.: (ST) the CCC
(MT) het CCC
(PE) de CCC

Noun-adjective agreement

Definition An adjective or noun from the MT output has been adjusted, with the objective of having it agree with the relevant adjective/noun.

e.g.: (ST) safe supply and use of these granules
(MT) een veilige levering en gebruik van deze korrels
(PE) een veilige levering en veilig gebruik van dit granulaat

Subject-verb agreement

Definition A verb from the MT output has been adjusted, with the objective of having it agree in number with its subject.

e.g.: (ST) However, as noted in recital (90), rather than the physical mobility of workers itself
(MT) Zoals in overweging (90) is opgemerkt, zijn het echter niet zozeer de fysieke mobiliteit van werknemers zelf
(PE) Zoals opgemerkt in overweging 90 is het echter niet zozeer de fysieke mobiliteit van werknemers zelf

Reference

Definition A referring expression has been adjusted or substituted.

e.g.: (ST) In their comments on final disclosure, the CCC remarked that
(MT) In haar opmerkingen over de mededeling van de definitieve bevindingen merkte de CCC op dat
(PE) In zijn opmerkingen over de mededeling van de definitieve bevindingen merkte de CCC op dat

Structural change

Nominalisation

Definition A verb phrase from the MT output has been substituted for a nominalisation of the main verb in that verb phrase.

e.g.: (ST) as if resources are being allocated in the PRC without following any market forces

(MT) alsof er middelen worden toegewezen in de VRC zonder dat daarbij marktkrachten worden gevolgd

(PE) alsof bij de toewijzing van middelen in de VRC de marktwerking volledig buiten beschouwing wordt gelaten

Reduction of nominalisation

Definition A noun phrase from the MT output has been substituted for a verb phrase with the nominalised verb in the noun phrase as its main verb.

e.g.: (ST) A Farm Sustainability Plan sets out the agricultural holding's strategy to contribute significantly to climate change mitigation by both reducing greenhouse gases (GHG) emissions and strengthening land carbon sinks (or, in case of saturation, maintaining land carbon stocks).

(MT) In een landbouwbedrijfsduurzaamheidsplan wordt de strategie van het landbouwbedrijf uiteengezet om aanzienlijk bij te dragen aan de mitigatie van klimaatverandering door zowel de uitstoot van broeikasgassen terug te dringen als de koolstofputten op land te versterken (of, in het geval van verzadiging, het in stand houden van koolstofvoorraden op land).

(PE) Een landbouwbedrijfsduurzaamheidsplan stelt de strategie van het landbouwbedrijf vast om substantieel bij te dragen aan de mitigatie van klimaatverandering door zowel broeikasgasemissies te reduceren als koolstofputten in de grond te versterken (of, in het geval van verzadiging, de koolstofvoorraden in de grond in stand te houden).

Active → passive

Definition An active construction in the MT output has been converted to a passive construction.

e.g.: (ST) this will help track down rogue traders and will protect online shoppers against illegal, counterfeit and unsafe products.

(MT) dit zal malafide handelaren helpen opsporen en onlinekopers beschermen tegen illegale, nagemaakte en onveilige producten.

(PE) daardoor zullen malafide handelaren beter kunnen worden opgespoord en zullen onlinekopers worden beschermd tegen illegale, nagemaakte en onveilige producten.

Passive → active

Definition A passive construction in the MT output has been converted to an active construction.

e.g.: (ST) A Farm Sustainability Plan sets out the agricultural holding's strategy to contribute significantly to climate change mitigation

(MT) In een landbouwbedrijfsduurzaamheidsplan wordt de strategie van het landbouwbedrijf uiteengezet om aanzienlijk bij te dragen aan de mitigatie van klimaatverandering

(PE) Een landbouwbedrijfsduurzaamheidsplan stelt de strategie van het landbouwbedrijf vast om substantieel bij te dragen aan de mitigatie van klimaatverandering

Other structural change

Definition All other types of structural changes to the MT output, particularly substitutions of relative clauses or post-nominal phrases for adjectives and vice versa.

e.g.:

(ST) It concerns the targets related to use and risk of pesticides, sale of antimicrobials, nutrient loss (1), area under organic farming (2), high diversity landscape features (3) and access to fast broadband internet.

(MT) Het gaat om de doelstellingen in verband met het gebruik en de risico's van pesticiden, de verkoop van antimicrobiële stoffen, het verlies van nutriënten (1), het areaal dat onder de biologische landbouw valt (2), landschapselementen met een grote diversiteit (3) en toegang tot snel breedbandinternet.

(PE) De streefdoelen betreffen het gebruik en de risico's van pesticiden, de verkoop van antimicrobiële stoffen, het nutriëntenverlies (1), het areaal biologische landbouw (2), de diversiteitsrijke landschapselementen (3) en de toegang tot snel breedbandinternet.

Morphological change

Comparative/superlative of an adjective/adverb

Definition The structure or form of the comparative or superlative of an adjective or adverb from the MT output has been changed.

Singular/plural noun

Definition A plural form of a noun in the MT output has been changed to a singular form or vice versa, excluding agreement changes.

e.g.:

(ST) The Dossier Submitter took into consideration various exposure scenarios related to the use of granules in synthetic turf pitches, by workers installing and maintaining the pitches and by individuals playing sports on them (professional and amateur outfield players and goalkeepers), and related to granules or mulches used in loose applications on playgrounds and in sport applications, where people and especially children may be exposed.

(MT) De Dossier Submitter hield rekening met verschillende blootstellingsscenario's in verband met het gebruik van granulaten in kunstgrasvelden, door werknemers die de velden installeren en onderhouden en door personen die sporten op hen uitoefenen (professionele en amateursporten en goalkekers), en die verband hielden met granulaten of mulches die worden gebruikt in losse toepassingen op speelplaatsen en in sporttoepassingen, waarbij mensen en met name kinderen kunnen worden blootgesteld.

(PE) De indiener van het dossier heeft in verband met het gebruik van granulaat rekening gehouden met verschillende blootstellingsscenario's, waaronder voor werknemers die kunstgrasvelden aanleggen en onderhouden en mensen die daarop een sport beoefenen (professionele en amateurbuitenveldspelers en -doelverdedigers), of voor mensen en met name kinderen kunnen worden blootgesteld aan granulaat dat of mulch die in losse vorm voor speelplaatsen en in sporttoepassingen wordt gebruikt.

Named entity

Definition A Dutch named entity (geographical location, name, company, organization, etc.) from the MT output has been altered: (partially) untranslated named entities

that have an official Dutch equivalent, (partially) translated named entities that do not exist in Dutch, etc.

e.g.: (ST) ADA
(MT) ADA
(PE) WTO-ADO

Other morphological change

Definition Any other type of morphological change to the MT output.

e.g.: (ST) A new EU approach is needed to ensure we have a strong, fair competitive and green industry that delivers for patients, and which draws on the potential of the digital transformation of health and care, driven by technological advances in fields such as artificial intelligence and computational modelling.

(MT) Er is een nieuwe EU-aanpak nodig om te zorgen voor een sterke, eerlijke, concurrerende en groene industrie die resultaten oplevert voor patiënten en die gebruikmaakt van het potentieel van de digitale transformatie van gezondheid en zorg, aangestuurd door technologische vooruitgang op gebieden als kunstmatige intelligentie en computermodellen.

(PE) Er is een nieuwe EU-aanpak nodig om te zorgen voor een sterke, eerlijk concurrerende en groene bedrijfstak die resultaten oplevert voor patiënten en die het potentieel van de digitale transformatie van gezondheid en zorg benut, aangestuurd door technologische vooruitgang op gebieden als kunstmatige intelligentie en computermodellering.

1.3.3 Style & Register Parameter

Definition Categorisation of a register change or stylistic change that has been made to the MT output by the post-editor.

Synonymy

Definition A word or phrase from the MT output has been substituted for a synonym of that word/phrase.

e.g.: (ST) However, the gap between agricultural and non-agricultural income has been growing since 2012.

(MT) De kloof tussen het landbouwinkomen en het niet-landbouwinkomen neemt echter toe sinds 2012.

(PE) Sinds 2012 neemt de kloof tussen het landbouwinkomen en het niet-landbouwinkomen evenwel toe.

Preposition

Definition A preposition from the MT output has been added, deleted, or substituted without that causing a meaning shift as opposed to the source text.

e.g.: (ST) The present investigation revealed nothing that would call those findings into question, with the CCC and the Yiguan group merely submitting that land allocation does not equate restriction or forbidding of commercial land use, as well as that all countries protect farmers and since agriculture represents a crucial part of any country economic activities that are closely linked to social stability and security.

(MT) Het onderhavige onderzoek heeft niets aan het licht gebracht dat deze bevindingen ter discussie zou stellen, aangezien de CCC en de Yiguan-groep enkel stellen dat de toewijzing van grond niet gelijkstaat aan beperking of verbod van commercieel grondgebruik, en dat alle landen landbouwers beschermen en aangezien landbouw een cruciaal onderdeel vormt van alle economische activiteiten van een land die nauw verband houden met sociale stabiliteit en sociale zekerheid.

(PE) Het onderhavige onderzoek heeft niets aan het licht gebracht dat deze bevindingen ter discussie zou stellen, aangezien de CCC en de Yiguan-groep enkel stellen dat de toewijzing van grond niet gelijkstaat aan een beperking van of verbod op commercieel grondgebruik en dat alle landen landbouwers beschermen omdat landbouw in alle landen een cruciaal onderdeel vormt van de economische activiteiten die nauw verband houden met sociale stabiliteit en veiligheid.

Lexical change

Definition A word or phrase from the MT output has been substituted for another word/phrase that is not a synonym of the MT output word/phrase, without that causing a meaning shift as opposed to the source text.

e.g.:

(ST) nitrous oxide (N2O) emissions, including from manure management (including storage and treatment of manure) as well as direct N2O emissions from managed soils (in accordance with IPCC reporting framework) or other agricultural sources;

(MT) emissies van distikstofoxide (N2O), onder meer door mestbeheer (met inbegrip van opslag en behandeling van mest) en directe N2O-emissies van beheerde bodems (overeenkomstig het IPCC-rapportagekader) of andere agrarische bronnen;

(PE) emissies van distikstofoxide (N2O), onder meer van mestbeheer (met inbegrip van opslag en verwerking van mest) en directe N2O-emissies van bodembeheer (overeenkomstig het IPCC-rapportagekader) of andere agrarische bronnen;

Register

Definition A word or phrase in the MT output has been altered so that it would still convey the same meaning but belong to a more formal/informal/... register or a regional variety of the language that is more suitable for the target audience.

e.g.:

(ST) which was not true

(MT) wat niet waar was

(PE) wat niet op waarheid berustte

Untranslated text in MT

Definition A word or phrase of which a Dutch translation exists but that had been left untranslated in the MT output (i.e. for which the MT system had copied the English word/phrase), has been replaced by its Dutch equivalent.

e.g.:

(ST) Based on those conclusions and evaluations, on 17 September 2018 the Netherlands (hereinafter 'the Dossier Submitter') submitted to the Agency an Annex XV dossier, proposing a restriction on eight PAHs in granules for use as infill material in synthetic turf pitches and of granules or mulches in loose form on playgrounds or in sport applications.

(MT) Op basis van die conclusies en evaluaties heeft Nederland (hierna: „Dossier Submitter") op 17 september 2018 bij het Agentschap een bijlage XV-dossier ingediend, waarin een beperking werd voorgesteld op acht PAK's in korrels voor gebruik als infill materiaal in kunstgrasvelden en van korrels of mulches in losse vorm op speelplaatsen of in sporttoepassingen.

(PE) Op basis van die conclusies en beoordelingen heeft Nederland (hierna de "indiener van het dossier" genoemd) op 17 september 2018 bij het Agentschap een bijlage XV-dossier ingediend waarin voor acht pak's in granulaat een beperking werd voorgesteld voor gebruik als instrooimateriaal voor kunstgrasvelden en van granulaat of mulch in losse vorm voor speelplaatsen of in sporttoepassingen.

Split sentence

Definition A (long) MT output sentence has been split up into multiple sentences.

e.g.:

(ST) They apply in the EU single market, without discrimination, including to those online intermediaries established outside of the European Union that target the single market.

(MT) Zij zijn zonder discriminatie van toepassing op de interne markt van de EU, ook voor buiten de Europese Unie gevestigde onlinetussenpersonen die zich richten op de eengemaakte markt.

(PE) De regels zijn zonder onderscheid van toepassing op de interne markt van de EU. Dat geldt ook voor onlinetussendiensten die buiten de Europese Unie zijn gevestigd en zich op de eengemaakte markt richten.

Merged sentences

Definition (Short) MT output sentences have been merged into one sentence.

Repetition

Definition The same or a very similar word or phrase to the previous occurrence of the word/phrase in the target text, has been repeated by the post-editor.

e.g.:

(ST) The recommendations are based on analysis of the state of play, and the needs and priorities for agriculture and rural areas in Belgium.

(MT) De aanbevelingen zijn gebaseerd op een analyse van de stand van zaken en de behoeften en prioriteiten voor de landbouw en de plattelandsgebieden in België.

(PE) De aanbevelingen zijn gebaseerd op een analyse van de stand van zaken, de behoeften en de prioriteiten van de landbouw en de plattelandsgebieden in België.

Reduction of repetition

Definition The same or a very similar word or phrase to the previous occurrence of the word/phrase in the MT output (possibly in a preceding sentence) has been deleted or substituted by the post-editor.

e.g.:

(ST) The recommendations address the specific economic, environmental and social objectives of the future Common Agricultural Policy and in particular the ambition and specific targets of the Farm to Fork Strategy and the Biodiversity Strategy for 2030.

(MT) De aanbevelingen hebben betrekking op de specifieke economische, ecologische en sociale doelstellingen van het toekomstige

gemeenschappelijk landbouwbeleid, en met name op de ambitie en specifieke streefdoelen van de „van boer tot bord" -strategie en de biodiversiteitsstrategie voor 2030.

(PE) Zij sluiten aan bij de specifieke economische, ecologische en maatschappelijke doelstellingen van het toekomstige gemeenschappelijk landbouwbeleid, en met name bij de ambitie en specifieke streefdoelen van de "van boer tot bord"-strategie en de biodiversiteitsstrategie voor 2030.

Coherence

Definition A change has been made relating to the coherence of the text: clarification of relationships, introduction of logical structure, etc.

e.g.:

(ST) The successful candidate selected will be offered a position as Information and Communication Officer - Dutch-language Media, at AD5 level, within the Representation of the European Commission in Belgium of the Directorate-General for Communication (DG COMM).

(MT) De geselecteerde kandidaat krijgt bij de vertegenwoordiging van de Europese Commissie in België van het directoraat-generaal Communicatie (DG COMM) een functie aangeboden als functionaris voor voorlichting en communicatie — Nederlandstalige media op AD5-niveau.

(PE) De geslaagde kandidaat die is geselecteerd, wordt bij de vertegenwoordiging van de Europese Commissie in België van het directoraat-generaal Communicatie (DG COMM) een functie aangeboden als communicatiemedewerker voor de Nederlandstalige media in rang AD5.

Conjunction

Definition A conjunction or linking word from the MT output has been altered, or an implicit relation in the source text – that may or may not have been present in the MT output – has been made explicit in the post-edited target text.

e.g.:

(ST) The Commission considers that the Italian legislation, in addition to the apparent inconsistency with EU law, creates legal uncertainty for beach tourism services, discourages investment in a sector that is key to the Italian economy and already hit hard by the coronavirus pandemic while also causing a potentially significant loss of income for local Italian authorities.

(MT) De Commissie is van mening dat de Italiaanse wetgeving, naast de kennelijke strijdigheid met de EU-wetgeving, rechtsonzekerheid creëert voor strandtoeristische diensten, investeringen ontmoedigt in een sector die van cruciaal belang is voor de Italiaanse economie en reeds hard door de coronapandemie is getroffen, maar ook tot een mogelijk aanzienlijk inkomensverlies voor de lokale Italiaanse autoriteiten leidt.

(PE) De Commissie is van mening dat de Italiaanse wetgeving, naast de kennelijke strijdigheid met de EU-wetgeving, leidt tot rechtsonzekerheid voor strandtoerismediensten, investeringen ontmoedigt in een sector die van cruciaal belang is voor de Italiaanse economie en reeds hard is getroffen door de coronapandemie, en tevens leidt tot een mogelijk aanzienlijk inkomensverlies voor de lokale Italiaanse overheden.

Other stylistic change

Definition Any other stylistic change or register change to the MT output.

1.3.4 Spelling & Punctuation Parameter

Definition Categorisation of a change relating to spelling or punctuation that has been made to the MT output by the post-editor.

Capitalisation

Definition A lower-case letter in the MT output has been substituted for a capital letter or vice versa.

e.g.: (ST) Be (1) part of the Media team of the Representation (2) (currently consisting of 1 additional person) and work closely with the other teams of the Representation;

(MT) Deel (1) uitmaken van het mediateam van de Vertegenwoordiging (2) (momenteel bestaat uit 1 extra personen) en nauw samenwerken met de andere teams van de vertegenwoordiging;

(PE) deel (1) uitmaken van het mediateam van de vertegenwoordiging (2) (dat momenteel bestaat uit één bijkomende persoon) en nauw samenwerken met de andere teams van de vertegenwoordiging;

Compound

Definition A compound has been altered (deletion or insertion of a space, deletion or insertion of a hyphen, etc.). A compound can be a noun as well as an adjective, a verb or a preposition.

e.g.: (ST) The economic situation of farmers varies, with the level of income differing substantially according to physical farm size, sector, and location.

(MT) De economische situatie van de landbouwers varieert, waarbij het inkomensniveau aanzienlijk verschilt naar gelang van de omvang van het landbouwbedrijf, de sector en de locatie.

(PE) De economische situatie van de landbouwers varieert, waarbij het inkomensniveau aanzienlijk verschilt naargelang van de fysieke omvang van het landbouwbedrijf, de sector en de locatie.

Punctuation

Definition A change has been made to the punctuation in the MT output; punctuation marks have been added, deleted or substituted.

e.g.: (ST) “socialist economy”

(MT) „socialistische economie”

(PE) “socialistische economie”

e.g.: (ST) As mentioned in recital (23)

(MT) Zoals vermeld in overweging (23)

(PE) Zoals vermeld in overweging 23

Numeral

Definition A numeral from the MT output written as a number has been written in full or vice versa.

e.g.: (ST) By the deadline for submission of applications, and in addition to the diploma and when applicable to the one year experience required above,

candidates must have at least 2 years' full time professional experience gained after obtaining the diploma required under Point 3.3.1 directly related to the duties as described under Point 1 above.

(MT) Uiterlijk op de uiterste datum voor de indiening van sollicitaties moeten kandidaten, naast het diploma en, indien van toepassing, de hierboven vereiste werkervaring van één jaar, ten minste 2 jaar voltijdse beroepservaring hebben opgedaan na het behalen van het in punt 3.3.1 bedoelde diploma dat rechtstreeks verband houdt met de in punt 1 beschreven taken.

(PE) Uiterlijk op de uiterste datum voor de indiening van sollicitaties moeten kandidaten, naast het diploma en, indien van toepassing, de hierboven vereiste werkervaring van één jaar, na het behalen van het in punt 3.3.1 bedoelde diploma, ten minste twee jaar voltijdse werkervaring die rechtstreeks verband houdt met de in punt 1 beschreven taken hebben opgedaan.

Other spelling change

Definition Any type of spelling change other than the above-mentioned changes, has been made to the MT output.

e.g.: (ST) management practices
(MT) beheerspraktijken
(PE) beheerpraktijken

2 Schematic Overviews and Annotation Codes

2.1 Translation Quality Parameter

Translation Quality Parameter				
Style guide changes	Consistency & intertextuality changes	MT errors	Preferential changes	Undesirable changes
SG	CON	MT ERROR	PREF	UNDES

2.2 Translation Norm Parameter

Translation Norm Parameter		
Adequacy	Acceptability	Adequacy + acceptability
AD	ACC	AD+ACC

2.3 Text-Linguistic Characteristics

Semantics Parameter	Addition	Addition of source text information missing from the MT output	ADD ST INFO
		Explicitation of ST and MT output	EXPLIC
		Hyponymy	HYPONYMY
		Other addition	OTH ADD
	Deletion	Deletion of ST and MT information	DEL LEX INFO
		Deletion of extra information in MT output	DEL EXTRA INFO
		Hyperonymy	HYPERONYMY
	Other meaning shift	Action	AC
		Time	T
		Place	P
		Agent	AG
		Modality	MOD
		Other meaning change	OTH MEAN CH
Syntax & Morphology Parameter	Order	Internal phrase order	PHRASE ORDER
		Internal sentence order	SENT ORDER
	Agreement	Article-noun agreement	ART-NOUN
		Noun-adjective agreement	NOUN-ADJ
		Subject-verb agreement	SUBJ-VERB
		Reference	REF
	Structural change	Nominalisation	NOM
		Reduction of nominalisation	RED OF NOM
		Active → passive	ACT-PAS
		Passive → active	PAS-ACT
		Other structural change	OTH STRUCT CH
	Morphological change	Comparative/superlative of an adjective/adverb	COMP/SUPERL
		Singular/plural noun	SING/PLUR
		Named entity	NAMED ENT
		Other morphological change	OTH MORPH CH
Style & Register Parameter	Synonymy		SYNONYMY
	Preposition		PREP
	Lexical change		LEX CH
	Register		REG
	Untranslated text in MT		UNTRANS IN MT
	Split sentence		SPLIT SENT

	Merged sentences	MERG SENT
	Repetition	REP
	Reduction of repetition	RED OF REP
	Coherence	COH
	Conjunction	CONJ
	Other stylistic change	OTH STYL CH
Spelling & Punctuation Parameter	Capitalisation	CAP
	Compound	COMP
	Punctuation	PUNCT
	Numeral	NUMERAL
	Other spelling change	OTH SPEL CH

APPENDIX B DECISION TREE FOR THE TEXT-LINGUISTIC ANNOTATION

Note: Proceed to the next parameter as soon as a parameter has been annotated or appears to be inapplicable to the post-edit in question.

Semantics Parameter

1. Does the post-edit bring about a difference in meaning between the target text and the MT output and/or source text?
   Yes: **Semantics Parameter** → go to 2
   No: Go to Syntax & Morphology Parameter (page 95)

2. Has a MT output word or phrase been substituted for a hyponym/hypernym of that word/phrase? Has, in other words, information been made more specific (hyponym) or more general (hypernym)?
   Yes, more specific: **Hyponymy**
   Yes, more general: **Hyperonymy**
   No: Go to 3

3. Has information been added, deleted or changed to the MT output by the post-editor?
   Added: **Addition** → go to 4
   Deleted: **Deletion** → got to 7
   Changed: **Other meaning shift** → go to 9

4. Is the added information source text information that was not present in the MT output?
   Yes: **Addition of source text information missing from the MT output**
   No: Go to 5

5. Is the added information source text information that is implicit in both the source text and the MT output? Has, in other words, implicit source text information been made explicit by the post-editor?
   Yes: **Explicitation of ST and MT output**
   No: Go to 6

6. Is the added information meaningful information that cannot be derived from the source text and was not expressed in the MT output?
   Yes: **Other addition**
   No: Go to Syntax & Morphology Parameter (page 95)

7. Has information been deleted from the MT output by the post-editor?
   Yes: **Deletion** → go to 8
   No: Go to Syntax & Morphology Parameter (page 95)
8. Was the deleted information a relational or other meaningful element that was present in both the source text and the MT output?
   Yes: **Deletion of ST and MT lexical information**
   No: **Deletion of extra information in MT output**

9. Which of the following meaningful elements has been changed?
   - **Action (Ac)**
   - **Time (T)**
   - **Place (P)**
   - **Agent (incl. Object) (Ag)**
   - **Modality (Mod)**
   - None of the above: **Other meaning change**
   - Multiple of the above → go to 10

10. Any of the following combinations of various meaningful elements is possible as well:
    - **Ac + T**
    - **Ac + P**
    - **Ac + Ag**
    - **Ac + T + P**
    - **Ac + T + Ag**
    - **Ac + T + P + Ag**
    - **T + P**
    - **T + Ag**
    - **T + P + Ag**
    - **P + Ag**
    - **Ac + T + Mod**
    - **Ac + P + Mod**
    - **Ac + Ag + Mod**
    - **Ac + T + P + Mod**
    - **Ac + T + Ag + Mod**
    - **Ac + T + P + Ag + Mod**
    - **T + P + Mod**
    - **T + Ag + Mod**
    - **T + P + Ag + Mod**
    - **P + Ag + Mod**

Syntax & Morphology Parameter

1. Has a syntactical change been made to the MT output?
   Yes: **Syntax & Morphology Parameter** → go to 2
   No: Go to Style & Register Parameter (page 97)

2. Has the MT sentence order been altered by the post-editor?
   Yes: **Order** → go to 3
   No: Go to 4
3. Has the word order within a single phrase (e.g. a verb phrase) in the MT output been altered?
   Yes: **Internal phrase order**
   No: **Internal sentence order**

4. Has a change in terms of agreement been made to the MT output?
   Yes: **Agreement** → go to 5
   No: Go to 9
5. Has an article from the MT output been substituted or has an article been added or deleted, without that causing a meaning shift?
   Yes: **Article-noun agreement**
   No: Go to 6
6. Has an adjective or noun from the MT output been adjusted with the objective of having it agree with the relevant adjective/noun?
   Yes: **Noun-adjective agreement**
   No: Go to 7
7. Has a verb or its subject been adjusted with the objective of having it agree in number with its subject?
   Yes: **Subject-verb agreement**
   No: Go to 8
8. Has a referring expression from the MT output been adjusted or substituted?
   Yes: **Reference**
   No: Go to 9

9. Has a structural change been made to the MT output?
   Yes: **Structural change** → go to 10
   No: Go to 15

10. Has a verb phrase from the MT output been substituted for a nominalisation of the main verb in that verb phrase?
    Yes: **Nominalisation**
    No: Go to 11
11. Has a noun phrase from the MT output been substituted by a verb phrase with the nominalised verb in that noun phrase as its main verb?
    Yes: **Reduction of nominalisation**
    No: Go to 12
12. Has an active construction in the MT output been converted to a passive construction?
    Yes: **Active → passive**
    No: Go to 13
13. Has a passive construction in the MT output been converted to an active construction?
    Yes: **Passive → active**
    No: Go to 14
14. Has any other structural change been made to the MT output?
    Yes: **Other structural change**
    No: Go to 15

15. Has a morphological change been made to the MT output?
    Yes: **Morphological change** → go to 16
    No: Go to Style & Register Parameter (page 97)
16. Has the structure or form of the comparative or superlative of an adjective or adverb from the MT output been changed?
    Yes: **Comparative/superlative of an adjective/adverb**
    No: Go to 17
17. Has a plural form of a noun in the MT output been changed to a singular form or vice versa, excluding agreement changes?
    Yes: **Singular/plural noun**
    No: Go to 18
18. Has a Dutch named entity (geographical location, name, company, organization, etc.) from the MT output been altered? This includes (partially) untranslated named entities that have an official Dutch equivalent, (partially) translated named entities that do not exist in Dutch, etc.
    Yes: **Named entity**
    No: **Other morphological change**

Style & Register Parameter

1. Has a register change or a stylistic change been made to the MT output?
   Yes: **Style & Register Parameter** → go to 2
   No: Go to Spelling & Punctuation Parameter (page 99)

2. Has a word or phrase from the MT output been substituted for a synonym of that word/phrase?
   Yes: **Synonymy**
   No: Go to 3

3. Has a preposition from the MT output been substituted or has a preposition been added or deleted, without that causing a meaning shift?
   Yes: **Preposition**
   No: Go to 4

4. Has a word or phrase from the MT output been substituted for another word/phrase that is not synonymous with the MT output word/phrase, without that causing a meaning shift?
   Yes: **Lexical change**
   No: Go to 5

5. Has a word or phrase in the MT output been altered so that it would still convey the same meaning but belong to a more formal/informal/... register or a regional variety of the language that is more suitable for the target audience?
   Yes: **Register**
   No: Go to 6

6. Has a word or phrase of which a Dutch translation exists but that had been left untranslated in the MT output (i.e. for which the MT system had copied the English word/phrase) been translated?
   Yes: **Untranslated text in MT**
   No: Go to 7

7. Has a (long) MT output sentence been split up into multiple sentences?
   Yes: **Split sentence**
   No: Go to 8

8. Have (short) MT output sentences have been merged into one sentence?
   Yes: **Merged sentences**
   No: Go to 9

9. Has the same or a very similar word or phrase to the previous occurrence of the word/phrase in the target text been repeated by the post-editor?
   Yes: **Repetition**
   No: Go to 10

10. Has the same or a very similar word or phrase to the previous occurrence of the word/phrase in the MT output been deleted by the post-editor?
    Yes: **Reduction of repetition**
    No: Go to 11
11. Has change been made relating to the coherence of the text: clarification of relationships, introduction of logical structure, etc.?
    Yes: **Coherence**
    No: Go to 12
12. Has a conjunction or linking word that had expressed a different relation in the MT output than in the source text been altered, or has an implicit relation in the source text – that may or may not have been present in the MT output – been made explicit in the post-edited target text?
    Yes: **Conjunction**
    No: **Other stylistic change**

Spelling & Punctuation Parameter

1. Has a change relating to spelling or punctuation been made to the MT output?
   Yes: **Spelling & Punctuation Parameter** → go to 2
   No: You have finished your annotation of this item ☺

2. Which of the following changes has been made?
   - **Capitalisation (Cap)**
   - **Compound (Comp)**
   - **Numeral (Num)**
   - **Punctuation (Punct)**
   - **Cap + Comp**
   - **Cap + Punct**
   - **Comp + Punct**
   - **Num + Punct**

APPENDIX C POST-EDIT ANALYSIS

Excel spreadsheet appendix

APPENDIX D LIST OF UNDESIRABLE CHANGES

Note: The whole of the post-edit analysis was carried out to the best of the annotator's ability, yet without (comprehensive) inter-annotator agreement nor the extensive knowledge possessed by DGT translators. A number of (especially undesirable) post-edits may therefore be considered at the very least dubious. For each of the post-edits categorised as undesirable change below, both the full source text and machine translated sentences have been included, as well as a brief substantiating comment.

1.	T1	(ST)	The CCC also **raised** the subject of the role of SOEs in certain Chinese markets – state presence and market distortions, of land use rights, of alleged discrimination by policies or measures and of labour costs.
		(MT)	Het CCC heeft ook **gewezen op** de rol van staatsondernemingen op bepaalde Chinese markten — aanwezigheid van de staat en marktverstoringen, grondgebruiksrechten, vermeende discriminatie door beleid of maatregelen en arbeidskosten.
		(PE)	De CCC **verwees** ook **naar** de kwestie van de rol van staatsondernemingen in bepaalde Chinese markten – overheidsdeelneming en marktverstoringen, grondgebruiksrechten, vermeende discriminatie door beleid of maatregelen en loonkosten;
	Comment		Change of ACTION: weaker form. Raise the subject = something definitely needed to be discussed, not simply referred to.
2.	T1	(ST)	In their comments on final disclosure, the CCC reiterated its arguments presented in recital (53), **submitting** that Article 2(6a) of the basic Regulation - unilaterally created by the EU - runs beyond the existing legal provisions of ADA which does not bear a corresponding provision.
		(MT)	In zijn opmerkingen over de mededeling van de definitieve bevindingen herhaalde het CCC zijn argumenten in overweging (53) en **stelde het dat** artikel 2, lid 6 bis, van de basisverordening — dat unilateraal door de EU is ingesteld — verder gaat dan de bestaande wettelijke bepalingen van de ADA, die geen overeenkomstige bepaling bevatten.
		(PE)	In zijn opmerkingen over de mededeling van de definitieve bevindingen herhaalde de CCC zijn argumenten uit overweging 53 en **stelde** daarbij **dat** artikel 2, lid 6 bis, van de basisverordening – dat unilateraal door de EU is ingesteld – verder gaat dan de bestaande wettelijke bepalingen van de WTO-ADO, waar geen overeenkomstige bepaling in opgenomen is.
	Comment		A grammatical error was introduced in the target text.
3.	T1	(ST)	In their comments on final disclosure, the CCC remarked that the Commission **considers** socialist market economy under which the state-owned economy takes a predominant position in China and the CCP

			reinforcing such position as a reason to apply a different methodology of determining normal value.
		(MT)	In haar opmerkingen over de mededeling van de definitieve bevindingen merkte de CCC op dat de Commissie een socialistische markteconomie **beschouwt** waarin de staatseconomie een overheersende positie inneemt in China en de CCP, als reden om een andere methode voor het bepalen van de normale waarde toe te passen.
		(PE)	In zijn opmerkingen over de mededeling van de definitieve bevindingen merkte de CCC op dat de Commissie het bestaan in China van een socialistische markteconomie, waarin de staatseconomie een overheersende positie inneemt die door de CCP verder wordt versterkt, als reden **beschouwd** om een andere methode voor het bepalen van de normale waarde toe te passen.
	Comment		Incorrect verb tense, should be: *de Commissie beschouwt X als reden*.
4.	T2	(ST)	In the framework of the structured dialogue on the preparation of the **CAP strategic plan**, this document contains the recommendations for the CAP strategic plan of Belgium.
		(MT)	In het kader van de gestructureerde dialoog over de voorbereiding van het strategisch **GLB-plan** bevat dit document de aanbevelingen voor het strategisch GLB-plan van België.
		(PE)	Dit document, dat wordt voorgelegd in het kader van de gestructureerde dialoog over het opstellen van de strategische **GLB-plannen**, bevat de aanbevelingen voor het strategisch GLB-plan van België.
	Comment		CAP strategic plan in singular is nowhere else is the TM translated by a plural form in Dutch, only plural by plural and singular by singular.
5.	T2	(ST)	Agriculture in Belgium is characterised by **a share of agricultural entrepreneurial income** which, at close to 60% of the whole economy's average wage for 2005-2018, is above the EU-average.
		(MT)	De landbouw in België wordt gekenmerkt door **een aandeel van het inkomen uit bedrijfsuitoefening** in de landbouw dat bijna 60 % van het gemiddelde loon van de gehele economie in de periode 2005-2018 boven het EU-gemiddelde ligt.
		(PE)	De landbouw in België wordt gekenmerkt door **een_inkomen uit de landbouwbedrijfsuitoefening** dat, met bijna 60 % van het gemiddelde loon van de gehele economie in de periode 2005-2018, boven het EU-gemiddelde ligt.
	Comment		Slight meaning change compared to the ST and MT output as a result of deletion of source text information.
6.	T2	(ST)	Similarly, while the agricultural factor income per **worker** is well above the EU average, its evolution is following a slightly negative trend because expenses are growing faster than revenue.

		(MT)	Ook al ligt het landbouwfactorinkomen per **werknemer** ruim boven het EU-gemiddelde, dan volgt de ontwikkeling ervan een licht negatieve trend omdat de uitgaven sneller stijgen dan de inkomsten.
		(PE)	Ook bij het agrarisch factorinkomen per **werkende**, dat nochtans ruim boven het EU-gemiddelde ligt, is er een licht negatieve trend omdat de uitgaven sneller stijgen dan de inkomsten.
	Comment		IATE: worker = *werknemer* Causes a slight meaning shift, or at least confusion.
7.	T3	(ST)	The Dossier Submitter took into consideration various exposure scenarios related to the use of granules in synthetic turf pitches, by workers installing and maintaining the pitches and by individuals playing sports on them (professional and amateur outfield players and goalkeepers), and related to granules or mulches used in loose applications on playgrounds and in sport applications, where people and especially children may be exposed.
		(MT)	De Dossier Submitter hield rekening met verschillende blootstellingsscenario's in verband met het gebruik van granulaten in kunstgrasvelden, door werknemers die de velden installeren en onderhouden en door personen die sporten op hen uitoefenen (professionele en amateursporten en goalkekers), en die verband hielden met granulaten of mulches die worden gebruikt in losse toepassingen op speelplaatsen en in sporttoepassingen, waarbij mensen en met name kinderen kunnen worden blootgesteld.
		(PE)	De indiener van het dossier heeft in verband met het gebruik van granulaat rekening gehouden met verschillende blootstellingsscenario's, waaronder voor werknemers die kunstgrasvelden aanleggen en onderhouden en mensen die daarop een sport beoefenen (professionele en amateurbuitenveldspelers en -doelverdedigers), of voor **mensen en met name kinderen_kunnen worden blootgesteld** aan granulaat dat of mulch die in losse vorm voor speelplaatsen en in sporttoepassingen wordt gebruikt.
	Comment		Relative pronoun *die* (= that) is missing.
8.	T5	(ST)	The proposal is well-positioned to set a global regulatory benchmark, as it sets the highest standards for effective intervention, for **due process** and the protection of fundamental rights online, preserves a balanced approach to the liability of intermediaries, and establishes effective measures for tackling illegal content and societal risks online.
		(MT)	Het voorstel is goed geplaatst om een mondiale benchmark voor regelgeving vast te stellen, aangezien het de hoogste normen vaststelt voor doeltreffend optreden, een eerlijke **rechtsgang** en de bescherming van de grondrechten online, een evenwichtige aanpak van de aansprakelijkheid van tussenpersonen handhaaft en doeltreffende maatregelen vaststelt om illegale online-inhoud en maatschappelijke risico's aan te pakken.
		(PE)	Het voorstel is goed geplaatst om een wereldwijde benchmark voor regelgeving te bepalen, aangezien het de hoogste normen vastlegt voor

			doeltreffend optreden, eerlijke **rechtsbedeling** en de bescherming van de grondrechten online, de aansprakelijkheid van tussendiensten evenwichtig benadert en doeltreffende maatregelen vaststelt om illegale online-inhoud en risico's voor de maatschappij aan te pakken.
	Comment		Van Dale: (due) process = (behoorlijke) *rechtsgang*. However, in IATE: both terms. → dubious
9.	T5	(ST)	At the same time, citizens will be able to notify illegal content or products they encounter and contest the decisions made by **online platforms** when their content is removed:
		(MT)	Tegelijkertijd zullen burgers illegale inhoud of producten die zij tegenkomen, kunnen melden en de beslissingen van **onlineplatforms** kunnen aanvechten wanneer hun inhoud wordt verwijderd:
		(PE)	Tegelijkertijd zullen burgers illegale inhoud of producten die zij tegenkomen kunnen melden en de beslissingen van **onlineplatformen** aanvechten wanneer hun eigen inhoud wordt verwijderd:
	Comment		Plural used in TM: *onlineplatforms.* Was changed to *onlineplatformen* in all PE.
10.	T5	(ST)	At the same time, citizens will be able to notify illegal content or products they encounter and contest the decisions made by online platforms when their content is removed:
		(MT)	Tegelijkertijd zullen burgers illegale inhoud of producten die zij tegenkomen**,** kunnen melden en de beslissingen van onlineplatforms kunnen aanvechten wanneer hun inhoud wordt verwijderd:
		(PE)	Tegelijkertijd zullen burgers illegale inhoud of producten die zij tegenkomen**_**kunnen melden en de beslissingen van onlineplatformen aanvechten wanneer hun eigen inhoud wordt verwijderd:
	Comment		Grammatical error: deletion of comma between two finite verbs.
11.	T5	(ST)	In addition, specific rules will be introduced for very large **platforms** given their systemic impact in facilitating public debate, economic transactions and the dissemination of information, opinions and ideas.
		(MT)	Daarnaast zullen specifieke regels worden ingevoerd voor zeer grote **platforms**, gezien hun systemische impact op het bevorderen van het publieke debat, economische transacties en de verspreiding van informatie, meningen en ideeën.
		(PE)	Voor zeer grote **platformen** zullen er daarnaast specifieke regels worden ingevoerd, gezien hun wezenlijke impact op het openbaar debat, economische transacties en de verspreiding van informatie, meningen en ideeën.
	Comment		Plural used in TM: *onlineplatforms.* Was changed to *onlineplatformen* in all PE.

12.	T5	(ST)	In addition, specific rules will be introduced for very large platforms given their systemic impact in **facilitating** public debate, economic transactions and the dissemination of information, opinions and ideas.
		(MT)	Daarnaast zullen specifieke regels worden ingevoerd voor zeer grote platforms, gezien hun systemische impact op **het bevorderen van** het publieke debat, economische transacties en de verspreiding van informatie, meningen en ideeën.
		(PE)	Voor zeer grote platformen zullen er daarnaast specifieke regels worden ingevoerd, gezien hun wezenlijke impact op het openbaar debat, economische transacties en de verspreiding van informatie, meningen en ideeën.
	Comment		Deletion of ST information.

13.	T6	(ST)	Represent the Commission **vis-à-vis** members of the press in press conferences and meetings on the basis of the instructions from the Spokesperson service or Directorates A and B of DG COMM, as appropriate;
		(MT)	De Commissie vertegenwoordigen **ten aanzien van** de leden van de pers op persconferenties en vergaderingen op basis van de instructies van de dienst woordvoerder of de directoraten A en B van DG COMM, naargelang het geval;
		(PE)	de Commissie vertegenwoordigen **ten overstaan van** de leden van de pers op persconferenties en vergaderingen op basis van de instructies van de Dienst van de woordvoerder of de directoraten A en B van DG COMM, naargelang het geval;
	Comment		Van Dale: *ten overstaan van* = non-standard equivalent to *ten aanzien van*.

14.	T7	(ST)	[3] For instance, the EU Sure Farm project defines resilience of a farming system as its "ability to ensure the provision of the system functions in the face of increasingly complex and **accumulating** economic, social, environmental and institutional shocks and stresses, through capacities of robustness, adaptability and transformability" (Meuwissen et al., 2019).
		(MT)	[3] zo definieert het EU Sure Farm project veerkracht van een landbouwsysteem als zijn vermogen om ervoor te zorgen dat het systeem functioneert in het licht van steeds complexere en **toenemende** economische, sociale, ecologische en institutionele schokken en stress, door middel van robuustheid, aanpassingsvermogen en transformability (Meuwissen et al., 2019).
		(PE)	[3] Zo definieert het EU-project SURE-Farm de veerkracht van een landbouwsysteem als het vermogen om ervoor te zorgen dat het systeem functioneert ten aanzien van steeds complexere en **opeenvolgende** economische, sociale, ecologische en institutionele schokken en spanningen door middel van robuustheid, aanpassingsvermogen en veranderbaarheid (Meuwissen et al., 2019).
	Comment		Meaning shift compared to ST information.

15.	T7	(ST)	Use of information technology for trading (for example, **direct** sales online, or platforms to connect suppliers with excess food with food banks or other organisations)
		(MT)	Gebruik van informatietechnologie voor de handel (bijvoorbeeld **rechtstreekse** onlineverkoop of platforms om leveranciers met voedseloverschotten in contact te brengen met voedselbanken of andere organisaties)
		(PE)	Gebruik van informatietechnologie voor de handel (bijvoorbeeld **directe** onlineverkoop of platforms om leveranciers met voedseloverschotten in contact te brengen met voedselbanken of andere organisaties)
	Comment		IATE: direct sales = *rechtstreekse verkoop.*