UNIVERSITAT ROVIRA I VIRGIL

POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION: SUBJECT-MATTER EXPERT EXPERTS VERSUS PROFESSIONAL TRANSLATORS

Özlem Temizöz

Dipòsit Legal: T.58-2014

UNIVERSITAT ROVIRA I VIRGILI
POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION: SUBJECT-MATTER EXPERTS VERSUS PROFESSIONAL TRANSLATORS
Özlem Temizöz
ISBN: DL:

ÖZLEM TEMİZÖZ

POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION:

Subject-Matter Experts versus Professional Translators

DOCTORAL THESIS

UNIVERSITAT ROVIRA I VIRGILI

Tarragona
2013

ISBN: DL:

UNIVERSITAT ROVIRA I VIRGILI
POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION: SUBJECT-MATTER EXPERTS VERSUS PROFESSIONAL TRANSLATORS
Özlem Temizöz
ISBN: DL:

UNIVERSITAT ROVIRA I VIRGILI
POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION: SUBJECT-MATTER EXPERTS VERSUS PROFESSIONAL TRANSLATORS
Özlem Temizöz
ISBN: DL:

ÖZLEM TEMİZÖZ

POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION:

Subject-Matter Experts versus Professional Translators

DOCTORAL THESIS

Supervised by Professor Anthony Pym

Department of English and German Studies

Intercultural Studies Group

UNIVERSITAT ROVIRA I VIRGILI

Tarragona
2013

ISBN: DL:

Abstract

The present research aims to explore the specific characteristics of the postediting processes performed by subject-matter experts as opposed to professional translators. We compare the ways engineers and professional translators postedit a technical text, in terms of processing speed, time spent on documentation and the number of changes. We also compare the texts postedited by engineers and professional translators with regard to quality, using LISA QA Model 3.1. Further, we explore which of the following workflows is faster and produces outputs of higher quality: Postediting MT output by engineers and revising the postedited text by professional translators, or vice-versa. We conducted two sets of experiments that were screen-recorded with BB Flashback. First, ten professional translators and ten engineers postedited a 482-word text on the dismantling of end-of-life vehicles; the text had previously been translated by Google Translate. Then they filled out post-assignment questionnaires. Second, each group revised the posteditings performed by the other group. The findings suggest that expertise and experience in the subject-matter are the main factors determining postediting quality. When the recurrent errors are penalized and the outlier participants were removed, the engineers' postediting of technical texts is of significantly higher quality than the translators'. When postediting, the translators' and the engineers' processing speed did not differ significantly. For technical texts, the quality improvement (added value) brought about by engineer-revision of translator-postediting is higher than vice-versa. However, this added value is significant only when the recurrent errors are penalized. Further, the quality of both the postedited texts and their revised versions (either performed by professional translators or engineers) changes significantly as a result of penalizing and unpenalizing recurrent errors.

Keywords

Postediting, machine translation, professional translators, subject-matter experts, processing speed, translation quality, recurrent errors, technical translations, LISA QA Model 3.1., Google Translate, documentation, self-revision.

UNIVERSITAT
ROVIRA I VIRGILI

Professor Anthony Pym
URV. Av. Catalunya 35
43002 Tarragona, Spain
anthony.pym@urv.cat

July 2, 2013

I hereby certify that the present study *Postediting Machine Translation Output And Its Revision: Subject-Matter Experts versus Professional Translators*, presented by Özlem Temizöz for the award of the degree of Doctor, has been carried out under the supervision of myself at the Universitat Rovira i Virgili and that it fulfills the requirements for the mention "International Doctorate".

Professor Anthony Pym
Intercultural Studies Group
Universitat Rovira i Virgili
Tarragona, Spain

President
European Society for Translation Studies

ISBN: DL:

ISBN: DL:

Acknowledgments

My most sincere thanks go to Professor Anthony Pym, for this PhD program which provided me with the elusive opportunity to take part in a dynamic learning environment enriched with the leading figures in Translation Studies and fellow students from around the world. My learning experience throughout the program have influenced me profoundly. Moreover, I was privileged to have Dr Pym as my superviser, who not only read the numerous drafts of the thesis and commented on every single detail but also gave continuous guidance and support during the entire program.

I wish to thank Dr Lynne Bowker for her support providing the literature and to Dr Tina Paulsen Christensen for devoting time to my questions on her research. I would also like to thank Esther Torres Simón and David Orrego Carmona for their kind assistance regarding formatting and printing the thesis.

I am grateful to my friend Christy Fung-Ming Liu for her willingness to help in getting access to the literature and being there whenever I need her support and encouragement. I am also grateful to Courtney Parkins for his encouragement and friendship during the hard times of this endeavor. My gratitudes also go to the participants for making this study possible.

Last but not least, my special thanks go to my husband, Ayhan, for sharing pain and pleasure of this study with me and to my dearest son, Alp, who started studying with me on this research from the very beginning of his life! Without their support and love, this work would not have existed.

Table of Contents

ISBN: DL:

ISBN: DL:

ISBN: DL:

List of Tables

ISBN: DL:

ISBN: DL:

ISBN: DL:

List of Figures

ISBN: DL:

1. Introduction

Globalization and rapid development in technology have drastically changed the way people live, communicate and work. Globalization forces enterprises to jump borders and do business with international partners to survive in competing markets. Breaking into new and remote markets urges them to supply more goods and services faster in various parts of the world simultaneously. Allen (2003: 300) notes that many successful software companies are now "working in a simultaneous shipping ('simship') mode whereby documentation in the source language plus a core group of target languages are all shipped at the same time."

Allen (2003: 299) points to the expansion of businesses of various sizes to the four corners of the earth and states that "it is no longer possible to rely on local business, nor to base one's commercial expectations on a single language as the sole medium of communication." Therefore, a company aiming to be an international actor has to remove language barriers in order to increase the accessibility of its goods and services to wider audiences. The road to such an internationalization process usually goes through using English as a lingua franca. However, using translation in international business has grown exponentially due to the rightful preference of people to receive information in their own languages.

Ever-increasing technological developments have not only improved work processes but they have also increased productivity. The Translation profession has also received its share of these developments. Translation is no longer an act carried out by translators confined to their tables and rendering paper-based source texts to create paper-based target texts using a pen or a typewriter. The real involvement of technology in translators' daily life dates back to the 1980s when the first personal computers came into use.

As of the 1990s, the Internet became widespread and the translation of digital content for web-pages appeared. The Internet affected the way translators communicate with their clients and colleagues, do business (receive and send their translations) and get access to information sources (on-line dictionaries, web-pages, forums, etc.). Besides, widespread use of the Internet worldwide has created a tremendous content which brings together a big potential for translation.

Due to the close interaction of the abovementioned factors, demand for

ISBN: DL:

translations with shorter turnaround time has grown. Human resources have fallen short of covering the demand. Therefore, the need to increase translation productivity by also ensuring acceptable level of quality has become an important issue. This has brought about the development and proliferation of "text reuse" technologies. It is no exaggeration to say that translation memories (TM) have broken new ground by enabling reuse of previously translated texts, thereby increasing productivity. More recently, machine translation (MT) has been incorporated into translator workstations as a complementary component to translation memories when parts of the source text cannot be translated using the available TM database. This hybrid medium of translation has affected the way translators interact with the texts, changing their role from translators to posteditors. Translators are increasingly asked to postedit the texts on TM/MT incorporated workstations instead of translating source texts from scratch.

For similar reasons, the demand for translations is usually more than can be covered by full-time translators alone. Moreover, texts to be translated often require expertise in a specific field. This creates a demand for specialists (translators specialized in a particular field) in the translation of specific type of text, which in turn exacerbates the problem of the scarcity of translators. Hence we find a growing tendency to incorporate field experts (engineers, physicians, etc.) or volunteers in NGOs, NPOs and in social network groups into the translation process.

The involvement of volunteers in the translation profession is not a new phenomenon. Translation is a relatively new profession, and in the past, demand for translation was covered by bilinguals who could function as mediators between languages and cultures. As the practice of translation became a profession some decades ago, those bilinguals (or multilinguals) were replaced by professional translators. Nowadays, when the Internet has become part of almost everyone's life, and social network sites such as Facebook and Twitter are embedded into societies so powerfully that they can even affect political systems, the involvement of volunteers into the translation practice has returned and seems to set to increase in the near future. The practice of translation by non-professionals is labeled in various ways. Pym (2011a) lists some of them as "volunteer translation", "community translation", "crowdsourcing", "collaborative translation, "untrained translation", "paraprofessional translation" and "user-generated translation" (for comments on these terms, see Pym 2011b).

ISBN: DL:

Within those non-professional translator groups, our focus here is on those who are specialized in one subject matter (i.e. engineering) and have received no formal training in translation. They do not carry out translation regularly; however, they may be requested to do so when the need arises either as part of their daily work or for a client requesting translation. For example, for texts requiring specialized field knowledge, language service providers may prefer to work with field experts proficient in the source and target languages. In such scenarios, engineers translate technical texts in the engineering domain, physicians translate medical texts and people with a law background are preferred to translate legal texts. The demand for translation in these and other specific fields may also be covered by professional translators specialized in one or more subject areas. However, as discussed above, subject-matter experts are also among those providing specialized translation. In some cases, these subject-matter experts revise the work of professional translators to polish the texts in terms of specialized knowledge or professional translators revise the texts translated by these subject-matter experts in order to improve linguistic features and style of the translations.

In order to increase throughput, project managers, companies or language service providers managing the translation projects may ask these subject-matter experts, volunteers or professional translators to postedit TM/MT output rather than translating the source texts from scratch. Although we can predict that the processes carried out by these different subjects will differ in various ways, we do not know how they differ in detail in terms of postediting process and end-product. The present research attempts to fill this gap.

With the incorporation of subject-matter experts and volunteers in the translation process, we find various types of workflows including but not limited to volunteers translating a text and professional translators revising it, subject-matter experts translating a text and professional translators revising it, or vice-versa. To our knowledge, no empirical research has been carried out so far to explore which process is faster and yields final products of higher quality. The present research will also explore this.

ISBN: DL:

2. Overview of Text-Reuse Technologies

2.1. Translation Memories (TM)

In addition to generating excessive amounts of work for translators, information technology provided new tools for increasing their productivity. The use of translation memories has made language service providers and translators change the process, pricing and productivity of translation services.

Basically, a translation memory is a tool that stores previous translations aligned with their source texts in segments on a database. It retrieves those target segments when exactly the same (full match) or similar (fuzzy match) source text segments are introduced into the system during the translation process. The segments can be composed of phrases, paragraphs or even sentences. Full matches correspond to 100 percent matches, while fuzzy matches may range from 1 to 99 percent. The source-text segments that are not matched with any target segments are called “no match”.

García (2009: 201) points out some of the characteristics of TMs:

> They incorporated terminology management systems, alignment, and terminology extraction tools, then quality control (QC) and project management features, and eventually the capacity for batch handling of multiple files and formats, and simultaneously using several memories and glossaries.

Translation memories started to appear on the market at the beginning of the nineties and found their real use when localization activities came on the scene. In basic terms, localization is the linguistic and cultural adaptation of content regarding a particular product or a service into a particular language/culture or a locale. It is not limited to translating the content of a product or a service. It also involves engineering and managerial aspects, but it has translation in its core. Since the localized content must be supplied in multiple languages simultaneously for the launch of a product on the market, a high level of speed and productivity in translation services is required. Thus, among other computer tools aiming to increase productivity, translation memories have held their place as the most distinctive tools for translators working in the domain of localization.

Over the course of nineties, TM technology improved and became more widespread. The extensive survey on translation memories conducted by Lagoudaki

ISBN: DL:
(2006: 15) reveals that 82.5 percent of people in the translation industry, including project managers, reviewers/proofreaders, terminologists, subtitlers, translators, interpreters and other translation professionals use TM systems in their professional life.

A number of companies developed TM software to meet the demand of the translation market. Lagoudaki (2006) lists some of the most widely-used translation-memory systems: TRADOS, Wordfast, SDL Trados 2006, Déjà Vu, SDLX, Star Transit, Alchemy CATALYST, Omega-T, Logoport, and so on. Although their interfaces and use differ from one another, the basic working principle of these TMs is the same.

In 1998, an important step was taken to standardize the exchange format of TM databases. Translation Memory eXchange format (TMX) was developed and maintained by OSCAR (Open Standards for Container/Content Allowing Re-use), a LISA (Localization Industry Standards Association) Special Interest Group. The main purpose of the TMX format is "to provide a standard method to describe translation memory data that is being exchanged among tools and/or translation vendors, while introducing little or no loss of critical data during the process" (LISA 2010).

As of the end of nineties, after the deployment of translation memories, commercial translators' workstations have been integrating full-scale machine translation (MT) systems. In addition to being used to translate "no matches" in the workstations where the first draft translation is provided by a translation memory, MT has also been used independently to produce rough translations for gisting purposes or to pre-translate texts to be further postedited.

2.2. Machine Translation (MT)

Machine Translation (MT) is an automatic translation system that processes a source text in one language and creates a target text in another language. On its home page, the European Association for Machine Translation (EAMT) provides the following description and explanation:

> Machine translation (MT) is the application of computers to the task of translating texts from one natural language to another. One of the very earliest pursuits in computer science, MT has proved to be an elusive goal, but today a reasonable number of systems are available which produce output which, if not

perfect, is of sufficient quality to be useful in a number of specific domains. (EAMT 2010)

In its "Best Practice Guide on Implementing Machine Translation", the Localization Industry Standards Association (LISA) defines MT as follows: "Machine translation is a method for translating something from one language to another automatically, without human intervention" (LISA 2010).

Machine translation has indeed a history dating back to 1950s, which is full of ups and downs. Arnold et al. (1994: 15) present this fluctuating movement in the MT history by humorously associating it with a roller coaster (see Figure 1). (see also "A Translation Automation Timeline" on TAUS web site for a chronological and brief information on the history of translation technology).

Figure 1. Machine Translation and the Roller Coaster of History

In the early years of MT, Fully Automatic High Quality Machine Translation (FAHQMT) was the initial aim of machine translation research (Bar-Hillel 1960: 93). However, due to the insufficient results, machine translation has usually been the subject of tough criticisms, serious disputes and humor. Soon afterwards, researchers concluded that FAHQMT was an elusive and unachievable goal under the prevailing conditions. Therefore, the focus shifted from FAHQMT to the systems that could provide output to be postedited by human translators or field experts to improve the productivity of the translation process.

ISBN: DL:

Today, it became evident that "the achievements of MT are real, even if they fall short of the idea of FAHQMT all the time – useful MT is neither science fiction, nor merely a topic for scientific speculation" (Arnold et al. 1994: 14). It is also widely accepted that MT has its own capabilities and limitations. In order to evaluate and judge MT justly it is necessary to make people aware of what the prevailing MT is capable of and what should be expected from its outcome. As Arnold et al. (1994: 6) suggest "the criticism that MT systems cannot translate Shakespeare is a bit like criticism of industrial robots for not being able to dance Swan Lake."

In the 21st century where globalization and ever-increasing technological developments affect every aspect of our lives, production of content in multiple languages has become one of the most significant aspects of communicating information. This brings about the requirement of a high level of speed and productivity in translation services. In order to survive in the competitive markets, companies are urged to increase their translation speed and productivity and decrease the cost of translation while keeping quality at acceptable levels. This is the main reason behind the increasing interest in using TM and/or MT with postediting, namely in text-reuse technologies as coined by García (2008).

The growing interest in machine translation can be attributed to factors such as globalization, widespread use of the Internet and the corresponding dissemination of information in multiple languages (Lopez: 2008: 2), which altogether pave the way for localization activities. Localization industry deals with translation of mostly technical and repetitive texts (repetitive within the same document and due to the frequent version updates). Translation memory systems have already been used in the Localization Industry to ensure the production of content in a faster and more consistent way. Recently, machine translation has also been integrated into the translation workflow to further streamline the translation process, thereby obtaining target texts at reduced costs, in shorter periods of time and without the loss of quality.

Fiederer and O'Brien (2009: 1) also list the reasons for the renewed interest in MT as follows: (1) growing demand for translations, (2) wish to penetrate new markets, (3) requirement to publish the translated material at the same time as source language material, (4) on-going requirement to reduce the cost of translation. Thus, the growing demand for the translation of higher volumes of texts into a number of languages in shorter periods of time and for lower rates has led translators and language service

ISBN: DL:
providers to take a more realistic and pragmatic stand towards MT. They started to look for ways to incorporate MT solutions into their translation processes and use MT in multi-component translation workstations in combination with other tools such as translation memories.

OPTIMALE project (Optimizing Pofessional Translator Training in a Multilingual Europe) is an Erasmus Academic Network involving 70 partners from 32 European countries. It aims to determine the current and emerging competence requirements within the European translation industry. The synthesis report presents employers' expectations of professional translators with regard to the use of tools and technologies. According to this report, the abilities to preedit or postedit machine translation outputs are not the mainstream requirements, although it also underlines that a significant 22 percent of the respondents "require MT postediting skills, reflecting the growing interest in the integration of MT in the translation process" (OPTIMALE 2012: 8-9).

Although the demand for MT solutions has increased, their integration into the translation processes has not been that smooth. The knowledge of grammar and words - and more importantly - world knowledge of "machines" are limited as compared to those of humans. Therefore, MT systems usually make more mistakes than humans and the outcome might be difficult to understand. Generally speaking, today's MT technology is far from providing the user with high quality translation without human intervention. However, if used properly in a specific domain, with certain text types and for particular purposes, MT offers substantial time and cost savings.

Despite its pros and cons as well as widely adopted misconceptions, MT seems to have entered a new era where it will be used exponentially by wider number of people with different profiles for different purposes. If we attempt to provide how the movement of MT for the 21st century can be reflected on the roller coaster of Arnold et al. (1994: 15), we can speculate an upward movement or at least a more stable line with regard to the previous years.

The earliest MT systems were simple dictionary-based ones which were based on "*lexical transfer,* or word-for-word (dictionary style) translation" (LISA 2004: 42). Soon afterwards, the need for more systematic ways of doing syntactic analysis led to the emergence of rule-based (transfer-based) MT. From the nineties onward, data-driven MT systems (statistical MT systems), which make use of large corpora of previous

ISBN: DL:
translations aligned with their source texts, came on the scene.

Recent applications involving machine-translated segments together with fuzzy matches from translation memories have blurred the distinctions between MT and TM technologies. Language service providers and companies tend to create hybrid tools involving MT output and TM matches to increase translation turn-around time and cost. With the proliferation of TMs from the 1990s, companies and institutions have developed large TM databases for their own content. They reuse these previously translated texts for new translation tasks. The exact and fuzzy matches are retrieved from TM databases, and the remaining non-matches are fed into machine translation systems. The resulting hybrid output from TM and MT systems are then postedited by human translators to various levels so as to produce the level of quality depending on the purpose of use of the final target text.

In an on-line survey of language service providers (LSPs), based on responses from some 75 companies in North America, Europe, Asia and South America, TAUS (2010: 13) points out that:

> in terms of engine type, the large majority (64 percent) are postediting raw content from both rule based (RBMT) and statistical (SMT) systems. However, SMT engines appear to have the larger overall footprint.

Giménez (2008: 83) states that statistical machine translation (SMT) is the dominant approach to empirical MT research. He emphasizes that SMT systems are "characterized by generating translations using statistical models whose parameters are estimated from the analysis of large amounts of bilingual text corpora" (Giménez 2008: 83). As compared to rule-based MT systems, it is easier and less time-consuming to build SMT systems. However, in order to construct an SMT system, large parallel corpora containing aligned phrases/sentences in source and target languages are needed. The first statistical MT system was developed by Brown et al. (1988, 1990, 1993 cit. in Giménez 2008) at the IBM TJ Watson Research Center, and worked from French into English. The system was called Candide and was trained on the parallel corpus of proceedings from the Canadian Parliament (Giménez 2008: 83).

Giménez (2008: 85) suggests that "SMT systems address the translation task as a search problem." The basic type of statistical MT occurs at the word level. Word selection and word ordering are "two separate but interrelated" aspects of word-level SMT. Word selection is about selecting the appropriate translation of the source words in the target language, whereas word ordering is about deciding the position of the translated word in the target sentence. Word translation models fail to take into account the context in which words occur, and translation does not occur on a word-for-word basis. Therefore, in order to improve SMT systems, a step was taken to extend the scope of the translation unit from words to phrases, which constitutes the fundamentals of phrase-based SMT. Although they provide better output than word translation models, phrase based SMT models have some weaknesses in that they do not make use of linguistic knowledge. Various extensions of phrase-based models were developed to overcome this limitation (Giménez 2008: 88-89). Further, syntax-based approaches were proposed to employ more advanced reordering models than phrase-based models and to overcome its deficiencies by taking into account syntactic knowledge. Giménez (2008: 92) underlines that the syntax-based SMT models were slightly more effective than the best phrase-based systems, particularly in distant language pairs such as Chinese and English which present significant differences in word ordering.

In his tutorial, overview of the state-of the-art in SMT, Lopez (2008: 1) notes that "SMT has made tremendous strides in less than two decades" and has become the focus in academic MT research. He explains that translation from one natural language into another is treated as a machine learning problem by SMT, that is to say, "by examining many samples of human-produced translation, SMT algorithms automatically learn how to translate" (Lopez 2008: 1). Lopez also points to the lack of research on machine translation from English into other languages. He further underlines that the current research focuses on translation from other languages into English, mostly due to government funding, and this may obscure the deficiencies of the current approaches. For instance, translating from morphologically rich languages such as German into English, which is relatively morphologically simple, may make some loss of meaning and nuance less marked. However, translating in the other direction requires much more attention (Lopez 2008: 40) and thus needs more empirical research.

2.2.1. Types of Machine Translation Systems

Basically, all MT systems rely on highly sophisticated analyses made by artificial intelligence or computers to render the source language items into the target language. However, in terms of methodology, the way computers translate varies. The Localization Industry Standards Association (LISA) provides five different approaches

ISBN: DL:
to machine translation in its "Best Practice Guide on Implementing Machine Translation" (LISA 2004). Below, we look at these different types of MT.

2.2.1.1. Simple Dictionary-Based MT

As the name suggests, simple dictionary-based machine translation is based on "*lexical transfer,* or word-for-word (dictionary style) translation. In this model, words are simply translated as they occur" (LISA 2004). Obviously, this type of very basic and literal machine translation does not meet the needs for the translation of complex sentences. Even in simple sentences, pure word-for-word translation may result in a word salad rather than producing meaningful units, because the meaning of the individual words may depend on the other components of a sentence. Further, differing grammatical structures and word order in different languages make it necessary to employ linguistic analysis of some kind in machine translation in order to match the source language elements with the corresponding target language elements properly.

2.2.1.2. Rule-Based (Transfer-Based) MT

Rule-based or transfer-based MT systems can be traced back to the beginning of MT research in the 1950s. They involve breaking the source sentences into their component elements and determining the grammatical role of each word in the source sentence. After parsing the source text into its component units, "a series of transfer rules are used to reorder words and otherwise alter the structure of the incoming sentence to produce a translation that is grammatically correct for the target language" (LISA 2004).

Since the rule-based MT systems involve mapping grammatical structures of both source and target languages and since they take the context in which words are used into consideration, they yield much better outputs than simple dictionary-based systems. However, for the same reasons, rule-based systems can be built for particular language pairs and directions. The advantage of using rule-based MT systems is that once established for a specific language pair and direction and when the system dictionaries and glossaries are updated regularly, they produce satisfactory results. The negative aspect of these systems is that it is very time-consuming to complete those sophisticated linguistic analyses and carry out the regular updates. Another point is that, once a rule-based MT system is built, such work is dedicated only to one language pair and direction, and it cannot be extended to another language pair and/or direction.

Most commercial systems are examples of transfer or rule-based MT systems; i.e.

ISBN: DL:

Systran, Metal, Promt and Logos.

2.2.1.3. Interlingual MT

We have mentioned that rule-based MT systems are created for specific language pairs. Since it is demanding to build and maintain (do the necessary glossary and dictionary updates regularly) rule-based MT systems for a number of language pairs, researchers have attempted at building an interlingua which will express the source sentence in a "common notation (an "interlingua") that can be used to generate sentences in any other language" (LISA 2004). In this way, it would be possible to offer a common MT system that might cover a number of language pairs and directions. The two main components of these systems are "an analyzer that builds an interlingual notation for an incoming sentence and a generator that produces sentences for any interlingual notations" (LISA 2004).

2.2.1.4. Data-Driven MT

Recently, data-driven MT systems have appeared as an outcome of the researchers' attempts to find out new and much more practical approaches to machine translation systems. The distinctive feature of data-driven systems is that they use statistical means to analyze source texts and generate translations unlike those rule-based systems that are based on parsing the source sentences and implementing a set of transfer rules to produce the target text.

In data-driven MT systems, the source language items are translated into the target language by calculating the probability that those source items correspond to the target language words and phrases, using a corpus of aligned sentences in the source and target languages.

Data-driven MT systems make use of large corpora of previous translations aligned with their source texts. Due to the proliferation of translation memories, most translation companies and professional translators as well as other organizations conducting international business have already built up their own corpora of translated and source texts aligned on a translation memory database. The possibility of exploiting these readily available corpora makes it easier and less time-consuming to create data-driven MT systems.

There are two main approaches in data-driven MT: Statistical Machine Translation (SMT) or Statistics-Based Machine Translation (SBMT), and Example-

ISBN: DL:
Based Machine Translation (EBMT). "These systems analyze a large number of original sentence/translated sentence pairs to discover which words or expressions in one language are most highly correlated with words or expressions in the other" (LISA 2004).

Nowadays, growing number of researchers are in favor of the idea of developing hybrid MT systems which aim at bringing the best aspects of the rule-based and data-based approaches together.

2.2.1.5. Hybrid MT

Emergence of new methodologies in machine translation has led researchers to think multidimensionally. Facing with positive and negative sides of the abovementioned approaches to MT, researchers have focused on combining the strong aspects of them, thereby creating a hybrid system.

Hybrid systems are envisaged to "use rule-based analysis and generation connected by data-driven transfer rules, or statistical systems that blur the distinctions between example-based and standard statistical approaches" (LISA 2004).

In its report on "Postediting in Practice", TAUS (Translation Automation Users Society 2010) which defines itself as a think tank for the translation industry, undertaking research for buyers and providers of translation services and technologies, mentions about an experiment which takes RBMT output and passes it through a SBMT system trained on postedited rule-based MT content so that a rule-based system can benefit from statistical learning in a post-processing task (TAUS 2010: 20). The report points out that:

> research will focus more on further ways to automate MT (to such extent that the RB/S/MT/TM distinctions may soon become blurred) than to provide more tools for human posteditor interfaces and productivity tools. (TAUS 2010: 20)

Attempting to combine the strengths of rule-based and statistical machine translation systems, hybrid MT systems constitute an attractive field of research.

2.2.2. Examples from Free Online MT Systems

Nowadays we have easy access to free on-line MT systems that provide user-friendly interfaces open to both translators and non-translators, as well as to individual and commercial users. This has brought a new dimension to the translation profession.

ISBN: DL:

Experiments to explore these systems have started to go beyond in-class tests and the systems are being used in real-life translation workflows.

Ramos (2010) presents an overview of the analysis carried out as preparatory work for the actual implementation of Google Translate in the translation/localization process from a language vendor's perspective. Ramos (2010: 3) states that before this analysis, they had been using Google Translate at Ocean Translations S.R.L. as the initial translation step for some of their clients (upon the client's request) and this practice yielded encouraging results. Posteditors working in the company were satisfied with the quality they achieved and the fees they received when using this MT system. They were postediting around 650 words per hour, which made around 5000 words per day, considering an average working day to be eight hours. It is interesting to note that the posteditors also reported being happy with having the draft translation when working with Google Translate instead of having a "blank page", as is the case when they translate the text without MT. Ramos (2010: 3) mentions an estimated cost savings of 40 percent for the client on a project of more than 90,000 words from English into Spanish, completed (postedited and revised) in nine calendar days by using Google Translate. She also explains how, on a request from a client, they analyzed the cost of moving from TM tools to Google Translator Toolkit for the localization of the Reproductive Health Library (which had to be updated every year) at the World Health Organization (WHO). As a result of this analysis, they drew up a report "intended to gain some enhancements and adjustments from the Google and WHO collaborative efforts" (2010: 4). The report involves the tasks performed, first to overcome the restrictions posed by Google Translator Toolkit for the implementation of the project, and second to provide a smooth transition to using it for their localization project. In addition, the report covers the company's insights regarding the use of Google Translate (Ramos 2010: 4). Some of these comments concern system requirements and others are addressed to Google Translate system developers, proposing minor adjustments to the system. Based on their analysis, Ramos (2010: 4) points out that they were satisfied with the initial findings and had very positive expectations for the real implementation phase of the project.

Inspired by Ramos's paper, we made a small test in order to compare the outputs from different online machine-translation systems. We translated a 130-word authentic text from Turkish into English with four MT systems: Microsoft Bing Translator,

ISBN: DL:

InterTran, WorldLingo and Google Translate. We selected these MT systems because they offered Turkish among the language pairs they work with. The text is taken from the website of the first speech-recognition system in Turkish called "Dikte", developed by the companies Yöndata and CTD Systems. The text involves some technical information about the "general" version of the speech recognition software. In terms of quality, the MT outputs from Microsoft and Google are close to each other and far better than those obtained from the other two systems, InterTran and WorldLingo.

Below, we present the first three sentences of the raw (unedited) MT output from Microsoft and Google Translate as well as a possible human translation of the Turkish source text in English:

(1) MT Output from Microsoft Bing Translator:

- General Standard is based on the overall Technology dictate Dictation.
- High capacity Dictation in General, the world's first and only Turkish speech recognition technology.
- The highest in the world has ever been to dictate the word capacity and capacity of the fastest speech recognition system.

(2) MT Output from Google Translate:

- General Standard Dictation, dictation technology is based on the general.
- Dictation General, the world's first and only high-capacity Turkish speech recognition technology.
- Dictation, the highest ever made in the world's fastest vocabulary speech recognition system according to the capacity and capability.

Human Translation of the Turkish Source Text in English:

- Dikte General Standard is based on Dikte General Technology.
- Dikte General is the world's first and only Turkish high-capacity speech-recognition technology.
- Dikte is the speech recognition system with the highest word capacity in the world and is the fastest considering its capacity.

ISBN: DL:

We analyzed the MT outputs from Microsoft and Google Translate using seven error categories of LISA Quality Model 3.1 (see 6.1.1.5 below). The categories are Mistranslation, Accuracy, Terminology, Language, Style, Country and Consistency. However, the MT outputs in our test contained errors only in Accuracy, Terminology and Language categories. Since the MT outputs from InterTran and WorldLingo were full of errors and contained correct translations of only some words and phrases, these translations were not analyzed with LISA QA Model. Instead, we counted the correctly translated words and phrases for InterTran and WorldLingo.

Table 1 shows the distribution of errors and total error points in the unedited MT output of a 130-word text translated from Turkish into English with Google Translate and Microsoft Bing Translator.

Table 1. Test on Google Translate and Microsoft Translator

	MT System			
	Microsoft Bing Translator		**Google Translate**	
Error Category	Minor	Major	Minor	Major
Accuracy	1	1	1	1
Terminology		1	1	
Language	5	4	5	2
Total Error Points	36		22	
Quality Percentage	21.74		50.00	

The results show that the output from Microsoft contained one minor and one major accuracy error, one major terminology error, and five minor and four major language errors. The MT output from Google Translate had one minor and one major accuracy errors, one minor terminology error, and five minor and two major language errors. The total error points (calculated by LISA) in Microsoft is 36, and in Google Translate, 22. The quality percentage of the output from Microsoft is 21.74 whereas that of the output from Google Translate is 50.

As for the other two systems, the extracts from the outputs from InterTran and WorldLingo are given below:

ISBN: DL:

(3) MT Output from WorldLingo:

GeneralTechnical Information• Dictate Dictate settled on the General Standard, Dictate the General Technology
Dictate General, the world's first and only high-capacity English speech recognition technologythat
Dictate the made so far in the world, with a capacity of the word, and the capacity to
dictate the most rapid speech recognition system

(4) MT Output from InterTran:

Dictation General Technique Informed • Dictation General Standart , Dictation General Technology about established. • Dictation General , world first and a single thing high capacity Turkish subject tanma teknolojisidir. • Dictation , world today as much as yaplmº topmost wordcapacity and kapasitesine according to most hzl subject tanma sistemidir.

MT output from WorldLingo contained correct translations of twelve words, while InterTran included the correct translations of only eight words. The rest of the outputs included either non-translated Turkish words (for InterTran) or incorrect translations of the source text. Moreover, the MT output from InterTran also contained incorrect Turkish words due to the system's failure to identify Turkish characters such as "ş, ğ, ı".

As is clear from the MT outputs from WordLingo and InterTran, it was not possible to apply LISA grid to analyze the outputs from these MT systems. Therefore, we made a very rough analysis for WordLingo and InterTran. However, this example gives an overview of the comparison of the unedited MT outputs from four different free online machine translation systems operating with Turkish-English language pair. According to the results derived from the data, MT outputs can be ranked from the best to the worst as follows: Google Translate, Microsoft Bing Translator, WorldLingo and InterTran. None of the outputs are flawless. However, Google Translate and Microsoft yielded quite similar data, which seem to be suitable for postediting. On the other hand, outputs from InterTran and WorldLingo are far from being useful even to get the gist of the source-text content. Therefore, the latter will be excluded from our main research.

ISBN: DL:

From among the first two MT systems, namely Google Translate and Microsoft Bing Translator, we decided to use Google Translate in our main research. It is widely used and also operates on a web service called Google Translator Toolkit (GTT) which allows its users to postedit translations generated by Google Translate.

Our small experiment indicated that Google Translate and Microsoft can produce usable MT output which is suitable for postediting. Google Translate has been used in empirical research conducted both in translation classes (see Izwaini 2006, Pym 2009, García 2010, Vaezian 2010) and other settings (Chang-Maedows 2008, Zuo 2010). It has also been incorporated into the in real-life translation workflows in industrial settings as is outlined by Ramos (2010) above.

All four MT systems used in our small experiment are currently offered free on the Internet. However, Google Translate seems to have further advantages. The main distinguishing feature of Google Translate is that it is offered in a translation memory workstation called Google Translator Toolkit (GTT). In GTT, the user can postedit the MT output and send the postedited data to the MT database, thus train the database to the required level.

To sum up, our initial analysis suggests that some MT systems are not worth studying further, but others are. Based on the results of this analysis, we decided to use Google Translate as the data-based machine translation system in the present study.

ISBN: DL:

3. Postediting

The traditional workflow of translation has been changed substantially due to the rapid automatization of the translation processes. The incorporation of translation memories and machine translation into the translation workflow has brought about new ways of dealing with the source and the translated texts.

In order to improve the final outcome, two types of editing are employed in the translation processes where TM and/or MT are used. These are pre-editing and postediting. Pre-editing involves applying "controlled language rules" on the source text, thereby making it more plain and standard in terms of its form and style. Here, the main aim is to obtain better raw TM or MT output by controlling the input. Postediting, on the other hand, is the process of editing the texts that are pre-translated using a machine translation or a translation memory system. Here, the aim is to improve the TM or MT output quality by postediting it to various levels depending on the purpose of the target text use. It is also possible to implement both pre and postediting together to obtain better results.

One of the most extensive definitions of postediting, given in the Draft of European Standard for Translation Services (2004) is as follows:

> Postediting is the examination and correction of the text resulting from an automatic or semi-automatic machine system (machine translation, translation memory) to ensure it complies with the natural laws of grammar, punctuation, spelling and meaning, etc. (Draft of European Standard for Translation Services prEN15038: Brussels, 2004)

Although the term "postediting" is generally used for editing both the translation memory and the machine translation output, it is sometimes used to refer only to editing the MT output. For example, Veale and Way (1997; cit. in Allen 2003: 297) describe postediting as "the term used for the correction of machine translation output by human linguists/editors". Allen (2003: 297) states that "in basic terms, the task of posteditor is to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)". Similarly, Schäfer (2003: 3) defines postediting as "the task of polishing up the raw MT output to an acceptable, end-user friendly text quality". Pym (2011b: 88) defines postediting as "the process of making corrections or amendments to automatically generated text, notably machine-

ISBN: DL:

translation output". According to Yngve (1954: 21), the task of posteditors is "to take the imperfect output from the machine and edit it into a polished or at least easily comprehensible document". In the present research, we also use the term postediting in this sense, namely referring to editing the MT output.

It is important to point out that postediting does not mean translating the source text from scratch. Rather, it involves "improving a machine-generated translation with a minimum of manual labor" (TAUS 2010: 7) to increase translation speed and gain advantage over full human translation in terms of translation turn-around time and cost. Therefore, while postediting; only the corrections, replacements and retranslations which are really necessary should be performed, and unless otherwise stated by the client, stylistic considerations should not surpass rendering the accurate meaning of the content.

The factors determining the extent of the changes to be performed, that is the level of postediting, are described in the next section.

3.1. Levels of Postediting Machine Translation

There are several factors which determine the level of postediting. These are raw output quality, demanded final translation quality and the purpose of the target text with regard to its use. According to Allen (2003: 301), the level of postediting to be performed on a text is dependent on the following factors:

> (1) the user/client, (2) the volume of documentation expected to be processed, (3) the expectation with regard to the level of quality for reading the final draft of the translated product, (4) the translation turn-around time, (5) the use of the document with regard to the life expectancy and perishability of information, (6) the use of the final text in the range from information gisting to publishable information.

When we look at the literature on postediting, we usually come across a binary classification depending on the level of postediting. These are light and full postediting (TAUS 2010: 8), rapid and full postediting (Wagner 1985: 213), fast and conventional as well as partial and complete postediting (Krings 2001: 45) The former is applied for producing understandable target texts to be used for gisting purposes while the latter is employed for generating target texts of publishable quality.

ISBN: DL:

3.1.1. Light Postediting

As far as ephemeral texts are concerned, MT output is sometimes used without any intervention, that is, with no postediting. However, even if the translation is to be used only for browsing and gisting purposes with a limited circulation, in most cases MT is followed by postediting.

Light postediting is also referred to as rapid postediting (RPE) or minimal postediting. It "involves taking the raw MT output and performing as few modifications as possible on the text to make it an accurate reflection of the source text content" (TAUS 2010: 8). In light-postediting, the priority is on making the output content understandable. Therefore, stylistic modifications are not considered. TAUS (2010: 8) lists the type of modifications appropriate for light postediting as follows:

> (1) replacing unknown words, (2) deleting superfluous translation alternatives generated by the machine, (3) repairing machine-induced meaning distortion (possibly the chief priority), (4) correcting the most salient word and grammatical errors, (5) partially or wholly rewriting some sentences.

Light postediting can be applied on the MT output to be used for assimilation or personal communication purposes where quality requirements are not high. Electronic mails, short-lived documents and any other texts to be used merely for information purposes or for restricted circulation can be given as examples. Light postediting is highly preferred in cases of urgent translation needs which are not possibly met by human translators.

The advantage of light or rapid postediting is that it is relatively easier and less time-consuming to apply it to the texts rather than translate them from scratch. Its disadvantage, on the other hand, lies in the difficulty to store the modifications on the TM or MT database. Senez (1998: 4) states that rapid postediting can be a perfectly viable option if (1) the customer needs the translation urgently, (2) the text is not destined for publication, but will serve some temporary use, and (3) the customer makes an informed decision to opt for postedited MT after considering the advantages of a faster service and the risk of loss of quality inherent in the rapidly postedited machine translation.

ISBN: DL:

3.1.2. Full Postediting

Basically, full-postediting means editing the TM or MT output to the level of quality which is expected from a human translator. It involves all the steps to be taken in the light postediting practice, plus more detailed modifications and stylistic changes to produce a target text that reads like a human translation. Full postediting is usually applied to the MT output which is intended for dissemination and required to be of publishable quality. According to the TAUS report (2010: 9) "the large majority of MT output in production settings needs to be postedited to the same quality level as that found in a standard human translation."

Translators may not always prefer postediting automated translation output to translating the source texts from scratch. However, it is noteworthy to point out the advantage of applying full postediting underlined by TAUS (2010: 8) maintaining that "the corrections to a machine text can (ideally) be recycled into the system as 'linguistic' upgrades, thereby improving the systems over time and increasing translation automation ROI" (Return On Investment, explanation added).

Researchers and organizations committed to using automatic translation solutions explore further ways to facilitate the postediting practice. Developing tools that can perform automatic pre-editing on the source text, thereby improving the input quality as well as other tools that can perform automatic postediting by carrying out especially the formal corrections to be made on the MT output can be given as examples of such efforts. The use of such applications can facilitate the posteditors' work by taking over the repetitive, small, formal modifications; and enabling them to concentrate on the semantic corrections.

ISBN: DL:

4. Literature Review

4.1. Previous Research on Translation Memories (TM)

Translation memories (TM) are among the most-widely used translation tools. The origins of TM date back to earlier times when suggestions were made regarding the storage of previously translated source and target texts for reuse in new translation projects in the same domain (Arthern 1979, cit. in Christensen and Schjoldager (2010). However, the commercialization of TM systems did not begin until the 1990s. Since then they have been adopted by many language service providers as their primary aid for translation.

Christensen and Schjoldager (2010) have compiled previous empirical research on translation memories. Since the use of TM expanded in the late 1990s, empirical TM research carried out since 2000 was presented chronogically. In the present research, we provide an overview of the previous research on translation memories given in Christensen and Schjoldager (2010) as well as other studies in the field, grouping them according to the research questions.

Christensen (2003) wrote her PhD dissertation on TM and legal translation. Since the original dissertation was in Danish, a language we do not know, our knowledge regarding this empirical study is limited to the summary of the project provided by Christensen and Schjoldager (2010) and our personal contact with the author through electronic mail. The study was based on the "assumption that, because of the complexity and culture specificity of legal communication, TM technology will be less useful for legal translation than for technical translation, for which it was first designed" (Christensen and Schjoldager 2010: 93). Therefore, the main purpose of the study was to "evaluate the usefulness of a TM for legal translation." In order to collect data, the author contacted Danish and German law companies and requested Danish company articles and their translations into German. The data involve two different corpora: The first one is a parallel corpus which comprises a TM of source text segments from eleven different Danish articles (obtained from eleven different companies) and their authentic German translations. No information was collected as regards whether the translations had been done by professional translators, lawyers or someone else (Christensen, personal contact through e-mail). The second corpus comprises a reference corpus

ISBN: DL:

which consists of one German model article taken from a standard German handbook on German company law. This model article was "assessed as functionally equivalent to a selection of source text segments in the TM" (Christensen and Schjoldager 2010: 94). The reference corpus was only used to evaluate the translation quality of the content of the texts in the parallel corpus. The texts used in the two corpora were approximately six to ten pages each. Translators Workbench from Trados was used as the TM suite. In order to measure the degree of repetition, the analysis feature in Trados was used. It was hypothesized that a TM would be able to retrieve matches for those legal segments which are semantically and functionally identical. In order to test this, one of the texts from the parallel corpus was aligned and imported into the TM. By means of test runs, the author investigated whether the TM found the identical segments in the other ten texts. Data analysis showed that the Danish company articles did not represent sufficient level of repetitiveness to be useful as source texts in a TM. Second, contrary to the initial expectations, it was found that the TM consisting of Danish articles and their functionally equivalent German translations did not identify segments that were functionally equivalent, thus failing to improve the usefulness of TM for legal translation. Further, Christensen (2003) tested whether the TM would improve if the texts to translate were pre-edited. To test this, the TM was used as an "authoring memory". It was found that the identification of functionally equivalent matches could be improved using "authoring memory", which involved standardization of source text segments by pre-editing them before aligning with target text segments (Christensen and Schjoldager 2010: 93-94).

In another PhD dissertation on translation memory systems, Dragsted (2004) empirically investigated cognitive segmentation and effects of integrating TM systems into the translation process. Dragsted (2006) presented the findings derived from her PhD study in an article describing cognitive reflections of the use of translation memory systems on the translation process. The study is based on the assumption that the sentence-based segmentation introduced by TM systems does not correspond to the natural segmentation translators usually perform. It was composed of three phases. The first phase aimed at observing how translators segment texts naturally. The second phase looked at the effect of integrating TM systems (with their sentence-based segmentation) into the translation process. The third phase aimed at making suggestions as to how TM systems can be optimized to conform better to the natural cognitive

ISBN: DL:

translation process and produce higher match values. A series of experiments were conducted with six professional translators (with at least two years of professional experience) and six Translation students at the Copenhagen Business School, Department of English (in their fifth and final year of studies). The subjects translated two short texts from Danish into English. Two texts with different levels of difficulty were used in this study. The first text was extracted from a relatively straight-forward business letter which contained 212 words. The second text was extracted from a legal contract which contained 127 words and certain terminological problems. "The easy text was used to reflect the 'normal' translation process, [...] whereas the translation of the difficult text *adds* to the normal translation process the handling of problematic items" (Dragsted 2004: 101, emphasis in the original). The texts were extracted from authentic material, although several adjustments were made to make them suitable for the purposes of the study (i.e. company names were omitted to avoid recognition, the first text was cleaned of problematic items and the second one was selected from the parts of the original text that would contain problematic items). For the first phase of the study, Dragsted (2004) studied human translation where the subjects' translation activity was recorded using the keystroke logging program Translog. For the second phase, the subjects were tasked with translating the same texts using the translation memory system Trados Translator's Workbench. After the first and the second phases of the experiments, the subjects were asked to verbalize their own perception of the translation tasks retrospectively. In her data analyses, Dragsted (2004) focused on production time, subjects' inclination to change the source sentence structure, and the way in which subjects segmented the source texts. Pauses between text segments were regarded as indicating "an approximate break between two cognitive units" (Dragsted 2004: 109) and thus constituting the basis of segmentation. Both production time and pauses were recorded with the keystroke logging tool Translog. Findings regarding the first phase of the experiments revealed differences in the segmentation behavior of professional translators and Translation students. In the translation of the easier text, students worked with short segments at word/phrase level processing mode at low speed, whereas professionals worked with longer segments, processed at sense-oriented clause level and at high speed. When it came to the translation of the more difficult text, some of the differences between professionals and students were reduced and professionals took over some of the characteristics of students' processing. As for the

ISBN: DL:

second phase of the study, the data showed that the use of TM affected the translation process of both groups (students and professionals) in that they focused more on sentences instead of the shorter segments that were processed when no TM was involved. Students spent less time on the tasks due to the sentence-level segmentation presenting one sentence at a time. On the other hand, professionals' production process became slower due to TM. Another finding showed that when working with TM, subjects tended to change the sentence structure less than they did while working without TM. Overall, the findings suggest that the sentence does not constitute a central cognitive processing unit without TM and the use of TM affects translators' cognitive segmentation behavior. In the third phase of the study, Dragsted (2004: 281) recommends that "TM systems be adjusted so that the focus is removed from the sentence, while at the same time segments below sentence level are retrieved in order to ensure reuse of previously translated text strings of any length."

Colominas (2008) focused on the sentence-based text segmentation of the current translation memory systems and explored the implications of sub-sentential segmentation taking noun-phrase chunks as sub-sentential translation units. Similar to Dragsted (2004), Colominas (2008: 343) argues that sentence-level segmentation has some drawbacks in that it automatically excludes sub-sentential segments from potential matching candidates, even though these smaller segments could prove useful for producing the target text. Based on this argument, Colominas (2008) carried out experiments to evaluate both recall and precision of sub-sentential segments: "Recall refers to the segment of the sentence to be translated that can be covered by the proposal from the system, while precision refers to the usability of these proposals" (Colominas 2008: 345). Two different parallel corpora in English-Spanish, aligned at sentence level, were used as translation memories. Unlike most of the studies on TM systems, the corpora did not contain technical texts. The first corpus was the Europarl corpus containing material extracted from the proceedings of the European Parliament (ca. 193000 words for each language). The second corpus was a bilingual corpus (English-Spanish) built up from United Nations documents regarding the protection of children (ca. 255000 words for each language). Test material was extracted from these corpora. Noun-phrase chunks were obtained by means of the Connexor Phrase Tagger and analyzed using the analyze function of the Trados system. In the first corpus (Europarl), test material contained 7043 noun phrase (NP) units out of 19924 NP units of TM

ISBN: DL:

material. In the second corpus (United Nations), NP chunk-segmented test material involved 4463 NP units out of 12549 NP units of TM material. Data analysis for recall showed that, in the first corpus, noun-phrase segmentation brought together a 19 percent match between the levels of 74 percent and 100 percent, and 11 percent intra-textual repetition. As for the second corpus, results were quite similar to those obtained with the first one. This time, 25 percent match between the levels of 74 percent and 100 percent was obtained by NP chunk segmentation, and the rate of repetitions was 14 percent. These results from the two corpora indicated that, as compared to sentence segmentation, better recall results were obtained with NP segmentation. When it comes to precision, data revealed that only 10 percent of the retrieved NP chunks were really useful. However, the author points out that there was a clear gain with respect to 'only' sentence level segmentation (Colominas 2008: 351). With respect to precision, the results showed that NP segmentation should be much more refined in order to be really useful. Nevertheless, the study implies that promising results might be obtained by sub-sentential segmentation in the form of noun phrase chunks. The author suggests that the findings of this study should be taken into consideration in the design of future TM systems. On the other hand, more research with other tools, text types, languages and other sub-sentential chunks (i.e. verb phrases, prepositional phrases) is deemed to be necessary.

Fulford and Granell-Zafra (2005) adopted a survey-based approach and reported on the background, methodology and findings of the first phase of a three-year project which is "an exploratory survey of the uptake of ICT by freelance translators in the UK" (Fulford and Granell-Zafra 2005: 2). The first stage was to outline freelancers' translation workflow components (document production, information search, translation creation, communication, etc.) together with some software applications that might be used to support these activities (word-processing software, terminology databank, translation memory, machine translation, etc.) (Fulford and Granell-Zafra 2005: 4). A questionnaire was prepared and tested before being sent out to 1400 UK-based translators. An e-mail list of these translators was procured from a professional body of translators in the UK. A total of 591 responses were obtained. However, 200 of them were discarded since the respondents reported that translation was not their principal activity or they were working as in-house translators. Therefore, a total of 391 valid responses from freelance translators in the UK were analyzed for the study. The

ISBN: DL:

respondents were generally experienced translators ranging from less than 5 to over 20 years of experience in the profession. The principal working languages were German-English (37 percent), French-English (37 percent), Spanish-English (16 percent) and the remaining 10 percent were composed of other language combinations such as English-German, Russian-English and Italian-English. The respondents' fields of specialization were business/commerce, technical translation and legal translation in a descending order. The emphasis in this survey was on exploring the range and types of software applications freelancers had already been using in their translation workflow rather than the degree of automation that they were willing to use. The findings indicate that "there has been widespread adoption of general purpose software applications to support a number of the activities involved in the freelance translator's workflow" (Fulford and Granell-Zafra 2005: 11). However, the uptake of special purpose translation-oriented tools such as terminology databases and translation memories were found to be limited as compared to those general purpose applications such as word processing software, desktop publishing software, etc. Only 28 percent of the respondents stated that they had been using TMs. The survey revealed a very low level of use of machine translation as well. Only 5 percent of the respondents replied that they used machine translation software in their professional activities. Moreover, 75 percent of the respondents were reported to be unfamiliar with MT systems. The authors suggest that non-adoption of translation tools was more due to translators' lack of awareness of and familiarity with these tools than an informed rejection of them. It was reported that a deeper investigation of the reasons why the adoption of ICT among UK-based freelance translators was low would be carried out in a follow-up phase of the research project.

In a similar vein, Dillon and Fraser (2006) conducted a survey-based study to provide a snapshot of how UK-based professional translators perceived translation memories as a tool in their working environment. Three hypotheses were tested: (1) Novice translators have a more positive general perception of TM than experienced translators, (2) Translators who use TM have a more positive general perception of it than translators who do not, (3) Translators' perception of the value of TM is not linked to their perceived IT proficiency. A questionnaire composed of two sections was drawn up. First, a demographic questionnaire was used to collect data on the subjects' profile (i.e. years of experience, contact with TM and so on). Second, an attitude questionnaire was set up to examine the subjects' perception of TM technology and the effect of this

ISBN: DL:

perception on their adoption of TM. The questionnaire involved 24 statements expressing various positive and negative attitudes towards TM (i.e. "The disadvantages of TM far outweigh the advantages" or "Translators who use TM get more work in my area than those who do not"). These statements were modified so that they would be suitable for both TM user and non-user respondents. Moore and Benbasat's (2001) instrument for measuring perceptions with regard to the adoption of an information technology innovation was adapted and used. An online survey form was prepared and the link was sent to the members of a professional body of translators via e-mail. In addition, the survey was distributed to the students and recent graduates of a postgraduate translation course in order to collect data on novice and/or less experienced translators. Different levels of experience were represented at three bands: 3 or fewer years, 4-9 years and 10 or more years (Dillon and Fraser 2006: 71). Of 285 translators, 59 provided usable responses, which represented a response rate of 21 percent. Data analysis showed that the first and the second hypotheses held, while the third hypothesis was disproved. As for the first hypothesis, it was found that novice translators had a more positive attitude towards TM, despite the fact that this group had the lowest level of contact with TM. Further research is suggested by the authors to find out if the attitude of these novice translators would change if and when they are exposed to TM more frequently in later periods of their career. The second hypothesis held that translators having more contact with TM and more professional experience in using it seemed to be approaching to TM more positively. As for the third hypothesis, contrary to what was initially hypothesized (Hypothesis 3), a correlation was found between translators' IT skills and their positive perception of TM. In line with the results of a previous study on UK-based freelance translators conducted by Fulford and Granel-Zafra (2005), the findings of this study also suggest that there is "a large section of the translation community that does not fully understand the benefits and limitations of TM use and cannot therefore make informed decisions on the usefulness of its adoption in their working environment" (Dillon and Fraser 2006: 76). This conclusion implies that, rather than the technology itself, it is lack of knowledge about the benefits and limitations of the TM technology that may lead to an automatic rejection or adoption of it.

Lagoudaki (2006) carried out a survey similar to that of Fulford and Granell-Zafra (2005) and Dillon and Fraser (2006) to investigate the adoption of TM technology

ISBN: DL:

among translation professionals (such as translators, terminologists, project managers, reviewers, subtitlers) and their attitude towards TM systems. Lagoudaki (2006) conducted a pilot study prior to designing the main survey. In line with the results obtained from this pilot study, the questionnaire was refined by making the questions simpler, clearer and less ambiguous. The pilot study included ten interviews and a focus group session involving eight translation professionals. Then, the seven-minute TM survey in English was launched online and made available to user groups fora, translation and localization companies, translators' associations, academic institutions providing translator training and public authorities having translation departments. Although it is similar to previous surveys on TM, Lagoudaki's survey addressed a larger number of people with various language pairs and with different roles in the translation process/business. A total of 874 translation professionals from 54 countries responded the survey over a two-month period. However, only 699 of them completed the questionnaire fully. Although the respondents represented various translation professionals, most of them (90 percent) were translators, and the majority (73 percent) of these were freelancers. Unlike the previous UK-based studies, this survey found a high rate (82.5 percent) of TM adoption among translation professionals. In line with Dillon and Fraser's (2006) findings, Lagoudaki's (2006) survey revealed a correlation between the skilled use of computers and TM adoption. Some 64 percent of the respondents reported a "good" level of computer skills, whereas 30 percent reported an "excellent" level (Lagoudaki 2006: 11). More than half the respondents (61 percent) stated that they were specialized in the translation of technical texts (Lagoudaki 2006: 12). In parallel with Fullford and Granel-Zafra's (2005) findings, Lagoudaki (2006) found a strong correlation between TM adoption and specialization in the translation of technical texts. This may be because technical texts contain high levels of terminology, standard expressions, simple sentence structures and repetitions. The relationship between high levels of repetition and TM use was also stated in Christensen's (2003 cit. in Christensen and Schjoldager 2010: 94) study where she found that legal texts in Danish did not involve sufficient amounts of repetitions to be useful as source texts for TM. Another important finding of Lagoudaki's study is that 71 percent of the respondents reported using TM systems out of a personal choice rather than being affected by external factors such as translation company requirements and/or client requests (Lagoudaki 2006: 19). When they were asked why they were voluntarily used a

ISBN: DL:

TM tool, 86 percent responded that it saved time, 83 percent said it improved terminological consistency and 70 percent pointed out TM's contribution to improving output quality. Among other benefits, 34 percent underlined cost savings and 31 percent suggested TM's efficiency in exchanging resources such as glossaries and TM databases. Although the previous studies on TM adoption revealed a correlation between the lack of knowledge on TM technology and its rejection, Lagoudaki's survey (2006: 19) showed that 51 percent of the TM users received no training in the use of TM systems.

In her pilot study, Ribas (2008) explored whether translation memories containing errors have the potential to propagate errors. She designed two separate experiments using different subjects, texts and tools, yet with the same general structure. The first experiment was carried out with six Masters students in Translation at the Universitat Rovira i Virgili in Spain. The subjects were asked to postedit a 302-word text containing the instructions for the installation of a bicycle carrier, from English into Spanish within two hours. They were provided with a TM database and a terminological database and were asked to do the task with the TM system DéjàVu X (DVX). Three types of errors were planted in the TM database: agreement, opposite meaning and number. Out of eighteen segments (nine exact and nine fuzzy matches) that had been aligned for the database, twelve of them had an error planted by the researcher. The subjects were not aware of the existence of errors in the TM database and they had access to the Internet while performing the task. After completing the postediting task, the students were required to fill out a post-assignment questionnaire which contained questions such as how long the postediting and the subsequent revision tasks took, whether they found the TM useful, and to what extent they were familiar with TM systems. The screen activities of the students were recorded by means of a screen recorder. Data from two out of six students could not be used, because these subjects had problems with their version of DVX. These two students were encouraged to translate the text without TM and they were both unable to finish the task within the time limit (two hours). This is given as an indication of TM systems' contribution to production speed and terminology search (Ribas 2008: 30). The overall findings of the experiment confirmed the hypothesis that TM databases containing errors can lead to error propagation. None of the students found all twelve errors, although they were tasked with a short text by having access to the Internet. In addition, three out of four

ISBN: DL:

students did not supply the new, corrected pairs to the TM database, which might lead to serious problems in real life scenarios when the resulting TM database would be used in follow-up or future translation projects. Qualitative data obtained from the questionnaire showed that, on average, the students completed the task in 71.41 minutes. All of them revised the output after exporting, and half of them did this revision both within the TM system and after they exported the text to MS Word. All subjects found the TM useful for completing the task, despite acknowledging the existence of several mistakes in the TM. All of them had a few months experience working with TM and three out of four felt somewhat comfortable working with TM. Although screen recording software was used, we could not find any analysis regarding screen recording data.

For the second experiment within the same pilot study, Ribas (2008) refined her experiment design based on the results of the first experiment conducted with students, and replicated it with four professional translators. A job offer was posted on the website proz.com to hire the professional translators. Although the translators might conclude from the course of actions (i.e. receiving the consent form, filling out a post-assignment questionnaire, etc.) that they were being hired for an experiment, they were not otherwise informed that their translation assignments would be used to collect data for research. They were paid for their translations, in keeping with the professional environment. They were also unaware of the existence of the errors planted in the TM database. A 598-word instruction sheet for a digital thermometer - a complete genuine text - was used as the experiment text to be translated by the professionals from English into Spanish. The task files containing the source text, TM database, instructions, etc. were sent to the professionals via e-mail and they were given four hours to complete the task. This time, professionals were free to use whatever TM they might like to use. All of them chose Trados 7 Freelance software as the preferred TM system. One participant could not open the database on the TM system, and proceeded without using it. Therefore, data from this participant was discarded. As is the case with the students, none of the three professionals found all the errors. More important, none of the subjects corrected errors in the TM database, even though they had been informed that the database they would send after the task would be used for future translation projects. The professionals completed the translation assignment in 75 minutes on average and devoted a third of their time to revising it. “This is greater than the proportion of time

ISBN: DL:

the students spent revising their translations" (Ribas 2008: 45). Two participants revised the output within the TM system and did not revise after exporting it. Only one of them performed revision both within the TM system and after exporting the text to MS Word. Two of the subjects found the TM database useful, however the third one said it was not accurate and therefore not useful. It is interesting to note that the translator with the most experience in the profession (15 years), yet having the least experience in the use of TM (a few months), was also the one who found the fewest errors (Ribas 2008: 47). Another important point about this study was the unprofessional conduct and abilities of the professional translators contacted through proz. One out of four of them said s/he could use Trados, however s/he did not use it due to his/her failure in opening .tmx database in Trados 7 Freelance. This translator did not inform the researcher about the situation and completed the task without using TM. Moreover, two other translators were not able to open .tmx database in Trados 7 Freelance, either and asked for the TM in another format. In addition, none of the professionals cared about correcting the errors in the TM database, which was important for preventing error propagation in future translation projects. Finally, Ribas (2008) underlines that although the experiments with both Translation students and professionals seemed to confirm the hypothesis of the study, further research with larger sample would be necessary.

Yamada (2011) studies how productivity is affected by the content of TM databases. A pilot study was conducted with eight student translators to investigate the impact of free versus literal TM databases on productivity. The subjects were Masters students at the Monterey Institute of International Studies, with diverse backgrounds (some came directly after graduating from undergraduate degrees, some had professional translating experience). Half the subjects were Japanese native speakers while the rest were English native speakers. Both groups are stated to be proficient in English and Japanese. Based on the same source text, two different types of TM database were prepared for the experiment. The first database involved free translations of the source segments (free-translation content, TM-F). The second database contained literal translations of the same source text (literal translation content, TM-L). It was hypothesized that "TM-L would correlate with faster translation speeds than TM-F" (Yamada 2011: 65). Subjects were divided into two groups as TM-F and TM-L and asked to translate a 500-word text taken from an anti-virus software manual from English into Japanese. The source text for both TM-L and TM-F databases were the

ISBN: DL:
same and contained identical match rates. No time restriction was imposed and the subjects were allowed to have access to the Internet and use their own computers. SDL Trados 2007 was used as the TM system, which was stated to have the largest market share (80 percent) in Japan. Postediting processes were recorded with the screen recorder BB Flashback and post-experiment commentaries were collected. Production speed was measured as word-per-minute. Initial results showed insignificant difference between TM-F and TM-L in terms of processing speed. Contrary to what was hypothesized, TM-F was marginally faster than TM-L. Further analysis of the data was made by subdividing segments into match-rate categories. This showed that "production speed with TM-L is equal to or higher than with TM-F in all fuzzy match categories" (Yamada 2011: 69). The biggest difference between the processing speed for TM-F and TM-L was recorded for the match rate of 90-99 percent. At this rate level, the speed for TM-L was nearly twice that of TM-F.

Lange and Bennett (2000) carried out a case study (six-month pilot project) at Baan Company to explore the advantages and disadvantages of combining translation memories and machine translation for both the translation process and the translators. Baan is a company that provides Enterprise Requirement Planning Systems. In order to make its products available to a wider audience, it offers multilingual support with translations of its software into a number of languages. The specific aim of the project was to explore how a combination of TM and MT could be used to "reduce throughput time for translations by 50 percent" (Lange and Bennett 2000: 204). After the project team evaluated various MT systems, they decided to run a pilot project in Hannover, Germany, for English-German translation using Logos as the MT and Transit as the TM system. Before the conduct of this pilot project, the company introduced standard Baan English (simplified English that minimizes ambiguities). The project was carried out in four phases: (1) Text analysis and Logos as a stand-alone system, (2) Logos combined with TM, (3) analysis and readjustment of the translation workflow, and (4) enhancement of Logos output. Standard on-line help files from Baan's Enterprise Requirements Planning System targeted at markets in Germany were used as experiment texts. As far as we could see, details such as the number of participants and the methods of data analysis were not stated in the paper. At the end of the pilot project, throughput times for manual (human), TM-based and TM/MT-based translations were compared. The findings showed that if everything ran smoothly and everybody

ISBN: DL:

cooperated, throughput times could be reduced by 50 to 60 percent with the combination of TM and MT. Lange and Bennett (2000: 208-209) also mentioned the human factor and underlined how important the positive attitude of the translators towards MT was for the success of the project. They concluded that the negative attitude of a translator may even lead to postediting time taking longer than translating the source text from scratch (Lange and Bennett 2000: 209) and that the envisaged increase in productivity could only be possible if translators were comfortable with their new role as posteditors of MT/TM output. After this pilot study, the company decided to move forward with MT implementation and began to provide training for the translators. Baan's experience showed that "motivation increased dramatically after individual training (including training of freelancers)" (Lange and Bennett 2000: 209).

O'Brien (2006a) studied eye-tracking and translation memory matches to test the usefulness of eye-tracking as a method to investigate the interaction between translators and technology. More specifically, the study focused on "the cognitive effort required from translators for different match types in a TM system, with a particular emphasis on comparing matches generated by MT systems with other match types" (O'Brien 2006a: 187). It was pointed out that the main motivation behind this study, which compared the cognitive load for different TM match types with the cognitive load for MT matches, was the assumptions made in the industry about "the effort required to process MT matches (and the amount a translator should consequently be paid)" (O'Brien 2006a: 187). Four professional translators from the IT Company Symantec, familiar with using SDL Trados, translated a 235-word text from English into German and French (two translators worked into German and the other two into French). The source text contained a message from an anti-virus system and it was a typical text type that the subjects were used to working on. The text was manipulated by the researcher to make sure that there were adequate examples of each match type (exact match, fuzzy match, no match and MT match). Translation memories containing matches of different types were provided by the company. As MT matches, matches from the rule-based MT system Systran were used. The subjects were asked to postedit the source text containing no matches, exact matches and fuzzy matches from Trados and MT matches from Systran. Their eye-gaze data were collected with Tobii 1750 eye-tracker. After the task was finished, the postediting process was replayed and the subjects' comments on their activity were recorded with the screen recording software Camtasia. Processing

ISBN: DL:
speed (words per second) and the percentage change in pupil dilation were used as measures of cognitive effort. The study revealed strong correlations between percentage change in pupil dilation and processing speed for different match types. O'Brien (2006a) found that translators exerted the least cognitive effort while processing exact matches and they spent the greatest effort when they were processing no matches. The findings of the experiment also demonstrated that the cognitive effort increased while fuzzy match value decreased. However, there seems to be a disruption in the correlation between the cognitive effort and fuzzy match value as far as the fuzzy match values of 80-74 percent are concerned. From the perspective of machine translation research, it is interesting that MT matches exerted similar cognitive effort as 80-90 percent TM fuzzy matches. O'Brien (2006a: 200) underlines that these findings need to be validated in further studies with more participants, text types and language pairs, etc.

Based on her findings in the abovementioned study, O'Brien (2008) investigated the relationship between fuzzy match value and cognitive effort in more detail, again using eye-tracking as part of the methodology. A hypothesis was formulated stating that "the relationship between cognitive load and fuzzy match value was not a straightforward linear one" (O'Brien 2008: 80). To test this, an experiment was carried out where fuzzy matches of values ranging from 52 to 99 percent were put into the translation memory system SDL Trados Translator's Workbench and postedited by five participants. The participants were either in their final year of an undergraduate degree in Translation Studies or on their first year of an MA in Translation Studies at Dublin City University. They had taken a module in technical translation from German to English and also completed a module on Translation Technology involving training on translation memories. They postedited 25 segments (348 words) from German into English using the TM database seeded with matches of different fuzzy values. The source text was taken from the eCoLoRe resources website and was originally created by the company SAP. The text content was technical and related to computer application. During the experiment, participants could not look up terminology with electronic or paper-based resources because this would have an impact on the processing speed for each match type. The participants' eye movements were recorded with a Tobii 1750 eye-tracker while they were postediting the TM matches. In order to find out if there was any correlation between quantitative data (eye-tracking) and qualitative data, the participants were provided with a paper-based survey once they

ISBN: DL:

completed the task. The survey involved the source texts and the fuzzy matches used in the experiment. It served for determining the participants' perceived postediting effort for each match by asking them to rate the TM matches between 1-5, where 1 meant "Read-only, no edits", while 5 meant "a lot of editing effort". In addition to the qualitative survey data, two types of quantitative data were analyzed as indicators of cognitive effort: processing speed per match value (words per second) and pupil dilation. Data analysis confirmed the hypothesis and showed that the relationship between cognitive load and fuzzy match value was not straightforward. Both the processing speed and the participants' perceived effort indicated that decreasing fuzzy match value might mean increasing cognitive effort. Pupil dilation data also pointed to an increasing cognitive effort until the fuzzy match value reached 60-69 percent. However, below this fuzzy match level (50-59 percent), the trend was interrupted and reversed. O'Brien suggests that the participants might have reached a baseline cognitive effort when they reached the 60-69 percent match class (O'Brien 2008: 89).

Key data on the previous empirical research on translation memories that have already been mentioned above and grouped according to the research questions are provided in Table 2 :

ISBN: DL:

Table 2. Key data on the Previous Empirical Research on Translation Memories

Research	Number of Participants	Participant Profile	Text-type	Number of Words	Language Direction	TM or MT System Used
Christensen (2003) cit. in Christensen and Schjoldager (2010)	Corpus-based study	Authentic data is procured directly from law companies. No information was collected on who translated the texts	Legal texts (articles)	2 corpora : (1) a parallel corpus of 11 articles from 11 law companies, (2) a reference corpus of model translation from a book on company law)	Danish-German	Trados as TM
Dragsted (2004)	12	6 professional translators, 6 Translation students	Business letter and legal contract	2 texts, 212 and 127 words	Danish-English	Trados as TM
Colominas (2008)	Not stated	Not stated	EU and UN texts	2 corpora Europarl and UN. From Europarl 7043 phrases, from UN 4463 phrases	English-Spanish	Trados as TM
Fulford and Granell-Zafra (2005)	391*	Freelance translators	No text	-	German, French, Spanish-English	No tool used
Dillon and Fraser (2006)	59*	Professional translators	No text	-	Various languages-English	No tool used
Lagoudaki (2006)	699*	Translation professionals	No text	-	Various languages	No tool used
Ribas (2008) 1.phase	6	MA students of Translation	Instructions	302 words	English-Spanish	Déjà Vu X as TM
Ribas (2008) 2.phase	4	Professional translators	Instructions	598 words	English-Spanish	Trados as TM
Yamada (2011)	8	MA students of Translation	Anti-virus software manual	500 words	English-Japanese	Trados as TM
Lange and Bennett (2000)	Not stated	Professional translators	On-line help files	Not stated	English-German	Logos as MT, Transit as TM
O'Brien (2006a)	4	Professional translators	Message from an anti-virus system	235 words	English-German/English-French	Trados as TM, Systran as MT
O'Brien (2008)	5	Final year UG or 1st year MA Translation students	Technical, computer application	348 words	German-English	Trados as TM

* Survey-based study. No experiment was carried out.

ISBN: DL:

4.2. Previous Research on Machine Translation (MT) and Postediting

In parallel with the rise of MT and the incremental integration of machine translated segments into the translation workflow as TM input, empirical research on MT gained momentum from the turn of the new millennium.

One of the most extensive studies on postediting was written by Krings (2001). The main research questions in this study were as follows:

(1) From a psycholinguistic perspective, how does the process of postediting machine-translated text work, either with or without source text by human translators? How great is the associated cost and effort compared to human translation? (2) To what extent are raw machine translations comprehensible even without knowledge of the source text? (3) How do translators approach the complex task of translating text from one language into another or postediting a machine-translated text, and which sub-processes can be observed?

In order to investigate the questions, data were collected in 48 two-hour experiments with a total of 52 subjects in thirteen different experimental groups, using the rule-based MT systems SYSTRAN and METAL. Think-aloud protocols were used to investigate the MT postediting process. Five main data sets were obtained and analyzed both qualitatively and quantitatively. The data comprised postediting MT outputs of English texts with source text, postediting the same MT outputs without source text, pure human translation of the same English source texts, postediting the MT outputs of French texts with source texts, and postediting the MT outputs of the French texts with source texts but without Thinking Aloud. In addition, German texts translated by METAL were postedited without source texts. Experiments were carried out in a television studio where the subjects could be filmed with two cameras which could record both the subjects' postediting process and their use of reference works. Subjects were the students enrolled in the technical translation program at the University of Hildesheim who were taught technical translation into two foreign languages and how to use machine translation systems. Thinking Aloud was the main method of data-collection in this study. The subjects' verbalizations were transcribed completely into TA protocols and their time was measured. Comparison of the data sets with and without Thinking Aloud showed that processing speed in the experiment without Thinking Aloud was an average of one-third faster than postediting speed in the

ISBN: DL:
experiments with Thinking Aloud. However, it is reported that over the course of the experiments, the slow-down effect of the Thinking Aloud decreased (Krings 2001: 526). Krings (2001: 529) also notes that "Thinking Aloud not only externalizes cognitive processes during text production, but at least partially also changes them."

According to Krings (2001: 532) "the time necessary to remove deficiencies from a machine-translated text and give it the quality of a human translation is often seen as a direct measure of all postediting effort." Therefore, "the time savings that occur (or should occur) in comparison to a pure human translation are accordingly the most important characteristic value in the calculation of the economic viability of machine translation" (Krings 2001: 532). In Krings' study, processing speed was recorded as the amount of source text processed per minute. The results showed that the processing speed of postediting without source text was higher than that of postediting with source text, which was in turn, higher than that of translation. The differences were minor as far as postediting on paper is concerned (seven percent less than human translation of the same texts). As for postediting on computer (which represents the prevalent translation practice in 2013), processing speed for postediting was even twenty percent higher than the comparative value for translation. In addition to processing speed, cognitive postediting effort was also among the criteria that described postediting effort. Cognitive postediting effort which cannot be observed directly included "the type and extent of the cognitive processes that must be activated in order to remove a given deficiency in a machine translation" (Krings 2001: 533). While analyzing the cognitive postediting effort, it was also found that working with the source text, machine translation and the target text led to an additional cognitive load when compared to the cognitive load deriving from human translation. In order to explore this situation, the author analyzed the frequency with which the subjects changed their focus of attention between the source text, MT output, target text and the reference works. The results of the analysis showed that cognitive postediting effort (measured by the attention focus changes) led to a considerable increase in the cognitive effort spent by the subject when compared to translation.

The study also analyzed postedited MT outputs (with and without source texts) and human translation in terms of linearity of processing. The author means "linearity" in terms of flow of physical writing process and considers it an indication of cognitive effort (Krings 2001: 538). As far as linearity is concerned, the results showed that cognitive effort was the least when a source text and a MT output were available during

ISBN: DL:
postediting; it was higher in the human translation process, and the highest when postediting is performed only with the MT output (Krings 2001: 538).

Krings (2001: 539) also touches upon the relationship between MT output quality and postediting effort. Interestingly, he found that "machine translations that received high values in the expert ranking were postedited much more quickly than those rated lower." Another interesting finding of the study was that postediting effort for medium-quality MT output was higher than for poor MT output. This is explained by the availability of a number of elements in medium-quality MT, which required extensive comparison between the MT output, source text and the target text. However, for postediting poor MT output, the process is close to the traditional human translation process that deals with only the source text and the target text. This in turn, decreases the level of cognitive load.

The overall results show that there are clear structural differences between the processes of translation, postediting with source text and postediting without source text. However, most of these differences are of a quantitative nature than of a qualitative nature. According to Krings (2001: 555), even if it is handled without source text, raw MT output provides information to some extent. However, repairing its defects requires high cognitive effort. The study showed that postediting with source text was very similar to human translation with regard to the types of cognitive processes and with regard to the frequency distribution of such processes. An interesting finding is that the most visible reduction in the effort brought about by postediting was in the area of reference work use. Thanks to MT output, the subjects required less reference work than in the case of human translation where no MT output is used. The study concludes that "practical use of machine translations is still to be considered fairly limited" (Krings 2001: 554).

Bruckner and Plitt (2001) carried out a study to evaluate the operational benefit of using machine translation output as translation memory input. They used the ISLE taxonomy (developed for the evaluation of machine translation) as the basis for their evaluation. Speed, quality and user acceptance constituted the primary features to be evaluated and the targeted first-hand beneficiaries of the outcome were localization managers and translators. In order to create a method for finding a suitable TM penalty value (or fuzzy match value) for MT segments, Bruckner and Plitt (2001) conducted an experiment where they had two teams of equally qualified professional translators translate four sentences of software documentation drawn from a real-world corpus. The source language was English and the target language was German. One group used TM without MT input, while the other group used TM with MT input. Customized Systran was used as the rule-based MT and Trados Translator's Workbench was used as the TM system. By default, Trados Workbench applied a penalty of 15 percent to MT segments to express divergence of MT-generated segments from the 'perfect' translation as a percentage rate. In other words, when an MT segment had a penalty of 15 percent applied, then it had the same score as an 85 percent fuzzy match in a translation memory output. Bruckner and Plitt (2001) set out to find out whether this 85 percent MT match proved to be better than 84 percent or a lower fuzzy TM match. The test sentences already contained fuzzy matches of different percentage rates (i.e. 89 percent, 53 percent, 76 percent, etc. for TM output and 85 percent MT output). In order to evaluate these fuzzy matches from TM against the MT output rated at 85 percent, the researchers counted word deletions, word insertions and word changes (both changes in the word position and changes in morphology) that were necessary to bring the MT output to the desired quality. Then they compared the result with the number of corresponding operations required to edit different levels of fuzzy matches from TM. The results of the experiment revealed that the penalty of 15 percent assigned by the Trados Workbench for MT-based segments was too low for the test suite. For example, a 76 percent fuzzy match from TM required less postediting than the corresponding 85 percent MT match. Bruckner and Plitt (2001: 4) underline that four example sentences were not enough for a thorough and reliable evaluation study. They acknowledged that more sentences of different length and complexity should be examined, together with the corresponding TM fuzzy matches and MT translations. However, they pointed to the difficulty of finding comparable test sentences with the fuzzy match range between 60 and 90 percent from real-world corpora in the localization industry. Although they mention quality and user-acceptance among the evaluation aspects, Bruckner and Plitt (2001) neither presented any data analysis nor drew any conclusion regarding these aspects. All the same, the study provides useful data which reveal the need to rely on empirical research when determining the suitable fuzzy match value that a MT output should have for an effective integration of MT segments into TM systems.

Bowker and Ehgoetz (2007) conducted a study to explore user acceptance of machine translation output, using quality, time and cost as three parameters for evaluation. They asked professors from the Faculty of Arts at the University of Ottawa, a bilingual institution in Canada, to evaluate three different target texts of the same

ISBN: DL:
source texts. A total of seventeen administrative memos, each of which were less than 200 words, were translated from French into English. Three different target texts of each source text were produced: raw MT output, postedited MT output and human translation. As the MT system, rule-based system Systran was used. Given the nature of the source texts (administrative memos which contained perishable information), the researchers used rapid postediting (RPE) as the type of postediting. In order to improve the raw MT output, first, they translated the source texts with the MT system, identified unknown terms and added entries for these terms to the MT dictionary. Then, they translated the same texts again and used the resulting output as the raw MT output. Once the sets of three different target versions of the seventeen source texts were procured, packages were prepared. Each package involved a cover letter providing information about machine translation, an invitation to participate in the recipient evaluation survey, three different versions of a single text, information about the time and cost required to produce each of the three versions and instructions about how to respond to the survey. These packages were sent to 121 professors who were asked to evaluate the target texts balancing the overall quality alongside the time and cost required to produce them. However, only 31 responses were received. "Of the 31 responses received, 21 (67.7 percent) chose the postedited MT version as the text that they would be most satisfied to receive when balancing the quality of the text against the time and cost required to produce it" (Bowker and Ehgoetz 2007: 220). Despite their satisfaction with the postedited MT output, some subjects suggested that these target texts should bear a note indicating that they were postedited MT outputs. According to them, unavailability of such a note might give the reader a wrong impression as regards the linguistic skills of the author of the source texts. Meanwhile, the rest of the subjects (10 professors, constituting 32.3 percent) stated that they would prefer high quality human translations even though they were more expensive and took longer to produce. The authors pointed out that the reason for this attitude might be the fact that many of the participants were language specialists of one type or another who might be more critical about quality than the clients would be (Church and Hovy 1993: 253; Wagner et al. 2002: 90-91; Hutchins 2001, cit. in Bowker and Ehgoetz 2007: 220). As for the raw MT outputs, the subjects found them "completely unacceptable", because they contained errors and it took too long for readers to work out the message. Therefore, it was concluded that, "any time saved by the translator in producing the target text was lost many times over by each reader" (Bowker and Ehgoetz 2007: 221). To sum up, this

ISBN: DL:
study showed that slightly more than two-thirds of the subjects preferred postedited MT output against human translation of the information-only types of texts (administrative memos) when they evaluated the target texts balancing the parameters quality, cost and time. The authors underlined that the translators in the university's in-house translation service indicated that they also preferred such short-lived documents to be machine translated and postedited so that they could invest more time on challenging and interesting texts. Finally, the researchers suggest that similar studies might be conducted with subjects having different profiles (i.e. subjects from the Faculty of Engineering) and in the opposite language direction (English to French) (Bowker and Ehgoetz 2007: 221).

O'Brien has carried out a number of studies exploring various aspects of postediting machine translation and combining MT with translation memories. In her paper on teaching postediting (O'Brien 2002), she emphasizes the necessity of providing proper training in postediting in translator-training programs. Moreover, O'Brien touches on the distinction between postediting and revision, which was also underlined by Schäfer (2003). After listing the skill sets required of a posteditor, O'Brien (2002) proposes content for a course in postediting with its theoretical and practical components.

O'Brien (2006a) also conducted a study which focused on "the cognitive effort required from translators for different match types in a TM system, with a particular emphasis on comparing matches generated by MT systems with other match types" (O'Brien 2006a: 187). MT matches were found to exert similar cognitive effort as 80-90 percent TM fuzzy matches. Further details were provided above in chapter 4.1.

Guerberof (2008) carried out research extending the abovementioned study by O'Brien (2006a) to explore the relationship between productivity and quality of the postediting of outputs from translation memories and machine translation in relation to texts translated with no aid. Quality was measured as the number of errors in the target text. The errors were defined and weighed according to the LISA QA Model 3.1. Processing speed was measured as the number of source words processed per minute. Guerberof (2008) asked eight professional translators to translate, from English into Spanish, a 791-word text containing new segments and to postedit other segments that were either machine-translated (using Language Weaver's statistical MT) or translation memory (using Trados) segments from the 80-90 percent value. The experiment text was taken from a supply-chain software product and contained 265 words of new

ISBN: DL:

segments, 264 words of TM segments and 262 words of machine-translated segments. The subjects did not know the origin of the segments and they were asked to complete a questionnaire after the task was finished. She found that translators had higher productivity and quality when using MT output as compared to using fuzzy matches from the translation memory (TM). Guerberof (2008: 43) explained this by underlining the relative advantage of postediting MT output over TM segments due to the explicitness of typical and clear MT errors as compared to the naturally flowing TM segments involving errors. She further underlined that translators' technical experience seemed to have an impact on productivity, but not on quality.

In her doctoral research, Guerberof (2012) extended the abovementioned project. She investigated the correlation between the machine-translated segments and fuzzy matches from TM in terms of productivity, final quality and experience. In order to test these variables, she conducted an experiment with twenty four professional translators using an on-line postediting tool and a customized Moses statistical-based machine translation engine with a BLEU score of 0.6. [BLEU (BiLingual Evaluation Understudy) is a method for automatic evaluation of MT output. The main idea behind this metric is that "the closer a machine translation is to a professional human translation, the better it is" (Papineni et al. 2002: 311)]. As experiment texts, she used help system and user interface strings from MicroStrategy, a company specialized in business intelligence technology. The translators were asked to translate 149 segments from English to Spanish, working on no-match, machine-translated and translation memory segments from the 85-94 percent value using a postediting tool, without actually knowing the origin of each segment, and to complete an on-line questionnaire after the task. The fuzzy matches were obtained using Trados. The translations were analyzed through a TER (Translation Edit Rate) score calculation and they were reviewed by three professional translators to assess the resulting quality of the assignment. The findings suggest that translators have higher productivity and quality when using machine-translated output than when translating on their own, and that this productivity and quality is not significantly different from the values obtained when processing fuzzy matches from translation memories in the range 85-94 percent. Furthermore, translators' experience and training seems to have an impact on the productivity and on the quality of the posteditings (Guerberof 2012).

Another study by O'Brien (2005) aims to investigate the correlation between postediting effort and the presence of negative translatability indicators in the source

ISBN: DL:

texts submitted to machine translation. The study also attempts to assess the potential of keystroke monitoring (Translog) and Choice Network Analysis for measuring the temporal, technical and cognitive effort involved in postediting machine translation output. Choice Network Analysis compares the renditions of a single string of translation by multiple translators in order to propose a network of choices that theoretically represents the cognitive model available to any translator for translating that string. The technique is favored over the think-aloud method, which is acknowledged as not being able to access automaticized (sic) processes (Campbell 2000: 215; cited in O'Brien 2005: 46). Before embarking on these methods, O'Brien (2005) first questions the suitability of Think Aloud Protocols (TAPs) for measuring the cognitive effort involved in MT postediting. After discussing the prominent characteristics of TAPs, she explains their inappropriateness for the purpose of her research by referring to Krings' (2001) findings indicating that TAPs increased the postediting time by one-third as well as the number of non-linear writing acts.

After deciding to move on with Translog and Choice Network Analysis, O'Brien (2005) took two English sentences and had them machine-translated into German (the MT system used is not specified). The sentences were then edited by four student translators. Within the context of translatability assessment, characteristics of the source sentences that might cause problems (negative translatability indicators) for MT were determined. A negative translatability indicator is a linguistic feature, either stylistic or grammatical, that is known to be problematic for MT (O'Brien 2005: 38). Cognitive effort spent on selecting from among translation alternatives was measured by Choice Network Analysis, while temporal (total processing time, total pause time) and technical efforts (number of words deleted/or inserted) involved were measured using Translog. Choice Network Analysis revealed that posteditors were faced with the greatest number of choices when it came to editing or leaving unedited some of these problematic elements (negative translatability indicators). The analysis of the Translog data revealed that the extended pauses all occurred before the subjects edited those parts of the source sentence that correspond to the difficult items as identified by Choice Network Analysis (O'Brien 2005: 51). This suggested that there was a correlation between the evidence provided by Choice Network Analysis on source-text difficulty and the evidence provided by Translog on cognitive processing, in the form of extended pauses. Both keyboard monitoring and Choice Network Analysis provided evidence of increased cognitive processing around one negative translatability indicator in the source

ISBN: DL:
sentence. However, O'Brien (2005: 55) suggests that further analysis was needed to see if the correlation between the source text difficulty as indicated by Choice Network Analysis and cognitive processing as indicated by Translog was a coincidence or a trend.

In a subsequent paper, O'Brien (2006b) focused on the analysis of pauses as indicators of cognitive effort in postediting machine translation output. In this study, she presented data from the abovementioned project (O'Brien 2005) which included the analysis of cognitive, temporal and technical efforts involved in postediting, triangulated with Choice Network Analysis method and Translog. Unlike the previous study, in this paper (O'Brien 2006b) the focus was on cognitive effort only. Nine subjects were asked to postedit machine translation output from English into German. The source text "contained a mixture of segments with negative translatability indicators (e.g. long noun phrases, gerund, passive voice) or with minimal occurrence of such indicators" (O'Brien 2006b: 7). Each posteditor's activity was recorded using Translog. In the Translog analysis, the amount of time spent pausing while processing each segment was taken into account. Data analysis showed that the pause behavior for postediting sentences with "minimal" negative translatability indicators was not significantly different from that of postediting sentences containing negative translatability indicators. Although this study did not demonstrate significant correlation between pauses and source-text difficulty, the author did not want to "dismiss pause analysis as a potential indicator of translation difficulty or cognitive rhythm" (O'Brien 2006b: 17) relying on the general acceptance of pauses as indicators of cognitive effort. Further research was called on to have better insight into the question.

Schäfer (2003) reports on a study based on a project conducted at SAP (a software developer) dealing with the task of postediting MT output. The project derived from the need for collaboration in postediting research and for developing common postediting guidelines which would not only provide translators with information on postediting but also encourage them to adopt a more open-minded attitude towards working with MT. The specific purpose of this project was threefold: (1) to define postediting as a linguistic task in its own right, similar to translating and proofreading in some ways, yet not quite the same as either; (2) to describe the steps into which the postediting task can be broken down in the light of the experience gained in the daily translation business at SAP; and (3) to present a general typology of MT errors that can be used as a standard benchmark for postediting MT output (Schäfer 2003: 3). Four

ISBN: DL:

rule-based translation engines working with several language combinations deployed at SAP were used in the project: Logos (English-French and English-Spanish), Promt (English-Russian and English-Portuguese), Metal (German-English) and Logovista (English-Japanese).

Schäfer (2003: 4) points out that the analysis of the samples of raw MT output coming from different MT systems showed that “the errors found had a lot in common. Some errors occurred in all language pairs, irrespective of the system used.”

In the context of defining postediting as a linguistic task, three steps were identified and explained: (1) general output check, which involves “checks to identify the most important defects in the target language such as incompleteness or words not translated by the MT system” (Schäfer 2003: 4); (2) editing the MT output, which represents the main task in the process and focuses on repairing the MT output sentence by sentence, and (3) proofreading, which comes after the target text has been completed and aims to ensure that the translation contains no semantic errors, is idiomatic and stylistically adequate in the target language. Schäfer (2003: 5) further provides a general typology of MT errors occurring regardless of language pair and MT system. The main error categories are lexical errors, syntactic errors (sentence/clause analysis, syntagmatic structures, word order), grammatical mistakes (tense, number and active/passive voice) and errors due to defective input. Schäfer (2003: 4) explains the rationale behind this classification as follows: (1) to make posteditors aware of the main types of error that can occur when using MT, and to help train them in this kind of task more efficiently, (2) to provide a systematic framework for continued research in the field of postediting, (3) to provide an overview of the necessary corrections and enhancements to be carried out in the corresponding MT system, especially in its MT dictionary.

Comparison of postedited machine translation with human translation has been an attractive topic in MT-related research. Guerra (2003) explored whether full postediting of a machine-translated text provided output faster than translating the same text from scratch without using MT. She set up an experiment to compare the time spent on fully human translation and full postediting of MT output. Promt XT (English to Spanish and vice-versa) was used as the rule-based machine translation system in this project. Fully human translation of two 841 and 994-word marketing brochures from English into Spanish were carried out by a professional Spanish translator. The same source texts were run through the MT system and then postedited by the researcher, Guerra, herself.

ISBN: DL:

For the human translation task, Guerra (2003) measured time taken to read the source text, search for terminology, translate the source text and revise the target text, whereas in full postediting task she measured the time taken to read the source text, run it through the MT system, identify unknown and mistranslated terms, search for their meaning, code entries in a user dictionary, reprocess with automatic translation, fully postedit the MT output and revise the target text. The results showed that considerable time could be saved when the source texts were first machine-translated and then postedited rather than being human translated. When compared to full human translation, postediting the MT output led to time savings of 41 percent for the first brochure and 57 percent for the second brochure (Guerra 2003: 58).

In addition to these first two texts, a third text (635 words), again a marketing brochure, was first human-translated and then machine-translated and postedited by Guerra herself. For this third text, Guerra (2003: 33) attempted to justify carrying out both the human translation and the postediting of the same text successively by the same person, stating that in the real industrial environments where MT was implemented, companies did not usually employ novice translators for MT postediting. She implied that postediting MT output of an already translated text was representative of the real world contexts where posteditors were chosen from among the experienced translators who were already familiar with the area they were working in (Guerra 2003: 33). Unlike the first and the second texts, automatic translation of this third text was performed with the addition of a dictionary to the MT system to examine "how the dictionary build-up process speeds up MT PE [machine translation postediting] compared with HT that is conducted under normal working conditions for a translator" (Guerra 2003: 55). However, she does not mention about any direct effect of using an MT dictionary on the results. As in the second text, when compared to human translation, 57 percent of the time was saved by postediting the MT output of this third text (Guerra 2003: 58). Although Guerra (2003: 33) presents her justification for first conducting human translation and then MT postediting of the same text by the same person (herself), in order to eliminate this "same person effect", she provides the resulting time saving in postediting by deleting the time spent (20 percent) on reading the source text in the human translation process. Nevertheless, she underlines that there was still a time saving of 37 percent in MT postediting (Guerra 2003: 58). Based on her findings, Guerra suggests that, compared to human translation, postediting MT output brings a pronounced increase in productivity. Although the study emphasizes the possibility of achieving similar high-quality translations by postediting the MT output rather than performing human translation (Guerra 2003: 59), no information on the final quality of the postedited and human-translated outputs is provided.

Fiederer and O'Brien (2009) also explore MT from the perspective of comparing it with human translation. They conducted a study to investigate if machine translation output is necessarily of lower quality than human translation. They designed a study in which 30 sentences from a user guide in English was both human-translated and machine-translated (and postedited) into German. IBM and WebSphere were used as MT systems. Eleven raters rated these 30 source sentences, three human-translated and three postedited versions of them for the parameters of clarity, accuracy and style. The postedited MT outputs were judged to be of higher clarity and accuracy, while the human-translations were judged to be of better style. When they were asked to pick their favorite translated sentences, the majority of the evaluators preferred human translations.

Oflazer and Durgar El-Kahlout (2007) investigated different representational units in English-Turkish statistical machine translation. Before introducing their study, they underlined the lack of large-scale parallel text resources in English and Turkish as a problem of practical significance for English-Turkish statistical MT. It was hypothesized that using lexical morphemes as units in parallel texts instead of surface morphemes might improve the statistical results by conflating statistics for seemingly different suffixes (Oflazer and Durgar El-Kahlout 2007: 26). An example was given to show how two words that looked very different on the surface level might be similar if they were represented at the morpheme level: "although the words 'evinde' (in his house) and 'masasında' (on his table) look quite different, the lexical morphemes except for the root are the same: ev+sH+ndA vs. masa+sH+ndA" (Oflazer and Durgar El-Kahlout 2007: 26). In order to test their hypothesis, the researchers set up an experiment with a parallel data consisting of 45.709 sentences in English and the same number of sentences in Turkish. The texts were composed of documents in international relations and legal documents from the sources such as the Turkish Ministry of Foreign Affairs and the EU. The authors do not mention which statistical MT system they used in this study. As for the methodology, they segmented the words in the Turkish corpus into lexical morphemes so that "the differences in the surface representations of morphemes due to word-internal phenomena were abstracted out to improve statistics during alignment" (Oflazer and Durgar El-Kahlout 2007: 27). Data analysis showed that

ISBN: DL:
employing a "language-pair specific representation somewhere in between using full word forms and fully morphologically segmented representations, and using content words as additional data provide a significant boost in BLEU scores" such that the score was increased from 20.22 BLEU points to 25.08 (Oflazer and Durgar El-Kahlout 2007: 31-32).

The availability of free on-line MT systems working with a large number of language pairs has made research on those systems interesting and attractive. Izwaini (2006) used three online MT systems – Google, Sakhr's Tarjim and Systran – to "diagnose the faults, find out the reasons, and suggest solutions, trying to shed light on the areas where the right translation solution is missed" (Izwaini 2006: 118). The study involved two directionalities: English-Arabic and Arabic-English. A total of 76 English texts were translated into Arabic using Google, Sakhr and Systran, and 46 Arabic texts were translated into English using the same MT systems. The experiment texts were taken from an introductory course in translation given to Abu Dhabi University translation major students. These texts were to be translated as an in-class activity and home assignments. Text size and complexity were designed to evolve along the course, starting with very short and simple sentences and developing into paragraphs of specialized nature of different areas. Literary texts were not included in the data sets. As part of the analysis, raw MT outputs were evaluated in terms of readability, informativeness and closeness to a reference translation by a human translator. The author suggested that Arabic-English and English-Arabic translation "have some common problems as well as mode-specific problems" (Izwaini 2006: 119). In both EN-AR and AR-EN language pairs, high levels of readability, informativeness and even grammaticality were stated to be achieved with MT in a number of texts. Sakhr took the lead, Google came second and Systran ranked last in terms of producing the correct output. The author focused on the linguistic problems faced by the three systems while processing the source texts by analyzing the output in terms of lexical and grammatical problems. However, we could not see how the three evaluation criteria (readability, informativeness and closeness to human translation) were addressed. According to Izwaini (2006), although some of the output of the three systems represented correct and readable translations going beyond just giving the gist of the subject-matter, translations in general had problems. Some faults or deficiencies in one (or more) phase(s) or component of the MT systems were given as the cause of incoherent translations (Izwaini 2006: 147). It was also stated that the lexical environment and collocations

ISBN: DL:
were very important to help decide the meaning and choose the right translation. The main problems posed by Google Translate were spelling, addition and deletion problems that needed further consideration by the system. Finally, the author drew attention to the too literal translations (untranslated words, mistranslation, awkward syntax, etc.) generated by the Systran system in comparison with the two other systems, Sakhr and Google.

Analysis of MT errors is a significant topic of research on MT, since the result of such analyses can help us detect the deficiencies of the MT systems. User feedback deriving from the results of such analyses can be useful for MT developers to fix the deficient parts of the MT systems, thereby improving them. Chang-Maedows (2008) reports on a study dealing with MT errors in Chinese to English MT systems, conducted to obtain user feedback on three MT systems working (among others) with this language pair. These systems are Google, Systran and Microsoft Translator. The first or first two paragraphs from 30 webpages in Chinese covering a wide range of fields were translated into English using the three MT systems. The raw MT output was analyzed for accuracy and readability. It was hypothesized that the poor output was due to complex structures overloaded with information. Data analysis suggested that Chinese-English MT systems, as represented by the data, produced acceptable translations while dealing with texts having simple parallel structures as well as texts on personnel, assets and services. "Acronyms/abbreviations, names (proper nouns), segmentations, missing information and punctuation" were listed among the most commonly identified MT errors (Chang-Maedows 2008: 5). The main syntactic problems these MT systems had while processing texts from Chinese into English were twofold. The first problem related to the translation of the modifiers involving "de", which resulted in misidentification of the subject of the noun phrase. The second problem resulted from the involvement of multiple noun phrases, especially when they appear concurrently with the modifier "de". Such errors resulted in "misidentification of the subject of noun phrases and mismatching of the relationship among the related noun phrases" (Chang-Maedows 2008: 9). Contrary to what was hypothesized at the beginning of the study, data analysis showed that these "errors may occur even when the syntactical structure is simple" (Chang-Maedows 2008: 9).

Zuo (2010) examined the performance of Google online translation service from the users' perspective. Translations of eleven UN documents (3 resolutions, 2 letters, 2 provisional agendas, 1 plenary verbatim, 1 report, 1 note by the Secretariat and 1

ISBN: DL:
budget) from English into Chinese were used to test and evaluate Google's performance in assisting Chinese professional translators at the United Nations. Findings suggest that Google Translate had a remarkable performance with English-Chinese language pair due to its powerful infrastructure and large database. It was also reported that Google yielded correct terminology and was "better able to make intelligent guesses more frequently when compared to other translation tools such as MS Bing" (Zuo 2010). However, MT output from Google in Chinese was reported to be far from intelligible, particularly at the sentence level due to problems with word order and sentence parsing. Instances of technical problems such as adding and omitting words and wrong translations of numbers were stated as the weaknesses of the system. Nonetheless, it was concluded that Google Translate offered savings in time spent on translating and typing for translators by providing the translators with the option of postediting draft translations. The final conclusion suggests that, when used with postediting in English-Chinese translation of the test documents, Google Translate may not save a lot of time; however, it proves to be "a beneficial labor saver" (Zuo 2010).

In a further study, Pym (2009) conducted in-class experiments using Google Translator Toolkit. The primary aim of the experiment was to encourage the students to discover aspects of their own translation processes and technology. The subjects of the study were nineteen second-year masters students (7 Chinese, 5 Korean, 1 Japanese, 6 French) from the Graduate School of Translation, Interpretation and Language Education at the Monterey Institute of International Studies in California. The students were divided into two groups (Group A and Group B) and asked to translate a 138-word text for publication in an online dictionary of technical terms. Group A was asked to translate just using Internet reference sources while Group B was required to feed the source text through Google Translate, then review and modify the output. Quantitative data was collected through determining the total time taken to produce the final translation. Qualitative data was gathered by a post-experiment evaluation of both MT-based and non-MT based translations by the students in language-specific groups. According to the general result of the experiment, there was no significant difference in the time taken to produce the output with MT and without it, no significant difference between the language groups, and no systematic difference between the qualities of the translations as assessed by the students. The results of this experiment show us how close postedited MT output might be to full human translation in terms of processing speed and quality.

ISBN: DL:

García (2010) also conducted a study to compare translations performed with and without Google Translate in terms of processing speed and quality. He designed and performed tests with fourteen (two groups of seven) trainee translators (educated bilinguals with an interest in translation) using Google Translator Toolkit (GTT) to translate four 250-word extracts (two on legal, two on medical topics) from English into Chinese, either entirely from the source text or after running the source text through Google Translate Engine. The performances were timed, and their quality was assessed by independent experienced markers following Australian NAATI test criteria. "Their results show that, while time differences were not significant, the machine translation seeded passages were more favorably assessed by the markers in thirty-three of fifty-six cases" (García 2010: 7). According to García (2010: 7):

> this indicates that, at least for certain tasks and language combinations — and against the received wisdom of translation professionals and translator trainers — translating by proofreading machine translation may be advantageous.

Vaezian (2010) conducted a study with Translation students to explore the reflections of the use of GTT on the process and the product of translation as well as the perception of this tool by the users. Experiments were carried out with GTT in a translation class of twenty-two senior undergraduate Translation students from the Department of Translation Studies at Allameh Tabataba'i University in Iran. The students were asked to edit translations of political texts from English into Persian generated by Google Translate by using Google Translator Toolkit. The overall aim of this study was to explore the students' response to this on-line tool and further its applicability in the context of translation from English into Persian. The study was conducted in three sessions, each lasting one and a half hours. The first sessions was allocated for the brief introduction of the basic features of Google Translate and the toolkit. During the second session, the groups shared their experience of working with GTT with each other and the teacher, and translated a new text on GTT. During the third session, the students were given a questionnaire on using GTT. The study yielded interesting results. The teacher's observations and the students' class discussions and answers to the questionnaire showed that most students reacted positively to this innovative technology. Some 86 percent of the students stated that they enjoyed working with GTT and 90 percent said that they would like to continue using it in the future (Vaezian 2010: 11). The students were also pleased with the side features of GTT

ISBN: DL:
such as the sharing option, whereby more than one user can edit, read and comment on a translation. Another feature that the students appreciated was the presentation of the source and target texts side by side on the editing interface of GTT, which was regarded by the participants as making translation a more interesting activity for them. The students were also pleased with some of the political terms and phrases offered by Google Translate. Some students stated that the draft translations offered by Google provided them with an overview of the text in the target language.

As for the quality of the MT output generated by Google Translate, the majority of the students (72.71 percent) stated that using GTT had either no or a negative effect on translation quality. The students' major complaint about the quality of Google translation had to do with the grammar and structure of Persian sentences. However, 87.17 percent agreed that translating with GTT can increase translation speed. Although most students generally assessed the quality of the MT output to be unacceptable, their overall reaction to using GTT was positive. The author suggests that incorporating this new technology into the translation class indirectly familiarized the students with the online translation market where translators usually use MT, work collaboratively and communicate by means of e-mail (Vaezian 2010: 12).

Almost a decade ago, Champollion (2001, emphasis original) pointed out the question: "Will the future of human translation be … *proofreading computer output*?" and he answered it: "The bad news is yes. It all boils down to how long it will be until computers produce decent translations." Despite this statement, he also emphasizes that computers do not have common sense which is made up of personal experience gained throughout one's lifetime, the wisdom acquired from education and instinctive inborn knowledge. According to Champollion (2001), expecting computers to act like humans is "dreaming of man recreating man through science" and it is unrealistic. Therefore, he suggests that expectations should be taken to a reasonable level. That is, we should stop expecting computers to understand and translate in the same way humans do. Instead, we should expect them to perform some human tasks with reasonable accuracy. In a similar vein, Guzmán (2007a: 6) underlines that "MT can provide greater benefits and less frustration as long as reasonable expectations are set and complementary strategies are put into practice."

Allen published a number of articles on MT (Allen 2000, 2004a, 2004b, 2004c, 2005a) and postediting of MT output (Allen 2001a, 2001b, 2003, 2005b). In one of those papers, he (2001a) describes postediting as modification and/or correction

ISBN: DL:

performed on a text pre-translated with MT. Allen (2001a: 1) mentions about the increasing implementation of MT, thus postediting into the translation processes due to factors such as "the constant forces in the push for globalization, SIMship strategies, process streamlining, faster turn-around of information, etc." He also defines various levels of MT postediting based on the extent of the modifications to be carried out on the MT output. These are MT with no postediting, rapid postediting, minimal postediting and full postediting. Allen (2001a: 3) underlines that the level of postediting needs to be based on the needs and expectations and may depend on the factors such as the input text (whether it was written with controlled language rules), the required translation turn-around time, the life-expectancy of the target texts and the customer's needs.

In another paper, Allen (2001b) elaborates on the postediting process as an integral part of the translation software program Reverso Pro developed by Softissimo Pro, and he reports on a test conducted with this translation software (Allen 2001b). Four marketing brochures were translated with Reverso Pro from French into English. A 200-word French-to-English marketing dictionary was created starting with the first brochure and was completed at the end of the fourth brochure. The brochures 1 - 4 contained 540, 483, 472 and 222 words respectively and postediting time for them were 22, 20, 20 and 8 minutes. In order to draw a clearer picture on the results, Allen (2001b) states that the processing time of translating each of the four brochures with MT plus postediting them is equivalent to the time necessary for a translator to only type a text of the same number of words at an average of 25-30 words per minute, excluding the time to perform the cognitive task of translating. This small test shows that postediting the MT output from Reverso Pro proves to be more productive than translating the same text from scratch. In addition, the author mentions about the by-products of the system which were created during the postediting process such as customized user dictionary and the possibility of saving the source and the target texts as aligned bilingual texts.

Upon the provision of the first free online translation web service Babel Fish by Alta Vista at the end of 1997, Allen (2000) presented a perspective on the value of such free Internet online translation portals and services to discuss the impact of them on the MT market. After the release of Babel Fish, the Internet translation portal business grew rapidly and a number of free translation portals appeared on the web (Allen 2000: 1). Allen (2000: 2) states that MT system developers made those systems on the Internet freely available "with a timeout feature that is either based on connection time of use or

ISBN: DL:

on file size." In this way, the system developer companies aimed to encourage people to use, test and publicize the systems so that they could create a market for computerized translation programs.

Vasconcellos (1986) reports on Pan American Health Organization's (PAHO) experience with the postediting of English output translated by their MT system, SPANAM, over the seven-year period. He offers a typology of the postediting strategies developed throughout this period and states that these strategies have existed along a "continuum that progresses from the reactive to the proactive" (Vasconcellos 1986: 134). The reactive strategies cover the technical abilities of the posteditor regarding the use of computer whereas the proactive strategies constitute the posteditor's involvement in building dictionaries. Vasconcellos (1986: 145) concludes that the development of postediting skills requires time and it is a "process of gradual growth." Experience of fifteen posteditors who worked at PAHO revealed that the initial judgments of posteditors tend to be reversed as they experience the task and they feel more comfortable with postediting after completing a hundred thousand words of postediting. It also showed that postediting was more relaxing and fun than translating a text from scratch. Further, those who were involved in the development of dictionaries found a new dimension in their work and enjoyed it (Vasconcellos 1986: 145).

Roberts (2007) presents the summary of a postediting research conducted by Linguistic Systems. A 316-word Dutch newspaper article was translated into English with the rule-based MT system, Language Weaver, and postedited at two different levels. The aim of the study was to determine the relationship between different levels of postediting and accurate comprehension of the outputs. Accurate comprehension and correct grammar were regarded as the indicators of quality. Three data points were collected from the 316-word Dutch newspaper article: the time required for (a) a human translation (HT), (b) a postedit using two-thirds of that time, and (c) a postedit using one-third of that time (Roberts 2007). Postediting within one-third of human translation time was called as a Brief Edit. Postediting within two-third of the time needed for HT was called as a Full Edit. Then, an analysis was made to determine the accurate comprehension quality of the two postediting results as a fraction of the human translation quality. The text was divided into elements and an element was defined as "a sentence, whole portion of a compound sentence, or an entire headline if it was a phrase" (Roberts 2007). Each element of the text was analyzed for subject, verb, object and syntax. Accuracy measurement of each item was done using a score of 1-10 and

ISBN: DL:

[Subject + Object] x [Verb/10 x Syntax/10] formula was used to determine the total score for an element. The results of the measurements were compared for raw MT output, Brief Edit, Full Edit and Human Translation. The author underlines that their method of quantifying accurate comprehensibility is basically subjective and should be corroborated by other researchers. The results showed a linear relationship between the amount of postediting and accurate comprehensibility. That is, among the 27 elements defined in the text, the average score for raw output was 0.397. For Brief Edit, this score was 0.894 and for Full Edit, the average score was 0.996. Although Full Edit and human translation were almost identical, HT was superior when the evaluation included style. Roberts (2007) points out that both human translation and postedits were carried out by one and the same translator. It is emphasized that this was important to ensure a valid comparison when measuring each postedit against human translation. However, we have not found any explanation on how he overcame the remembering effect while the same translator was both translating and postediting the same text successively. The author also suggests that a larger investigation with more subjects, more texts and different languages is necessary.

In order for the MT postediting to prove useful and cost-beneficial, it is important to decrease time and effort spent on correcting the raw MT output. This has led the researchers to develop automatic postediting tools which would carry out the initial postediting and correct the most frequent and typical MT errors, thereby reducing time and effort to be spent by the human posteditor.

Guzmán (2007b) emphasizes the potential spread of MT postediting in the future of translation profession, which will bring together the possible increase in the need for this practice. In order to reduce manual posteditings and costs involved, he proposes to automate MT postediting by making use of the potential of regular expressions to search and replace linguistic patterns in raw MT outputs. Experiments with English to Spanish rule-based MT system were carried out using software instruction texts. Nine linguistic categories were created: Misspellings, punctuation, articles, prepositions, grammatical agreement (singular vs. plural), word order, reflexive pronoun (-se), style and redundancies. Within these categories, linguistic patterns were identified based on common recurring linguistic errors and regular expressions were assigned to solve these problems. As a result, the recurring linguistic errors were fixed automatically, therefore they do not have to be fixed by the posteditor. Guzmán (2007b) also states that, in order to test the automatic postediting, he used a rule-based MT system where linguistic patterns were more likely to occur consistently. He underlines that the results might be different in case of statistical-based MT systems and his approach might not be useful as far as SMT systems are concerned.

In a further paper on automatic MT postediting, Guzmán (2008: 1) describes "how regular expressions can be used to search for ambiguous text in the source segments and decide if their translation in the MT output is wrong, in which case it will be automatically postedited." Ambiguous texts include "words, phrases and full sentences that are either missing or have more than one meaning" (Guzmán 2008: 1). Guzmán (2008) analyzed rule-based MT outputs of software instructions translated from English into Spanish. He focused on three linguistic error patterns which were mistranslations of –ing words, subordinate clauses and verbs with several meanings. In order to automatically postedit these mistranslations with regular expressions, one needs to disambiguate them by using linguistic context. To this end, the author suggests to align both the source text segments and their MT translations in a translation memory text format and open this text file in a text editor that supports regular expressions. Then, the text editor might search for the typical error patterns in the translated text that may be caused by an ambiguity in the source text. Finally, the suggested system will replace and correct the error in the MT output. Since most of the recurring ambiguity patterns in the source text are often mistranslated in the same or similar way by the rule-based MT system, this type of automatic postediting for eliminating ambiguities in the MT output seems promising. However, the author warns that such work should be done carefully to prevent possible side effects (Guzmán 2008: 8).

Hogan and Allen (2000) make an analogy between the tasks of the machine translation posteditor and the tasks of the controlled language writer and editor. They claim that "postediting is a task that can be considered as a controlled language processing stage" and propose an automatic postediting module for MT output. Due to the varying postediting criteria, the ongoing reinvention of postediting rules and the cognitive load placed on the posteditors who have to fix minor, but repeated MT errors; Hogan and Allen (2000) decided to develop a prototype automated postediting (APE) module. The main idea behind the APE module was that MT errors were typical and likely to occur repeatedly under the same conditions. An APE module would correct such numerous errors in the MT output, thereby relieving the posteditor from correcting such errors recurring within and across documents as well as over time. In addition to correcting the MT output automatically, this APE module would automatically learn

ISBN: DL:
from previously postedited documents. It was based upon corpus-based analyses that were conducted on sets of parallel "tri-texts (source text, raw MT output and postediting version)". Hogan and Allen (2000) carried out an experimental research on tri-text sets taken from the database of the Translation Services of the European Commission. The initial experiments were performed on 8 sets of English to French tri-text files and 17 sets of French to English tri-text files from the same database. The ultimate aim of the research was to demonstrate that production turn-around time can be affected by using new tools in production environments where MT is used. Detailed analysis of the results of this experiment was not mentioned in the paper. However, the initial results showed that the APE is a viable concept and that "with a significant amount of additional training and test, it would be possible to reduce the number of postediting fixes that human posteditors must perform on raw MT output without changing the internal rules of the MT system itself" (Hogan and Allen 2000: 7).

The amount of postediting to be employed to keep the balance between acceptable quality and minimum time and effort has usually been a controversial issue. In order to take the advantage of postediting over traditional human translation, time and/or effort spent on postediting should be less than those spent on human translation. Guzmán (2007a) discusses this issue and suggests narrowing down the scope of postediting as much as possible. He points out that corrections on the MT output should be performed for the sake of clarity and correct meaning, and stylistic considerations should be avoided. He also provides examples of unnecessary postediting on software instruction texts translated from English into Spanish (Guzmán 2007a). According to Guzmán (2007a), writing the source texts in an MT friendly manner (using controlled language); automating the correction of many typical grammatical and stylistic errors (automatic postediting) can be the steps to decrease the time and effort spent on manual postediting.

Aikawa et al. (2007) investigated the relationships among controlled language (CL), MT quality and postediting. They hypothesized that the use of controlled language would lead to greater productivity or reduced postediting effort. The second aim of the research was to determine the types of CL rules that have the greatest cross-linguistic impact on MT quality. They created two sets of data containing a total of 520 English sentences in technical domain, one which violated the CL rules and the other which conformed to them. Both sets of sentences were translated from English into four typologically different languages - Dutch, Chinese, Arabic and French using MSR-MT,

ISBN: DL:

a statistical machine translation system developed at Microsoft, and postedited by localizers. The MT system was trained on the data from the IT domain. Human evaluation and BLEU scores were used to measure the impact of the CL rules on MT output. For both types of evaluation, the MT output for each sentence in the two different data sets was compared to the corresponding postedited version of that output. The authors examined whether the use of CL improves productivity in terms of reduced postediting effort, using character-based edit-distance (difference between the postedited version and the raw MT output in terms of characters). The results of the experiments supported the hypothesis that the use of CL improves PE productivity as well as MT quality. The study also revealed the three CL categories with the greatest impact on MT quality were formal style, spelling and capitalization. The authors underline that the impact of CL rules on MT quality may vary depending the types of MT systems, which is an area requiring further research.

Allen (2004a, 2005b) describes the use of MT technologies for translating both pre-sales marketing and post-sales software deployment documentation in the software product division of Mycom International. The software used for the implementation was Reverso Expert (v5) which was based on the PROMT v5 MT kernel (Allen 2004a: 2). Two texts were used in the experiment. The first text was a 1494-word pre-sales marketing document. It was translated from English into French with an MT system and then postedited. The second text was an 8364-word customer acceptance test validation plan (a post-sales acceptance document). It was translated from French into English with the same MT system. The implementation was led by an experienced bilingual MT user who was a subject-matter expert in the field and had worked for several years as a translator and translator trainer. Output quality was measured by this subject who was "fully competent to appropriately measure, in both languages, the level of textual quality that is necessary for complete end-to-end care of Mycom customers." Allen (2004a: 3) takes the "translation of 2400 words per day" as the average human translation speed. This rate was based on a previous study (Allen 2004c) which made use of a questionnaire that had been sent to various translation agencies, government translation bureaus, translation departments of international organizations, freelance translators, etc. The entire translation process for the first text lasted three hours. The process involved tasks such as identifying the potential ambiguous terms, deciding on the desired translated form for each of them with another subject-matter expert, dictionary entry coding and testing on the entire document to produce high quality draft

ISBN: DL:

with no postediting. The MT output was postedited by the sales representative who was also the requester of the translation.

The second document was translated in two sessions in a week. The translation process included file preparation, separation, text analysis, dictionary building, identification of poorly translated terms, testing new dictionary entries and postediting. The author underlines that an average human translator would produce 2400 words per day which would mean that it might take a translator 3-4 days to do the translation of the second text from scratch without MT. However, the experiment suggested that a bilingual subject-matter expert with professional translation experience and mastery of an MT system could complete the entire translation workflow (including dictionary building and postediting) in about 11 hours. Thus, with the given MT tool, described subject, texts and for French-English language pair, it would be possible to attain the same level of productivity within 25-30 percent of the time required for traditional human translation (Allen 2004a: 5). The results are promising in that "these types of translation implementation efforts could enhance the ability for the company to write pre-sales marketing and post-sales technical documents in one language, and later translate, adapt, localize them into another language for other customers" (Allen 2004a: 5) In addition, Allen (2004a: 5) points out that, during the translation tasks, a specific custom dictionary for the technical terminology of the documentation of the company was created, and this dictionary could be used for the translation of similar software products.

In a subsequent paper, Allen (2004b: 6-11) provides an overview of different translation approaches according to the purpose of the use of MT output, i.e. inbound vs. outbound translation. He presents the range of different levels of human-MT interaction with respect to the purpose of target text use. Inbound translation (translation for assimilation) is described as translation for browsing information and/or content gisting. This approach can be used with no postediting or with rapid postediting for ephemeral texts to be used for information only purposes. The quality requirement of the output of outbound translation (translation for dissemination), thus the level of postediting to be performed can differ based on the purpose of target text use. According to Allen (2004b: 9-11), outbound translation can be postedited on three different levels: MT with no postediting, MT with minimal (rapid, partial) postediting and full postediting (2004b: 20). Allen (2004b: 20) finally notes that the failure of implementation of automatic translation systems might have been caused by "the

ISBN: DL:

mismatch between real user needs and what the translation system/software was designed to be used for."

Key data on the previous empirical research on machine translation that have already been mentioned above and grouped according to the research questions are provided in Table 3:

ISBN: DL:

Table 3. Key data on the Previous Empirical Research on Machine Translation and Postediting

Research	Number of Participants	Participant Profile	Text Type	Number of Words	Language Direction	TM or MT System Used
Krings (2001)	52	Students in a technical translation program	Instructions	EN: 1352 words total, FR: 762 words total, GE: 325 words total	English-German, French-German, German-English	SYSTRAN and METAL as MT
Bruckner and Plitt (2001)	2 teams (nr not stated)	Professional translators	Software documentation	4 sentences	English-German	Systran as MT, Trados as TM
Bowker and Ehgoetz (2007)	31	Professors at the Faculty of Arts	17 administrative memos	Each text less than 200 words	French-English	SYSTRAN as MT
O'Brien (2006a)	4	Professional translators	Message from an anti-virus system	235 words	English-German, English-French	Trados TM, Systran MT
Guerberof (2008)	8	Professional translators	Supply-chain software product	791 words	English-Spanish	Language Weaver as MT, Trados as TM
Guerberof (2012)	24	Professional translators	Help system and user interface strings in business intelligence technology	149 segments	English-Spanish	Moses as SBMT, Trados as TM
O'Brien (2005)	4	Translation students	Not stated	2 sentences	English-German	MT- Not stated, Trados as TM
O'Brien (2006b)	9	Not stated	Not stated	Not stated	English-German	Not stated
Schäfer (2003)	Not stated	Not stated	Not stated	Not stated	English- French, English-Spanish, English Russian, English-Portuguese, English-Japanese, German-English	Logos, Promt, Metal and Logovista as MT
Guerra (2003)	2	1 professional translator, 1 MA student of translation (researcher)	Marketing brochures	841, 994 and 635 words	English-Spanish	Promt X as MT
Fiederer and O'Brien (2009)	Nr of translators not stated, 11 raters	Not stated	User guide	30 sentences	English-German	IBM and Websphere as MT
Oflazer and Durgar El-Kahlout (2007)	Not stated	Not stated	Documents in international relations, legal texts	45.709 sentences in English and Turkish	English-Turkish	Not stated

ISBN: DL:

Research	Number of Participants	Participant Profile	Text Type	Number of Words	Language Direction	TM or MT System Used
Izwaini (2006)	Not stated	Translation students	Texts from a translation course book	76 texts in English, 46 texts in Arabic	English-Arabic, Arabic-English	Google, Sakhr's Tarjim, Systran as MT
Chang-Maedows (2008)	Not stated	Not stated	Wide range of fields	first/first two paragraphs from 30 webpages	Chinese-English	Google, Systran and Microsoft Translator as MT
Zuo (2010) Abstract				11 UN documents	English-Chinese	Google Translate as MT
Pym (2009)	19	MA students of translation	From an on-line dictionary of technical terms	138	English-Korean, English-Japanese, English-French	Google Translator Toolkit as MT
García (2010)	14	Trainee translators	2 medical, 2 legal texts	250-word 4 texts	English-Chinese	Google Translator Toolkit as MT
Vaezian (2010)	22	Senior undergraduate Translation students	Political texts	Not stated	English-Persian	Google Translator Toolkit as MT
Allen (2001b)	Not stated	Translators	4 marketing brochures	540, 483, 472, 222 words respectively	French-English	Reverso-Pro as MT
Vasconcellos (1986)	15	Posteditors at PAHO	Texts postedited over seven years	Not stated	Not stated-English	SPANAM as MT
Roberts (2007)	1	Posteditor	Newspaper article	316 words	Dutch-English	Language Weaver as MT
Guzman (2007a)	-	Automatic Postediting System	Software instructions	Not stated	English-Spanish	Rule-based MT system (name is not provided)
Guzman (2008)	-	Automatic Postediting System	Software instructions	Not stated	English-Spanish	Rule-based MT systems (names are not provided)
Hogan and Allen (2000)	-	Automatic Postediting System	75 texts from EC Translation Services database	Not stated	English-French, French-English	Not stated
Aikawa (2010)	Not stated	Localizers	Technical texts	520 sentences	English into Dutch, Chinese, Arabic and French	MSR as MT
Allen (2004a)	1	Bilingual MT user, subject matter expert, experienced translator and translator trainer	2 texts (1pre-sales marketing doc., post-sales acceptance doc.	1494 and 8364 words respectively	English-French, French-English	Reverso Expert (v5) as MT

ISBN: DL:

4.3. Subject-Matter Experts in the Translation Process

4.3.1. Explanation of the Term "Expert"

The term "expert" is a contentious one in that it may be used to represent different phenomena. Pym (1996: 139) discusses ideologies of the "expert" in discourses on translator training and suggests that "critical and self-critical questioning might prove to be the most effective antidote to recent outbreaks of expertise." The term "expert" is sometimes used to refer to people who enter organizations as outsiders, usually invited to the institutions to support the changes they want to introduce (Sullivan 1990 cit. Pym 1996: 141). Holz-Mänttäri (1984: 62-68, cited in Pym 1996: 143) approaches expertise from a different perspective by developing the theory of "expert distance". According to Holz-Mänttäri, "translators are experts in the sense that they have no personal, emotional or immediately intuitive involvement in the communication situation" (Pym 1996: 143). Kaiser-Cooke (1994: 135) defines the term expert as "someone who possesses a high level of competence in a given domain which results from interaction between structure and processing abilities, expert performance being characterized by rapid access to an organized body of conceptual knowledge." In this sense, expertise also refers to competence in a specific domain or a subject-matter (i.e. medicine, engineering, biology, educational sciences). As far as translation competence is concerned, "expert" is also used to refer to the experienced professionals who employ some "cognitive routines" occurring "automatically" (Kaiser-Cooke 1994: 137).

In the present study, we use the term "expert" (or "subject-matter expert") to refer to a person with specialized knowledge in one particular field (other than translation) as a result of having received undergraduate education and training in that particular field. In the present study, the "subject-matter" is engineering, thus subject-matter experts are engineers. Professional translators can also be specialized in one or more fields such as technical or legal translation or more specifically, wine translation, etc. No matter whether they are specialized in one particular field or not, translators are also "experts in intercultural communication" (Holz-Mänttäri 1984). However, in order to avoid ambiguity, in the present research, the term "expert" will only refer to "subject-matter expert", as described above.

ISBN: DL:

4.3.2. Previous Research on the Role of Subject–Matter Experts in the Translation Process

The technological improvements leading to the production of personal computers and the World Wide Web have been milestones in the use of machine translation. According to García (2012: 5) "the advent of Free Online Machine Translation can truly be considered epochal: it supplanted personal computer-based machine translation and exponentially expanded the use of MT." This has made MT accessible to any user of the web, for various purposes such as translating electronic mails, getting the gist of a text or translating it to be further postedited by a human translator or subject-matter expert.

Yngve (1954: 21) stated that a "posteditor is a person skilled in the output language but who may be entirely ignorant of the input language", although he pointed out that the posteditors could perform better if they know the source language as well. He also observed that "the posteditor is better able to do his job if he is an expert in the particular field of knowledge."

Although limited in number, the use of subject-matter experts or comparison of them with professional translators has been the focus of interest since the early efforts on postediting/translation research. Orr and Small (1967) rated the usability and comprehensibility of Russian texts in physics, earth sciences and engineering in three versions: raw MT output, postedited MT output and traditional human translation. The participants were students in physics, earth sciences and engineering. Direct answer, paraphrase and inference were used as the methods to measure comprehensibility and accuracy. The results showed that human translation performed better than postedited MT and postedited MT was better than raw MT output.

Van Slype (1979) submitted a report to the EU Commission after the adoption of Systran as its MT system. In this report, raw MT output, its postedited version and the revised human translation were compared with the source text in terms of the intelligibility of the target text. The postediting was performed by two groups: professional translators and revisers, and bilingual engineers and end-users. The first group worked slower. They required twenty-two minutes to process 100 words, while the second group required eight minutes. The final intelligibility level of both groups was the same with a rating of 98 percent. Professional translators and revisers made more corrections (36 percent) than subject-matter experts and end-users (31 percent). Although it is not great, the difference between both groups' correction rate shows that professional translators tend to adopt a much more critical approach to MT than subject-matter experts do, while the latter group benefit from MT assistance much more than professional translators do.

Vasconcellos and Léon (1985) report on the machine translations systems SPANAM and ENGSPAN employed in the Pan American Health Organization. They comment on Kay's (1982: 74) suggestion that the posteditor "would not have to be a translator and could possibly be drawn from a much larger segment of the labor pool." They conclude that "experience at PAHO environment suggests that this conclusion would be valid only for technical experts working on a text for information purposes" (Vasconcellos and Léon 1985: 125). According to the authors, users, even technical experts may misinterpret MT output and assign an incorrect meaning while correcting it. Thus, they argue that the professional translators should be the posteditors underlining that "only an experienced translator will be aware of the words whose variable meanings are dependent on extra-linguistic context" (Vasconcellos and Léon 1985: 125). On the other hand, they also emphasize the significance of the combined background containing training, experience and "good knowledge of the subject-matter vocabulary in both languages, and a technical understanding of what is meant by the text" (1985: 125).

In the context of cooperation between subject-matter experts and translators, Way (2004) reports on a project that united the work of final-year undergraduate Translation and Law students in solving a practical case of international private law. The aims of the project were twofold: (1) To allow the Translation students to work in a real life situation with experts from another field, thereby increasing their interpersonal competence, (2) to introduce the Law students to the possibilities that working with a translator could offer them (Way 2004: 584). The task of the Law students was to resolve a case of international private law applying the knowledge they gained throughout their legal training whereas the Translation students' task was to provide research skills, summaries and translations of material in languages other than Spanish. A total of 78 final year Law students (16 of them were Erasmus students) and 90 final year Translation students participated in the project. The students were divided into 15 groups, one for each practical case to be resolved. The resolution of the practical case formed part of the Law students' final assesment and the translation of the resolution into the Law students' B language (English, French or German) by the Translation students formed part of the Translation students' final assessment. In order to evaluate

ISBN: DL:
the success of the project, two different quaestionnaires were prepared (one for the Law students and one for the Translation students), distributed to the participants and completed anonymously. The degree of satisfaction with the project was assessed using a Likert scale. Additional comments by the participants were also invited. In general, 50 percent of the Law students considered their participation to have been very positive, while the remaining 50 percent considered it to have been positive. Among the Translation students, 10.5 percent considered their participation to have had a very positive effect, 58 percent a positive effect and 31.5 percent neither positive nor negative (Way 2004: 590). When the students were asked if the project contributed to their training, 25 percent of the Law students found the project to have been very useful and 75 percent useful. On the ther hand, 6 percent of the Translation students found the project to have been very useful, 64 percent useful, 15 percent were indifferent and 15 percent considered it of little use in their training. On the whole, the project yielded positive results in that it provided the students with a real-life example of an interdisciplinary cooperation. This project point to the positive outcomes that can come from cooperation between subject-matter experts and translators. The results also show that both parties may benefit from the mutual exchange of knowledge and interdisciplinary cooperation.

In her proposal for a programme to train professional translators, Snell-Hornby (1992) addresses the question of specialization versus all-round competence. She also deals with the question of whether the translator is only a language specialist or needs to be an expert in other fields, too. She suggests that the job profiles of the future will require combinations of various types of qualifications. This applies particularly to multilingual communities such as those in Europe where “we will need multilingual legal experts and export managers beside translators with different kinds of subject area competence” (Snell-Hornby 1992: 15). She further points to the importance of flexibility for future translation programmes, i.e. someone wishing to specialize in legal language could attend a separate course in the Faculty of Law or work on special translation assignments within the Translation degree course. Interdisciplinary cooperation (between the subject area and translation) is also important in cases where students work on theses where different areas intersect (i.e. supervision of a thesis on the terminology of acupuncture by the translation department in cooperation with the Faculty of Medicine) (Snell-Hornby 1992: 17). She concludes that “an effective degree programme should offer a representative cross-section of subject areas (e.g. technology,

ISBN: DL:

medicine, economics, law)", and the training of the future gradutes requires both interdisciplinary cooperation and multifaceted qualifications of the teaching staff.

Künzli (2007) carried out an empirical study of revision where ten professional translators were asked to think aloud while revising three texts that had been translated from French into German. The quality of the revised translations was assessed by a subject-matter expert for each text. The results were analyzed in terms of the changes introduced by the revisers in the draft translations or their failures to make the necessary changes. Among the three texts, Künzli (2007) presented the results from the revision of the legal text only. The analysis shows that when revising draft translations, professional translators do not justify most of the changes they make and the proportion of uncorrected errors in the draft translation is quite high. The percentage of hyper-revisions and over-revisions by the translators reached 30 percent. It is reported that nearly one change in three is either unnecessary or even consists in introducing an error into the draft translation and some of them are serious errors in meaning (Künzli 2007: 124).

Carroll (1966) used subject-matter experts in the evaluation of translations. He carried out experiments for evaluating the quality of both human and machine translations to "lay the foundations for a systematic procedure that could be applied to any scientific translation" (Carroll 1966: 55) He selected two groups of raters for evaluation of translations including outputs from various MT systems as well as various human translations (i.e. published HT, rapid HT, etc.) of scientific texts from Russian into English. The evaluation was based on the parameters of intelligibility and informativeness and the raters recorded the time taken to read and evaluate each sentence. The first group of raters were composed of eighteen male students in their third year of various majors such as chemistry, biology, physics, astronomy or mathematics. They did not know Russian at all and they were selected on the basis of their high verbal intelligence, interest and knowledge in science. Since they did not know the source language, these raters (also called "monolinguals") evaluated the translations by comparing them to their translation performed by two members of the Department of Slavic Languages and Literatures of Harvard University, one of whom was an experienced professional translator of scientific Russian. The second group of raters were composed of eighteen males who were graduate students of Russian or teachers of Russian. They were selected on the basis of their expertise in reading Russian (generally scientific Russian). While performing evaluations, the second group

ISBN: DL:

(also called "bilinguals") compared the translations with their Russian originals. All raters were native speakers of English. Each rater evaluated 144 sentences on three separate occassions. Participants recorded the time it took to read and rate each sentence. They paced themselves and took, on the average, about ninety minutes per session. Monolinguals attained greater reliability in their ratings than did the Russian-reading subjects, while for both groups, the inter-rater variance was smaller for the intelligibility scale than it was for the informativeness scale. Bilingual raters tended to rate all translations a little higher in intelligibility than did the monolingual raters. This is explained with the instructions given to bilinguals to use their knowledge of Russian to divine the meaning of the translations (Carroll 1966: 62). The rankings of the translations in terms of informativeness were complementary to the rankings on intelligibility. There was a negative correlation between the average reading time and the average intelligibility ratings, while there was a positive correlation with the informativeness ratings, results and times. Bilingual readers recorded slightly (but significantly) longer task times than monolinguals. According to the author, the reason might be that the bilinguals' knowledge of Russian impelled them to study the translations more carefully or their lack of scientific knowledge slowed down their comprehension of the texts. The research thus showed that monolinguals could differentiate quality better and faster than bilinguals.

The synthesis report presenting the results of the OPTIMALE project carried out to determine the current and emerging competence requirements within the European translation industry refers to specialization in the translation profession. According to this report, which collected data from 685 respondents (employers, partners in the translation industry), "translation of materials in one or more highly specialized domains" is among the most important and essential competences that the professional translators should possess to ensure optimal quality (for 89 percent of the respondents) (OPTIMALE 2012: 12). Further, the report states that "domain specialization is almost unanimously required by employers, with almost 90 percent giving it priority status". It also emphasizes that, according to the workshop discussions, technical translation is among the dominant market segments (OPTIMALE 2012: 8). In the "additional employer comments" part of the same report, respondents underline the importance of specialist knowledge within a certain field (e.g. medical, technical, legal, etc.), experience and knowledge in the subject-matter. The report also states that the best combination is someone with a law/economics/technical/medical degree and/or relevant

ISBN: DL:

work experience plus linguistic training. According to the additional employer comments presented at the end of the report, translators with an MA in Linguistics will never reach the in-depth subject field knowledge that those subject-matter experts have (OPTIMALE 2012: 17)

ISBN: DL:

5. Background, Aims and Hypotheses

5.1. Background

As we have seen, rapid developments in technology, widespread use of the Internet and globalization have led to an increase in the volume of material to be translated. These factors have paved the way for language service providers to look for ways of increasing efficiency in the translation process. The use of technology (translation memories and machine translation systems) is the most common way of increasing efficiency.

In order to meet the growing demand and to increase their throughput rate, language service providers tend to employ TM/MT systems into their workflow. Therefore, professional translators have increasingly been using MT/TM systems and postedit outputs of MT/TM systems. As a result, a considerable part of the translation profession has been carried out by using MT systems to produce the first draft of translation.

Incorporation of people who are not professional translators into the process is another method which has been employed to meet the global demand. The non-professional translators include volunteers who are members of social networks and NGOs and subject-matter experts.

5.1.1. The Role of Volunteers in the Translation/Postediting Process

Recent examples of volunteers translating interfaces of social networks can be seen in Facebook and Twitter. Facebook is now available in over 70 languages in 2013, thanks to a framework that allows the user community to translate the texts. In this way, members of the Facebook community both function as the translators of the site content and enjoy the benefits their translation can bring. Through its Translation Center, Twitter (2013) also invites its users to translate its content into their local languages as well as asking for feedback for the translations provided by the users.

Google provides automatic translation through its statistical-based translation service Google Translate. Based on output from Google Translate, Google can automatically translate a web site which is not written in the user's native language. Next to each translated web site, Google invites the users to "contribute a better

ISBN: DL:
translation", thereby improving the MT database with users' alternative translations. Recently, a new feature was added to Google Translate to provide the user with alternative translations for each phrase in the translated text. The user can click on the translated phrase and come across a pop-up menu of possible alternatives for that phrase below the original phrase. The user, then, can choose a better alternative and click on the option "use". With such systems, Google collects feedback from the users on more proper translation alternatives. Although relatively indirect, these are also methods of using volunteer translation.

Another example of volunteer translation or cooperation can be seen through the use of Google Translator Toolkit (GTT) which is an automatic and free (in 2013) on-line translation portal introduced by Google in 2009 involving a data-driven MT system Google Translate and a postediting tool. When a source text is uploaded to GTT, the system produces the draft translation (raw MT output) based on the aligned corpora of source and target texts stored on the global database of the MT system. The users can postedit the MT output provided by the MT system and share their postedited texts with other users of GTT. In this way, the MT database improves more rapidly with the continued use by multiple users than it would be by a single user.

Abekawa et al. (2010) report on a translation environment where collaboration takes place between experienced translators and volunteer translators serving for NGOs or NPOs. The authors constructed a translation corpus that contained source texts, their draft translations in several versions and their final translations (called the SDF corpus) through the publicly available translation hosting site Minna no Hon'yaku (MNH). MNH offers a group-based translation environment where inexperienced volunteer translators translating for NGOs or NPOs perform draft translations. Other users with more translation experience can edit these draft translations and save the edited versions on MNH. Finally, experienced translators can produce final versions of the draft or edited translations, save them and provide feedback to inexperienced volunteers. MNH provides an integrated translation aid environment, which enables translators to look up dictionaries and Wikipedia as well as to search on Google. As new users keep logging onto MHN, the SDF corpus is constructed. Inexperienced volunteers can display draft translations and their edited final versions in the same interface where deletions and additions are highlighted in different colors. This system is an example of collaboration between volunteer and professional translators. The resulting corpus and MNH can be useful for self-training of volunteers whereas experienced translators or translator

ISBN: DL:
trainers can use the corpus of draft translations for educational purposes (i.e. to see what type of mistakes inexperienced volunteer translators tend to make). Further, the authors state that accumulation of such a corpus would be useful for improving MT output, although they do not elaborate on this aspect (Abekawa et al. 2010).

5.1.2. The Role of Subject-Matter Experts in the Translation/Postediting Process

In addition to the use of technology, producing higher-quality translations in shorter periods of time requires specialization in at least a few fields. However, not many translators can choose one or two areas to be specialized in. Even if this could be possible for some, the need to receive subject-matter experts' ideas still maintains its importance. This has made subject-matter experts and/or cooperation with them crucial in the translation/postediting and revision processes.

With the incorporation of subject-matter experts in the translation/postediting processes and in the context of collaboration between subject-matter experts and professional translators, we observe that the first draft of the translation/the postedited MT output is either created by the subject-matter experts and revised by the professional translators or the professional translators create the first draft and the output is revised by the subject matter-experts. Examples of this phenomenon can be observed in the translation and localization industry. SDL Language Services mentions that they ensure quality and consistency by employing "more native-speaking subject-matter experts than any other company, meaning excellent standards of language and cultural awareness." Another language service provider SimulTrans mentions about cooperation between linguists and subject-matter experts in the revision of translations as a significant component of their internal quality checks during the conduct of a project. They note the following points in this regard:

> 1. After the glossary is created, researched, and translated by the lead linguist, all members of the linguistic team review the glossary, in cooperation with appropriate subject-matter experts.
>
> 2. After the translation of each component (software, help, documentation), a subject-matter expert technically reviews the translations to ensure the terminology accurately reflects the contents of the glossary, the industry customs, and the expectations of the target customers. (This step is often called "technical review" or "editing").

ISBN: DL:

3. After the subject-matter review of each component (software, help, documentation), a second professional linguistically reviews all translated materials to ensure complete, correct, and consistent translations. (This step is often called "linguistic review" or "proofreading").

ENLASO is another company providing translation and localization solutions. Its web-site carries an article describing the best practices for the localization of the certification exams to key target languages (Soloduk 2011). To this end, phases of the process are outlined. In the phase of provision of translation instructions and terminology glossaries, Soloduk (2011) advises that "bilingual glossaries and updates to existing glossaries be reviewed for acceptance by your subject-matter expert to help reduce change requests towards the end of the project." In the translation and editing phase, she suggests that "the accurate translation of technical documentation, such as certification exams, demands the efforts of experts with appropriate linguistic and subject-matter knowledge." She further recommends that "an in-country review by a subject-matter expert be scheduled upon completion of the editing phase" and the "qualified candidates are typically volunteers of the association/institution who have already passed the certification exam and who are native speakers of the target language that they are reviewing."

In an article published on the web-site of Tcworld, a magazine for international information management, Muegge (2011) emphasizes the requirement to "employ subject-matter experts to check the suitability of those translated terms on the client side."

Language Scientific, a US-based translation and localization company, goes further and asks its employees to be both translators and subject-matter experts. The company states its translation-quality management principles and methods on its web-site, underlining that its employees should:

1. hold an advanced degree in the subject-matter to be translated (MD, PhD, MSEE, etc.).

2. have extensive work experience within the relevant field of science, engineering, medicine or law.

3. have at least 7 years of experience translating technical documents in the language pair and subject to be translated.

ISBN: DL:
HISPAWORDS is a language service provider which reports to "ensure access to specialized translators and subject-matter experts with the superior skill levels that the pharmaceutical industry expects." They state that their "pharmaceutical linguists hold PhD's or Masters degrees in Pharmaceutical Sciences to guarantee the highest levels of quality." They also note that medical texts are "professionally translated by a specialized linguist and then proofread by an expert in the field, followed by a Quality Assurance process that ensures top accuracy, consistency and quality."

The language service provider McElroy Translation explains its three-step quality assurance (QA) process to provide its clients with the highest-quality output for their technical translations. They note that "all translations are sent to a highly qualified native-speaking subject-matter expert for the initial translation. After that, the three-step QA process begins:

> The project is sent for review and editing to a subject-matter expert in the target language. The initial translator is notified of any questions, omissions or corrections altering the meaning of the translation and a dialogue between translator and editor commences until both are satisfied. Documents are then processed in-house where standard formatting styles are applied, or the document is formatted according to the client's special instructions, then sent to a proofreader in step 2 of our QA process. In step 3, the project undergoes a final quality control check in-house to verify that all the text is present and all the special instructions have been met.

More interestingly, while explaining its translator selection criteria, McElroy Translation alternates "a university degree in translation" to "a minimum of five years of professional experience in area(s) of expertise."

In terminology management, Le Néal (2001: 651) points to the key role played by subject-matter experts "to check the accuracy of the technical information contained in the entries and to comment on usage or on any terms" proposed by the bilingual terminologist in creating terminology.

Although we are aware of the fact that the quality standards and the quality assurance processes declared by the translation service providers on media might describe the ideal or optimum conditions rather than reflecting the real quality processes implemented, it is obvious that the involvement of subject-matter experts in the translation/postediting processes is regarded as a significant component of quality

ISBN: DL:
assurance.

5.2. Aims of the Study

To our knowledge, there has been no research carried out so far to explore the specific characteristics of the postediting processes performed by subject-matter experts as opposed to professional translators. What features does each process have? How do the postediting processes by subject-matter experts vs professional translators compare in terms of processing speed, time spent on documentation and amount of corrections? How do the end-products compare in terms of quality? Further, we are interested in exploring which of the following workflows that have been practised in the translation market is faster and produces outputs of higher quality: Postediting MT output by subject-matter experts and revising the output by professional translators, or vice-versa.

With these considerations in mind, we intend to carry out a study with the following aims:

1. To investigate the processes of postediting MT output by subject-matter experts and professional translators.
2. To compare MT postediting by subject-matter experts and professional translators in terms of:
   a. processing speed of postediting the MT outputs (based on words processed per minute, w/pm),
   b. time spent on documentation,
   c. amount of changes performed,
   d. quality of the postedited MT outputs (based on LISA QA Model 3.1.),
3. To compare the following workflows in terms of speed and end-product quality:
   a. Postediting the MT output by subject-matter experts and subsequent revision by professional translators, and
   b. Postediting the MT output by professional translators and subsequent revision by subject-matter experts.

Speed will be measured by the words processed per minute in both postediting and revision task of each participant. Quality will be measured by evaluating the postedited/revised products using LISA QA Model 3.1.

ISBN: DL:

5.3. Hypotheses

The present study tests the following hypotheses:

Part 1: Postediting

For technical texts:

1. Subject-matter experts and professional translators postedit MT output differently in terms of processing speed, time spent on documentation and number of changes.
    a. Professional translators complete the postediting task faster than subject-matter experts do.
    b. Professional translators spend more time on documentation than subject-matter experts do.
    c. Professional translators perform more changes than subject-matter experts do.
2. The quality of postediting by subject-matter experts is higher than postediting by professional translators.

Part 2: Revision

1. When professional translators revise postediting by subject-matter experts and when subject-matter experts revise postediting by professional translators, the speed and end-product quality differ.

5.3.1. Operationalization of Variables

Producing translations of consistent quality in shorter period of time has been the main impetus for using postedited MT in preference to full human translation. The TAUS report (2010: 6) underlines that "the overall aim of any translation automation solution is to accelerate throughput at consistent quality levels (and where possible reduce costs)." Therefore, processing speed and quality are the main variables in the present study.

5.3.1.1. Processing Speed

Processing speed is measured by dividing the total number of words processed by the total processing time. In other words, it is the number of source words processed per minute (w/pm).

We measured the processing speed by means of the screen-recording software

ISBN: DL:
BB Flashback. Four measures were taken: (1) the processing speed of subject-matter experts' postediting MT outputs, (2) the processing speed of professional translators' postediting MT outputs, (3) the processing speed of revision of subject-matter experts' posteditings by professional translators, (4) the processing speed of revision of professional translators' posteditings by subject-matter experts.

For the first two items (1 and 2) above, more specifically, we also measured and compared time spent by each group (subject-matter experts or professional translators) on documentation during the postediting process. For the latter items (3 and 4), we identified the speed of both processes by measuring their processing speed. Further, in order to find which work cycle is faster, comparison was made between the processing speed of subject-matter expert postediting + professional translator revision and professional translator postediting + subject-matter expert revision.

5.3.1.2. Quality

In order to determine the quality of the postedited MT outputs and their revised versions, we used the LISA QA Model 3.1 (see 6.1.1.5 below).

We identified errors and categorized them in line with LISA error categories. Then, we entered the errors in each category in the section where its level of severity was indicated on the LISA QA Model 3.1 interface. The interface automatically counts the number of errors as we enter them. In addition, it calculates the total error points and quality percentages and provide a project review report containing the quality evaluation information (Appendix H).

We compared the quality of the posteditings performed by subject-matter experts and professional translators to identify the differences. Quality of the revised end-products (subject-matter expert postediting + professional translator revision and professional translator postediting + subject-matter expert revision) was also identified using the same grid.

ISBN: DL:

6. Methodology

6.1. Experiment Design

6.1.1. Setting

The study is composed of two parts, each part involving two sets of experiments:

PART 1: Postediting MT output of a technical text by:

1. subject-matter experts (engineers), and
2. professional translators.

PART 2: Revision of the postedited MT outputs

1. Revision by professional translators of the texts postedited by subject-matter experts, and
2. Revision by the subject-matter experts of the texts postedited by professional translators.

In Part 1, the texts were first machine-translated from English into Turkish; they were then postedited by subject-matter experts and by professional translators. The two groups both worked on the same MT output. The aim was to identify the differences between the way subject-matter experts and professional translators postedit the MT output (with regard to speed, documentation and changes) on a subject matter that the subject-matter experts know well. In addition, we identified the quality differences between the final products of both groups.

In Part 2, the subject-matter experts and the professional translators revised one another's posteditings. The aim was to compare the two groups' revision processes with regard to speed and to evaluate the revised products with regard to quality.

Before we conducted the main study, we carried out a pilot study with two subject-matter experts and two professional translators, in order to test the methodology and detect possible flaws in our research design. After analyzing the results of the pilot study, we revised our research design and as soon as we made the necessary corrections, we conducted the main study with the total of twenty subjects (ten subject-matter experts and ten professional translators).

In the first experiment of Part 1 (subject-matter expert postediting), we had ten subject-matter experts (engineers) postedit a 482-word technical text in Turkish pre-

ISBN: DL:

translated from English with Google Translate.

In the second experiment (professional translator postediting) of Part 1, we had ten professional translators postedit the same 482-word technical text in Turkish pre-translated from English with Google Translate.

Before they began the experiment, all subjects were provided with a research participant release form (Appendix A) to be signed by the participants.

The posteditors were asked to postedit the MT output (provided as a Word.doc) using their own computers on which the screen recording software BB Flashback had been set up before. Before they carry out the experiments, we sent them instructions on how to upload BB Flashback on their computers and carry out the task (see Appendix B). The subjects would have access to the source text (provided as a Word.doc), the Internet and on-line dictionaries during the postediting task, however, they were not allowed to use any translation memory. They were asked to do the postediting by performing the changes they want to introduce on the MT output provided for them instead of creating a separate target text. In order to neutralize as many variables as possible, we imposed a time restriction. Each subject was given forty minutes to postedit the MT output. However, in case they might not finish the task in forty minutes, they were set free to take any extra time.

At the end of the experiments in Part 1, we gave post-assignment questionnaires to the participants to gather further data on their profiles and their perception of the process. Thus, after completing Part 1, we had twenty research participant release forms, twenty different posteditings of the same MT output, twenty post-assignment questionnaires and twenty screen-recordings.

The speed of each postediting performance was captured with the screen recordings. Time spent on documentation (minutes) and the amount of changes performed was also identified from the screen recordings.

The quality of the target texts were analyzed by the researcher using LISA QA Model 3.1. While analyzing the quality of the postedited texts, we used the reference translation of the test text from English into Turkish which was performed by a professional translator and an engineer in cooperation.

In Part 2, we conducted two sets of experiments. For the first experiment, we asked ten professional translators revise the ten subject-matter expert posteditings collected in Part 1. The professional translators were provided with the source text (as a Word.doc) as well for reference. The revision tasks were carried out on the participants'

ISBN: DL:
own computers at their usual work places or in their houses and they were recorded using BB Flashback. The subjects were allowed to use the Internet during the task, however they were not allowed to use any translation memory. They were asked to do the revision by performing the changes they want to introduce on the postedited text provided for them instead of creating a separate target text. For the second experiment in Part 2, we asked ten subject-matter experts to revise the ten professional translator posteditings collected in Part 1. The subject-matter experts were provided with the source text (as a Word.doc) as well for reference. These revisions were also carried out using the postedited texts (not on separate target texts) on the participants' own computers in their houses (In private companies, engineers are not allowed to upload external programs such as BB Flashback on their computers at work). Further, revision tasks were recorded using BB Flashback and the subjects were allowed to use the Internet. No translation memory is allowed to be used.

The subject-matter experts and the professional translators doing the revision tasks were the same with those who carried out the postediting tasks. In other words, subject-matter expert posteditors and professional translators postedited the same MT output. Then, both groups of revisers that were composed of the same participants in the postediting task revised each others' postedited texts. This might result in their learning of the text that they have already postedited. However, since revising a text in which the reviser has already had experience is very natural and common in real life, this learning factor does not pose a problem for our research.

6.1.1.1. Research Participant Release Form

All the subjects participated in the postediting and the revision tasks were asked to sign a research participant release form (Appendix A) which allowed us to use and publish anonymously the findings of the study to be gathered from the participants' answers to the questionnaires as well as their individual postediting/revision protocols.

6.1.1.2. Subjects

The pilot study was conducted with four subjects: Two subject-matter experts and two professional translators. After we completed the pilot study and revised our research design accordingly, we conducted the main study with twenty subjects (ten for subject-matter expert postediting, ten for professional translator postediting and the same participants for revision) having the same profile with those subjects in the pilot study.

ISBN: DL:

The subject-matter experts were graduates of various engineering departments who had been working at international automotive companies in Turkey for at least five years. Turkish is their mother tongue, and they are all proficient in English. Due to the international structure of the companies, engineers have long been using English as part of their daily work. None of them were working as translators or had professional experience in translation.

Professional translators were freelancers with translation experience of at least five years. We did deliberately not selected translators who were specialized in the translation of technical and/or engineering texts, because we aimed to find out how postediting of a technical text by subject-matter experts in engineering and professional translators who did not hold expertise in engineering compare. As with the subject-matter experts, Turkish is the translators' mother tongue and English is the primary foreign language. None of them was working as an engineer or had experience in engineering.

6.1.1.3. Task Instructions

We prepared realistic task instructions and a brief for postediting (see Appendix B) so as to provide the subjects with basic task descriptions. The instructions were given to the subjects before the postediting experiment (Part 1), and they involved information such as the sources where the text was taken from and what the output would be used for. Further, information on how to download and use the screen recorder BB Flashback, information on the target audience, quality requirements and the resources that could be used during the tasks were specified.

The instructions as regards the experiments in revision (Part 2) were made available to the subjects via electronic mail.

6.1.1.4. The Source Text

As an experiment text, we used one machine-translated 482-word technical text in Turkish on dismantling end-of-life vehicles. It was translated from English into Turkish with Google Translate. The original English text contains 587 words and was taken from the International Dismantling Information System (IDIS) (see Appendix C for the original English text and Appendix D for its MT ouput in Turkish). IDIS contains data on the environmental sound pre-treatment and dismantling of End-of-Life Vehicles. It was initially developed by the European Automotive Industry and the project became

ISBN: DL:

worldwide in January 1999. Since then, all IDIS development is controlled by an international group of manufacturers, containing 60 members. The IDIS software currently supports 30 languages and it is updated twice a year.

The experiment text was selected by the researcher from among the dismantling information instructions (from IDIS) belonging to various automotive brands by keeping the brand name anonymous. The participants were provided with the MT output in Turkish to be postedited as well as with the source text in English.

6.1.1.5. LISA QA Model 3.1

As a tool for measuring quality in the present study, we used the LISA QA Model 3.1, developed and released by the Localization Industry Standards Association (LISA) in 2007 as a result of an extensive collaboration between LISA members, localization service providers, software and hardware developers and end-users.

LISA QA Model 3.1 is a quality metric which can be used to collect, track and manage information about the quality of translation projects. The LISA grid runs on an interface involving a single control panel (see Appendix G) which is the only window the user will see. The user can select the source and target languages as well as specifying the client, project, translator and the reviewer for each quality assessment tasks. The quality of any document starts at 100 and it decreases as the reviewer starts logging in error data. The reviewer has to specify the size of the document (type the number of words) in a given sample text and determine a minimum quality percentage (i.e. 75) which will indicate the minimum acceptable quality for the document under review. For carrying out a review, errors occurring in the translation need to be categorized and their degree of severity should be assigned. LISA QA metric has seven error categories with three severity levels assigned to each error category. The error categories are Mistranslation, Accuracy, Terminology, Language, Style, Country and Consistency, which can be explained as follows:

> Mistranslation refers to the incorrect understanding of the source text; Accuracy to omissions, additions, not reflecting the source text properly; Terminology to glossary adherence (consistent, accurate and unambigous use of terminology) and context; Language to grammar, semantics, punctuation, spelling; Style to the use of consistent, user-friendly, informative style and adherence to style guides, register/tone and language variants; Country to country standards (time, date, currency, measurements, etc.) and local suitability (reflecting the cultural

ISBN: DL:
elements of the relevant country); Consistency to coherence in terminology across the project (LISA QA Model 3.1 User's Manual).

Errors in each category are assigned a degree of severity at three levels. The three levels are Minor, Major and Critical; which weigh 1, 5 and 10 respectively. The figures may change depending on the number of words in the document under review. However, their weights remain the same. For example, in a document of 500 words, when the error is minor, then it receives 0.2 point. If it is major, it is weighted as 1 point; and if it is classified as critical, it is graded as 2 points.

The total error points determine the percentage of text quality. If the percentage of final text quality is below the quality threshold established by the reviewer at the beginning of the reviewing task, then the document gets "Fail". If the final quality is above (or equal to) the minimum quality level, then, the document under review gets "Pass". When the quality levels of the texts are presented, numerical evaluation results (percentage values) are also retained instead of providing the results only as "Pass" or "Fail".

6.2. Methods of Data Collection

In the present research, we used both qualitative and the quantitative methods of data collection, as explained below.

6.2.1. Google Translate

We used Google Translate as the data-based machine translation engine. Google Translate operates on a postediting tool called Google Translator Toolkit (GTT). It is an automatic and free (currently in 2013) on-line translation portal introduced by Google in 2009 involving a data-driven MT system and a postediting tool.

In order to obtain MT output from Google Translate, the user simply has to upload the source text to the system and select the source and target languages. When a source text is sent to Google Translate, the system produces the draft translation (MT output) based on the aligned corpora of source and target texts stored on the global database of the MT system.

In the present research, we provided the subjects with both the source text and the raw MT output from Google Translate to be postedited on their computers. The subjects

ISBN: DL:

postedited the text in word documents, not on GTT. This is because if they had postedited the MT output on GTT, then the resulting target text would have affected the following tasks to be carried out by other subjects.

In our initial trials with Google Translate, we realized that the MT database in Turkish was relatively poor. Therefore, for the experiment, we fed the general MT database with the postedited versions of a few texts that belong to the same genre as the experiment text. These source texts were not the same texts as the one we used in the experiment. We did the translations on Google Translator Toolkit (GTT) with access to the Internet, so as to train the engine. Bowker and Ehgoetz (2007: 217) applied a similar method to improve MT output. In their study of user acceptance of MT output, they first translated the source text using the MT system, identified unknown terms and added 65 entries for these terms to the MT dictionary. Then they translated the same texts again and used the resulting output as the raw MT output (Bowker and Ehgoetz 2007: 224). It is also possible to improve the database by adding an external TM and/or glossary. This was done by Guerberof (2008: 22) in her research project where she provided Language Weaver (a statistical-based MT engine) with a TM containing 1.1 million words and a core glossary to create a customized engine. Guerberof (2012) implemented similar MT database training in her doctoral project. In the present research, however, we did not plan to supply the MT database with such a large volume of TM, since our aim is to investigate the differences between the ways MT output is postedited by subject-matter experts and professional translators, rather than how good the output is or can become.

6.2.2. Screen Recording

Screen recording software records the whole process on the computer screen and enables one to replay the recorded data by showing the exact time each action has been performed on the screen. It offers valuable qualitative and quantitative data on processes, yet the data analysis can be extremely time-consuming. However, in the present study, we focus on processing time to be recorded by the screen recording software, which is not very time-consuming.

In the present research, we used the screen recording software BB Flashback. It records all the screen activities (i.e. translation carried out, Internet searches, dictionary look ups), which provide in depth understanding of the processes involved. Recording the exact time on task provides quantitative data, while the real-time recording of the

ISBN: DL:

screen activities constitutes the qualitative data with regard to the process in question.

The screen recorder also involves the option of recording the user's keystrokes and voice during the process. This enables the researcher to record keystrokes and concurrent thinking-aloud, in addition to registering screen activities, without using an additional recording (i.e. video or audio recording). It is also possible to record the subjects' voice while they are monitoring the replay of their translation and simultaneously making retrospective commentaries on them.

In the present research, we recorded the screen activities of the posteditors as they are postediting the MT outputs from Google Translate as well as their revision processes. The subjects were allowed to use on-line dictionaries and the Internet sources. Thus, the processing speed of each task was determined and detailed information as regards the postediting and revision processes was gathered. Previous studies (Krings 2001: 527, Jakobsen 2003: 93) have shown that Think-Aloud Protocols (TAPs) slow down the translation processes, thus affect the measurement of time on task. Therefore, we decided not to have the subjects comment on their processes during the tasks.

6.2.3. Post-Assignment Questionnaires

At the end of the postediting tasks, post-assignment questionnaires — Appendix E for translators and Appendix F for engineers, some parts adapted from the post-assignment questionnaires by Guerberof (2012) — were used to collect data about the participants as well as qualitative data to see if the perceived process components were in line with what is obtained from quantitative data.

ISBN: DL:

7. Pilot Study - Data Analysis and Results

7.1. Postediting

7.1.1. Speed, Documentation and Changes in Postediting

After we completed data collection for the pilot study conducted with four subjects (two profesional translators and two engineers), we first analyzed the postediting process (Part 1) in terms of speed (procesing speed, words processed per minute) documentation and changes using BB Flashback data.

It was easy to obtain the processing speed data, because BB Flashback provides the exact time of each task in "hour.minute.second.split second" format. In order to determine the time spent on different phases of the postediting process (only postediting and self-revision), we watched each subject's screen recording data, recorded the exact time when the subject completed the postediting task and started self-revision. Further, we calculated the words processed per minute by dividing the total number of words processed by the total time spent on the postediting task. We made this calculation to determine the processing speed both including and excluding self-revision process.

In order to collect data on documentation, we watched the screen recording data and recorded when the subject started and finished doing a search on the Internet or when they used an on-line dictionary. Then, we obtained data including the number of look-ups, the exact time (hour.minute.second.split second) when the look-up started and finished, the source used and the item of the text that was searched. By using these data, we calculated the total time each subject spent on documentation as well as details about the sources used and the items searched.

The most time-consuming part of BB Flashback analysis was carried out to find the number and type of changes performed by the posteditor. While watching the screen recordings, we stopped them at each change performed by the subject and recorded the changes using six categories: content changes, formal changes (capitals and punctuations), omissions, additions and reorganizations.

Content changes mean changing the word or phrase in the MT output into a completely different one in a way that the content represented by the previous phrase or word is changed (e.g. changing "flylead" in the MT output into "bağlantı kablosu"

ISBN: DL:
(connection wire) or changing "yerleştirin" (locate/place/put) in the MT output into "bulun" (find the location).

Formal changes are divided into two subcategories as capitals and punctuations. Capitals represent the changes performed regarding the use of capitals (e.g. changing "cam" (glass) in the MT output into "Cam" (Glass). Punctuations denote to changing punctuation marks in the MT output (e.g. changing ", araçta" (, on the vehicle) at the beginning of a sentence in the MT output into "Araçta" (On the vehicle). Due to their successive and relatively automatic occurrence, corrections such as deleting the unnecessary comma and one space and changing the lower case "a" into capital "A" was counted as one change under one category).

Omissions mean omitting the word/word group/suffix in the MT output, that the posteditor deems unnecessary or wrong (e.g. changing "talimatları" (instructions) in the MT output into "talimatı" (instruction) or changing "dokunmak yok" (no touch) into "dokunmayın" (do not touch).

Additions represent the opposite case of omissions. They are used for adding the word/word group/suffix in the MT output, that the posteditor deems missing (e.g. changing "elektrik" (electricity) into "elektrik-li" (electrical) or changing "düğmeleri bırakın" (release/leave the buttons) in the MT output into "düğmelere basmayı bırakın" (stop pressing the buttons).

Reorganizations mean changing the wrong word order in the MT output so as to correct the phrase (e.g. changing "OFF (Lock) kontak anahtarını" (OFF (Lock) ignition switch) into "kontak anahtarını OFF (Lock)" (ignition switch to OFF (Lock). Another example could be changing "hemen nötralizasyon sonra" (immediately neutralization after) into "nötralizasyondan hemen sonra" (immediately after neutralization).

After counting all the changes performed by the subjects under the categories explained above, we calculated the total number of changes performed. In section 7.1.1.1, we provide data regarding speed (processing speed, words processed per minute), documentation and changes for subject-matter experts (engineers) and in 7.1.1.2, for professional translators.

Before moving onto the next section, we would like to give basic information on the subjects' background on the basis of the data collected by means of post-assignment questionnaires (Appendix E and F). Translator 1 was between 31-35 years old and had 9-10 years of experience in translating. She held a BA, MA and PhD degrees in Translation. She had neither professional experience in engineering nor received training in engineering. As part of her job, Translator 1 usually translates technical, legal texts and user manuals from English into Turkish and vice-versa.

Translator 2 was between 36-40 years old and had more than 10 years of experience in translating. She held a BA degree in Translation and MA in Communication Sciences. She had neither professional experience in engineering nor received training in engineering. As part of her job, Translator 2 usually translates texts on social sciences, and sometimes technical, legal texts and user manuals from English into Turkish and vice-versa.

Engineer 1 was between 31-35 years old and had 9-10 years of experience in engineering. He held a BSc degree in Mechanical Engineering and MBA. He had neither professional experience in translation nor received training in translation and/or postediting. As part of his job, Engineer 1 has to translate electronic mails, minutes of meetings, reports, technical documents and Power Point presentations from English into Turkish and vice-versa.

Engineer 2 was between 26-30 years old and had 6-8 years of experience in engineering. He held BSc and MSc degrees in Mechanical Engineering. He had neither professional experience in translation nor received training in translation and/or postediting. As part of his job, Engineer 2 has to translate electronic mails, reports and some simple technical documents from English into Turkish and vice-versa.

7.1.1.1. Speed, Documentation and Changes in Postediting by Subject-Matter Experts

In order to neutralize as many variables as possible, we decided to impose a time limit for the postediting task and asked the subjects to complete the postediting task in forty minutes. However, they were also told to use extra time if they could not complete the task within this time limit (see Appendix B for Task Instructions for Part 1). The two subject-matter experts participated in the pilot study exceeded this time limit. Therefore, for the main study, we decided to increase this limit to fifty minutes.

When we analyzed their screen recording data, we found that - unlike translators - engineers completed the postediting task without performing any self-revision. The mean task duration for subject-matter experts was 50.21 minutes (see Table 4) which was more than the specified time limit, 40 minutes.

In order to identify the speed of the subjects' performances, we calculated the number of words each participant processed per minute. The average number of words processed per minute by engineers was 9.6 (See Table 4).

ISBN: DL:

There is difference between both engineers' documentation behavior. Engineer 1 spent 2.76 minutes (5.2 percent of the total postediting duration) on searching for the meanings of the words in the text, while Engineer 2 spent only 20 seconds (0.33 minute) for this process which equals to 0.7 percent of his total task period. On average, engineers' documentation process amounts to 3.07 percent of their total postediting duration.

Table 4. Speed, Documentation and Changes in PE – Engineers

Subjects	Time on PE Excluding Self-revision (minute)	Time on Self-revision (minute)	Total Task Time (PE and Self-rev) (minute)	w/pm (words processed/ total task time)	Time Spent on Documentation (minute)	Documentation as % of the Total Task Time	Number of Changes
En 1	52.96	00.00	52.96	9.1	02.76	5.2	155
En 2	47.45	00.00	47.45	10.1	00.33	0.7	107
Mean	**50.21**	**00.00**	**50.21**	**9.6**	**01.55**	**3.07**	**131**

Detailed analysis of documentation data (Table 5) confirms the differences between both engineers' documentation behavior. Engineer 1 looked-up the Internet sources 10 times, while Engineer 2 did such a search only once. On average, subject-matter experts searched for 3.5 text-related items in 93 seconds using 2 Internet sources 5.5 times.

Table 5. Documentation in PE – Engineers

Subjects	Number of Look-up	Time spent (min.)	Number of Sources Used	Number of Items Searched
En 1	10	2.76	3	6
En 2	1	20	1	1
Mean	**5.5**	**93**	**2**	**3.5**

Table 6 presents more detailed information about the documentation data belonging to Engineer 1.

ISBN: DL:

Table 6. Detailed Documentation Data in PE - Engineer 1

Number of Look-up	Time spent on documentation (min.sec.split sec)			Source Used	Item Searched
	Start	Finish	Time used		
1	01.51.4	02.14.1	00.22.7	Seslisozluk online dict.	deployment
2	02.53.7	03.04.4	00.10.7	Seslisozluk online dict.	onboard
3	17.13.4	17.24.7	00.11.4	Seslisozluk online dict.	harness
4	23.17.9	23.37.1	00.19.3	Seslisozluk online dict.	pretensioner
5	23.37.1	23.58.8	00.21.6	Zargan online dictionary	pretensioner
6	24.47.0	24.58.4	00.11.4	Zargan online dictionary	multiplug
7	27.28.3	27.43.5	00.15.2	Zargan online dictionary	flylead
8	27.43.5	27.53.0	00.09.5	Seslisozluk online dict.	flylead
9	27.53.0	28.03.6	00.10.6	Seslisozluk online dict. forum	flylead
10	28.07.1	28.41.1	00.34.0	Google (images, web pages)	flylead
Total			**02.46.4 (2.76 min.)**		

The difference between the engineers' documentation data is also reflected in the changes they made on the text while postediting. Table 7 presents the number and the distribution of the changes both engineers carried out. The biggest difference between the numbers of changes made by both engineers is seen in content changes. Engineer 2 carried out almost half as much content changes as Engineer 1. On average, engineers performed 62.5 content changes, 13 formal changes (4.5 capitals, 8.5 punctuations), 15 omissions, 32 additions and 8.5 reorganizations. The mean number of total changes for engineers is 131 (see Table 7).

Table 7. Changes in PE – Engineers

Subjects	Content Changes	Formal Changes		Omission	Addition	Reorganization	Total
		Capitals	Punctuation				
En 1	83	5	12	12	33	10	155
En 2	42	4	5	18	31	7	107
Mean	**62.5**	**4.5**	**8.5**	**15**	**32**	**8.5**	**131**

7.1.1.2. Speed, Documentation and Changes in Postediting by Professional Translators

We used the same method to analyze the professional translators' postediting data. The same time limit, forty minutes, for the postediting task was imposed for them as well. They were also told to use extra time if they cannot complete the task within this time limit.

ISBN: DL:

The mean postediting duration for translators is 45.48 minutes (see Table 8) which is shorter than engineers' mean total processing time (see Table 4). However, unlike engineers, both translators performed self-revision right after they finished postediting. Translator 1 spent 5.28 minutes on self-revision, while Translator 2 spent 16.28 minutes for this (Table 8). The mean self-revision duration of translators is 10.78 minutes that equals to 19.1 percent of the mean total task duration of the translators. When we add self-revision durations to postediting durations, we find the translators' total task durations. The difference between both translators' total task duration is relatively high. Translator 1 completed the whole postediting process (including self-revision) in 44.98 minutes whereas Translator 2 completed the whole task in 67.53 minutes. The mean total task duration for translators including self-revision is 56.26 minutes and it is 45.48 minutes excluding self-revision (see Table 8).

In order to find the level of translators' speed, we calculated the number of words processed per minute. The mean number of words processed per minute is 8.9. For translators, we also calculated the number of words processed per minute by excluding the time spent on self-revision. We did this, because none of the engineers performed self-revision and we wanted to see how translators and engineers compare when we included and excluded self-revision. When we did not include self-revision, we found that the translators' mean number of words processed per minute is 10.8 (Table 8).

Table 8. Speed, Documentation and Changes in PE – Translators

Subjects	Time on PE Excluding Self-revision (minute)	Time on Self-revision (minute)	Total Task Time (PE and Self-rev) (minute)	w/pm (words processed/ total task time)	w/pm (excluding self-revision)	Time Spent on Document. (minute)	Document. as % of the Total Task Time	Number of Changes
Tr 1	39.70	5.28	44.98	10.7	12.1	7.51	16.7	150
Tr 2	51.25	16.28	67.53	7.1	9.4	7.51	11.1	135
Mean	**45.48**	**10.78**	**56.26**	**8.9**	**10.8**	**7.51**	**13.3**	**143**

As for documentation, it is interesting that both translators spent exactly the same amount of time (7.51 minutes) to search for data using the Internet sources (see Table 8). On average, both translators spent 13.3 percent of their total task time on documentation.

ISBN: DL:

Unlike engineers, both translators behave similarly in terms of documentation (Table 9).

Table 9. Documentation in PE – Translators

Subjects	Number of Look-up	Time spent (minute)	Number of Sources Used	Number of Items Searched
Tr 1	16	7.51	3	14
Tr 2	15	7.51	2	13
Mean	**16**	**7.51**	**3**	**14**

Table 10 and Table 11 present more detailed information about the documentation data of Translator 1 and Translator 2, respectively.

Table 10. Detailed Documentation Data in PE - Translator 1

Number of Look-up	Time spent on documentation (min.sec.split.sec)			Source Used	Item Searched
	Start	**Finish**	**Time used**		
1	00.49.9	01.10.3	00.20.4	Tureng online dictionary	onboard
2	01.23.43	01.51.9	00.28.4	Tureng online dictionary	pyrotechnic
3	01.51.9	02.12.9	00.21.0	Tureng online dictionary	deployment
4	05.44.4	05.55.3	00.10.8	Tureng online dictionary	deployment noise
	05.55.3	07.52.1	01.56.9	Google (web pages, images, International Journal of Crashworthiness	deployment noise
5	12.15.2	12.29.6	00.14.4	Tureng online dictionary	harness
6	15.34.7	15.59.1	00.24.4	Tureng online dictionary	pretensioner
7	16.36.5	16.48.3	00.11.8	Tureng online dictionary	multiplug
8	19.42.0	19.53.6	00.11.6	Tureng online dictionary	flylead
9	19.53.6	20.57.6	01.03.9	Google (images, web pages	flylead terminal, flylead
10	21.07.3	21.21.9	00.14.6	Tureng online dictionary	dual stage
11	21.30.1	22.10.9	00.40.9	Google (images, web pages)	dual stage/dual stage device
12	22.10.9	22.31.2	00.20.3	Tureng online dictionary	stage
13	25.12.0	25.22.0	00.10.0	Tureng online dictionary	multiplug
14	32.35.3	33.04.6	00.29.3	Tureng online dictionary	switch
15	33.40.4	33.43.4	00.03.0	Tureng online dictionary	switch
16	44.26.0	44.35.3	00.09.3	Tureng online dictionary	ensure
Total			**07.31.0 (7.51 min.)**		

ISBN: DL:

Table 11. Detailed Documentation Data in PE - Translator 2

Number of look-up	Time spent on documentation (min.sec.split sec)			Source Used	Item Searched
	Start	Finish	Time used		
1	16.42.2	17.28.0	00.45.8	Tureng online dic.	pretensioner harness, pretensioner
2	18.27.5	18.42.1	00.14.6	Tureng online dic.	multiplug
3	26.59.4	27.20.4	00.21.0	Tureng online dic.	multiplug
4	33.15.6	33.30.5	00.14.9	Tureng online dic.	power up
5	36.14.8	36.33.9	00.19.1	Tureng online dic.	check button
6	41.36.4	41.52.3	00.15.9	Tureng online dic.	deploy
7	47.15.9	47.33.3	00.17.4	Tureng online dic.	deploy
8	54.52.8	55.32.0	00.39.2	Google	gerdirici kabloları
9	55.34.1	55.56.1	00.22.1	Tureng online dic.	pretensioner
10	55.58.6	57.48.3	01.49.6	Google	ön gerilim kablo/ön gerilim tertibatı kabloları
11	57.52.8	58.42.6	00.49.8	Tureng online dic.	pretensioner
12	58.42.8	59.11.4	00.28.6	Google	"ön gerilim veren kablolar"
13	01.00.14	01.00.36	00.22.0	Tureng online dic.	locate
14	01.00.58	01.01.07	00.08.8	Tureng online dic.	locate
15	01.02.34	01.02.56	00.22.9	Tureng online dic.	battery
Total			**07.31.7 (7.51 min.)**		

Although both translators' documentation duration is the same, there are differences between the changes they made on the text while postediting. However, as compared to engineers, this difference in the translators' data is less striking. Table 12 presents the number and distribution of the changes both translators carried out. On average, both translators performed a total of 143 changes with 61 content changes, 12 formal changes (6.5 capitals and 5.5 punctuations), 17 omissions, 41 additions and 11.5 reorganizations.

Table 12. Changes in PE – Translators

Subjects	Content Changes	Formal Changes		Omissions	Additions	Reorganization	Total
		Capitals	Punctuation				
Tr 1	62	5	7	20	46	10	150
Tr 2	60	8	4	14	36	13	135
Mean	**61**	**6.5**	**5.5**	**17**	**41**	**11.5**	**143**

7.1.2. Quality in Postediting

Although data regarding speed, documentation and the changes performed give us information about the postediting process, this information is not complete without the knowledge of the quality of the postedited products.

ISBN: DL:
We analyzed the quality of the postedited texts of the four participants (two professional translators and two engineers) in our pilot study with LISA QA Model 3.1. (See 6.1.1.5 above). First, we activated the user interface of LISA QA Model 3.1 (see Appendix G) which had been set up on our computer previously. Then, we selected both the source and the target language as Turkish, because we used the MT output of an English text pre-translated into Turkish with Google Translate as the source text and the target text was the postedited version of that Turkish MT output. After that, we named the subjects anonymously such as Translator 1 or Engineer 1 and entered these names on the interface before starting the evaluation of the relevant participant's postediting. Afterwards, we entered the number of words of the document under review and set the minimum acceptable level of quality as 75 percent for it. For carrying out the review, we used the postedited target text and the reference translation of the test text from English into Turkish which was performed by a professional translator and an engineer in cooperation. As we read the postedited text, we determined the errors, categorized them in line with seven LISA error categories (Mistranslation, Accuracy, Terminology, Language, Style, Country and Consistency) by assigning their degree of severity as Minor, Major and Critical and started entering the errors into the LISA interface.

Before starting the review, the quality percentage was 100 by default. As we entered each error, the interface registered it, calculated the error point and the quality percentage diminished from 100 in line with this error point. When the quality percentage reached a point below 75, which was the minimum level of acceptable quality, the interface labels the postediting/translation with "Fail", yet the reviewer can go on reviewing the text. Any level of quality between 100 and 75 percent is labeled as "Pass." After the review was completed, we could export the review data using the "Project Review Report" option (see Appendix H).

Since each posteditor/translator might translate the same text differently (even the same translator may translate the same text in slightly different ways at different times), when determining errors in the postedited texts, we did not look for exactly the same words or expressions occurring in the reference translation. Since the test text was a technical text and not open to interpretations, we did not have to deal with translations that were very different from the reference translation.

ISBN:
DL:

7.1.2.1. Quality in Postediting by Subject-Matter Experts

The quality of the texts postedited by both subject-matter experts is very different from each other. The quality percentage of Engineer 1 is 91.29 and that of Engineer 2 is 68.46. Obviously, Engineer 1 produced a high-quality text and received "Pass" from the LISA interface. However, Engineer 2 ended up with a low-quality text and received "Fail" (see Table 13). The mean error point for engineers is 97 and the mean quality percentage is 79.88.

Table 13. Quality in PE – Engineers

Subjects	Total Error Points	Postediting Quality %	Result
En 1	42	91.29	Pass
En 2	152	68.46	Fail
Mean	**97**	**79.88**	**Pass**

The main reason for this difference might be personal differences between the two participants. Further data analysis may also help us understand this gap.

Table 14 presents the distribution of errors according to LISA categories. Since we did not find any errors in the "style" and "country" categories, we omitted them in the tables throughout the thesis. Similarly, we did not find any errors with "critical" level of severity. The total number of errors of Engineer 1 is relatively small and most of these errors are minor ones, therefore, his postediting received high quality (91.29 percent).

An overview of Engineer 2's error distribution does not seem very different from Engineer 1's (see Table 14). However, the difference between their postediting qualities is explicit (91.29 percent versus 68.46 percent). The main reason for this difference is the high number of major terminology errors made by Engineer 2. When we watch BB Flashback videos of each participant for quality analysis, we realize that this difference is mainly caused by one recurrent term. The word "deployment" in "deployment of airbag" was incorrectly translated by the MT system into Turkish as "dağıtılması" which literally means "scattering". A simple dictionary look up for the word "to deploy" gives both subjects the translation "dağıtmak" (scatter) which confirms the translation provided by the MT system. However, doing a more careful search on the Internet and using his subject-matter knowledge, Engineer 1 changes this translation into "patlatılması" or "patlatmak" meaning "explosion" (of the airbag) or "to explode"

ISBN: DL:

(an airbag) in English which is the correct term for the test text. Translation/postediting of the word "deployment" incorrectly into Turkish as "dağıtılması", in other words, failure to change the term "dağıtılması" in the MT output into "patlatılması" was classified as a major terminology error. Since this term is recurring throughout the text, Engineer 2's failure to correct it while postediting brought him low level of quality.

Table 14. Distribution of Errors - PE – Engineers

Subjects	Mistranslation			Accuracy			Terminology			Language			Consistency		
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**
En 1		2		5			8	1		11			3		
En 2	1	2		1			4	24		15			1		
Mean	**1**	**2**		**3**			**6**	**13**		**13**			**2**		

7.1.2.2. Quality in Postediting by Professional Translators

The difference between the two participants in the subject-matter experts' postediting quality is also seen in the professional translators' data. The quality percentage is 88.59 for Translator 1 and 62.66 for Translator 2. As a result, Translator 1's postediting received "Pass" from the quality metric LISA Q.A Model 3.1 and Translator 2's postediting received "Fail" (see Table 15). The mean error point for translators is 118 and the mean quality percentage is 75.63.

Table 15. Quality in PE – Translators

Subjects	Total Error Points	Postediting Quality %	Result
Tr 1	55	88.59	Pass
Tr 2	180	62.66	Fail
Mean	**118**	**75.63**	**Pass**

In order to understand the reasons behind this difference between the product qualities of these two translators having similar profiles, we looked at the distribution of their errors. As we see in engineers' data, the most evident difference between both translators' errors can be seen in the major terminology category (see Table 16). A careful analysis of the subjects' BB Flashback data reveals similar results with engineers' data. As explained above, the main reason of this difference is Translator 2's failure to correct the recurrent word "dağıtmak" ("to deploy" in the English source text) in the MT output which is incorrectly translated by the MT system.

ISBN: DL:

Table 16. Distribution of Errors - PE – Translators

Subjects	Mistranslation			Accuracy			Terminology			Language			Consistency		
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**
Tr 1		1		1			19	5		5				0	
Tr 2	1	1		5			12	27		2				4	
Mean	**1**	**1**		**3**			**16**	**16**		**3.5**				**2**	

7.1.3. Post-assignment Questionnaire Data

In order to collect qualitative data regarding the postediting process, we had the subjects fill out post-assignment questionnaires. Below, we first present data from these questionnaires and then comment on the questionnaire data by relating them with the quantitative results of the postediting experiment. Basic information on the participants' background was provided in section 6.1.1.2, therefore it will not be repeated here.

Before participating our study, Translator 1 used Google Translate for getting the gist of the texts in a language that she did not know and to understand pages she comes across while surfing the Internet. Translator 1 reported her mean daily throughput without any translation aid as between 3100 and 5000 words per day.

When asked about her thoughts on Google Translate, she said “I think Google's translations are far better than other translations known as “machine translation”. It may not produce very bright outputs in English-Turkish pair. All the same, I think it is successful in terms of offering a general idea on the text. I understand this especially when I use Google Translate from a language I do not know at all into English. Of course, I do not come across “a very professional” translation, yet it definitely gives an idea on the content of the text.”

Translator 1 had no experience in using Google Translator Toolkit. Except for the postediting task carried out for the present study, she had no professional experience in postediting TM/MT output. Besides translating, Translator 1 does revision and proofreading. She did not receive any training on postediting. Translator 1 found the MT output in the experiment quite understandable. She said as she postedited, she often went back and forth and rechecked her postediting before going to the next sentence. She also added she always reviewed the text after completing postediting. Her BB Flashback recording showed signs of this attitude. In the screen recording of her postediting task, she revised the text as she postedited and performed self-revision at the end of the task.

ISBN: DL:

Based on her postediting experience (in the present experiment), she sometimes preferred MT postediting to translating from scratch. When she was asked to compare the effort spent when translating and postediting, she stated that postediting required less effort than translating.

As advantages of MT postediting, Translator 1 stated that the MT output provided a draft translation and it reduced the frequency of looking up a dictionary with the useful meanings for words and phrases. She added that the MT output decreased the time and effort spent on translating the text. As the disadvantage of postediting MT output, Translator 1 stated that the MT system sometimes offers incorrect and misleading translations. If the translator fails to correct and accept such incorrect translations, the postedited output might be erroneous. When she was asked to mention the problems she came across during the task, Translator 1 stated the following, which underlines the importance of reference to the source text when postediting MT output: “It is easy to distinguish the absurd translations the MT system offers. However, sometimes the MT system may offer misleading translations. If you read only the target text, then the text might sound fluent, but the meaning might be incorrect. For example, in the translation task I have just made, the sentence “Turn the ignition switch OFF” means “Ateşleme düğmesini kapalı konuma getirin (Turn off the ignition switch). But the MT system translates this as “OFF (Lock) Kontak anahtarını açın” (Turn on the OFF (Lock) ignition switch). It probably took “turn” as “turn on”. Such translations may sound correct if the user does not pay attention to the source text. For this reason, it is crucial to refer to the source text. Besides, due to poor translations, it seems easier to translate some sentences from scratch, yet for some others, there is almost no point to be corrected. In fact, it depends on the text type and the richness of the MT system's database.”

Translator 2 did not use any TM or MT system before participating our study. She reported her average daily throughput without any translation aid as more than 5100 words per day. She had no professional experience in postediting either TM or MT output. Besides translating, Translator 2 rarely revises translations and proofread translated texts. She also carries out terminology work depending upon the subject and the content of the text. She did not receive any training on postediting.

Translator 2 found the MT output in the experiment understandable. She stated as she postedited, she often went back and forth and rechecked her postediting before going to the next sentence. She also added she always revised the text after completing

ISBN: DL:

postediting. Her BB Flashback recording showed signs of this attitude. In the screen recording of her postediting task, she revised the text as she postedited and performed self-revision at the end of the task. She added that she usually examined the source text and searched for similar texts on that topic in order to use the correct terminology. However, in this postediting task, she was not able to do this due to limited time. So, she only used Google and on-line dictionaries.

Based on her postediting experience (in the present experiment), she rarely preferred MT postediting to translating from scratch. She said "Normally, I do not prefer to postedit MT output to translate. However, since there might be some exceptions, I selected 'rarely'." When she was asked to compare the effort spent when translating and postediting, she stated that postediting required the same effort as translating. She also added that this depends upon the text, subject and terminology. Depending upon these factors, postediting might require less or more effort than translating or it might require the same effort as translating.

As advantages of MT postediting, Translator 2 stated that (a) the MT output provides more or less a draft of what you will translate. It becomes easier to polish the draft and make complete sentences, (b) It depends on the content of the text. If I consider this text, thanks to the draft translation offered by the MT system, I can say that I did less terminology search than usual. I usually searched for information to determine the best Turkish equivalent rather than finding out what the word or phrase meant. So, it became easier to polish the draft and make complete sentences. As for the disadvantages of postediting MT output, Translator 2 stated that it took more time to postedit when she needed to follow two different texts (the source text and the MT output). She also added correcting an existing text requires much more attention, because she needed to be careful of a possible existing mistake or skipping any item.

Engineer 1 stated that he used Google Translate for translating texts he came across surfing the Internet, providing draft translations and translating some technical documents. Although he said he made use of Google Translate, he thought it should be improved. With this statement, he meant the database in Turkish should be improved, because he said "Google Translate has not been able to provide successful translations in Turkish." When he used MT, he postedited the output by himself. He has not used Google Translator Toolkit before. Engineer 1 found our test text (raw MT output in Turkish) understandable. He stated as he postedited, he sometimes went back and forth and rechecked his postediting before going to the next sentence. He also added he rarely

ISBN: DL:

prefers to revise the text after completing postediting. His BB Flashback recording affirms this statement. In the screen recording of his postediting task, he revised the text as he postedited and performed no self-revision after completing the postediting task.

Based on his postediting experience (either previous or from our experiment), he sometimes preferred MT postediting to translating from scratch. When he was asked to compare the effort spent when translating and postediting, he stated that postediting required less effort than translating.

As an advantage of MT postediting, Engineer 1 stated that the MT output reminded him of some words and phrases which he could not remember at once when he saw the word. He also thought that postediting MT output rather than translating from scratch saved time. As a disadvantage of MT postediting, he stated that when the draft translation was very poor, it became more difficult and slower to rewrite the sentence than translating from scratch.

Engineer 2 said he used Google Translate for translating electronic mails, reports and sometimes simple technical documents. Although he said he made use of Google Translate, he thought "its translations from English into Turkish were difficult to understand. Sometimes, it became more logical to translate each word individually and make the translation by yourself."

When he used MT, he postedited the output by himself. He has not used Google Translator Toolkit before. Engineer 2 found our test text (raw MT output in Turkish) slightly understandable. He stated that as he postedited, he often went back and forth and rechecked his postediting before going to the next sentence. He also added he sometimes preferred to revise the text after completing postediting. However, his BB Flashback recording showed that he did not revise the text at all after completing postediting. In the screen recording of his postediting task, he revised the text as he postedited and performed no self-revision at the end of the task.

Based on his postediting experience (either previous or from our experiment), he sometimes preferred MT postediting to translating from scratch. When he was asked to compare the effort spent when translating and postediting, he stated that postediting required the same effort as translating. While he did not list any advantage of postediting, he stated that "it is necessary to understand the subject-matter and the logic of the sentences" as the disadvantage of postediting. He might mean that he could not accept the translation provided by the MT system as it stood, rather he needed to approach it in a more critical way (i.e. scrutinize the meaning).

ISBN: DL:

Although the proximity among the participants' profiles which can be seen from the questionnaire data does not match with the differences among their quantitative data, we can see that the positive attitude of the participants towards postediting MT output is reflected in the quality of their postedited texts (Engineer 1's and Translator 1's attitude to postediting is more positive than Engineer 2 and Translator 2). Lange and Bennett (2000: 208-209) also mentioned the human factor and underlined how important the positive attitude of the translators towards MT was for the success of the project.

It is also interesting to underline Translator 2's confession that she did not carry out the appropriate preparation and documentation before and during the postediting task as she would do in her usual work atmosphere. She says the reason was that she had limited time for this task. Although all the participants were told to use extra time in case they could not complete the task in the specified time (40 minutes) and although she did so (Translator 2 completed the postediting task in 51.25 minutes excluding self-revision and 67.53 minutes including self-revision), it seems that she did not want to spend much more time on the task.

Post-assignment questionnaire data presented above reveal that the participants' profiles are close to each other. All the same, they are not identical persons. When we look at the translators' profiles, we can see that Translator 1, who obtained more successful results during the experiment in terms of speed and quality, was more familiar with MT systems and had more positive attitude towards the use of MT and postediting when compared to Translator 2. On the other hand, Translator 2 is more experienced in translating than Translator 1. However, this does not affect the speed and the quality of her work positively. Of course, it is not possible to generalize that experience in translation is less significant than positive attitude towards postediting in the speed and the quality of the final product. This is what our small-scale pilot data implies. On the other hand, exploring the situation with more subjects might reveal other factors affecting the processing speed and output quality.

As we explained in Section *6.1.1.2. Subjects*, we did deliberately not selected translators who were specialized in the translation of technical and/or engineering texts, because we aimed to find out how postediting of a technical text by subject-matter experts in engineering and professional translators who did not hold expertise in engineering compare. Nevertheless, this does not mean the subjects have had no experience in technical translations at all. Post-assignment questionnaire data reveal that

ISBN: DL:

both translators have translated technical texts before.

Both translators' replies to the questionnaire and their screen recording reveal that they performed self-revision after completing the postediting task. Translator 2 spent 16.28 minutes on self-revision, while Translator 1 spent 5.28 minutes for this (Table 8). According to the results of this pilot study, we did not find any correlation between the time spent on self-revision and the quality of the postedited text.

Despite their similar profiles which can be observed from the questionnaire data, engineers' posteditings were also different from each others' in terms of speed and quality. As in translators, both engineers have used Google Translate output before our experiment. However, Engineer 1, whose quality result in the postediting experiment is better than Engineer 2 has a more positive view on the MT output from Google Translate and postediting. This finding is supported in both engineers' and translators' data, yet it needs to be further tested.

7.2. Revision

In the second part of this pilot study, we aim to compare the speed and the end-product quality of two different workflows which have been implemented in the translation market: (1) Postediting by professional translators and revision of the postedited output by subject-matter experts, (2) Postediting by subject-matter experts and revision of the postedited output by professional translators. To this end, we hypothesized that when professional translators revise postediting by subject-matter experts and when subject-matter experts revise postediting by professional translators, the speed and end-product quality differ. In order to explore this hypothesis, we had subject-matter experts revise the posteditings performed by translators and we also had professional translators revise the posteditings performed by subject-matter experts. Both groups were asked to record their revision processes using BB Flashback screen recording software. They were free to use the Internet sources during the task.

Although we are aware that our small-scale pilot study conducted with two subject-matter experts and two professional translators does not involve sufficient data to test the relevant hypothesis, with this pilot study, we aim to test our methodology for the main study to be conducted with more participants.

ISBN: DL:

7.2.1. Speed in Revision

We evaluated the revision process in terms of speed and the revised end-products in terms of quality. As in Part 1 of this pilot study, speed was measured with two types of data. First, we used processing speed captured with the screen recording software BB Flashback. Second, we calculated the number of words processed per minute. Both data were recorded for only the revision tasks (only translators' revision or only engineers' revision) as well as for the whole task involving postediting and revision (engineers' postediting + translators' revision and translators' postediting + engineers' revision).

7.2.1.1. Speed in Revision by Subject-Matter Experts

As is seen in their postediting tasks, there are differences between both engineers' revision task results (see Table 17). The mean revision duration for engineers is 33.58 minutes. In order to find the total task time which involve postediting and revision, we added the duration of postediting by translators and the duration of revision by engineers. The mean total task time (translators' postediting + engineers' revision) is 89.83 minutes (Table 17).

In addition to recording the processing speed, we calculated the number of words processed per minute as a measure of speed. First, we calculated the words processed per minute by each reviser during only the revision task. The mean number of w/pm for both engineers is 16.4 (Table 17). Second, in order to find the speed of the whole process of translators' postediting and engineers' revision and compare it with engineers' postediting and translators' revision, we calculated the w/pm during the total task duration (postediting by translators + revision by engineers). The mean number of the words processed per minute during the total task duration is 11.7 (See Table 17).

Table 17. Speed in Revision – Engineers

Subjects (Revisers)	Time on REV Task (minute)	Task Time on PE by Posteditors (minute)	Total Task Time (PE+REV) (minute)	Words processed per minute PE+REV (words processed/total task time)	Words processed per minute only REV (words processed/REV time)
En 1	24.73	Tr 1 - 44.98	69.71	14.2	20.5
En 2	42.41	Tr 2 - 67.53	109.94	9.1	12.2
Mean	**33.58**	**56.26**	**89.83**	**11.7**	**16.4**

Below, we will have a look at the revision results of professional translators. After analyzing the data, we will compare the two cycles involving translators'

ISBN: DL:
postediting + engineers' revision and engineers' postediting + translators' revision.

7.2.1.2. Speed in Revision by Professional Translators

Professional translators revised the texts postedited by subject-matter experts (engineers). This revision process was evaluated in terms of speed and the revised end-products were evaluated in terms of quality. Speed was measured with processing speed and words processed per minute. Revision time (processing speed) was captured with BB Flashback screen recording software. The quality of the final texts was measured with LISA QA Model 3.1.

On average, translators completed the revision process in 22.56 minutes (see Table 18) which is 11.02 minutes faster than engineers' mean revision duration (33.58 minutes, see Table 17). In order to calculate the duration of the total task (engineer postediting + translator revision), we added the processing speed of engineers' postediting and translators' revision of them. The mean total task duration in engineer postediting + translator revision is 72.77 minutes (see Table 18), which is 17.06 minutes faster than the total task duration of translator postediting and engineer revision (89.83 minutes, see Table 17). So, the mean processing speed data show that translators' revision of engineers' postediting is faster than engineers' revision of translators' postediting.

In addition to recording the processing speed, we calculated the number of words processed per minute as a measure of speed. First, we calculated the words processed per minute by each reviser during only the revision task. The translators' mean number of words processed per minute in the revision task is 22.3 (Table 18). If we compare translators' revisions with engineers' revisions using the words processed per minute data, we find parallel results with the processing speed data. That is, translators' revision of engineers' postediting is faster than engineers' revision of translators' posteditings: 22.3 vs. 16.4 (Table 18 and Table 17).

Second, in order to find the speed of the whole process, that is, engineers' postediting + translators' revision, we calculated the words processed per minute during the total task duration (postediting by engineers + revision by translators). The mean number of words processed per minute is 13.5 for engineers' posteditings + translators' revision of them (Table 18). So, in terms of w/pm data, engineers' postediting and translators' revision (13.5 w/pm, see Table 18), seems slightly faster than translators' postediting and engineers' revision (11.7 w/pm, see Table 17) with 1.8 more words

ISBN: DL:

processed per minute in the former process.

Table 18. Speed in Revision – Translators

Subjects (Revisers)	Time on REV Task (minute)	Task Time on PE by Posteditors (minute)	Total Task Time (PE+REV) (minute)	Words processed per minute PE+REV (words processed/total task time)	Words processed per minute only REV (words processed/REV time)
Tr 1	26.21	En 1 - 52.96	79.17	12.2	18.4
Tr 2	18.91	En 2 - 47.45	66.36	14.7	26.1
Mean	**22.56**	**50.21**	**72.77**	**13.5**	**22.3**

7.2.2. Quality in Revision

The figures in *7.2.1.* give us an overview of the speed of the revision by professional translators of the texts postedited by subject-matter experts and vice-versa as measured by the processing speed and the number of words processed per minute. Analysis of such processes cannot be complete without the analysis of the quality of the revised product. Therefore, we measured the quality of the final revised products with LISA QA Model 3.1 in the same way we measured the quality of the postedited texts. However, for revision, the minimum acceptable level of quality was set as 85 percent which was above the quality threshold set for the posteditings (75 percent). The revised texts are the final versions of the posteditings, thus their quality requirement is higher.

First, we compared the end-products revised by professional translators and subject-matter experts. Second, in order to determine how the subject-matter knowledge of translators/engineers as revisers affect the quality of engineers'/translators' postediting, comparison was made between the quality of the postedited texts and their final revised versions.

7.2.2.1. Quality in Revision by Subject-Matter Experts

Engineer 1 revised the text postedited by Translator 1. In this revision, the total error point is 24 and the quality percentage is 95.02 which corresponds to the "Pass" status in LISA (Table 19).

Table 19. Quality in Revision – Engineers

Subjects	Total Error Points	Quality of the Revised Text %	Result
En 1	24	95.02	Pass
En 2	171	64.52	Fail
Mean	**97.5**	**79.77**	**Pass**

ISBN: DL:

If we look at the text postedited by Translator 1 (Table 15), we can see that its quality is already high and revision of this postediting by Engineer 1 led to an increase in the quality by 6.43 percent (95.02 vs. 88.59).

Table 20 presents the distribution of errors in the revised product according to LISA categories.

Table 20. Distribution of Errors - Revision – Engineers

Subjects	**Mistranslation**			**Accuracy**			**Terminology**			**Language**		
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**
En 1		1		1			10	1		3		
En 2				2			3	32		6		
Mean		**1**		**1**			**6. 5**	**16.5**		**4.5**		

When we compare error distribution of the texts postedited by translators (see Table 16) with that of the texts revised by engineers (Table 20), we can see that the number of minor and major terminology errors and the number of minor language errors in the Translator 1’s postediting decreases with Engineer 1’s revision. In this case, we are faced with a high-quality postediting by Translator 1 which is further improved by Engineer 1’s revision.

Engineer 2 revised the text postedited by Translator 2. In this revision, the total error point is 171 and the quality percentage is 64.52 which corresponds to the “Fail” status in LISA (Table 19). If we consider that the quality of the posteditings by both Translator 2 and Engineer 2 is not high, it is not a surprise to us to end up with a low quality final product postedited by Translator 2 and revised by Engineer 2. The quality of the revised end-product by Engineer 2 is 1.86 percent higher than the quality of the text postedited by Translator 2 (64.52 vs. 62.66) (see Table 19 and Table 15).

The comparison of the error distribution of the texts revised by engineers (Table 20) with that of the texts postedited by translators (Table 16) shows that, unlike the pair of Translator 1 and Engineer 1, there are more major terminology errors and minor language errors in the text revised by Engineer 2 than the text postedited by Translator 2 (32 vs. 27 for terminology and 6 vs. 2 for language errors). On the other hand, in other categories, we can see a decrease in the number of errors in the revised text.

To sum up, revision by Engineer 1 and Engineer 2 of the posteditings by Translator 1 and Translator 2 led to an increase in the quality by 6.43 and 1.86 percent

ISBN: DL:

respectively (see Table 15 and Table 19). On average, by reaching the quality percentage of 79.77; engineer revision improved the quality of translator posteditings (having the mean quality percentage of 75.63) by 4.14 percent.

7.2.2.2. Quality in Revision by Professional Translators

Translator 1 revised the text postedited by Engineer 1. In this revision, the total error point is 32 and the quality percentage is 93.36 which corresponds to the "Pass" status in LISA (Table 21). If we look at the text postedited by Engineer 1 (Table 13), we can see that its quality is already high and revision of this postediting by Translator 1 led to an increase in the quality by 2.07 percent (93.36 vs. 91.29).

Table 21. Quality in Revision – Translators

Subjects	Total Error Points	Quality of the Revised Text %	Result
Tr 1	32	93.36	Pass
Tr 2	162	66.39	Fail
Mean	**97**	**79.88**	**Pass**

Translator 2 revised the text postedited by Engineer 2. In this revision, the total error point is 162 and the quality percentage is 66.39 which corresponds to the "Fail" status in LISA (Table 21). If we consider that the quality of the posteditings by both Engineer 2 and Translator 2 is not high, it is not a surprise to us to end up with a low quality final product postedited by Engineer 2 and revised by Translator 2. However, this time, revision by Translator 2 of the postediting by Engineer 2 led to a 2.07 percent decrease in the quality (68.46 vs. 66.39) (see Table 13 and Table 21).

Table 22 presents the distribution of errors in the revised product according to LISA categories.

Table 22. Distribution of Errors - Revision – Translators

Subjects	Mistranslation			Accuracy			Terminology			Language			Consistency		
	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.
Tr 1				2				3		1					
Tr 2	1	1		2			6	29		14			3		
Mean	**1**	**1**		**2**			**3**	**16**		**7.5**			**1.5**		

Two figures in Table 22 are interesting. These are 29 major terminology errors and 14 minor language errors made by Translator 2 during revision. When we watch the

ISBN: DL:

screen recording data, we can easily detect that the reason for the availability of that much major terminology errors is the repetative erroneous revision of the word, "dağıtmak" (meaning "to scatter"), which is "to deploy" in the English source text and should be translated/revised as "patlatmak" (to explode).

As for 14 minor language errors, both the screen recording data of revision and the revised output show that Translator 2 does not revise the text very carefully and makes some minor punctuation or typing errors. Unexpectedly, Engineer 2 made more major terminology errors (32 vs. 29) and less language errors (6 vs. 14) than Translator 2 while performing revision (Table 20 and Table 22).

Since the mean quality percentage of engineers' postediting and translators' revision of them are the same - both 79.88 percent (See Table 13 and Table 21), with this pilot study, it is not possible to say if translators' revision improves or reduces the quality of engineers' postediting. On the other hand, comparison of the quality percentages and error points of translators' postediting and engineers' revision of them reveals slight improvement in the quality of posteditings when they are revised by engineers. In other words, when revised by engineers, we can observe 4.14 percent improvement in the quality percentage of translators' posteditings. Although this finding requires further testing with more participants, it shows that subject-matter expert revision improves the quality of translators' postediting.

ISBN: DL:

8. Conclusions of the Pilot Study

8.1. Postediting

In Section 7.1, we presented the preliminary results of Part 1 of our pilot study which included postediting of MT output carried out with four subjects (two engineers and two professional translators). Obviously, this small-scale pilot study is not aimed to draw conclusions but to test our methodology for the main research to be conducted with a total of twenty subjects (ten engineers and ten professional translators). Nevertheless, below, we will interpret what the pilot study data mean with reference to our hypotheses.

We hypothesized that subject-matter experts and professional translators postedit MT output differently in terms of processing speed, time spent on documentation and the number of changes performed during the postediting process. Under this hypothesis, we formulated three sub-hypotheses. The first sub-hypothesis suggests that professional translators complete the postediting task faster than subject-matter experts do. This may sound to contradict with the second sub-hypothesis suggesting that translators spend more time on documentation than engineers and with the third one stating that translators make more changes than engineers. One may expect the task to be more time-consuming if the posteditor carries out more documentation and makes more changes. However, our logic behind the first sub-hypothesis lies under professionalism. Although we hypothesize that the professional translators perform more documentation and changes on the technical text they postedit, we expect them to complete the task faster than engineers, because the translators are the professionals of translating or postediting while engineers are not.

Both engineers participating in the pilot study completed the postediting task without performing any self-revision while both translators did self-revision right after finishing the task. If we compare the processing speed of both groups excluding translators' self-revision, then this first sub-hypothesis is supported. Table 23 provides comparison of engineers' and translators' data on the speed of the postediting process.

ISBN: DL:

Table 23. Speed in PE - Translators and Engineers

Subjects	Time on PE Excluding Self-revision (minute)	Time on Self-revision (minute)	Total Task Time (PE and Self-rev) (minute)	Words processed per minute (words processed/ total task time)	Words processed per minute (excluding self-revision)	Time Spent on Document. (minute)	Document. as % of the Total Task Time	Number of Changes
Tr 1	39.70	05.28	44.98	10.7	12.1	07.51	16.7	150
Tr 2	51.25	16.28	67.53	7.1	9.4	07.51	11.1	135
Mean	**45.48**	**10.78**	**56.26**	**8.9**	**10.8**	**07.51**	**13.3**	**143**
En 1	52.96	0	52.96	9.1	0	02.76	5.2	155
En 2	47.45	0	47.45	10.1	0	00.33	0.7	107
Mean	**50.21**	**0**	**50.21**	**9.6**	**0**	**01.55**	**3.07**	**131**

On average, engineers completed the task in 50.21 minutes and translators finished postediting in 45.48 minutes (Table 23) However, the results change when we include translators' self-revision duration. In that case, translators complete the whole task in 56.26 minutes. We need to conduct the study with more subjects to understand if self-revision is a routine part of many translators' postediting and to see if it is always disregarded by engineers. We can, then, decide which duration (including or excluding self-revision) we will take into account. Nevertheless, if we currently compare both groups, it seems more logical to do this comparison by considering translators' self-revision. This is because we made the rest of the evaluations regarding other factors such as documentation, changes performed and quality on the basis of the final products of each subject. Obviously, translators' final products are the postedited texts which involve their self-revision process. Based on this information, it is more convenient to say that the first sub-hypothesis is not supported and engineers completed the task faster than translators (50.21 minutes vs 56.26 minutes, Table 23). This might be because engineers think and act in a more practical and short-cut way and translators approach the text in a more scrupulous way. Translators' preference to perform self-revision versus engineers' preference to quit the process right after they finish postediting might be an indicator of translators' scrupulousness as compared to engineers.

The main idea behind registering the processing speed of both groups is to compare their speed. A more precise way of measuring speed is to determine the number of words processed per minute. Table 23 shows that the mean number of words processed per minute for engineers is 9.6, while this figure is 8.9 for translators. Although the difference is small, both processing speed data and the number of words

ISBN: DL:

procesed per minute indicate that engineers work faster than translators.

Our second sub-hypothesis under Hypothesis 1 states that professional translators spend more time on documentation than subject-matter experts. Our data support this sub-hypothesis. On average, engineers spent 3.07 percent of their total task time on documentation, while translators spent 13.3 percent of their task time on searching for information (see Table 23). When we look at documentation data in detail, we can see that both groups used similar number of Internet sources (translators 3 vs engineers 2, see Table 24). However, translators searched for information almost three times more than engineers (translators 16 times vs engineers 5.5 times). Moreover, translators spent almost 5 times more time on documentation (translators 7.51 min. vs engineers 1.55 min). Finally, translators searched four times more items than engineers (translators 14 items vs engineers 3.5 items). This difference can be explained both with the engineers' knowledge on the subject-matter and the translators' scrupulousness. Although Engineers 2's use of very little documentation affects the results, comparison of the detailed documentation data of Engineer 1 (see Table 6 in Section 7.1.1.1) and those of Translator 1 and Translator 2 (see Table 10 and Table 11 in Section 7.1.1.2) implies that this difference results from engineers' knowledge on the subject-matter rather than translators' scrupulousness. This is because Engineer 1 searched for less items more times than translators (Engineer 1 searched for 6 items 10 times, whereas translators looked up 14 items 16 times), which implies that Engineer 1 might be more scrupulous than translators (see Table 24 as well).

Table 24. Documentation in PE - Translators and Engineers

Subjects	Number of Look-up	Time spent (minute)	Number of Sources Used	Number of Items Searched
Tr 1	16	7.51	3	14
Tr 2	15	7.51	2	13
Mean	**16**	**7.51**	**3**	**14**
En 1	10	2.76	3	6
En 2	1	0.33	1	1
Mean	**5.5**	**1.55**	**2**	**3.5**

Our third sub-hypothesis under Hypothesis 1 suggests that professional translators perform more changes than subject-matter experts do. This sub-hypothesis is slightly supported by our data in the pilot study, yet the difference between both groups

ISBN: DL:

is not striking. While postediting, translators made 143 changes and engineers made 131 changes (see Table 25). The logic behind this hypothesis was the assumption that translators would approach the MT output in a more critical way and try to improve it, while engineers would make do with what reflects the correct meaning. Although the number of changes by translators is a little more than the number of changes by engineers, detailed data on the changes performed by each group (see Table 25) do not support this assumption.

Table 25. Changes in PE - Translators and Engineers

Subjects	Content Changes	Formal Changes		Omissions	Additions	Reorganization	Total
		Capitals	Punctuation				
Tr	61	6.5	5.5	17	41	11.5	**143**
En	62.5	4.5	8.5	15	32	8.5	**131**

Our second hypothesis suggests that, for technical texts, the quality of postediting by subject-matter experts is higher than postediting by professional translators. The result of our quality analysis with LISA QA Model 3.1 supports this hypothesis. The total error point of engineers is 97 and it is 118 for translators. The mean percentage of postediting quality is measured as 79.88 for engineers, while it is 75.63 for translators (Table 26). Although the difference is small and our small-scale pilot study does not present enough data to draw conclusions, this seems to be a slight difference between the qualities of both groups' work in favor of subject-matter experts.

It is important to remind that none of the engineers participating in this study is currently working as a translator or has work experience as a translator. The quality result is interesting, because, although they may not have as much subject-matter knowledge as engineers, one may also expect that professional translators should be more meticulous in translating and/or postediting a text which would finally represent the quality of their professional work. Translators do spend more time on documentation and the postediting task itself. Unlike engineers, they perform self-revision which is expected to improve the final product. However, the extra time and effort spent by translators on documentation and self-revision do not lead to higher-quality products.

The subject-matter knowledge of the engineers might be a factor in the superiority of the quality of their work. On the other hand, quality analysis of the

ISBN: DL:

posteditings shows that one participant in each group did a better job than the other (see Table 26). If we take Engineer 1 and Translator 1 into consideration, we can see that both received a "Pass" from LISA and the quality percentage difference between their works is not that sharp. Likewise, when we compare the quality results of Engineer 2 and Translator 2, we can see that both received a "Fail" from LISA and their quality percentages are close to each other. This makes it hard to associate quality directly with subject-matter expertise of engineers. In order to clarify this point, we need to replicate the study with more participants.

Table 26. Quality in PE - Translators and Engineers

Subjects	Total Error Points	Postediting Quality %	Result
Tr 1	55	88.59	Pass
Tr 2	180	62.66	Fail
Mean	**118**	**75.63**	Pass
En 1	42	91.29	Pass
En 2	152	68.46	Fail
Mean	**97**	**79.88**	Pass

To sum up, in Part 1 of this pilot study, we compared the postediting of technical texts by subject-matter experts and professional translators in terms of documentation, changes performed, speed and quality. Although our data is too small to draw conclusions, the preliminary results of this pilot study show that in the postediting of technical texts, subject-matter experts (in our case engineers) perform fewer changes on the raw MT output while postediting it. Engineers carry out less documentation than translators during the postediting task and they produce higher-quality texts in shorter period of times, thus work faster than translators (See Table 27).

Table 27. Summary of Speed and Quality Data in PE

Subjects	Mean Total Task Time (minute)	Mean Words processed per minute (words processed/total task time)	Mean Total Error Points	Mean Postediting Quality %
Tr	56.26	8.9	118	75.63
En	50.21	9.6	97	79.88

ISBN: DL:

8.2. Revision

In Section 7.2, we presented the preliminary results of Part 2 of our pilot study regarding revision carried out with four subjects (two engineers and two professional translators). Obviously, this small-scale pilot study is not aimed to draw conclusions, but to test our methodology for the main research to be conducted with a total of twenty subjects (ten engineers and ten professional translators). Nevertheless, we will try to interpret what the pilot study data mean with reference to our hypothesis on revision.

For Part 2 of this pilot study, we formulated a general hypothesis stating that when professional translators revise postediting by subject-matter experts and when subject-matter experts revise postediting by professional translators, the speed and end-product quality differ.

Speed was measured with the processing speed and the number of words processed per minute. Both data were recorded for the revision tasks (only translators' revision or only engineers' revision) as well as for the whole task involving postediting and revision (engineers' postediting + translators' revision and translators' postediting + engineers' revision).

Processing speed data show that, while revising, translators spent 11.02 minutes less time than engineers. If we take the whole process including postediting and revision into consideration, we find that engineers' postediting + translators' revision is 17.06 minutes faster than the total task duration of translator postediting and engineer revision (see Table 28).

Table 28. Speed in Revision - Translators and Engineers

Subjects (Revisers)	Time on REV Task (minute)	Task Time on PE by Posteditors (minute)	Total Task Time (PE+REV) (minute)	Words processed per minute PE+REV (words processed/total task time)	Words processed per minute only REV (words processed/REV time)
Tr	22.56	En - 50.21	72.77	13.5	22.3
En	33.58	Tr - 56.26	89.83	11.7	16.4
Difference	**11.02**	**06.05**	**17.06**	**1.8**	**5.9**

If we compare translators' revisions with engineers' revisions using the words processed per minute data, we find results which are parallel with the results of the processing speed data. That is, on average, while revising engineers' postediting, translators process 5.9 more words than engineers' revision of translators' posteditings.

ISBN: DL:

However, the difference declines when we look at the whole processes of postediting + revision. In engineer postediting + translator revision, the number of w/pm is only 1.8 more than that in translator postediting + engineer revision (Table 28). Nevertheless, our pilot study shows that translators' revision of engineers' postediting is faster than engineers' revision of translators' postediting (Table 28).

Our hypothesis formulated for the revision part of this study suggests that when professional translators revise postediting by subject-matter experts and when subject-matter experts revise postediting by professional translators, the speed and end-product quality differ. Above, we presented data analysis regarding speed. Now, we will look at the quality aspect in both translators' and engineers' revision processes. Although the speed data are in favor of translators' revision of engineers' postediting rather than engineers' revision of translators' postediting, quality data do not reveal such a difference in favor of either process. Table 29 presents the mean quality results of engineers' and translators' revised products measured by LISA QA Model 3.1. Engineers' total error point is very close to translators' total error point, thus the quality percentages of their final revised texts are almost the same. This may be caused by the existence of one successful translator and one successful engineer pair (Translator 1 and Engineer 1) and one unsuccessful translator and one unsuccessful engineer pair (Translator 2 and Engineer 2) in our small subject group.

Table 29. Quality in Revision - Translators and Engineers

Subjects (Revisers)	Total Error Points	Quality of the Revised Text %	Result
Tr	97	79.88	Pass
En	97.5	79.77	Pass

Analysis of the screen recordings of the postediting experiments in Part 1 of this pilot study reveals that similar errors were made both by Engineer 2 and Translator 2 during postediting. During postediting, both participants failed to correct the translation of the repetitive word "to deploy" (to explode) which was incorrectly translated by the MT system as "dağıtmak" (to scatter). Not surprisingly, the same participants failed to correct the same error in one another's postediting. From this result, we can deduce that searching for the correct information by means of documentation may be more important than subject-area knowledge (supposing that Engineer 2 had the subject-area knowledge). On the other hand, while postediting, Translator 1 and Translator 2 spent

ISBN: DL:

exactly the same amount of time (7.51 minutes) on documentation. However, Translator 1 correctly postedited the term "dağıtmak" (scatter), while Translator 2 failed to correct it. This shows us that time spent on documentation may not be a determinant factor on the quality of the final product.

When we compare the quality of the posteditings with the quality of their revised versions, we find that engineer revision brought about 4.14 percent improvement in the quality of the texts postedited by translators (See Table 30). Engineer revision also decreased the total error point of translator postediting by 20. On the other hand, translators' revision does not lead to any changes in the quality of the texts postedited by engineers. Thus, with this pilot study, it is not possible to say if translators' revision improves or reduces the quality of engineers' postediting. Obviously, the pilot study should be extended to include more participants to obtain more meaningful results.

Table 30. Quality Improvement Brought About by Revision

Task Type	Engineers / Revision	Translators / Postediting	Quality Difference
Total Error Point	97.5	118	**20**
Quality of the Revised Text	79.77%	75.63%	**4.14%**
Task Type	Translators / Revision	Engineers / Postediting	Quality Difference
Total Error Point	97	97	**0**
Quality of the Revised Text	79.88%	79.88%	**0**

To sum up, when we consider the whole process involving postediting + revision, translators' revision of engineers' postediting is faster than engineers' revision of translators' postediting (see Table 28), but the quality of the end-products of both processes is almost the same (see Table 29).

On the other hand, when we compare the postedited texts with their revised versions, engineers' revision of translators' postediting brings quality improvement, while translators' revision of engineers' posteditings brings no quality improvement (see Table 30). Thus, although the end-product quality of translators' postediting + engineers' revision and that of engineers' postediting + translators' revision is almost the same, comparison of the postedited texts with the revised texts reveals that engineers' revision brings quality improvement to translators' postediting, yet it brings no advantage over translators' revision of engineer postediting in terms of speed. Since

ISBN: DL:

speed is not meaningful unless it is combined with quality, it is possible to hypothesize that engineer revision of translator postediting brings better results than translator revision of engineer postediting.

UNIVERSITAT ROVIRA I VIRGILI
POSTEDITING MACHINE TRANSLATION OUTPUT AND ITS REVISION: SUBJECT-MATTER EXPERTS VERSUS PROFESSIONAL TRANSLATORS
Özlem Temizöz
ISBN: DL:

ISBN: DL:

9. Main Study - Data Analysis and Results

9.1. Postediting

9.1.1. Speed, Documentation and Changes in Postediting

After we completed and analyzed the results of the pilot study, we conducted the main study with ten subject-matter experts (engineers) and ten professional translators. We used the same text (one machine-translated 482-word technical text in Turkish on dismantling end-of-life vehicles). It was translated from English into Turkish with Google Translate and we used the same methods as in the pilot study to identify speed. Thus we measured speed in two ways: capturing the processing speed with screen recording and calculating the words processed per minute. By means of the screen recording software BB Flashback, we determined the time spent on the different phases of the postediting process (only postediting and self-revision). We watched each subject's screen recording data, noting the exact time when the subject completed the postediting task and started self-revision. We calculated the words processed per minute by dividing the total number of the words processed by the total time spent on the postediting task. We made this calculation to determine the processing speed both including and excluding the self-revision process.

The only difference between the experiment designs of Part 1 of the pilot study and the main study was with respect to the postediting duration. In the pilot study, the participants were given forty minutes to complete the task. However, they were also told to use extra time if they were not able to complete the task within this time frame. Since the mean task time in the pilot study was 50 minutes for engineers and 56 minutes for translators, in the main study, we decided to set the time limit for postediting at 50 minutes. As in the pilot study, the participants were free to use extra time if they were unable to complete the task within this time frame. Most participants exceeded this time limit and completed the task within 1 hour or more.

On average, the translators completed the postediting task in 54.68 minutes when we exclude the time spent on self-revision. This figure is 51.18 minutes for engineers (see Table 31). In the pilot study, the translators self-revised, while the engineers did not. When we look at the results of the main study, we can see a parallel trend. While

ISBN: DL:

postediting, 60 percent of the translators made self-revision for an average of 8.03 minutes. On the other hand, only 30 percent of the engineers performed self-revision, for an average of 4.48 minutes. This suggests that translators may be more prone to self-revise their posteditings than are engineers. When we realized this trend, we decided to include self-revision durations in the total task time in order to obtain more realistic data on speed. When we include self-revision, the translators completed the postediting task in an average of 62.70 minutes and engineers completed it in an average of 55.66 minutes (see Table 31). Including self-revision time, the translators processed 8.3 words per minute and the engineers processed 9.1 words per minute (Table 32).

Table 31. Speed in PE 1 - Translators and Engineers

Subjects	Time on PE Excluding Self-revision (minute)	Time on Self-revision (minute)	Total Task Time (PE and Self-rev) (minute)	Time Spent on Documentation (minute)	Documentation as % of the Total Task Time
Tr 1	39.70	05.28	44.98	07.51	16.7
Tr 2	51.25	16.28	67.53	07.51	11.1
Tr 3	45.36	00.00	45.36	02.65	5.8
Tr 4	37.45	22.71	60.16	13.08	21.7
Tr 5	92.06	15.68	107.75	28.00	25.9
Tr 6	38.73	00.00	38.73	05.00	12.9
Tr 7	54.28	14.46	68.75	08.76	12.8
Tr 8	62.66	00.00	62.66	04.56	7.3
Tr 9	60.03	00.00	60.03	08.86	14.7
Tr 10	65.23	05.85	71.08	15.08	21.2
Mean	**54.68**	**08.03**	**62.70**	**10.75**	**16.1**
En 1	52.96	00.00	52.96	02.76	5.2
En 2	47.45	00.00	47.45	00.33	0.7
En 3	46.00	00.00	46.00	03.70	8.0
En 4	65.61	00.00	65.61	03.93	6.0
En 5	60.96	00.00	60.96	01.51	2.5
En 6	35.28	09.23	44.51	00.96	2.1
En 7	61.70	00.00	61.70	00.48	0.7
En 8	61.05	23.25	84.30	09.35	11.0
En 9	95.04	00.00	35.40	01.30	3.6
En 10	45.43	12.30	57.73	02.91	5.0
Mean	**51.18**	**04.48**	**55.66**	**02.71**	**4.8**
p-value	**0.5860**	**0.3498**	**0.3598**	**0.0078**	**0.0002**

ISBN: DL:

Table 32. Speed in PE 2 - Translators and Engineers

Subjects	Words processed per minute	Words processed per minute (excluding self-revision)	Number of Changes
Tr 1	10.7	12.1	150
Tr 2	7.1	9.4	135
Tr 3	10.6	10.6	80
Tr 4	8.0	12.8	147
Tr 5	4.5	5.2	148
Tr 6	12.4	12.4	58
Tr 7	7.0	8.8	139
Tr 8	7.6	7.6	172
Tr 9	8.0	8.0	158
Tr 10	6.7	7.3	110
Mean	**8.3**	**9.4**	**130**
En 1	9.1	9.1	155
En 2	10.1	10.1	107
En 3	10.4	10.4	97
En 4	7.3	7.3	129
En 5	7.9	7.9	124
En 6	10.8	13.6	71
En 7	7.8	7.8	128
En 8	5.7	7.8	165
En 9	13.6	13.6	119
En 10	8.3	10.6	140
Mean	**9.1**	**9.8**	**124**
p-value	**0.4195**	**0.7157**	**0.6709**

In order to test the significance of these results, we used an unpaired t-test. First we calculated the time-on-task in seconds. This includes the time on postediting and the time on self-revision, and thus the time on the total task (postediting + self-revision). Additionally, we used the number of the words processed per minute, both including and excluding self-revision. Before conducting the t-tests, we first checked whether our data had parametric (normal) distribution by using the Kolmogorov-Smirnov test. We found that most of our data are normally distributed. In cases when the data are normally distributed, we calculated the p-value using unpaired t-tests. In cases when the data are not normally distributed, we applied SPSS to calculate the p-value. The p-value for the engineers' and the translators' postediting time is 0.5860 and it is 0.3498 for the two groups' time spent on self-revision. The p-value for the engineers' and the translators' total task time involving both postediting and self-revision is 0.3598. It is 0.4195 for the comparison of both groups in terms of the number of the words processed per minute and this figure is 0.7157 when we exclude the self-revision duration (Table 32). The unpaired t-test results for all of the speed indicators listed

ISBN: DL:

above indicate that there is no significant difference between translators and engineers in terms of the speed of postediting a technical text. This finding is not in line with what we found in the pilot study and it refutes our first sub-hypothesis (hypothesis 1a), which posited that professional translators postedit technical texts faster than engineers.

In order to collect data on documentation, we watched the screen recording data and noted when the subjects started and finished doing a search on the Internet or when they used an on-line dictionary. Then we elicited data including the number of look-ups, the exact time (hour.minute.second.split second) when the look-up started and finished, the source used and the item of the text that was searched. By using these data, we calculated the total time each subject spent on documentation as well as details about the sources used and the items searched.

Although the time spent on documentation changes depending on the participant, all of the translators and the engineers performed documentation while postediting. However, there is a clear difference between translators and engineers in the amount of documentation performed (Table 31). On average, translators performed documentation for 10.75 minutes, while engineers carry out documentation for 2.71 minutes (see Table 31). So documentation took 16.1 percent of the translators' total task duration, while it took only 4.8 percent of engineers' total task duration (Table 31).

The p-value for the time spent on documentation is 0.0078. Since this is much lower than the threshold 0.05, we can say that this finding is statistically significant. When we compare the percentiles for the documentation duration of both groups, we found the p-value of 0.0002 (see Table 31), which represents a high degree of statistical significance. This result is parallel to what we found in the pilot study regarding documentation and it clearly affirms our second sub-hypothesis (Hypothesis 1b) that professional translators perform more documentation than engineers while they are postediting a technical text.

A more detailed look at the data reveals difference between the translators' and engineers' documentation behaviour as well. Translators look up the meaning of the words and phrases 2.3 times more than engineers (19 vs 8) and the number of the items they search is 3 times more than is the case for the engineers (21 vs 7, see Table 33). The p-value for the number of look-up is 0.0156 and it is 0.0050 for the number of items searched (see Table 33). These figures affirm that our findings regarding the number of Internet searches and that of the text-items searched for are statistically significant. This difference can be explained by the engineers' advantage over

ISBN: DL:

translators in terms of subject-matter expertise. However, both groups used a very limited number of sources on the Internet while searching for information. Translators used only three sources and engineers used two sources on the Internet (Table 33). The p-value for the number of the Internet sources used during documentation is 0.2448, which represents that the difference between the two groups is not statistically significant. It is interesting to find that the translators and engineers use similar and a very limited number of Internet sources although other factors regarding documentation indicate statistically significant difference between the two groups.

Table 33. Documentation in PE - Translators and Engineers

Subjects	Number of Look-up	Time spent (minute)	Number of Sources Used	Number of Items Searched
Tr 1	16	07.51	3	14
Tr 2	15	07.51	2	13
Tr 3	8	02.65	3	9
Tr 4	15	13.08	6	19
Tr 5	41	28.00	4	44
Tr 6	11	05.00	1	11
Tr 7	21	08.76	2	31
Tr 8	6	04.56	2	7
Tr 9	25	08.85	2	29
Tr 10	30	15.08	3	31
Mean	**19**	**10.75**	**3**	**21**
En 1	10	02.76	3	6
En 2	1	00.33	1	1
En 3	11	03.70	3	7
En 4	9	03.93	3	8
En 5	6	01.51	1	5
En 6	2	00.96	1	3
En 7	2	00.48	1	2
En 8	23	09.35	4	21
En 9	4	01.30	1	4
En 10	13	02.91	3	13
Mean	**8**	**02.71**	**2**	**7**
p-value	**0.0156**	**0.0078**	**0.2448**	**0.0050**

Individual data on the documentation reveal that these limited numbers of Internet sources are composed of some popular online dictionaries in English-Turkish and Turkish-English and, not surprisingly, Google (Table 34).

ISBN: DL:

Table 34. Detailed Documentation Data in PE - Translator 9

Number of Look-up	Time Spent on Documentation (min.sec.split sec.)			Source Used	Item Searched
	Start	Finish	Time used		
1	01.56.3	02.30.8	00.34.5	Tureng online dictionary	onboard
2	06.37.8	06.43.4	00.05.6	Tureng online dictionary	battery cable terminal
3	07.30.1	07.35.9	00.05.9	Tureng online dictionary	ignition switch
4	07.58.8	08.20.3	00.21.5	Tureng online dictionary	minus terminal, negative terminal
5	09.23.1	09.39.3	00.16.1	Tureng online dictionary	overview table, overview
6	10.55.7	11.21.9	00.26.1	Tureng online dictionary	glass scattering, scatter
7	12.05.00	12.16.14	00.11.1	Tureng online dictionary	deployment
8	13.30.07	13.52.93	00.22.9	Tureng online dictionary	deployment
9	15.43.1	15.50.1	00.07.1	Tureng online dictionary	deployment
10	17.28.4	17.45.0	00.16.6	Tureng online dictionary	procedure
11	20.22.9	20.40.1	00.17.3	Tureng online dictionary	deployment harness, harness
12	24.05.4	24.24.5	00.19.1	Tureng online dictionary	tool
13	25.07.1	25.17.3	00.10.1	Tureng online dictionary	operational
14	27.39.3	27.48.3	00.09.0	Tureng online dictionary	harness
15	28.50.9	29.02.7	00.11.8	Tureng online dictionary	pretensioner
16	29.35.5	30.00.5	00.25.0	Tureng online dictionary	multiplug
17	31.23.3	31.51.1	00.27.9	Tureng online dictionary	wire
18	32.50.9	33.22.5	00.31.6	Tureng online dictionary	wire
19	33.36.7	34.02.5	00.25.8	Tureng online dictionary, Google	insulation wire
20	35.26.8	36.21.2	00.54.4	Tureng online dictionary, Google	flylead, fly lead, patch cord
21	36.59.4	37.18.4	00.19.0	Tureng online dictionary	terminal
22	38.17.3	39.27.7	01.10.4	Tureng online dictionary, Google	stage device, dual stage
23	40.16.0	40.19.0	00.03.0	Tureng online dictionary	gas generator
24	48.30.1	48.55.5	00.25.4	Tureng online dictionary	battery leads, lead
25	53.43.2	53.56.5	00.13.3	Tureng online dictionary	sequence
Total			**08.50.5**		

In order to identify the number and type of changes performed by the posteditors, we watched the screen recordings, stopped them at each change and recorded these changes using six categories: content changes, formal changes (capitals and punctuations), omissions, additions and reorganizations (see *7.1.1.* above for an explanation of these categories).

After counting all the changes performed by the subjects under these categories, we calculated the total number of changes. Table 35 presents the changes performed by the translators and engineers while postediting. In each category, the mean numbers of changes made by translators and engineers are very close to each other. The mean total

ISBN: DL:
number of changes made by the translators is 130; this figure is 124 for the engineers. The p-value for the number changes performed is 0.6709, which is far from being statistically significant (see Table 35). Thus, this result in the main study does not support our third sub-hypothesis (under hypothesis 1) suggesting that while postediting a technical text, professional translators perform more changes than subject-matter experts do (see Table 35). This finding is not parallel with the findings of the pilot study (see Table 25).

The logic behind our third sub-hypothesis was the assumption that translators would approach the MT output in a more critical way and try to improve it, while engineers would make do with anything reflecting the correct meaning. Although the number of changes by translators is a little higher than the number of changes by engineers, detailed data on the changes performed by each group (see Table 35) do not support the hypothesis. This shows us that the type and number of changes are more text-dependent than subject-dependent. That is to say, the changes are the results of the amendments required by the machine-translated text rather than subjects' preferences. The technical nature of the test text plays a role in this result: since the text was rarely open to different interpretations, the posteditors did not have much room for subject-specific changes. Rather, they made only the changes that were necessary to improve the MT output to a level that was close to a human translation.

ISBN: DL:

Table 35. Changes in PE - Translators and Engineers

Subjects	Content Changes	Formal Changes		Omissions	Additions	Reorganization	Total
		Capitals	Punctuation				
Tr 1	62	5	7	20	46	10	150
Tr 2	60	8	4	14	36	13	135
Tr 3	29	8	1	13	19	10	80
Tr 4	82	4	2	10	38	11	147
Tr 5	82	3	5	18	30	10	148
Tr 6	10	0	0	12	26	10	58
Tr 7	65	8	8	18	27	13	139
Tr 8	76	8	5	14	58	11	172
Tr 9	93	8	4	12	30	11	158
Tr 10	41	4	5	18	29	13	110
Mean	**60**	**6**	**4**	**15**	**34**	**11**	**130**
En 1	83	5	12	12	33	10	155
En 2	42	4	5	18	31	7	107
En 3	32	5	2	12	35	11	97
En 4	81	5	1	15	14	13	129
En 5	76	5	4	11	19	9	124
En 6	31	1	2	8	24	5	71
En 7	63	7	3	8	32	15	128
En 8	74	10	10	21	39	11	165
En 9	53	5	3	17	28	13	119
En 10	53	4	4	15	50	14	140
Mean	**59**	**5**	**5**	**14**	**31**	**11**	**124**
p-value	**0.9089**	**0.6692**	**0.7228**	**0.4933**	**0.4874**	**0.7157**	**0.6709**

9.1.2. Quality in Postediting

After obtaining the results on speed, documentation and the changes performed during the postediting process, we analyzed our data to find out the quality of the postedited products using LISA QA Model 3.1. (See 6.1.1.5 above). Detailed information on how LISA QA Model 3.1 user interface works and how we analyzed the product quality is given in the pilot study, Section *7.1.2. Quality in Postediting*. In the main study, we used the same tool and method as in the pilot study.

In the pilot study, we had four participants: two engineers and two translators. When we completed the quality analysis in the pilot study, we find that there was a gap between the postediting quality of the participants *within* each group. Further analysis showed that this difference resulted from the failure to correct the same terms recurring throughout the text - terms that were incorrectly translated by the MT system. Due to

ISBN: DL:

the recurrent nature of the errors, the number of errors, and thus the quality of the posteditings as measured by LISA QA Model 3.1, change when we penalize or unpenalize the recurring errors. In an article presenting the results of a study benchmarking eleven translation quality evaluation models, O'Brien (2012: 10) states that:

> only three of the QE models give instructions on how to deal with recurring errors. In two cases, the model specifically rules out the counting of repeated errors. In the third case, whether or not an error is counted more than once depends on the nature of the error: if the error results from translator negligence or lack of grammatical knowledge, the error is counted each time it occurs. If, on the other hand, the error is not the fault of the translator (e.g. the term was not included in the glossary), it is counted only once.

Since counting or uncounting the recurring errors are both ways of quality analysis employed in the translation market, the difference between the results of each method, if any, could be significant in terms of the validity of the QE models. This has led us to approach the quality analysis in two ways. First, we carried out the quality analysis by taking into consideration the recurring errors. Second, we conducted the quality analysis by disregarding them. For the former analysis, we counted each error each time it occurred throughout the text. For the latter, namely, when we disregarded the recurring errors, we counted each error only once and disregarded the recurrent versions of the same error. Below we present the results of both types of quality analysis.

9.1.2.1. Quality in Postediting - Recurring Errors Penalized

Table 36 presents the error distribution of each translator and engineer when we penalize all errors, including the recurrent ones. The mean error count shows that the most striking difference between translators and engineers is in the terminology errors. The mean number of minor terminology errors for translators is 18.9, while this figure is 5.7 for engineers. The mean number of major terminology errors for translators is 25.6, while this figure is 14 for engineers. Translators make three times more minor terminology errors and twice as many major terminology errors than do engineers. In order to determine the significance of the findings, we used unpaired t-test for each error category. According to the unpaired t-tests, only the result for the minor

ISBN: DL:

terminology errors is statistically significant. The p-value for the minor terminology errors is 0.0025, which represents a high degree of statistical significance. For the major terminology category, the p-value is 0.0774, which is not quite significant, although it is rather lower than the results of other categories. The difference between both groups in terms of the terminology errors can be explained by the advantage that the engineers possess in terms of subject-area knowledge. On the other hand, in *9.1.1. Speed, Documentation and Changes in Postediting,* we mentioned that translators spent nearly four times as long on documentation than engineers (10.75 vs 2.71 minutes, see Table 33) during the postediting task. The results show that although translators spent more time searching for information on the Internet, they were unable to compensate for their lack of subject-area knowledge.

ISBN: DL:

Table 36. Distribution of Errors - Translators and Engineers - Recurring Errors Penalized

Subjects	Mistranslation			Accuracy			Terminology			Language			Consistency		
	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.
Tr 1	0	1		1	0		19	5		5	0		0	0	
Tr 2	1	1		5	0		12	27		2	0		0	4	
Tr 3	0	7		4	0		14	34		21	1		9	0	
Tr 4	0	2		1	1		46	1		2	0		6	1	
Tr 5	0	4		0	1		11	33		20	1		0	3	
Tr 6	1	6		3	1		19	35		36	0		0	3	
Tr 7	0	0		4	0		12	34		7	0		4	0	
Tr 8	3	1		1	0		21	28		7	0		1	0	
Tr 9	0	1		3	0		18	26		14	0		2	2	
Tr 10	1	0		1	4		17	33		5	0		3	0	
Mean	**0.6**	**2.3**		**2.3**	**0.7**		**18.9**	**25.6**		**11.9**	**0.2**		**2.5**	**1.3**	
En 1	0	2		5	0		8	1		11	0		3	0	
En 2	1	2		1	0		4	24		15	0		1	0	
En 3	0	1		1	0		20	33		26	1		3	0	
En 4	0	3		2	1		11	0		12	0		7	1	
En 5	0	3		1	0		1	5		23	1		5	1	
En 6	0	2		3	0		7	34		32	2		3	0	
En 7	0	4		1	0		1	4		11	0		2	0	
En 8	0	0		2	0		4	4		13	0		6	4	
En 9	0	5		3	0		1	34		9	4		0	0	
En 10	0	1		5	0		0	1		16	0		2	1	
Mean	**0.1**	**2.3**		**2.4**	**0.1**		**5.7**	**14**		**16.8**	**0.8**		**3.2**	**0.7**	
p-value	**0.1373**	**0.537**		**0.8932**	**0.121**		**0.0025**	**0.0774**		**0.2599**	**0.261**		**0.5647**	**0.3566**	

ISBN: DL:

The second error category where, based on the mean values, translators and engineers performed quite differently was language errors. Table 36 shows that translators make fewer language errors than engineers. On average, translators made 11.9 minor language errors and 0.2 major language errors, while engineers made 16.8 minor language errors and 0.8 major language errors. However, the unpaired t-tests show no significant difference between translators an engineers in terms of their language errors (see Table 36).

So far, only the results in terminology errors run parallel to our expectations. In other words, we already expected engineers to perform better than translators in terminology. However, it is surprising for us to see that the translators' performances are not superior to the engineers' in the other categories.

Although the mean figures for the language category looks as if translators perform better than engineers, unpaired t-test result for the language category (0.2599 for the minor errors and 0.261 for the major errors) does not affirm this assumption. We expected translators to be more successful with respect to the language, because we presumed that their knowledge of and attention to language-related matters such as correct grammar and punctuation are superior to engineers.

Similarly, we expected translators would be more successful than engineers in mistranslation, accuracy and consistency. Style might also be included in these categories. However, in the present study, we did not find any style errors as such. This can be explained by the technical nature of our test text, which involves instructions; not much stylistic work (i.e. register, tone or language variants) by the posteditor is necessary.

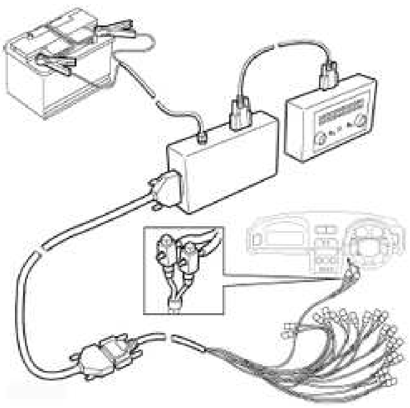
As for mistranslaton, our data show that translators make 0.6 minor mistranslation errors, while engineers make 0.1 errors in this category. Both groups made 2.3 major mistranslation errors. The p-value for minor mistranslation errors is 0.1373 and 0.537 for the major ones. The results are not significant. Contrary to our expectations, translators did not performed better than engineers in terms of "correct understanding of the source text", which is the explanation of the "mistranslation" category in the LISA QA Model. Further, the errors in the mistranslation category have nothing to do with "subject-area knowledge." Rather, they result from misunderstanding some words or phrases in the source text. Let us try to explain what types of errors are classified under this category.

For instance, the MT system translated the phrase "the pyrotechnic device

ISBN: DL:

multiplug" as "piroteknik cihazı çoklu" (the pyrotechnic device multi-). In this phrase, the term "multiplug" is incorrectly translated as "çoklu". The correct translation of the term "multiplug" should be "çok uçlu soket" or "çoklu soket" and it should procede "piroteknik cihazı" (pyrotechnic device) and become "piroteknik cihazı çoklu soketi" (pyrotechnic device multiplug)" so that the "multiplug" is part of the pyrotechnic device, as in the phrase "the door handle". However, some translators postedited the phrase "piroteknik cihazı çoklu", which is the translation made by the MT system, as "çok uçlu piroteknik cihazı" (multi-[pin] pyrotechnic device). In this case, the word "plug" (soket), which is omitted by the MT system, is not added by the posteditor and only the translation of "multi-pin" (çok uçlu) or "multi-" (çoklu) is given. More importantly, the word order is incorrectly changed and made "çok uçlu piroteknik cihazı" (multi-pin pyrotechnic device) instead of "piroteknik cihazı çok uçlu soketi" (pyrotechnic device multiplug). When "multiplug" or "multi-pin" precedes the phrase "pyrotechnic device", it becomes an adjective and qualifies the pyrotechnic device representing a non-existent object. However, when it procedes the phrase "pyrotechnic device", it becomes a noun and constitutes a part of that device. In our text, the latter is correct, that is, "multiplug" is a noun and it is a part of the pyrotechnic device. However, mistranslation by the posteditors made the word "multiplug" an adjective, giving an incorrect meaning to the sentence it is used.

Figure 2. Pyrotechnic Device Multiplug



We give this example to explain that mistranslation errors are caused by incorrect understanding of the source text rather than lack of terminological or subject-area knowledge. Since they are professionals, we expected translators to be more competent in the correct understanding of the source text where the cause of the error is not related

ISBN: DL:

to subject-area knowledge. However, the results (see Table 36) reveal that the translators did not meet this expectation and they did not perform better than engineers in understanding the text correctly.

When it comes to accuracy, similarly, there is not a big difference between translators and engineers. On average, translators made 2.3 minor accuracy errors and engineers made 2.4 errors in this category. As for major accuracy errors, this figure is 0.7 for translators and 0.1 for engineers. The unpaired t-test result does not indicate any significant difference, either ($p = 0.8932$ for the minor and 0.121 for the major accuracy errors, Table 36). In the LISA QA Model, accuracy refers to omissions, additions, not reflecting the source text properly. For example, the phrase “onboard deployment” is incorrectly translated by the MT system as “yerleşik dağıtım” (settled scattering). The correct translation should be “araç üzerinde patlatma” (“onboard” means “araç üzerinde” - on the vehicle - and “deployment” means “patlatma” - explosion). While postediting, Translator 8 omitted the term “onboard” (yerleşik) in the MT output. Errors such as omissions are expected to be made less frequently by professional translators, whose job theoretically requires them to be more meticulous in translating/postediting a text. Moreover, once they are made, such errors are expected to be corrected by the subject during the self-revision process. Although translators are more prone to self-revise their posteditings (60 percent of the translators made self-revision for an average of 8.03 minutes, while only 30 percent of engineers performed self-revision for 4.48 minutes, see Table 31), they did not identify and correct these accuracy errors. According to the results (Table 36), there was no significant difference between the groups in terms of their accuracy errors and the number of these errors is relatively small.

As for consistency, the translators’ performances are not much better than the engineers’. On average, translators made 2.5 minor consistency errors, while engineers made 3.2 errors in this category. The situation is just the opposite as far as major consistency errors are concerned. The translators made 1.3 major consistency errors and the engineers made 0.7 errors in this category. The unpaired t-test results do not indicate any significant difference ($p = 0.5647$ for the minor and 0.3566 for the major consistency errors, Table 36). We expected translators to be more careful in keeping terminological consistency, because this is among the basic requirements of their job and it usually has priority in clients’ list of expectations. The recent survey carried out to identify the competences that employers look for when employing new staff showed

ISBN:
DL:

that 69 percent of the respondents considered the ability to extract and manage terminology as one of the basic skills required of applicants for positions in translation (OPTIMALE 2012). However, our results do not confirm this expectation: the translators do not perform better than engineers in keeping terminological consistency while postediting a technical text.

All in all, based on the p-value (Table 36) of the error categories except for the minor terminology, we cannot mention any statistically significant difference between both groups' error distribution data obtained through LISA QA Model 3.1 interface. The results indicate that when postediting a technical text, engineers perform better than translators in keeping the correct terminology. On the other hand, as for the other categories, namely, language, mistranslation, accuracy and consistency, the two groups' performances are similar (Table 36).

The error distribution data give us detailed information on the subjects' error counts. However, we need more precise information on the quality of the posteditings. To this end, using the LISA QA Model interface (see Appendix G), we calculated the total error points and obtained the quality percentages. For this calculation, we entered the errors in each category in the section where their level of severity (minor, major or critical) was indicated on the LISA QA Model 3.1 interface. The interface automatically counts the number of errors as we enter them and it calculates the total error points and quality percentages. Before starting the review, the quality percentage was 100 by default. As we entered each error, the interface registered it, calculated the error point, and the quality percentage diminished from 100, in line with this error point. When the quality percentage reached a point below 75, which was the minimum level of acceptable quality for posteditings set for this research, the interface labels the postediting with "Fail", yet the reviewer can go on reviewing the text. Any level of quality between 100 and 75 percent is labeled as "Pass." After the review was completed, we exported the "Project Review Report" (see Appendix H), which contains error counts, the total error points and the quality percentage.

A total of 60 percent of the engineers received a "Pass", while this percentage is only 20 for the translators. The mean total error points for engineers is 118 and the quality percentage is 75.58. For translators, these figures are 187 and 61.27, respectively (Table 37). Although the mean values for the error points and the quality percentages seem to indicate that engineers performed better than translators, the unpaired t-test results indicate that the difference between the quality of both groups' performances is not quite significant. The p-value for both the error points and the quality percentages is 0.0636 which is above the threshold 0.05.

Table 37. Quality in PE - Translators and Engineers - Recurring Errors Penalized

Subjects	Total Error Points	Postediting Quality %	Result
Tr 1	55	88.59	Pass
Tr 2	180	62.66	Fail
Tr 3	258	46.47	Fail
Tr 4	80	83.40	Pass
Tr 5	241	50.00	Fail
Tr 6	284	41.08	Fail
Tr 7	197	59.13	Fail
Tr 8	178	63.07	Fail
Tr 9	182	62.24	Fail
Tr 10	212	56.02	Fail
Mean	**187**	**61.27**	
En 1	42	91.29	Pass
En 2	152	68.46	Fail
En 3	225	53.32	Fail
En 4	57	88.17	Pass
En 5	80	83.40	Pass
En 6	235	51.24	Fail
En 7	55	88.59	Pass
En 8	65	86.51	Pass
En 9	228	52.70	Fail
En 10	38	92.12	Pass
Mean	**118**	**75.58**	
p-value	**0.0636**	**0.0636**	

When we look back at the error distribution data, we can see that both groups' results are similar for mistranslation, accuracy and consistency. For language, the mean value for translators' errors is lower than engineers', yet the difference between the groups is not statistically significant. As for the terminology, both the mean values and the t-test results reveal that the engineers make fewer terminology errors than translators. However, according to the t-tests, the quality of the posteditings by engineers is not significantly higher than translators. We can interpret this result in two ways: (1) For technical texts, postediting quality does not change depending on whether the posteditor is a subject-matter expert (engineer) or a professional translator, (2) For technical texts, the quality of postediting is related with factors other than the posteditor's being an engineer or a translator. In order to elaborate on this, we analyzed the post-assignment questionnaire data that contain information on the subjects' profile,

ISBN: DL:

their approach to postediting in general and to the present postediting task in particular. Before we present the post-assignment questionnaire data, we will look at how the quality results are affected when we do not penalize the recurrent errors.

9.1.2.2. Quality in Postediting - Recurring Errors Unpenalized

In this part, we present engineers' and translators' data comparatively. However, our emphasis will be on the comparison of one group with itself under both conditions (recurrent errors penalized or unpenalized) in order to see how penalizing or unpenalizing the recurrent errors affects the results.

Table 38 presents the error distribution of each translator and engineer when we do not penalize the recurrent errors.

ISBN: DL:

Table 38. Distribution of Errors - Translators and Engineers - Recurring Errors Unpenalized

Subjects	**Mistranslation**			**Accuracy**			**Terminology**			**Language**			**Consistency**		
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**
Tr 1	0	1		1	0		5	2		2	0		0	0	
Tr 2	1	1		5	0		2	4		2	0		0	2	
Tr3	0	6		4	0		4	4		19	1		3	0	
Tr 4	0	1		1	1		5	1		2	0		4	1	
Tr 5	0	3		0	1		1	5		20	1		0	2	
Tr 6	1	3		3	1		3	4		34	0		0	2	
Tr 7	0	0		4	0		6	4		7	0		2	0	
Tr 8	1	1		1	0		5	3		7	0		1	0	
Tr 9	0	1		3	0		5	3		14	0		1	2	
Tr 10	1	0		1	1		4	3		5	0		1	0	
Mean	**0.4**	**1.7**		**2.3**	**0.4**		**4.0**	**3.3**		**11.2**	**0.2**		**1.2**	**0.9**	
En 1	0	2		5	0		3	1		8	0		3	0	
En 2	1	2		1	0		2	3		10	0		1	0	
En 3	0	1		1	0		6	3		24	1		3	0	
En 4	0	2		2	1		1	0		8	0		1	1	
En 5	0	2		1	0		1	5		15	1		4	1	
En 6	0	2		3	0		3	3		28	2		2	0	
En 7	0	2		1	0		1	2		9	0		2	0	
En 8	0	0		2	0		2	1		13	0		2	1	
En 9	0	3		3	0		1	3		9	3		0	0	
En 10	0	1		3	0		0	1		11	0		1	1	
Mean	**0.1**	**1.7**		**2.2**	**0.1**		**2.0**	**2.2**		**13.5**	**0.7**		**1.9**	**0.4**	
p-value	**0.131**	**0.432**		**0.8848**	**0.131**		**0.0135**	**0.0803**		**0.5719**	**0.261**		**0.2448**	**0.275**	

ISBN: DL:

When we unpenalize the recurrent errors, the mistranslation data show that translators made 0.4 minor and 1.7 major errors. These figures were 0.6 and 2.3 when we penalize the recurrent errors. Engineers made 0.1 minor and 1.7 major mistranslation errors when the recurrent errors are unpenalized. These figures were 0.1 and 2.3 when we penalize the recurrent errors. In both cases, the p-value does not indicate significant differences between the translators' and engineers' with regard to mistranslation errors, (Table 36 and Table 38), and there is no noticeable difference within one group under the recurring errors penalized and unpenalized conditions. Thus, for this category, we cannot talk about a big difference caused by unpenalizing the recurring errors.

As for accuracy, for the translators, the number of errors in the minor category does not change (2.3), while there is a slight difference in the major error category (0.4 vs 0.7). For engineers, the number of minor accuracy errors is 2.2 when the recurrent errors are not counted, and it is 2.4 when these errors are counted. The number of errors in the major category does not change (0.1 when we penalize or unpenalize the recurrent errors). The p-value does not indicate significant differences between the translators and engineers with regard to accuracy errors (Table 36 and Table 38) when the recurring errors are penalized or unpenalized, and there is no noticeable difference within the groups between the penalized and the unpenalized conditions.

The biggest difference between the translators' and engineers' error distribution results under both conditions (recurring errors penalized or unpenalized) is seen in the terminology category. When the recurrent errors are unpenalized, the mean number of minor terminology errors for translators is 4, while this figure is 18.9 when each recurrent error is counted. The mean number of major terminology errors for translators is 3.3 when the recurrent errors are unpenalized, while this figure is 25.6 when the recurrent errors are penalized. Thus, according to the mean values in our data, for the minor terminology category, penalizing the recurrent errors increases the number of errors by almost five times (4 vs 18.9), and for the major terminology category, it increases the mean number of errors by almost eight times (3.3 vs 25.6). When the recurrent errors are not penalized, the p-value for the difference between the translators' and the engineers' minor terminology errors is 0.0135, which means that the result is highly significant. For the major terminology category, the p-value is 0.0803 which is not quite significant, although it is closer to the threshold (0.05) as compared to the p-value for other categories. Both under the penalized and the unpenalized conditions, the difference between the translators' and engineers' minor terminology errors is

ISBN: DL:

statistically significant, while the difference between their major terminology errors is not quite significant. Although the t-test data - when the recurrent errors are penalized or unpenalized - are parallel with each other, it is obvious that the difference between the number of terminology errors under both conditions will affect the quality percentages.

Unlike the terminology category, penalizing and unpenalizing the recurrent errors does not affect the quantity of the language errors. This is mainly caused by the nature of errors. When a term recurring throughout the text is translated/postedited incorrectly, the quality of the product will change to a great extent whether or not we decide to penalize the erroneous translation/postediting of that recurring term. However, language errors are not usually recurring errors. For example, in Turkish, the sentences start with a capital letter. When the posteditor fails to correct the letter "a" into capital "A" in the sentence "aracı örtün" (cover up the vehicle), this is categorized as a minor language error. This type of error is unlikely to recur throughout the text. Similar types of capitalization errors may occur, yet since they are not the recurring versions of each other, they are regarded as distinct individual errors and counted each time they occur. This is the reason why, in the language category, penalizing or unpenalizing the recurrent errors does not change the quality of the text to a great extent.

When the recurrent errors are not penalized, translators make 11.2 minor and 0.2 major language errors. These figures are 13.5 and 0.7 for engineers (Table 38). Table 36 shows that when the recurring errors are penalized, translators make 11.9 minor and 0.2 major language errors and engineers make 16.8 minor and 0.8 major language errors. Based on the mean number of errors, the difference worth mentioning occurs when we penalize or unpenalize engineers' minor language errors (16.8 vs 13.5). The t-test data of the penalized and the unpenalized conditions do not reveal significant difference between translators and engineers with regard to their language errors, yet the p-value of both conditions are not close to each other. The p-value of the difference between both groups' minor language errors is lower when the recurrent errors are penalized than when they are unpenalized (0.2599 vs 0.5719) (Table 36 and Table 38). As we mentioned above, most of the language errors occur only once and they are not of the recurrent nature. However, there are a few exceptions. For example, the translation of the word "ignition" is made by the MT system as "atesleme". Although this seems to be the correct translation, there is a small mistake. The MT system used "s" instead of "ş", which is a Turkish character (s with cedilla). The correct translation should be

ISBN: DL:

"ateşleme". This word occurs in the text three times and most of the engineers did not correct this error, while the translators usually corrected it. This is why the number of engineers' minor language errors changes when the recurrent errors are penalized or unpenalized.

The number of consistency errors also changes when the recurrent errors are penalized or unpenalized. Translators make 1.2 minor and 0.9 major consistency errors when we unpenalize the recurrent errors. These figures are 2.5 and 1.3 respectively when we penalize the recurrent errors (Table 36 and Table 38). Engineers make 1.9 minor and 0.4 major consistency errors when we unpenalize the recurrent errors. These figures are 3.2 and 0.7 respectively when we penalize the recurrent errors (Table 36 and Table 38). Although these figures represent small number of error points, they indicate that penalizing the recurring errors may lead to the calculation of almost twice the consistency errors. The t-test data under both conditions do not reveal significant differences between translators and engineers (Table 36 and Table 38).

We calculated the total error points and obtained the quality percentages when the recurrent errors are unpenalized (see Table 39). Based on mean values, the total error points (hence the quality percentages) change to a great extent when we penalize or unpenalize the recurrent errors. Unlike when all errors are penalized, all of the translators and engineers received "Pass" from the LISA QA Model 3.1 when the recurrent errors are not penalized. When the recurring errors are unpenalized, the mean total error point for translators is 52 and the quality percentage is 89.30. When the recurrent errors are penalized, these figures were 187 and 61.27, respectively (Table 37). For engineers, when the recurring errors are unpenalized, the mean total error point is 45 and the quality percentage is 90.73. In case when the recurrent errors are penalized, these figures were 118 and 75.58 respectively (Table 37).

Based on the mean values, the quality of the engineers' postediting seems higher than the translators', no matter whether we penalize or unpenalize the recurrent errors. When we penalize the recurrent errors, the difference between both groups' product quality is 14.31 percent (75.58 percent for engineers vs 61.27 percent for translators, see Table 37). However, when we unpenalize the recurring errors, this difference diminishes by almost 90 percent and becomes 1.43 (90.73 for engineers vs 89.30 for translators, see Table 39).

ISBN: DL:

Table 39. Quality in PE - Translators and Engineers - Recurring Errors Unpenalized

Subjects	Total Error Points	Postediting Quality %	Result
Tr 1	23	95.23	Pass
Tr 2	45	90.66	Pass
Tr 3	85	82.37	Pass
Tr 4	32	93.36	Pass
Tr 5	81	83.20	Pass
Tr 6	91	81.12	Pass
Tr 7	39	91.91	Pass
Tr 8	35	92.74	Pass
Tr 9	53	89.00	Pass
Tr 10	32	93.36	Pass
Mean	**52**	**89.30**	
En 1	34	92.95	Pass
En 2	40	91.70	Pass
En 3	54	88.80	Pass
En 4	32	93.36	Pass
En 5	66	86.31	Pass
En 6	71	85.27	Pass
En 7	33	93.15	Pass
En 8	29	93.98	Pass
En 9	58	87.97	Pass
En 10	30	93.78	Pass
Mean	**45**	**90.73**	
p-value	**0.4711**	**0.4708**	

According to the t-tests, there is no significant difference between the quality of engineers' and translators' postediting a technical text (the quality being measured by LISA QA Model 3.1), no matter whether we penalize or unpenalize the recurrent errors. However, the degree of the statistical significance change noticeably when the recurring errors are penalized (p = 0.0636, Table 37) or unpenalized (0.4708, Table 39).

We ran unpaired t-tests in order to see if there are significant differences between each groups' postediting quality under the recurring errors penalized and unpenalized conditions. When we unpenalize the recurring errors, the mean value for the engineers' total error points decreases from 118 to 45 and the quality percentage increases from 75.58 to 90.73. The p-value for the difference between the engineers' quality results under the recurring errors penalized and unpenalized conditions is 0.0140 (Table 40).

ISBN: DL:

Table 40. Engineers' Quality Results in PE - Recurring Errors Penalized and Unpenalized

Subjects	Total Error Points	Postediting Quality %	Result
	Recurring Errors Penalized		
En 1	42	91.29	Pass
En 2	152	68.46	Fail
En 3	225	53.32	Fail
En 4	57	88.17	Pass
En 5	80	83.40	Pass
En 6	235	51.24	Fail
En 7	55	88.59	Pass
En 8	65	86.51	Pass
En 9	228	52.70	Fail
En 10	38	92.12	Pass
Mean	**118**	**75.58**	
	Recurring Errors Unpenalized		
En 1	34	92.95	Pass
En 2	40	91.70	Pass
En 3	54	88.80	Pass
En 4	32	93.36	Pass
En 5	66	86.31	Pass
En 6	71	85.27	Pass
En 7	33	93.15	Pass
En 8	29	93.98	Pass
En 9	58	87.97	Pass
En 10	30	93.78	Pass
Mean	**45**	**90.73**	
p-value	**0.0140**	**0.0140**	

As for the translators, the mean value for the total error points decreases from 187 to 52 when the recurring errors are not penalized. The quality percentage increases from 61.27 to 89.30. The p-value for the difference between the translators' quality results under the recurring errors penalized and unpenalized conditions is 0.0001 (Table 41). Thus, the quality results of both the engineers and the translators change significantly under the recurrent errors penalized and unpenalized conditions. This is striking when we consider the fact that both penalizing and unpenalizing recurring errors are methods employed in the translation industry (O'Brien 2012).

ISBN: DL:

Table 41. Translators' Quality Results in PE - Recurring Errors Penalized and Unpenalized

Subjects	Total Error Points	Postediting Quality %	Result
	Recurring Errors Penalized		
Tr 1	55	88.59	Pass
Tr 2	180	62.66	Fail
Tr 3	258	46.47	Fail
Tr 4	80	83.40	Pass
Tr 5	241	50.00	Fail
Tr 6	284	41.08	Fail
Tr 7	197	59.13	Fail
Tr 8	178	63.07	Fail
Tr 9	182	62.24	Fail
Tr 10	212	56.02	Fail
Mean	**187**	**61.27**	
	Recurring Errors Unpenalized		
Tr 1	23	95.23	Pass
Tr 2	45	90.66	Pass
Tr 3	85	82.37	Pass
Tr 4	32	93.36	Pass
Tr 5	81	83.20	Pass
Tr 6	91	81.12	Pass
Tr 7	39	91.91	Pass
Tr 8	35	92.74	Pass
Tr 9	53	89.00	Pass
Tr 10	32	93.36	Pass
Mean	**52**	**89.30**	
p-value	**0.0001**	**0.0001**	

To sum up, our study shows that neither professional translators work faster than engineers nor engineers produce higher-quality texts than professional translators when they postedit technical texts. However, engineers are stronger than professional translators with respect to terminological knowledge when postediting a technical text. They are more successful than translators in keeping the correct terminology throughout the text. Moreover, there is a significant difference between both groups' documentation behavior. When postediting a technical text, translators carry out significantly more documentation than engineers for longer periods. They also tend to self-revise their posteditings more than engineers do, yet the difference is not significant. Our data showed no correlation between quality and the time spent on documentation and/or self-revision. Both groups perform similar number of changes while postediting a technical text. Further, we found that, while assessing the quality of a postediting, penalizing the recurrent errors affects the quality of the product significantly. Although we found no significant difference between the engineers' and translators' postediting quality, the degree of statistical significance decreases when the recurrent errors are unpenalized

ISBN: DL:

(see Table 42 for the summary data).

Table 42. Summary of Data in PE - Speed and Quality

Subjects	Total Task Time (minute)	Words processed per minute	Time Spent on Documentation (minute)	Minor Terminology Errors	Major Terminology Errors	Total Error Points	Postediting Quality %
		Recurring Errors Penalized (Mean Values)					
En	55.66	9.1	02.71	5.7	14	118	75.58
Tr	62.70	8.3	10.75	18.9	25.6	187	61.27
p-value	**0.3598**	**0.4195**	**0.0078**	**0.0025**	**0.0774**	**0.0636**	**0.0636**
		Recurring Errors Unpenalized (Mean Values)					
En	55:66	9.1	02.71	2	2.2	45	90.73
Tr	62.70	8.3	10.75	4	3.3	52	89.30
p-value	**0.3598**	**0.4195**	**0.0078**	**0.0135**	**0.0803**	**0.4711**	**0.4708**

In order to look at the results from a different point of view, we listed the quality results (under the penalized condition) of both groups in an ascending order (Table 43). We also added speed data to this order, but could not see any meaningful pattern between the quality and speed results. The data in Table 43 suggest that the quality of the posteditings might be related with factors other than being a professional translator or an engineer. In order to elaborate on this new order, we analyzed the post-assignment questionnaires.

Table 43. Speed and Quality in PE - Translators and Engineers (Quality in Ascending Order)

Subjects	Total Error Points	Postediting Quality % in Ascending Order	Quality Result	Time on PE Excluding Self-revision (minute)	Time on Self-revision (minute)	Total Task Time (PE and Self-rev) (minute)	Words processed per minute	Words processed per minute (excluding self-revision)	Time Spent on Documentation (min:sec)	Documentation as % of the Total Task Time	Number of Changes
Tr 6	284	41.08	Fail	38.73	00.00	38.73	12.4	12.4	05.00	12.9	58
Tr 3	258	46.47	Fail	45.36	00.00	45.36	10.6	10.6	02.65	5.8	80
Tr 5	241	50.00	Fail	92.06	15.68	107.75	4.5	5.2	28.00	25.9	148
En 6	235	51.24	Fail	35.28	09.23	44.51	10.8	13.6	00.96	2.1	71
En 9	228	52.70	Fail	35.40	00.00	35.40	13.6	13.6	01.30	3.6	119
En 3	225	53.32	Fail	46.00	00.00	46.00	10.4	10.4	03.70	8.0	97
Tr 10	212	56.02	Fail	65.23	05.85	71.08	6.7	7.3	15.08	21.2	110
Tr 7	197	59.13	Fail	54.28	14.46	68.75	7.0	8.8	08.76	12.8	139
Tr 9	182	62.24	Fail	60.03	00.00	60.03	8.0	8.0	08.86	14.7	158
Tr 2	180	62.66	Fail	51.25	16.28	67.53	7.1	9.4	07.51	11.1	135
Tr 8	178	63.07	Fail	62.66	00.00	62.66	7.6	7.6	04.56	7.3	172
En 2	152	68.46	Fail	47.45	00.00	47.45	10.1	10.1	00.33	0.7	107
En 5	80	83.40	Pass	60.96	00.00	60.96	7.9	7.9	01.51	2.5	124
Tr 4	80	83.40	Pass	37.45	22.71	60.16	8.0	12.8	13.08	21.7	147
En 8	65	86.51	Pass	61.05	23.25	84.30	5.7	7.8	09.35	11.0	165
En 4	57	88.17	Pass	65.61	00.00	65.61	7.3	7.3	03.93	6.0	129
En 7	55	88.59	Pass	61.70	00.00	61.70	7.8	7.8	00.48	0.7	128
Tr 1	55	88.59	Pass	39.70	05.28	44.98	10.7	12.1	07.51	18.9	150
En 1	42	91.29	Pass	52.96	00.00	52.96	9.1	9.1	02.76	5.2	155
En 10	38	92.12	Pass	45.43	12.30	57.73	8.3	10.6	02.91	5.0	140

ISBN: DL:

9.1.3. Post-Assignment Questionnaire Data

In order to collect qualitative data, we asked the participants to fill out post-assignment questionnaires right after they complete the postediting task. Although the general content of the questionnaires used for engineers and translators is the same, we prepared two different questionnaires for two groups due to the slight difference in the questions we needed to ask them (See Appendix E and Appendix F).

The questionnaires were partly adapted from those prepared by Guerberof (2012) and contained personal data such as the participants' age, sex and educational details. They also cover information on the participants' professional experience, experience with regard to translation/postediting and the use of machine translation. One section contains the participants' opinions on the comparison of engineers and translators with regard to the various aspects of the postediting process. The final part of the questionnaire was dedicated to their opinions on the postediting task.

9.1.3.1. Professional Translators' Post-Assignment Questionnaire Data

All of the translators were female. We did not select female translators on purpose, yet we did not pay particular attention to selecting equal number of male and female participants, because the participants' sex is not a significant variable for this research. We do not know if there is any research on the sex of the translators in Turkey, however according to our experience, most of the professional translators are women. On the contrary, engineering is usually performed by men (this information is also based on our experience). For example, for the present research, we worked with engineers working for various international automotive factories in Turkey. Similar to the selection of translators, we did not select male engineers on purpose, yet all of the engineers participated this research are male. However, as we pointed above, this does not pose a problem for the present research.

The translators' ages range from 26 to less than 40. Only one of them (Translator 2) is 36-40 and three of them (Translators 3, 5 and 6) are 26-30. The rest of them are 31-35 years old. Although they all do translation as a professional activity, not all of them hold a degree in Translation. Translator 3 holds a BA degree in English Language and Culture and Translators 6 and 7, in English Language Teaching. Others hold either a BA, MA, PhD or two/three of them in Translation. Translators 6 and 3 who have the lowest quality scores do not have a degree in Translation. Translator 7 does not have

ISBN: DL:

any degree in Translation and she ranked sixth among the translators in terms of quality, yet she has more than 10 years of professional experience which might compensate for the degree in Translation to some extent.

The level of professional experience in translation ranges from 3-5 years to more than 10 years. Two participants have translation experience of more than ten years (Translators 2 and 7). Four of them (Translators 1, 4, 8 and 10) have 9-10 years of experience, one (Translator 9) has 6-8 years of experience. The rest (Translators 3, 5 and 6) have 3-5 years of translation experience.

None of the translators has had a training and/or professional experience in engineering. They are not highly specialized in a specific field, that is, they usually translate any type of texts that circulate in the market. However, when we asked them to pick up the most common text types they translate, they reported that they usually translate educational, academic, literary and legal texts and the texts on social sciences. Translators 1 and 4, who are the only translators receiving a “Pass” from the quality analysis, also selected the option of “technical texts”. Translators 2 and 8 said that they do technical translations from time to time. This is meaningful for us, because although they received a “Fail” from the quality analysis, Translators 8 and 2, respectively, follows Translators 1 and 4 in the quality results (see Table 43). This implies us that, in the quality of the postedited text, the experience in the subject-matter is far more important than the years of experience in the translation profession or a degree in Translation.

None of the translators use translation memory and/or machine translation in their professional activities. Their average daily throughput without TM and MT ranges from 3100 to more than 5100 words. Translator 2 mentioned that she translated more than 5100 words per day and Translator 10 said she translated around 5000 w/pd. The rest reported to translate between 3100 to 5000 w/pd. Half of the translators (Translators 1, 3, 4, 5 and 10) reported to have used Google Translate before. Translator 10 said she used Google Translate just out of curiosity, while others pointed that they use it for a general understanding of a text and getting the gist of a text. Translators 1 and 4 use it to translate the pages they come across while surfing the Internet and only Translator 5 said she sometimes use its outputs to compare with her own translations. They agree with the usefulness of Google Translate to get the gist of the source texts. However, they also agree that its translations are not as “sound, successful and meaningful” as

ISBN: DL:

human translations. Translator 10 stated that:

> In particular situations, it may be useful. However, it usually focuses on the most general meaning of the word, therefore it is difficult to get a meaningful translation with Google Translate. Moreover, even if we get a meaningful text using Google Translate, the result may be misleading, because the MT system cannot comprehend the context. Translation requires comprehension in addition to proficiency in the source and target languages. For this reason, it is not possible to compare machine translation with human translation. Google Translate usually focuses on the most general meaning of a source word or phrase and since it cannot comprehend the context, the result might be misleading.

Translator 4 stated that "it could be useful in understanding the gist of the text in language pairs that are lexically close to each other. I am not sure if it could be useful in professional translation. I have never tried." Translator 1 said:

> I think Google's translations are far better than other translations known as "machine translation". It may not produce very bright outputs in English-Turkish pair. All the same, I think it is successful in terms of offering a general idea of the text. I understand this especially when I use Google Translate from a language I do not know at all into English. Of course, I do not come across "a very professional" translation, yet it definitely gives an idea on the content of the text.

The translators participated in our research have neither used Google Translator Toolkit before nor they have had professional postediting experience. Six out of ten (Translators 1, 2, 5, 6, 9 and 10) stated that they did revision professionally besides translating. Five of them (Translators 1, 2, 7, 9 and 10) said they also did proofreading. Three of them (Translators 2, 7 and 9) said they also did terminology work. Translator 10 carries out academic research, teaching and consecutive interpreting besides translating, revision and proofreading, while Translator 8 teaches English and does revision in addition to translating. Only Translator 3 said she did not do any other tasks besides translating. None of the translators has had a previous training on postediting.

With a view to learning how translators compare translator and engineer postediting in terms of speed and quality, we presented six statements on engineers' and translators' postediting and asked the translators to state whether or not they agree (using Likert scale) with these statements. Five out of ten translators (Translators 2, 3,

ISBN: DL:

4, 5 and 9) think that translators translate/postedit technical texts faster than engineers, while four of them (Translators 1, 6, 7 and 8) pointed that they are neutral to this statement. Translator 10 said "it depends on the expertise in the relevant field and proficiency in the relevant languages."

All of the translators agreed (Translators 5 and 6 strongly agreed) that "while translating/postediting technical texts, translators spend more time on documentation." According to our empirical data, on average, translators spent 10.75 minutes, while engineers spent 2.71 minutes on documentation. It is also affirmed in a statistically significant way ($p = 0.0078$) that translators spend more time on documentation than engineers (see Table 33).

A total of seven out of ten translators (Translators 1, 2, 4, 6, 8, 9 and 10) agreed that (Translators 2, 8 and 9 strongly agreed) when translators translate/postedit technical texts, the result is more readable (of higher language quality) than when engineers translate/postedit. Translators 3 and 7 were neutral to this statement and Translator 5 disagreed with it. This assumption is not affirmed with our empirical data, because we did not find significant difference between engineers' and translators' posteditings in terms of language qualities as measured by LISA (p-value for minor language errors is 0.2599 and for the major 0.261, see Table 36).

Similarly, a total of seven out of ten translators (Translators 1, 2, 5, 6, 7, 9 and 10) agreed that (Translators 2 and 5 strongly agreed) when engineers translate/postedit technical texts, the result is of higher terminological quality than when translators translate/postedit. Translators 3 and 4 were neutral to this statement and Translator 8 disagreed with it. Our empirical data confirm this opinion of the majority of translators. We found significant difference between engineers' and translators' posteditings in terms of the terminological quality as measured by LISA (p-value for minor terminology errors is 0.0025 and for the major 0.0774, see Table 36).

We dedicated a section in our questionnaire to obtain the translators' opinions on the postediting task. When we asked them how understandable the MT output was, six out of ten translators (Translators 3, 5, 6, 7, 9 and 10) said they found the text slightly understandable. Translators 2, 4 and 8 found it understandable and only Translator 1 found it quite understandable. Interestingly, this is also parallel with the order of the participants in the quality data. In the quality results, Translator 1 ranked first among translators and she found the MT output quite understandable. Translators 4, 8 and 2

ISBN: DL:

who found the text understandable follow Translator 1 in the ascending results regarding quality (Table 43). This implies us that there is a correlation between the translators' perception of the understandability of the MT output and the quality of the postedited text.

The questionnaire also contained two statements with a view to getting the participants' opinions about the method of their own postediting. The first statement was: "As I postedit, I go back and forth and I recheck my postediting before going to the next sentence." Half of the translators (Translators 1, 2, 5, 9 and 10) said they often do that, while the other half said they sometimes go back and forth and recheck their postediting before going to the next sentence. However, screen recording analyses reveal that most translators do not behave in that way, that is, they postedit one sentence usually referring to the source text and they do not go back to the previous sentence(s) throughout the task.

The second statement was: "Immediately after I finish postediting the whole text, I go back to all sentences and review them one by one again." Translators 1, 2, 4 and 7 said they always do that. Translators 5, 9 and 10 said they often carry out such revision. Translator 3 said she rarely does self-revision, while Translators 6 and 8 said they sometimes go back to all sentences immediately after they finish postediting. The translators' replies to this question overlap with the screen recording data to a great extent. According to the screen recording data; Translators 3, 6, 8 and 9 did not do any self-revision during the task. So, only in the case of Translator 9, the screen recording data conflicted with the questionnaire data, because she said she often carried out self-revision, yet did not behave that way in our experiment.

As additional comments on the postediting task, Translator 2 said:

> I usually examine the source text and search for similar texts on that topic in order to use the correct terminology. However, in this postediting task, I haven't been able to do this due to the limited time. I only used Google and on-line dictionaries.

Translator 10 who holds a BA, MA and PhD in Translation said:

> The more technical the text, the more effort I spend for being stick with the source text. It is the same for translation and interpreting. Since I do not understand a technical text deeply and comprehensively, I prefer a more

ISBN: DL:

> superficial translation, because it is safer. If I have time, I do research on the text and the subject and study its terminology. The more understandable the text to me, the freer I feel as a translator. Of course, the basic thing is the function, the purpose and the context of the translation. Textual and functional aspects determine my approach in translation.

We asked the participants if they preferred postediting to translating based on their postediting experience (either previous or from this experiment). Since none of them has had a professional experience on postediting, they presented their opinions based on their experience in the present task. Only Translator 6 said she often prefers postediting to translating. Translators 1, 4, 5 and 9 said they sometimes prefer postediting to translating. Translator 4 who was the second among the translators in terms of the quality ranking said:

> Especially in academic texts, the sentence types might be too complex. Instead of postediting the machine translation of such texts, I prefer translating by myself. The same applies to the literary texts. I am not sure how MT output will work in literary texts. However, MT could be useful in the translation of user manuals and technical texts containing simple sentence types.

Translators 2 and 8 said they rarely prefer postediting to translating. Translator 2 said “Normally, I do not prefer postediting MT output to translating. However, since there might be some exceptions, I selected “rarely”.” Translators 3, 7 and 10 replied that they would never prefer postediting to translating. Translator 10 further added:

> It is not possible to trust the MT output. In such a technical text, if I translated myself, I would try to stick with the source text as much as possible. Therefore, the MT output did not negatively affect my work. It was even useful for me. All the same, it is crucial to postedit and check the output.

We asked the participants to evaluate their postediting experience with regard to the effort they spent. Translators 1, 4, 5, 6, 8, 9 and 10 said postediting requires *less* effort than translating. Translator 4 added: “This might change depending on the text.” Translators 2 and 3 said postediting requires *the same* effort as translating. Translator 2 added: “I think it depends on the text, subject and terminology. Depending on these factors, three of the above items might be possible.” Only Translator 7 thinks that postediting requires more effort than translating. She also added: “It is hard to answer

ISBN: DL:

this question. In fact, it depends on the length of the text. All the same, it is usually easier for me to translate a text than edit it." In reply to this question, Translator 10 said:

> It completely depends on the textual features. For the text I have just postedited, I would have spent much more effort if I had translated it from scratch. That is, for this text, postediting requires less effort than translating. However, I am sure it would be vice-versa if it were a literary text or a social science text. Then, it would be easier to translate from scratch than to postedit.

Further, we asked the participants to share the advantages and disadvantages of the postediting process. Although the answers vary, the participants agree on the following points:

1. Postediting saves time and effort.
2. It reduces the frequency of looking up a dictionary.
3. The MT system provides a draft translation. It is easier to polish it than translating from scratch, yet this depends highly on the text type.
4. When you postedit the MT output, you already know what the text is about in broad terms before you start.

One of the participants said she usually finds the alternatives the MT system proposes useful, especially when the translation is from L1 into L2.

As for the disadvantages of postediting, the translators agree on the following points:

1. The translations offered by the MT system might be misleading. They may sound correct, yet they might be incorrect.
2. Working with two texts (source text and the MT output) may require extra time and attention.
3. The MT system usually chooses the most general or wrong equivalent of the source word.
4. The MT output may not be useful at all in complex sentences.

Translator 7 states that, when she edits an already translated text, she is usually not satisfied with the output as much as when she translate it herself. It is interesting to note that Translator 5 mentions about the possibility of machine translation to substitute human translators while listing the disadvantages of working with MT output.

Finally, we asked the participants to state the problems they encountered during the postediting task. All of the translators who have low levels of postediting quality

ISBN: DL:

mention about the terminological problems, while those who have higher levels of quality results do not address any problem. Translator 6 whose postediting has the lowest level of quality said she could not find the meaning of those technical words in online dictionaries, therefore she had to postedit some parts of the text without even understanding them well. Translator 3, who has the second lowest level of quality among the translators, said lexical and grammatical errors in the MT output cause loss of time and the terminology was problematic for her. Translator 5 pointed out that it was hard to decide and choose from among the alternatives the MT system proposed and the ones she had in mind. Translator 7 complained about the loss of time caused by postediting. She said:

> I felt I lost time. For instance, if I had translated this text myself, I would have finish it in shorter period of time. I lost time especially when correcting the inverted sentences, punctuation and the terminological errors. Further, the technical terms in the text made it harder for me to edit it.

Translator 10 said the incorrect sentence structures and incorrect terminology made it harder to understand the text which was already technical and difficult for her. Translator 9 left this section empty, while Translator 2 said she did not come across a serious problem except for some technical terms such as “deployment”, which was the cause of most recurring errors made by most of the translators (except for Translator 1 and 4) and by some of the engineers. Translator 8, who ranked third among the translators according to the postediting quality results, also mentioned about the terminological problems, yet she did not complain about them as much as the others who have lower levels of quality results:

> Although it was not a very big problem, the erroneous translation of some terms made it more difficult for me to understand the text. Somehow, it seems as if you do the same work twice, yet it is not true to say that postediting is completely disadvantageous.

Translator 4, who ranked second among the translators in terms of quality results, said she did not experience any serious problem, while Translator 1 who produced the highest-quality text among the translators did not mention any terminological problems. However, she pointed to the importance of referring to the source text while editing the MT output. She also stressed that the problems you might come across while

ISBN: DL:

postediting depends on the text type and the richness of the MT system's database:

> It is easy to distinguish the absurd translations the MT system offers. However, sometimes the MT system may offer misleading translations. If you read only the target text, then the text might sound fluent, but the meaning might be incorrect. For this reason, it is crucial to refer to the source text. It seems easier to translate some sentences from scratch, yet for some others, there is almost no point to be corrected. In fact, it depends on the text type and the richness of the MT system's database.

9.1.3.2. Subject-Matter Experts' (Engineers') Post-Assignment Questionnaire Data

All of the engineers were male. We did not select male engineers on purpose, yet as we stated in the previous section, engineering is usually performed by men (this information is based on our experience), thus - incidentally - all of the engineers participated this research are male.

The engineers' ages range from 26 to 40. Engineer 2 is 26-30 and only Engineer 5 is 36-40. The rest of them are 31-35.

Engineers 1, 2, 4, 5, 7 and 10 hold a degree in Mechanical Engineering; Engineers 6 and 9 in Metallurgy and Materials Engineering, Engineer 3 in Electrical Engineering and Engineer 8 in Electrical and Electronics Engineering. Engineers 2, 3, 4, 5, 6, 8 and 9 hold both a BSc and MSc in Engineering, while Engineer 1 hold a BSc in Engineering and an MBA. Engineers 7, 9 and 10 had only BSc in Engineering.

The level of professional experience in engineering ranges from 3-5 years to more than 10 years. Two participants have engineering experience of more than ten years (Engineers 5 and 10). Four of them (Engineers 1, 4, 7 and 8) have 9-10 years of experience, one (Engineer 2) has 6-8 years of experience. The rest (Engineers 3, 6 and 9) have selected 3-5 years of engineering experience in the questionnaire.

None of the engineers has had a training and/or professional experience in translation. Except for Engineer 3, all of the engineers participated in the study stated that they had to do translation as part of their daily work as engineers. When we asked them the type of texts they had to translate, Engineer 10 said he had to translate electronic mails, technical documents and reports. Engineers 1, 5 and 7 said they had to translate electronic mails, reports, minutes of meetings and technical documents. Engineer 1 said he also had to translate Power Point presentations and Engineer 4 said

ISBN: DL:

he had to translate electronic mails, minutes of meetings and technical documents. Engineer 8 stated that he had to translate or prepare in English the documents containing the electrical-electronical descriptions of the modifications to be implemented on the vehicles exported to the services abroad. It is interesting to note that the abovementioned engineers (Engineers 10, 1, 7, 4, 8 and 5, respectively) all passed the quality test and they received high quality-scores in descending order. All of them have to do technical translation as part of their daily work. The questionnaire results of other engineers affirm this trend, that is, Engineer 2, who follows Engineer 5 (the one with the lowest passed quality-score among the engineers) in the quality results said he translated technical documents from time to time, while Engineer 3 who comes after Engineer 2 said he did not do any translation as part of his work. Engineers 9 and 6 who had the lowest quality-scores among the engineers respectively (see Table 43) said they translated only electronic mails as part of their work. According to our questionnaire data, engineers having to do technical translation as part of their daily work carry out higher-quality posteditings than those who either do not translate technical documents or do not have to translate at all as part of their professional activities.

Except for Engineer 5 and 9, all of the engineers participated in the study said they used machine translation before. Engineer 4 said he used “Babel Fish” while surfing the Internet and for translating some simple texts. Engineers 1, 2, 3, 6, 7, 8 and 10 reported that they used Google Translate. Those who used machine translation stated that they use it to translate electronic mails, to get the gist of a text and while surfing the Internet. A total of three out of ten engineers (Engineers 4, 5 and 9) stated that they had not used Google Translate before.

We asked the engineers their opinions about Google Translate. Engineer 1 said “It should be improved. It has not yet been able to provide successful translations in Turkish.” Engineer 2 said “Generally, its translations from English into Turkish are difficult to understand. Sometimes, it becomes more logical to translate each word individually and make the translation by yourself.” Engineer 3 said “the system is constantly improved. Its translations are getting more successful day by day.” According to Engineer 6:

> It provides a general idea of what a text is about. However, when there are more than one meaning of a word, it usually chooses the incorrect alternative amongst them. It provides a poor translation from German into Turkish, yet the same

ISBN: DL:

system usually performs far better when it translates from German into English. When the source and the target languages are close to each other, the output becomes much better.

Engineer 7 is very optimistic about the future of Google Translate stating that:

I am sure that it will produce much more successful translations in 10 years' time. Especially, its cross multilanguage support and real-time web translations are very useful. Although its translations are not at a professional level, it is a masterpiece which can help people - who do not know a language at all - translate in seconds. It is also vey helpful for those who do not know any foreign language for surfing the Internet.

Engineer 8 said he uses Google Translate as a general support while preparing or translating the documents. He added "it helps me understand the basics of a text. I use it considering it may contain errors and mistranslations."

The engineers participated in our research have neither used Google Translator Toolkit before nor they have had professional postediting experience. Five out of ten (Engineers 3, 4, 6 and 10) stated that they used raw MT output, while four out of ten (Engineers 1, 2, 7 and 8) said they postedited MT output themselves.

With a view to learning how engineers compare translator and engineer postediting in terms of speed and quality, we presented six statements on engineers' and translators' postediting and asked the engineers to state whether or not they agree (using Likert scale) with these statements. Engineers 2, 4 and 6 agreed that translators translate/postedit technical texts faster than engineers, while Engineer 7 strongly agreed with this statement. Engineer 6 added that:

Translators may translate technical texts faster than engineers, however, sometimes in such translations, some terms are incorrecly translated i.e. the incorrect use of the terms "heat" and "temperature". If the reader is a technical person, such translations might be understood, yet they don't reflect the correct terminology.

He further presents the following examples for this explanation:

ISBN: DL:

English term	Literal translation	The correct translation
Heat capacity	Sıcaklık kapasitesi	Isı sığası
Brake lining	Fren astarı	Fren balatası
Thermal expansion	Isı genişlemesi	Isıl genleşme
Smelting	Eritme (for sugar)	Ergitme (for ice)
Material strength	Madde gücü	Malzeme dayanımı

Engineers 3, 8, 9 and 10 disagreed with this statement and Engineer 1 said he is neutral. Engineer 5 replied that it depends on the engineer or the translator, thus cannot be generalized.

Engineers 2, 3, 4, 8 and 9 agreed that "while translating/postediting technical texts, translators spend more time on documentation." According to our empirical data, on average, translators spent 10.75 minutes while engineers spent 2.71 minutes on documentation. It is also affirmed in a statistically significant way ($p = 0.0078$, Table 33). Engineers 6 and 7 disagreed with this statement, while Engineers 1 and 10 remained neutral. Engineer 5 chose the option "I do not know".

Half of the engineers (Engineers 3, 4, 7, 8 and 9) disagreed (Engineer 7 strongly disagreed) that when translators translate/postedit technical texts, the result is more readable (of higher language quality) than when engineers translate/postedit. Engineers 1, 2 and 10 were neutral to this statement and only Engineer 6 agreed with it. Engineer 5 said it depends on the engineer or the translator, thus cannot be generalized. It is interesting to note the difference between the translators' and engineers' point of view of this statement. A total of seven out of ten translators agreed (one of them strongly agreed) that when translators postedit technical texts, the language quality of the output would be higher than when engineers did the postediting. On the other hand, only one of the engineers agreed with this statement. Our empirical data do not represent significant difference between engineers' and translators' posteditings in terms of language qualities as measured by LISA (p-value for minor language errors is 0.2599 and for the major 0.261, see Table 36).

A total of eight out of ten engineers (Engineers 1, 2, 3, 6, 7, 8, 9 and 10) agreed that (Engineers 1, 2, 7, 9 and 10 strongly agreed) when engineers translate/postedit technical texts, the result is of higher terminological quality than when translators translate/postedit. Engineer 4 was neutral to this statement and Engineer 5 said it

ISBN: DL:

depends on the engineer or the translator, thus cannot be generalized. Unlike the previous statement, most of the translators (7 out of 10) and engineers (8 out of 10) agree on this statement, that is, they think that engineers' postediting would be of higher terminological quality than translators' postediting. Our empirical data also confirm this opinion of the participants. We found significant difference between engineers' and translators' posteditings in terms of terminological quality of the final texts as measured by LISA (p-value for minor terminology errors is 0.0025 and for the major 0.0774, see Table 36).

We dedicated a section in our questionnaire to obtain the participants' opinions on the postediting task. When we asked the engineers how understandable the MT output was, four out of ten (Engineers 2, 4, 8 and 9) said they found the text slightly understandable. Engineers 1, 5, 6, 7 and 10 found it understandable and Engineer 3 chose the option "I do not know." Unlike the case of the translators, in the engineers' questionnaire data, the level of perceived understandability of the MT output does not correlate with the postediting quality. When we look at ascending order of the engineer postediting quality results, we can see that Engineer 6 who did the lowest-quality postediting found the MT output understandable, while Engineer 4, who ranked fourth among other engineers with regard to postediting quality, found the text slightly understandable (Table 43).

The questionnaire also contained two statements with a view to getting the participants' opinions about the method of their own postediting. The first statement was: "As I postedit, I go back and forth and I recheck my postediting before going to the next sentence." Engineers 2, 6 and 7 said they often do that while Engineers 1, 3, 4, 5, 8, 9 and 10 said they sometimes go back and forth and recheck their postediting before going to the next sentence. However, as is the case of translators, screen recording analyses reveal that most engineers do not behave in that way, that is, they postedit one sentence usually referring to the source text and they do not go back to the previous sentence(s) throughout the task.

The second statement was: "Immediately after I finish postediting the whole text, I go back to all sentences and review them one by one again." Engineers 7 and 10 who produced high-quality posteditings said they always do that. On the other hand, Engineers 1 and 4 who also produced high-quality outputs and Engineer 9 having poor-quality postediting said they rarely go back to all sentences and review them one by one

ISBN: DL:

again. Engineer 9 added that “in order to give the correct meaning of the words which do not have a Turkish equivalent or the ones having multiple meanings, translation of the whole sentence or the page might be necessary.” Engineers 6 and 8 said they often carry out such revision. Engineer 8 stated that, instead of reviewing them one by one, he preferred to review them faster once or twice.” Engineers 2, 3 and 5 said they sometimes do such revision. The screen recording data also affirm what the engineers said they did with regard to self-revision. According to the screen recordings, only Engineers 6, 8 and 10 did self-revision after they completed the postediting, which is in line with what they said in the questionnaire. The only point where screen recordings and the questionnaire data conflict is the case of Engineer 7. He stated that he always did self-revision after he completed the task, yet did not do any self-revision.

We asked the participants if they preferred postediting to translating based on their postediting experience (either previous or from this experiment). Since none of them has had a previous experience on postediting, they presented their opinions based on their experience in the present task. Engineers 6, 8 and 9 said they rarely preferred postediting to translating. Engineers 1, 2, 3, 4, 5, 7 and 10 said they sometimes prefer postediting to translating.

We also asked the participants to evaluate their postediting experience with regard to the effort they spent. Engineers 1, 4, 5, 7, 8 and 10 said postediting requires *less* effort than translating. It is interesting to note that all of these engineers received a “Pass” from the LISA QA Model. On the other hand, Engineers 2, 3, 6 and 9, who think that postediting requires *the same* effort as translating, all received a “Fail” degree from the LISA QA Model. This implies us that there is an inverse relationship between the quality and the perceived effort spent on the postediting task as compared to translation.

Further, we asked the participants to share the advantages and disadvantages of the postediting process. The engineers agree on the following points. The same points are also shared by the translators:

ISBN: DL:

1. Postediting saves time and effort.
2. It reduces the frequency of looking up a dictionary.
3. The MT system provides a draft translation. It is easier to polish it than translating from scratch.
4. When you postedit the MT output, you already know what the text is about in broad terms before you start.

The only difference between the engineers' and the translators' point of view of the postediting is that the translators said "the easiness the MT system brings by providing a draft translation depends highly on the text type", while engineers did not touch upon this point.

As for the disadvantages of postediting, the engineers agree on the following points which were also shared by the translators:

1. The incorrect translations offered by the MT system might lead to misunderstandings and semantic errors in complex sentences.
2. The proposal of the wrong equivalent of a word, i.e. for the translation of "to book" (meaning "to reserve"), the MT system may propose the translation of "a book" (the one we read).

Further, Engineer 8 states that, while postediting, one might not elaborate on some sentences (they might just read and accept them), thus corrections/editings required in such parts might be skipped. Similarly, Engineer 10 lists the incorrect translations that sound correct when half-read, among the disadvantages of postediting the MT output. Engineer 9 complains about the extra work caused by the inverted sentences in the MT output. Engineer 6 and 9 point to some technical terms that remain untranslated. Engineer 5 and 7 state that they always suspect the accuracy of machine translation and find it unreliable. Engineer 1 says when the draft translation is very poor, it becomes more difficult and slower to edit the sentence than translating from scratch.

Finally, we asked the participants to state the problems they encountered during the postediting task. Most of the problems are the same with the disadvantages. Engineer 1 did not mention any problems, while Engineer 2 complained that the MT system did not translate some technical terms as they usually use them, therefore the translations of those terms proposed by the MT system were incomprehensible. Engineer 3 said the MT output was sometimes misleading and Engineer 4 said the

ISBN: DL:

numbers in the titles seemed to him as if they were within the sentences. Engineer 5 complained about the unavailability of a proper technical dictionary. Engineer 6, who produced the lowest-quality postediting among the engineers, said he could not find the equivalent of some terms in the dictionary and had to postedit them without understanding well. Engineer 7 emphasized the ambiguity caused by the incorrect equivalent of the terms the MT system proposes and the semantic errors in complex sentences. Engineer 8 regarded the requirement to use a dictionary as a problem. Engineer 9 stated that the points he mentioned among the disadvantages were the problems he came across during the postediting task: (1) Ambiguity, (2) Extra work caused by the inverted sentences in the MT output, (3) The technical terms that remain untranslated. Engineer 10 said trying to follow both the source text and the MT output made the task a little more difficult for him.

9.1.4. Findings for Part 1 (Postediting):

The following hypotheses were tested in Part 1 of the present study:

For technical texts:

1. Subject-matter experts and professional translators postedit MT output differently in terms of processing speed, time spent on documentation and number of changes.

   a. Professional translators complete the postediting task faster than subject-matter experts do.

   b. Professional translators spend more time on documentation than subject-matter experts do.

   c. Professional translators perform more changes than subject-matter experts do.

2. The quality of postediting by subject-matter experts is higher than postediting by professional translators.

Both quantitative and qualitative data were collected to test these hypotheses. As quantitative data, we analyzed subject-matter experts' (engineers') and professional translators' processing speed, words processed per minute (w/pm), time spent on documentation and self-revision, the number and type of changes, distribution of errors, error points and the percentages of quality derived from the LISA QA Model 3.1. As qualitative data, we analyzed post-assignment questionnaires that the subjects filled out after completing the postediting task.

ISBN: DL:

First, we tested if speed and quality in postediting a technical text change depending on the posteditor's being an engineer or a professional translator. So speed and quality are the main dependent variables, while profession (being an engineer or a translator) is the independent variable in this study. Besides, we analyzed the subjects' documentation and self-revision behavior as well as the number and type of changes while postediting a technical text.

The quantitative data did not reveal a significant difference between both groups' (translators and engineers) total task time, words processed per minute, number of changes and the time spent on self-revision (Table 31 and Table 32). However, there was a significant difference between the time translators and engineers spent on documentation while postediting a technical text ($p = 0.0078$, Table 31). We also compared both groups in terms of the time spent on documentation as the percentage of each participant's total task time and found that the translators spent significantly more amount of their total task time on documentation than engineers ($p = 0.0002$) (see Table 31). Thus, hypotheses 1a and 1c were not affirmed by the data, while hypothesis 1b stating that professional translators perform more documentation than engineers while postediting a technical text was affirmed.

Hypothesis 2 stating that - for technical texts - the quality of postediting by subject-matter experts is higher than postediting by professional translators was not affirmed at the first stage, however this hypothesis was affirmed later through further analysis. At first, the mean values for the error points and the quality percentages seemed to indicate that engineers performed better than translators. However, the unpaired t-test results indicated that the difference between the qualities of both groups' performances was not quite significant. The p-value for both the error points and the quality percentages was 0.0636 which was above the threshold 0.05 (Table 37).

In the pilot study, we found that there was a gap between the postediting qualities of the participants *within* each group. Further analysis showed that this difference resulted from the failure to correct the same terms recurring throughout the text - terms that were incorrectly translated by the MT system. Due to the recurrent nature of the errors, the number of errors, and thus the quality of the posteditings as measured by LISA QA Model 3.1, changed when we penalized or unpenalized the recurring errors. Therefore, we decided to look at the quality results by unpenalizing the recurrent errors as well. However, we found that there was no significant difference between the quality

ISBN: DL:

of engineers' and translators' postediting a technical text when we unpenalize the recurrent errors, either. Under the unpenalized condition, the p-value for the error points was 0.4711 and it was 0.4708 for the quality percentages (Table 39). However, the degree of the statistical significance (p-value) changed noticeably when the recurring errors were penalized (p-value is 0.0636 both for the error points and the quality percentages, Table 37) and unpenalized (0.4711 and 0.4708, respectively for the error points and the quality percentages, Table 39). This is striking when we consider the fact that both penalizing and unpenalizing the recurring errors are methods employed in the translation industry (O'Brien 2012).

With a view to obtaining more detailed information on quality, we also analyzed the error points of each error category of the LISA QA Model 3.1. The categories are "mistranslation, accuracy, terminology, language, style, country and consistency." However, we did not find any errors in the "style" and "country" categories. Among the rest, (mistranslation, accuracy, terminology, language and consistency), we found noticeable difference between the translators and engineers in terminology in both cases when the recurrent errors are penalized and unpenalized (Table 36 and Table 38). Table 42 shows the summary of the data we mentioned so far with regard to the conclusions derived from Part 1.

In order to look at the results from a different point of view, we listed the quality results (under the penalized condition) of both groups in an ascending order (Table 43). We also added speed data to this order, but could not see any meaningful pattern between the quality and speed results. The data in Table 43 suggest that the quality of the postediting might be related with factors other than being a professional translator or an engineer. There was extreme variation within each group in terms of the postediting quality. For example, the quality score of Translator 10's postediting was 56.02 while that of Translator 4's was 83.40. A similar gap was found between Engineer 1 with a score of 91.29 and Engineer 2 with 68.46. In order to explain the reasons of this variation and elaborate on the ascending order of the quality results in Table 43, we analyzed the post-assignment questionnaires. The post-assignment questionnaire data proved very clear information with regard to these variations. In order to reduce this variation, we decided to remove the outliers (two participants with the lowest quality in each group) from both the engineers (Engineer 6 and 9) and translators (Translators 3 and 6) for the analyses in Part 1.

ISBN: DL:

The removal of the outliers was based on the post-assignment questionnaire data, where the translators concerned had both of the following features:

1. 3-5 years of professional experience in translation,
2. No degree in Translation.

The outlier engineers had the following features:

1. 3-5 years of professional experience in engineering,
2. No experience in the translation of technical texts as part of their jobs as engineers,
3. Occupation as Metallurgy Engineers while others are Mechanical or Electrical Engineers.

Below, we present what happens to postediting speed and quality when the outliers are removed.

9.1.4.1. Quality in Postediting - Outliers Removed

After removing the outliers, we looked at the quality results both when the recurrent errors are penalized and unpenalized. Table 44 shows the distribution of errors in LISA error categories under the outliers removed and the recurrent errors penalized condition. For most of the categories, there is no significant difference between the translators' and engineers' error distribution data. However, both groups perform differently in significant terms with regard to the terminology errors. Translators make more terminology errors than engineers. The p-values for the minor and major terminology errors are 0.007 and 0.039, respectively when the outliers are removed. These figures are 0.0025 and 0.0774 when the outliers are not removed from the data (Table 44). Another significant value is with the minor language errors. Translators do less minor language errors than engineers. The p-value for the minor language errors is 0.0163 under the recurrent errors penalized and outliers removed condition.

When the recurrent errors are not penalized (Table 45), the only significant difference between the translators' and engineers' is with the minor terminology errors, in both cases when the outliers are removed and included (p-value is 0.0325 and 0.0135, respectively).

ISBN: DL:

Table 44. Distribution of Errors - Outliers Removed - Recurring Errors Penalized

Subjects	Mistranslation			Accuracy			Terminology			Language			Consistency		
	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.	Min.	Maj.	Cri.
Tr 1	0	1		1	0		19	5		5	0		0	0	
Tr 2	1	1		5	0		12	27		2	0		0	4	
Tr 4	0	2		1	1		46	1		2	0		6	1	
Tr 5	0	4		0	1		11	33		20	1		0	3	
Tr 7	0	0		4	0		12	34		7	0		4	0	
Tr 8	3	1		1	0		21	28		7	0		1	0	
Tr 9	0	1		3	0		18	26		14	0		2	2	
Tr 10	1	0		1	4		17	33		5	0		3	0	
Mean	**0.6**	**1.3**		**2.0**	**0.8**		**19.5**	**23.4**		**7.8**	**0.1**		**2.0**	**1.3**	
En 1	0	2		5	0		8	1		11	0		3	0	
En 2	1	2		1	0		4	24		15	0		1	0	
En 3	0	1		1	0		20	33		26	1		3	0	
En 4	0	3		2	1		11	0		12	0		7	1	
En 5	0	3		1	0		1	5		23	1		5	1	
En 7	0	4		1	0		1	4		11	0		2	0	
En 8	0	0		2	0		4	4		13	0		6	4	
En 10	0	1		5	0		0	1		16	0		2	1	
Mean	**0.1**	**2.0**		**2.3**	**0.1**		**6.1**	**9.0**		**15.9**	**0.3**		**3.6**	**0.9**	
p-value-outliers excluded	**0.2266**	**0.211**		**0.7808**	**0.239**		**0.007**	**0.039**		**0.0163**	**0.535**		**0.1563**	**0.6186**	
p-value-outliers included	**0.1373**	**0.537**		**0.8932**	**0.121**		**0.0025**	**0.0774**		**0.2599**	**0.261**		**0.5647**	**0.3566**	

ISBN: DL:

Table 45. Distribution of Errors - Outliers Removed - Recurring Errors Unpenalized

Subjects	Mistranslation			Accuracy			Terminology			Language			Consistency		
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**
Tr 1	0	1		1	0		5	2		2	0		0	0	
Tr 2	1	1		5	0		2	4		2	0		0	2	
Tr 4	0	1		1	1		5	1		2	0		4	1	
Tr 5	0	3		0	1		1	5		20	1		0	2	
Tr 7	0	0		4	0		6	4		7	0		2	0	
Tr 8	1	1		1	0		5	3		7	0		1	0	
Tr 9	0	1		3	0		5	3		14	0		1	2	
Tr 10	1	0		1	1		4	3		5	0		1	0	
Mean	**0.4**	**1.0**		**2.0**	**0.4**		**4.1**	**3.1**		**7.4**	**0.1**		**1.1**	**0.9**	
En 1	0	2		5	0		3	1		8	0		3	0	
En 2	1	2		1	0		2	3		10	0		1	0	
En 3	0	1		1	0		6	3		24	1		3	0	
En 4	0	2		2	1		1	0		8	0		1	1	
En 5	0	2		1	0		1	5		15	1		4	1	
En 7	0	2		1	0		1	2		9	0		2	0	
En 8	0	0		2	0		2	1		13	0		2	1	
En 10	0	1		3	0		0	1		11	0		1	1	
Mean	**0.1**	**1.5**		**2.0**	**0.1**		**2.0**	**2.0**		**12.3**	**0.3**		**2.1**	**0.5**	
p-value-outliers excluded	**0.264**	**0.146**		**1.000**	**0.264**		**0.0325**	**0.1395**		**0.1235**	**0.535**		**0.1309**	**0.492**	
p-value-outliers included	**0.131**	**0.432**		**0.8848**	**0.131**		**0.0135**	**0.0803**		**0.5719**	**0.261**		**0.2448**	**0.275**	

ISBN: DL:

As for the quality percentage, which is the main indicator of quality, when the recurrent errors are not penalized, there is no significant difference between the quality of the translators' and engineers' postediting of a technical text in both cases when the outliers are included and excluded. Table 46 presents both groups' quality data with the mean values and the p-value under the recurrent errors unpenalized and the outliers removed condition as well as the p-value under the outliers included (not removed) condition for comparison of the significance of the results in both cases. There is no significant difference between both groups' postediting qualities under the unpenalized condition when the outliers are removed (p = 0.7331) and not removed (p = 0.4708).

Table 46. Quality in PE - Outliers Removed - Recurrent Errors Unpenalized

Subjects	Total Error Points	Postediting Quality %	Result
En 1	34	92.95	Pass
En 2	40	91.70	Pass
En 3	54	88.80	Pass
En 4	32	93.36	Pass
En 5	66	86.31	Pass
En 7	33	93.15	Pass
En 8	29	93.98	Pass
En 10	30	93.78	Pass
Mean	**40**	**91.75**	
Tr 1	23	95.23	Pass
Tr 2	45	90.66	Pass
Tr 4	32	93.36	Pass
Tr 5	81	83.20	Pass
Tr 7	39	91.91	Pass
Tr 8	35	92.74	Pass
Tr 9	53	89.00	Pass
Tr 10	32	93.36	Pass
Mean	**43**	**91.18**	
p-value-outliers excluded	**0.7335**	**0.7331**	
p-value-outliers included	**0.4711**	**0.4708**	

When the recurrent errors are penalized, on the other hand, the engineers' postediting of a technical text proved to be of significantly higher quality than the translators' (p = 0.0339) under the outliers removed condition (see Table 47). When the outliers are not removed, the p-value is 0.0636 which is also near significant.

ISBN: DL:

Table 47. Quality in PE - Outliers Removed - Recurrent Errors Penalized

Subjects	Total Error Points	Postediting Quality %	Result
En 1	42	91.29	Pass
En 2	152	68.46	Fail
En 3	225	53.32	Fail
En 4	57	88.17	Pass
En 5	80	83.40	Pass
En 7	55	88.59	Pass
En 8	65	86.51	Pass
En 10	38	92.12	Pass
Mean	**89**	**81.48**	
Tr 1	55	88.59	Pass
Tr 2	180	62.66	Fail
Tr 4	80	83.40	Pass
Tr 5	241	50.00	Fail
Tr 7	197	59.13	Fail
Tr 8	178	63.07	Fail
Tr 9	182	62.24	Fail
Tr 10	212	56.02	Fail
Mean	**166**	**65.64**	
p-value outliers excluded	**0.0339**	**0.0339**	
p-value outliers included	**0.0636**	**0.0636**	

The post-assignment questionnaire analysis revealed three factors that affect the translators' postediting quality (Table 48):

1. expertise in the subject-matter
2. perceived understandability of the MT output
3. addressing terminology as a problem.

When we look at the translators listed according to their postediting quality (Table 48), we can see that only Translators 1 and 4 passed the LISA quality test (where the postediting quality threshold is set as 75 percent). The questionnaire data reveal that the technical texts are among those that these participants translate. Although they were unable to pass the LISA QA Model, Translators 8 and 2 are similar to Translators 1 and 4 as far as postediting quality is concerned (Table 48). According to the questionnaires, Translators 8 and 2 also carry out technical translations from time to time. None of these translators are specifically specialized in technical translations, yet "technical" texts are among the text types they translate and the frequency they said they translated technical texts correlates with the quality score they received in the postediting experiment. That is, postediting quality correlates with reported frequency of translation experience with the specific text-type concerned.

Table 46. Post-assignment Questionnaire Data Concerning Conclusions – Translators

Subjects	Postediting Quality % in Ascending Order	Experience in Translation	Education	Type of Texts Translated Mostly	Opinion on the Postediting task	Any problems encountered during the postediting task
Tr 6	41.08	3-5 years	BA in English Language Teaching	Literary	Slightly understandable	I could not find the equivalent of some terms in the dictionary and had to postedit them without understanding well.
Tr 3	46.47	3-5 years	BA in English Language and Culture	Other: Texts on social sciences	Slightly understandable	Lexical and grammatical errors in the MT output cause loss of time. The terminology was problematic for me.
Tr 5	50.00	3-5 years	BA in Translation, PhD in Political Sciences	Educational	Slightly understandable	I had to retranslate some sentences. I had problems regarding the terminology and had to search for some terms using Google. No matter the source text was an MT output or not, I would do these searches. In some terms, it was hard to decide and choose from among the alternatives the MT system proposed and the ones I had in my mind.
Tr 10	56.02	9-10 years	BA in Translation, MA and PhD in Translation	Legal, literary, yet mostly texts on social sciences	Slightly understandable	The incorrect sentence structures and incorrect terminology made it harder to understand this text which was already technical and difficult for me.
Tr 7	59.13	More than 10 years	BA in English Language Teaching	Literary. Other: Medical and pharmacology	Slightly understandable	I felt I lost time. For instance, if I had translated this text myself, I would have finish it in shorter period of time. I lost time especially when correcting the inverted sentences, punctuation and the terminological errors. Further, the technical terms in the text made it harder for me to edit it.
Tr 9	62.24	6-8 years	BA in International Relations, MA and PhD in Translation	Literary	Slightly understandable	
Tr 2	62.66	More than 10 years	BA in Translation, MA in Communication Sciences	Other: Mostly texts on social sciences, yet from time to time technical, legal texts and user manuals	Understandable	I haven’t come across a serious problem except for some technical terms, i.e. deployment.

ISBN: DL:

Tr 8	63.07	9-10 years	BA in Translation	Other: Educational, social, from time to time technical documents	Understandable	Although it was not a very big problem, the erroneous translation of some terms made it more difficult for me to understand the text. Somehow, it seems as if you do the same work twice, yet it is not true to say that postediting is completely disadvantageous.
Tr 4	83.40	9-10 years	MA in Translation, PhD in Translation	Technical, legal, Other: academic texts on social sciences and education	Understandable	I haven't come across a serious problem.
Tr 1	88.59	9-10 years	BA, MA and PhD in Translation	Technical, legal, texts and user manuals	Quite understandable	It is easy to distinguish the absurd translations the MT system offers. However, sometimes the MT system may offer misleading translations. If you read only the target text, then the text might sound fluent, but the meaning might be incorrect. For this reason, it is crucial to refer to the source text. It seems easier to translate some sentences from scratch, yet for some others, there is almost no point to be corrected. In fact, it depends on the text type and the richness of the MT system's database.
r-value btw. quality and experience		**0.50**				

ISBN: DL:

According to the questionnaires, there is also a correlation between the level of perceived understandability of the MT output and the postediting quality. Translator 1, who produced the highest-quality postediting among the translators in the study, found the MT output quite understandable and the subsequent translators (Translator 4, 8 and 2) also found it understandable, while others producing lower-quality postediting thought the MT output was slightly understandable. Although there are other factors such as the participants' background knowledge and expectations of the MT system that might affect their point of view on the understandability of the MT output, according to our data, the level of perceived understandability correlates with expertise in the subject-matter. That is to say, the more expertise the translators have with technical texts and terms, the better they understand the MT output containing technical terms. We also analyzed the questionnaire data to see if the level of perceived understandability is related with the translators' experience in revision and/or proofreading, however, we could not find any significant correlation between those factors.

As with perceived understandability of the MT output, the third factor (addressing terminology as a problem) can also be associated with the first. In other words, the participants' perception of terminology as a problem is seen to be related with their level of expertise in the subject-matter. Thus, in general, for the translators, subject-matter knowledge stands out as the main factor affecting the quality.

Our qualitative data do not reveal a strong correlation between the postediting quality and a degree (BA, MA or PhD) in Translation. Three out of ten translators do not hold a degree in Translation. These are Translators 3, 6 and 7. Others hold either a BA, MA, PhD or two/three of them in Translation. Translators 6 and 3 have the lowest quality-scores. Translator 7 ranked sixth among the translators in terms of quality, yet she has more than 10 years of professional experience, which might compensate for the degree in Translation to some extent. Translator 10, who ranked seventh among the translators with regard to the postediting quality, holds a BA, MA and PhD in Translation, and Translator 5 who is the eighth, has a BA in Translation. On the other hand, when we look at the most successful translators (Translators 1 and 4), we see that they both hold an MA and PhD in Translation. Translators 8 and 2, who follow Translator 1 and 4 in terms of postediting quality, also hold a BA in Translation (see Table 48). Table 48 shows that these translators (Translators 1, 4, 8 and 2) have more than 9 years of experience in the translation profession and they also have experience in the translation of technical texts. Thus, according to our data, a degree in Translation per se is not directly correlated with postediting quality unless it is in combination with professional experience and/or subject-matter knowledge.

Our data reveal that a strong correlation exists (r = 0.50, Table 48) between postediting quality and experience in the translation profession, though it is not so direct. Translators 1 and 4, who passed the LISA quality test, have 9-10 years of translation experience. Translators 8 and 2, who ranked third and the fourth in the quality test - yet did not pass it -, have 9-10 and more than ten years of translation experience respectively. Translators 9, 7 and 10, who follow Translators 1, 4, 8 and 2 with regard to quality; have 6-8, more than ten and 9-10 years of experience respectively (see Table 48). The rest of the translators (Translators 5, 3 and 6) have 3-5 years of experience. This means that the translators with 3-5 years of experience produced relatively lower-quality postediting, while translators with 9-10 years of experience produced higher-quality postediting. However, Translators 7 and 10, who received low scores from the LISA QA Model, but had more than ten and 9-10 years of experience respectively, contravene the ascending order of the years of translation experience in relation to quality. This can be explained with Translator 7 and 10's lack of experience in the translation of technical texts.

According to the results derived from the post-assignment questionnaires, three factors stand out as affecting the engineers' postediting quality (Table 49):

1. experience in the translation of technical texts as part of their job as engineers
2. the level of professional experience in engineering
3. perceived effort spent on postediting the MT output in comparison with translating.

Except for Engineer 3, all of the engineers participated in the study stated that they had to do translation as part of their daily work as engineers. Engineers 10, 1, 7, 4, 8 and 5 all passed the LISA quality test and they received high quality-scores in descending order. It is interesting to note that all of them have to do technical translation as part of their daily work (see Table 49). The questionnaire results of other engineers affirm this trend, that is, Engineer 2, who is the first engineer that failed the quality test, said he carried out technical translation as part of his job only from time to time, while Engineer 3 said he did not do translation as part of his job. Engineers 6 and 9, who had the lowest quality-scores among the engineers (see Table 49), said they only translate electronic mails as part of their work and those were not of technical nature

ISBN: DL:
(Table 49). According to our data, engineers who have to do technical translation as part of their daily work produced higher-quality postediting than those who either do not translate technical documents or do not have to translate at all as part of their professional activities.

The correlation between the engineers' level of professional experience in engineering and their postediting quality is very strong (r = 0.87, Table 49). Engineers 5, 8, 4, 7, 1 and 10 passed the LISA QA grid in ascending order (Table 49). The questionnaire data show that all of these engineers have more than nine years of professional experience in engineering and they all do the translation of technical texts as part of their job. Further, all of these engineers think that postediting MT output requires less effort than translating from scratch. On the other hand, Engineers 6, 9, 3 and 2, who failed the quality test in ascending order, have 3-5 years (Engineers 6, 9 and 3) or 6-8 years (only Engineer 2) of professional experience in engineering.

ISBN: DL:

Table 49. Post-assignment Questionnaire Data Concerning Conclusions – Engineers

Subjects	**Postediting Quality % in Ascending Order**	**Education/Level**	**Experience in Engineering**	**Do you have to do translation as part of your professional activity as an engineer?**	**Type of texts translated mostly as an engineer**	**Based on your postediting experience (either previous or from this experiment), which of the options below best represents your experience?**
Eng 6	51.24	BSc and MSc in Metallurgy and Materials Engineering	3-5 years	Yes	e-mails	Postediting, for me, requires the *same* effort as translating
Eng 9	52.70	BSc in Metallurgy and Materials Engineering	3-5 years	Yes	e-mails	Postediting, for me, requires the *same* effort as translating
Eng 3	53.32	BSc and MSc in Electrical Engineering	3-5 years	No	NA	Postediting, for me, requires the *same* effort as translating
Eng 2	68.46	BSc in Mechanical Engineering, MSc in Mechanical Engineering	6-8 years	Yes	e-mails, reports. Other: from time to time simple technical documents	Postediting, for me, requires the *same* effort as translating
Eng 5	83.40	BSc and MSc in Mechanical Engineering	More than 10 years	Yes	e-mails, minutes of meetings, reports, technical documents	Postediting, for me, requires *less* effort than translating
Eng 8	86.51	BSc and MSc in Electrical Electronics Engineering	9-10 years	Yes	Reports, Other: I have to translate or prepare in English the documents containing the electrical-electronical descriptions of the modifications to be implemented on the vehicles exported to the services abroad.	Postediting, for me, requires *less* effort than translating
Eng 4	88.17	BSc in Mechanical Engineering, MSc in Automotive Engineering	9-10 years	Yes	e-mails, minutes of meetings, technical documents	Postediting, for me, requires *less* effort than translating
Eng 7	88.59	BSc in Mechanical Engineering	9-10 years	Yes	e-mails, minutes of meetings, reports, technical documents	Postediting, for me, requires *less* effort than translating
Eng 1	91.29	BSc in Mechanical Engineering, MBA	9-10 years	Yes	e-mails, minutes of meetings, reports, technical documents and PP presentations	Postediting, for me, requires *less* effort than translating
Eng 10	92.12	BSc in Mechanical Engineering	More than 10 years	Yes	e-mails, technical documents, reports	Postediting, for me, requires *less* effort than translating
r-value btw. quality and experience		**0.87**				

ISBN: DL:

9.1.4.2. Changes, Documentation, Self-Revision and Speed in Postediting - Outliers Removed

After removing the outliers from engineers and translators, we recalculated the p-values that belonged to the data on the changes, documentation, self-revision and speed in order to see if the removal of outliers will change the significance of the results.

The changes performed by each group while postediting was among the parameters we looked at in the quantitative data analysis. We did not find any significant difference between the changes performed by the translators and the engineers while postediting a technical text before and after we remove the outliers (Table 50)

Table 50. Changes in PE - Translators and Engineers - Outliers Removed

Subjects	Content Changes	Formal Changes		Omissions	Additions	Reorganization	Total
		Capitals	Punctuation				
Tr 1	62	5	7	20	46	10	150
Tr 2	60	8	4	14	36	13	135
Tr 4	82	4	2	10	38	11	147
Tr 5	82	3	5	18	30	10	148
Tr 7	65	8	8	18	27	13	139
Tr 8	76	8	5	14	58	11	172
Tr 9	93	8	4	12	30	11	158
Tr 10	41	4	5	18	29	13	110
Mean	**70**	**6**	**5**	**16**	**37**	**12**	**145**
En 1	83	5	12	12	33	10	155
En 2	42	4	5	18	31	7	107
En 3	32	5	2	12	35	11	97
En 4	81	5	1	15	14	13	129
En 5	76	5	4	11	19	9	124
En 7	63	7	3	8	32	15	128
En 8	74	10	10	21	39	11	165
En 10	53	4	4	15	50	14	140
Mean	**63**	**6**	**5**	**14**	**32**	**11**	**131**
p-value-outliers excluded	**0.4349**	**0.7266**	**0.9355**	**0.4472**	**0.3627**	**0.8149**	**0.1868**
p-value-outliers included	**0.9089**	**0.6692**	**0.7228**	**0.4933**	**0.4874**	**0.7157**	**0.6709**

As for the documentation carried out by both groups, the removal of the outliers did not change significance of the difference between translators and engineers, either.

ISBN: DL:

Both when we include and remove the outliers, we found significant difference between both groups' documentation behavior in four out of five parameters: the number of on-line dictionary look-up, the time spent on documentation, documentation as the percentage of the total task time and the number of items searched while carrying out documentation (Table 51). As a result, no matter whether we include or exclude the outliers, translators spent significantly more time on documentation than engineers while postediting technical texts. In addititon, they use on-line dictionaries more and search for more items for longer periods of time. However, both groups use a very limited number of sources on the Internet while performing documentation.

Table 51. Documentation in PE - Translators and Engineers - Outliers Removed

Subjects	Number of Look-up	Time spent (minute)	Documentation as % of the Total Task Time	Number of Sources Used	Number of Items Searched
Tr 1	16	07.51	16.7	3	14
Tr 2	15	07.51	11.1	2	13
Tr 4	15	13.08	21.7	6	19
Tr 5	41	28.00	25.9	4	44
Tr 7	21	08.76	12.8	2	31
Tr 8	6	04.56	7.3	2	7
Tr 9	25	08.85	14.7	2	29
Tr 10	30	15.08	21.2	3	31
Mean	**22**	**13.06**	**17.2**	**3**	**24**
En 1	10	02.76	5.2	3	6
En 2	1	00.33	0.7	1	1
En 3	11	03.70	8.0	3	7
En 4	9	03.93	6.0	3	8
En 5	6	01.51	2.5	1	5
En 7	2	00.48	0.7	1	2
En 8	23	09.35	11.0	4	21
En 10	13	02.91	5.0	3	13
Mean	**9**	**03.18**	**5.2**	**2**	**8**
p-value-outliers excluded	**0.0216**	**0.0086**	**0.0004**	**0.3547**	**0.0066**
p-value-outliers included	**0.0156**	**0.0078**	**0.0002**	**0.2448**	**0.0050**

The removal of the outliers did not change the results on the self-revision, either. Both when the outliers are included and removed, there is not a significant difference between the time spent by the translators and the engineers on self-revision (Table 52).

Similarly, the results on the translators' and the engineers' speed while they are postediting a technical text did not change when we remove the outliers. None of the parameters with regard to speed yielded significant difference between the translators

ISBN: DL:

and the engineers when the outliers are included or removed (Table 52).

Table 52. Self-Revision and Speed in PE - Translators and Engineers - Outliers Removed

Subjects	Time on PE Excluding Self-revision (minute)	Time on Self-revision (minute)	Total Task Time (PE and Self-rev) (minute)	Words processed per minute	Words processed per minute (excluding self-revision)
Tr 1	39.70	05.28	44.98	10.7	12.1
Tr 2	51.25	16.28	67.53	7.1	9.4
Tr 4	37.45	22.71	60.16	8.0	12.8
Tr 5	92.06	15.68	107.75	4.5	5.2
Tr 7	54.28	14.46	68.75	7.0	8.8
Tr 8	62.66	00.00	62.66	7.6	7.6
Tr 9	60.03	00.00	60.03	8.0	8.0
Tr 10	65.23	05.85	71.08	6.7	7.3
Mean	**57.83**	**10.03**	**67.86**	**7.5**	**8.9**
En 1	52.96	00.00	52.96	9.1	9.1
En 2	47.45	00.00	47.45	10.1	10.1
En 3	46.00	00.00	46.00	10.4	10.4
En 4	65.61	00.00	65.61	7.3	7.3
En 5	60.96	00.00	60.96	7.9	7.9
En 7	61.70	00.00	61.70	7.8	7.8
En 8	61.05	23.25	84.30	5.7	7.8
En 10	45.43	12.30	57.73	8.3	10.6
Mean	**55.15**	**04.45**	**59.60**	**8.3**	**8.9**
p-value-outliers excluded	**0.6946**	**0.2127**	**0.2995**	**0.3014**	**0.9806**
p-value-outliers included	**0.5860**	**0.3498**	**0.3598**	**0.4195**	**0.7157**

9.1.4.3. The Relation between Changes, Documentation, Self-Revision, Speed and Quality within the Groups

In order to see whether there is any correlation between postediting quality and (1) changes, (2) documentation, (3) self-revision and (4) speed within the groups, we arranged the relevant data in ascending order: (1) the number of the changes performed, (2) the time spent on documentation, (3) the time spent on self-revision, and (4) words processed per minute. Then we matched the data on these four parameters with the postediting quality under the recurring errors penalized condition. Table 53 shows the number of changes - in ascending order - performed by the translators and their postediting quality. Translator 6 and Translator 3, who have the lowest-quality posteditings among the translators, made the fewest number of changes while postediting the MT output. On the other hand, Translator 5, who is the third translator

ISBN: DL:

with the lowest quality-score, made 148 changes. Translators 1, 4, 8 and 2, having the highest quality-scores (in the descending order) among the translators, made 150, 147, 172 and 135 changes respectively. We ran a Pearson correlation test on the data and found that a positive correlation exists between the changes performed by translators and their postediting quality (r = 0.63).

Table 53. Changes in PE in Ascending Order – Translators

Subjects	Number of Changes in Ascending Order	Postediting Quality %
Tr 6	58	41.08
Tr 3	80	46.47
Tr 10	110	56.02
Tr 2	135	62.66
Tr 7	139	59.13
Tr 4	147	83.40
Tr 5	148	50.00
Tr 1	150	88.59
Tr 9	158	62.24
Tr 8	172	63.07
r-value	**0.63**	

Table 54 presents the number of changes - in ascending order - performed by engineers. The engineers who made fewer than 120 changes (Engineers 6, 3, 2 and 9) are those with the lowest quality-scores, while those performing between 124 and 165 changes have higher quality-scores (Table 54). The Pearson test shows that a very strong correlation exists between the quality of engineer postediting and the number of changes performed (r = 0.81).

ISBN: DL:

Table 54. Changes in PE in Ascending Order - Engineers

Subjects	Number of Changes in Ascending Order	Postediting Quality %
En 6	71	51.24
En 3	97	53.32
En 2	107	68.46
En 9	119	52.70
En 5	124	83.40
En 7	128	88.59
En 4	129	88.17
En 10	140	92.12
En 1	155	91.29
En 8	165	86.51
r-value	**0.81**	

Table 55 shows the number of words the translators processed per minute (w/pm) in ascending order, along with their postediting quality. The Pearson test shows that there is no correlation between the translators' postediting quality and the number of the words they processed per minute (r = - 0.07). However, interpretation of the data provides us with a different perspective on the relationship between quality and the w/pm. Translator 1 is the most successful and fastest translator, with the highest quality-score and speed rate among the translators (88.59 percent for quality and 10.7 w/pm for speed). Except for Translator 1, the translators producing 10 or more w/pm had the lowest quality-scores among the translators (Translators 3 and 6). Except for Translator 1, the translators with the highest quality-scores (Translators 4, 8, 2 and 7) are those with the speed rate of 7-8 w/pm (Table 55), whereas Translators 5 and 10, producing lower quality-scores, had the lowest speed rate of 4.5 and 6.7 w/pm respectively. Consequently, the data show that the slowest and the fastest translators had lower quality-scores while those with the mean values of speed produced higher-quality outputs.

ISBN: DL:

Table 55. Words Per Minute Data in PE in Ascending Order – Translators

Subjects	Words processed per minute in ascending order	Postediting Quality %
Tr 5	4.5	50.00
Tr 10	6.7	56.02
Tr 7	7.0	59.13
Tr 2	7.1	62.66
Tr 8	7.6	63.07
Tr 4	8.0	83.40
Tr 9	8.0	62.24
Tr 3	10.6	46.47
Tr 1	10.7	88.59
Tr 6	12.4	41.08
r-value	**- 0.07**	

When we look at the engineers' speed data (Table 56), we see that those who processed more than 10 w/pm (Engineers 3, 6 and 9) are also those with the lowest quality-scores, while engineers producing between 5.7 and 10.1 w/pm had higher quality-scores. The data show that, as compared to the translators, engineers produce higher-quality posteditings with higher speed. The Pearson test shows that there is a strong negative correlation between the engineers' postediting quality and the number of the words they processed per minute (r = - 0.80). In other words, for engineers, working very fast brings about lower quality. This is no surprise, since one might expect that the quality could be disregarded when the output is produced too fast. However, it is interesting to note that, for the translators, postediting the MT output very slowly (Table 55) ends up with lower quality texts, too. A previous pilot study also revealed that "translations done under time pressure tend to be of better quality than those done without any time constraints" (Khalzanova 2008: 66). Although our subjects did not work under time pressure, this is an intriguing result, pointing out that spending more time on the task does not always guarantee higher-quality output.

ISBN: DL:
Table 56. Words Per Minute Data in PE in Ascending Order – Engineers

Subjects	Words processed per minute in ascending order	Postediting Quality %
En 8	5.7	86.51
En 4	7.3	88.17
En 7	7.8	88.59
En 5	7.9	83.40
En 10	8.3	92.12
En 1	9.1	91.29
En 2	10.1	68.46
En 3	10.4	53.32
En 6	10.8	51.24
En 9	13.6	52.70
r-value	**- 0.80**	

The present study has revealed a weak correlation between documentation and the postediting quality (r = 0.30). The engineers with the highest quality scores (Engineers 10, 1 and 7) and the ones with the lowest (Engineers 6 and 9) spent less than three minutes on documentation (Table 57). This shows us that, for the engineers, the time spent on searching for information for postediting technical texts was not an important factor affecting the level of postediting quality.

Table 57. Time on Documentation in PE in Ascending Order – Engineers

Subjects	Time Spent on Documentation (minute) in Ascending Order	Postediting Quality %
En 2	00.33	68.46
En 7	00.48	88.59
En 6	00.96	51.24
En 9	01.30	52.70
En 5	01.51	83.40
En 1	02.76	91.29
En 10	02.91	92.12
En 3	03.70	53.32
En 4	03.93	88.17
En 8	09.35	86.51
r-value	**0.30**	

The result is similar for the translators. The translators who spent the longest time on documentation, namely Translators 5 and 10 (28 and 15.08 minutes respectively), had relatively low quality-scores of 50.00 and 56.02 percent respectively. The

ISBN: DL:
translators with the highest quality-scores (Translators 1 and 4) spent 7.51 and 13.08 minutes on documentation, whereas the ones with the lowest quality-scores (Translators 6, 3 and 5) spent 5, 2.65 and 28 minutes on documentation respectively (Table 58). When we look at the time each translator spent on documentation, we see that there is no correlation between the quality and the time spent on documentation (r = - 0.02)

Table 58. Time on Documentation in PE in Ascending Order – Translators

Subjects	**Time Spent on Documentation (minute) in Ascending Order**	**Postediting Quality %**
Tr 3	02.65	46.47
Tr 8	04.56	63.07
Tr 6	05.00	41.08
Tr 2	07.51	62.66
Tr 1	07.51	88.59
Tr 7	08.76	59.13
Tr 9	08.86	62.24
Tr 4	13.08	83.40
Tr 10	15.08	56.02
Tr 5	28.00	50.00
r-value	**- 0.02**	

Our data indicate that the correlation between the time engineers spent on self-revision and quality is negligible (r = 0.19). Most of the engineers did not do any self-revision (Table 59). Engineer 10, who had the highest quality-score, spent 12.3 minutes on self-revision (Table 59), while Engineers 1 and 7, who produced the second and the third highest-quality postediting among the engineers, did not do any self-revision. On the other hand, Engineer 6, who had the lowest quality-score among the engineers, spent 9.23 minutes on self-revison (Table 59).

ISBN: DL:

Table 59. Time on Self-revision in PE in Ascending Order – Engineers

Subjects	Time on Self-revision (minute) in Ascending Order	Postediting Quality %
En 9	0	52.70
En 3	0	53.32
En 2	0	68.46
En 5	0	83.40
En 4	0	88.17
En 7	0	88.59
En 1	0	91.29
En 6	09.23	51.24
En 10	12.30	92.12
En 8	23.25	86.51
r-value	**0.19**	

As compared to the engineers, the translators carried out almost twice as much self-revision, although the difference between the groups is not statistically significant ($p = 0.3498$, Table 31). A total of six out of ten translators did self-revision (Table 60). Translator 1, with the highest quality-score among the translators, spent 5.28 minutes on self-revision, while Translator 4, who had the second highest quality-score among the translators, spent 22.71 minutes on self-revision. Translator 5, with a 50 percent quality-score, spent 15.68 minutes on this process. The Pearson test indicates a moderate correlation between the time translators spent on self-revision and their postediting quality ($r = 0.37$).

Table 60. Time on Self-revision in PE in Ascending Order – Translators

Subjects	Time on Self-revision (minute) in Ascending Order	Postediting Quality %
Tr 6	0	41.08
Tr 3	0	46.47
Tr 9	0	62.24
Tr 8	0	63.07
Tr 1	05.28	88.59
Tr 10	05.85	56.02
Tr 7	14.46	59.13
Tr 5	15.68	50.00
Tr 2	16.28	62.66
Tr 4	22.71	83.40
r-value	**0.37**	

ISBN: DL:

9.2. Revision

In Part 2 we compared the speed and product quality of two different workflows: (1) postediting by professional translators and revision of the postedited output by subject-matter experts, (2) postediting by subject-matter experts and revision of the postedited output by professional translators. We hypothesized that, for technical texts, when professional translators revise postediting by engineers and when engineers revise postediting by professional translators, speed and product quality differ. In order to explore this hypothesis, we had ten engineers revise the posteditings performed by professional translators and we also had ten professional translators revise the posteditings performed by engineers. Both groups were asked to record their revision processes using BB Flashback screen-recording software. They were free to use Internet sources during the task. We evaluated the revision process in terms of speed and the revised end-products in terms of quality.

9.2.1. Speed in Revision

Speed was measured with two types of data: processing speed captured with screen recording software and the number of words processed per minute (w/pm). Both sets of data were recorded for the revision tasks only (only translators' revisions or only engineers' revisions) as well as for the whole task involving postediting and revision (engineers' postediting + translators' revision and translators' postediting + engineers' revision). When calculating the w/pm for the postediting task, we used the word-count of the raw MT output (482) for all the posteditors. However, for revision, the number of words each reviser processes depends on the number of words produced by the posteditor. That is, the word-count of the raw MT output is changed during postediting with additions and deletions made by the posteditor. Thus, for the calculation of w/pm for each revised text, we used the different word-counts produced by each posteditor. In order to calculate the w/pm for postediting + revision tasks, we considered both the word count of the raw MT output (for w/pm of postediting) and that of the postedited text (for w/pm of revision).

Table 61 presents the speed of translator postediting in comparison with its revision by engineers, the speed of engineer postediting in comparison with its revision by translators as well as the total task time comprising postediting and revision. Both groups completed the revision task in almost half as much the time they completed the

ISBN: DL:

postediting task. On average, translators spent 62.70 minutes for postediting and 27.40 minutes for revision (Table 61). On the other hand, engineers spent 55.66 minutes for postediting and 33.31 minutes for revision. Unlike the mean duration of postediting where engineers worked faster than translators (55.66 minutes against 62.70 minutes), the mean duration of the engineers' revision is longer than that of translators' (33.31 minutes against 27.40 minutes). In revision, engineers processed, on average, 18.6 w/pm, while translators processed 29.2 w/pm. Although the mean values indicate that the translators revise faster than the engineers, according to the t-test, there is no significant difference between translators and engineers with regard to speed in revising one another's postediting. The p-value for the comparison of both groups' revision duration is 0.4640 and it is 0.2727 for their w/pm (Table 61).

In addition to comparing the speed of only the revision tasks, we compared the speed of postediting + revision processes. The mean duration of engineer postediting + translator revision is thirteen minutes shorter than translator postediting + engineer revision. On average, the former process took 83.06 minutes, while the latter took 96.01 minutes. The mean value of w/pm for engineer postediting + translator revision is 13.2 and this figure is 10.7 for translator postediting + engineer revision (Table 61). Although the mean values for the total task time comprising postediting + revision indicate that engineer postediting + translator revision is faster than translator postediting + engineer revision, the t-tests do not show any significant difference between both processes with regard to speed (p-value for postediting + revision task time is 0.3021 and it is 0.1611 for w/pm, Table 61).

ISBN: DL:

Table 61. Speed in Revision - Translators and Engineers

	Translator Postediting			**Engineer Revision of Translator Postediting**			
Subjects	**PE Task Time incl. self-revision (minute)**	**Word processed per minute**	**Subjects**	**REV Task Time (minute)**	**Word processed per minute**	**Translator PE+Engineer Revision Task Time (minute)**	**Translator PE+Engineer Revision w/pm**
Tr 1	44.98	10.7	**En 1**	24.73	20.5	69.71	14.2
Tr 2	67.53	7.1	**En 2**	42.41	12.1	109.94	9.1
Tr 3	45.36	10.6	**En 3**	54.73	8.7	100.10	9.6
Tr 4	60.16	8.0	**En 4**	12.70	39.1	72.86	13.4
Tr 5	107.75	4.5	**En 5**	14.50	32.2	122.25	7.8
Tr 6	38.73	12.4	**En 6**	22.66	21.5	61.40	15.8
Tr 7	68.75	7.0	**En 7**	28.93	16.8	97.68	9.9
Tr 8	62.66	7.6	**En 8**	58.91	8.2	121.58	7.9
Tr 9	60.03	8.0	**En 9**	38.48	12.9	98.51	10.0
Tr 10	71.08	6.7	**En 10**	35.03	13.7	106.11	9.1
Mean	**62.70**	**8.3**		**33.31**	**18.6**	**96.01**	**10.7**
	Engineer Postediting			**Translator Revision of Engineer Postediting**			
Subjects	**PE Task Time incl. self-revision (minute)**	**Word processed per minute**	**Subjects**	**REV Task Time (minute)**	**Word processed per minute**	**Engineer PE+Translator Revision Task Time (minute)**	**Engineer PE+Translator Revision w/pm**
En 1	52.96	9.1	**Tr 1**	26.21	18.3	79.18	12.2
En 2	47.45	10.1	**Tr 2**	18.91	26.1	66.36	14.7
En 3	46.00	10.4	**Tr 3**	24.93	20.0	70.30	13.9
En 4	65.61	7.3	**Tr 4**	32.53	14.7	98.15	9.8
En 5	60.96	7.9	**Tr 5**	40.26	11.6	101.23	9.4
En 6	44.51	10.8	**Tr 6**	16.50	29.9	61.01	16.0
En 7	61.70	7.8	**Tr 7**	18.40	27.9	80.10	12.4
En 8	84.30	5.7	**Tr 8**	75.28	6.8	159.58	6.3
En 9	35.40	13.6	**Tr 9**	04.55	105.0	39.95	24.0
En 10	57.73	8.3	**Tr 10**	16.40	31.5	74.13	13.5
Mean	**55.66**	**9.1**		**27.40**	**29.2**	**83.06**	**13.2**
p-value	**0.3598**	**0.4195**		**0.4640**	**0.2727**	**0.3021**	**0.1611**

ISBN: DL:

9.2.2. Quality in Revision

We measured the quality of the final revised products with LISA QA Model 3.1 in the same way we measured the quality of the postedited texts. However, for revision, the minimum acceptable level of quality was set as 85 percent, which was above the quality threshold set for the posteditings (75 percent). The reason for this 10-percent increase in the quality threshold is that the revised texts are the final versions of the posteditings, thus their quality requirement is higher.

We evaluated the quality of revision in three different ways. First, we measured whether the same participant performed differently when postediting and revising the same text. This was to see how the quality of each group's postediting compares with the quality of the same group's revision, which has already been postedited by the other group (i.e. how Translator 1's postediting compares with Translator 1's revision, which has already been postedited by Engineer 1). Second, we compared the quality of revisions performed by professional translators and subject-matter experts (engineers). Third, we compared the quality of the postedited texts and their final revised versions in order to determine how the subject-matter knowledge of translators/engineers as revisers affect the quality of engineers'/translators' postediting.

As is the case in postediting, due to the recurrent nature of the errors, the number of errors, thus the quality of revision as measured by LISA QA 3.1 might change when we penalize or unpenalize the recurring errors. Therefore, we decided to look at the quality results by penalizing and unpenalizing the recurrent errors.

9.2.2.1. Quality in Revision - Recurring Errors Penalized

The revised texts were compared with the reference translation of the source text (translated by a professional translator and an engineer in cooperation) with a view to detecting errors. The errors were classified in line with LISA error categories. As is the case in postediting, there were no errors in "style" and "country" categories, so they were removed from the error analysis.

First we measured whether the same participant performed differently when postediting and revising the same text. Postediting and revising the same text might result in the participants' learning of the text that they have already postedited. However, since revising a text in which the reviser has already had experience is very natural and common in the industry, this learning factor does not pose a problem for the

ISBN: DL:

present research. Table 62 presents the quality percentages of each engineer's postediting and revision. When they did revision, all of the engineers except 2 and 6 produced higher-quality texts than when they postedited. The mean quality percentage of engineer postediting is 75.58, while it is 81.05 for engineer revision. The mean difference between the quality percentage of engineer revision and engineer postediting is 5.47. Although the mean numbers indicate an increase in the quality of engineers' revision with respect to their postediting, there is no significant difference between the two (p = 0.5004, Table 62).

Table 62. Engineer Postediting and Engineer Revision - Quality Compared - Recurring Errors Penalized

Subjects	Postediting Quality %	Revision Quality %	Difference btw REV. and PE %
En 1	91.29	95.34	4.05
En 2	68.46	66.54	-1.92
En 3	53.32	72.11	18.79
En 4	88.17	98.78	10.61
En 5	83.40	86.21	2.81
En 6	51.24	47.65	-3.59
En 7	88.59	93.59	5.00
En 8	86.51	96.09	9.58
En 9	52.70	59.35	6.65
En 10	92.12	94.80	2.68
Mean	**75.58**	**81.05**	**5.47**
P-value of the difference between Engineer PE and REV			**0.5004**

Table 63 presents the quality percentages of each translator's postediting and revision. When they did revision, all of the translators except Translator 9 produced higher-quality texts than when they postedited. When we compare the mean quality percentages of translator postediting (61.27) and translator revision (79.68), we find that there is a significant difference (p = 0.0232) between the end-product quality of the two processes.

ISBN: DL:

Table 63. Translator Postediting and Translator Revision - Quality Compared - Recurring Errors Penalized

Subjects	Postediting Quality %	Revision Quality %	Difference btw REV. and PE %
Tr 1	88.59	93.31	4.72
Tr 2	62.66	67.73	5.07
Tr 3	46.47	56.55	10.08
Tr 4	83.40	99.59	16.19
Tr 5	50.00	83.80	33.80
Tr 6	41.08	57.75	16.67
Tr 7	59.13	92.31	33.18
Tr 8	63.07	93.54	30.47
Tr 9	62.24	56.16	-6.08
Tr 10	56.02	96.09	40.07
Mean	**61.27**	**79.68**	**18.42**
P-value of the difference between Translator PE and REV			**0.0232**

The difference between the engineers' revision and postediting (Table 62) is very small and insignificant as compared to the difference between the translators' revision and postediting (Table 63). In Part 1, we found that the quality of engineers' postediting is higher than that of translators' (p = 0.0636, Table 37). When the outlier participants are removed, this difference becomes significant (p = 0.0339, Table 47). The quality difference between the translators' revision and postediting (whose quality is lower than engineers' postediting) can be explained by the higher quality of the posteditings they revised. Similarly, the insignificant difference between the engineers' revision and postediting (whose quality is higher than translators' postediting) can be explained by the lower-quality posteditings they revised. Thus, the comparison of the quality of each group's postediting and revision reveals that revision quality is not only dependent on the qualifications of the reviser but also on the quality of the text to be revised.

Second, we compared the revision processes performed by translators and engineers. Table 64 shows the distribution of errors in the texts revised by translators and engineers when the recurrent errors are penalized. In both groups, most of the errors in revision are in the language and terminology categories. According to the mean values, both the translators and the engineers made more major terminology errors than errors in other categories. On average, translators made 14.4 major terminology errors while engineers made 11.8 errors in this category. As for the language category, translators made, on average, 9.2 minor language errors while engineers made 9.6 errors in this category. The mean values for language and terminology errors is 26.8 for translators and 26.7 for engineers. The mean total number of errors of the translators' is

ISBN: DL:

31.8 and it is 32.3 for engineers. The p-values for each error category and for the total number of errors indicate that there is no significant difference between the translators' and engineers' distribution of errors when revising each other's posteditings (Table 64). At this point, it is important to note that the total number of errors is different from the total error points. The former is obtained by simply adding up the number of errors, whereas the latter is calculated by the LISA QA Model upon entering the error count to the system.

ISBN: DL:

Table 64. Distribution of Errors - Revision - Recurring Errors Penalized

Subjects	**Mistranslation**			**Accuracy**			**Terminology**			**Language**			**Consistency**			**Number of Language and Terminology Errors**	**Total Number of Errors**
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**		
Translator Revision of Engineer Postediting																	
Tr 1	0	1		2	0		6	3		1			3	0		10	16
Tr 2	1	0		2	0		0	29		14			0	0		43	46
Tr 3	3	2		1	0		16	34		19			0	0		69	75
Tr 4	0	0		0	0		0	0		2			0	0		2	2
Tr 5	0	5		1	1		2	3		13			0	3		18	28
Tr 6	1	2		2	0		1	34		26			0	0		61	66
Tr 7	0	4		1	0		0	3		3			1	0		6	12
Tr 8	0	1		1	0		4	4		4			0	0		12	14
Tr 9	0	5		1	0		2	33		6			1	2		41	50
Tr 10	0	1		1	0		1	1		4			0	1		6	9
Mean	**0.5**	**2.1**		**1.2**	**0.1**		**3.2**	**14.4**		**9.2**			**0.5**	**0.6**		**26.8**	**31.8**
Engineer Revision of Translator Postediting																	
En 1	0	1		1	0		10	1		3			0	0		14	16
En 2	0	0		2	0		3	32		6			0	0		41	43
En 3	0	4		3	0		6	13		16			0	6		35	48
En 4	0	0		2	0		0	0		4			0	0		4	6
En 5	0	1		0	0		11	5		13			0	2		29	32
En 6	1	5		2	0		15	35		28			0	2		78	88
En 7	0	0		2	0		3	4		3			4	0		10	16
En 8	1	1		1	0		4	0		10			0	0		14	17
En 9	0	4		1	0		1	26		8			0	8		35	48
En 10	0	0		0	0		0	2		5			0	2		7	9
Mean	**0.2**	**1.6**		**1.4**	**0.0**		**5.3**	**11.8**		**9.6**			**0.4**	**2.2**		**26.7**	**32.3**
p-value	**0.549**	**0.5703**		**0.5906**	**0.3306**		**0.3610**	**0.704**		**0.9134**			**0.358**	**0.265**		**0.9925**	**0.9654**

LISA QA Model 3.1 Error Categories (Minor/Major/Critical)

ISBN: DL:

In order to gain a clearer view of both groups' revision qualities, we looked at the total error points (calculated by LISA QA Model based on the errors entered) and the quality percentages of translators' and engineers' revision (Table 65). According to the mean values, the engineers' total error points is 94 and translators' is 101. On average, the quality percentage of engineers' revision is 81.05 while that of translators' is 79.68. A total of six out of ten engineers passed the quality threshold set for revision (85 percent) while a total of five out of ten translators passed this threshold. Although the mean values indicate that engineer revision is of slightly higher quality than translator revision, the t-test shows no significant difference ($p = 0.8689$ for the total error points and 0.8685 for the quality percentages, Table 65). However, there was a significant difference between both groups' postediting qualities in that the engineers' postediting was of significantly higher quality than translators'. This decrease in the quality difference between translators and engineers in revision can be explained by the fact that the higher quality of engineer postediting affected the quality of translator revision positively while the relatively lower quality of translator postediting affected the quality of engineer revision negatively. This affirms what we have found above: revision quality is not only dependent on the qualifications of the reviser but also on the quality of the text to be revised.

ISBN: DL:
Table 65. Quality in Revision - Translators and Engineers - Recurring Errors Penalized

Subjects	Total Error Points	Revision Quality %	Result
En 1	24	95.34	Pass
En 2	171	66.54	Fail
En 3	140	72.11	Fail
En 4	6	98.78	Pass
En 5	64	86.21	Pass
En 6	256	47.65	Fail
En 7	32	93.59	Pass
En 8	21	96.09	Pass
En 9	200	59.35	Fail
En 10	25	94.80	Pass
Mean	**94**	**81.05**	
Tr 1	32	93.31	Pass
Tr 2	162	67.73	Fail
Tr 3	219	56.55	Fail
Tr 4	2	99.59	Pass
Tr 5	76	83.80	Fail
Tr 6	210	57.75	Fail
Tr 7	40	92.31	Pass
Tr 8	34	93.54	Pass
Tr 9	210	56.16	Fail
Tr 10	21	96.09	Pass
Mean	**101**	**79.68**	
p-value	**0.8689**	**0.8685**	

Cooperation between professional translators and subject-matter experts is expected to bring higher-quality texts. However, as is seen above, not all types of cooperation bring higher quality. The quality of the end-products revised by one group is affected by the strong and weak aspects of the postediting by the other group. However, we can test which cycle brings more added value in terms of quality of the end product: translator postediting + engineer revision or engineer postediting + translator revision. By "added value", we mean the quality increase brought about by revision. Thus, the third and the most important aspect of Part 2 is to determine the effect of the reviser's subject-matter knowledge on the quality of the postedited text. Table 66 compares the total error points and the quality percentages of translator postediting with engineer revision. An overview of the data shows that - except for Engineer 9 – engineer revision increased the quality of translator postediting. With engineer revision, the mean total error points in translator postediting decreased from 187 to 94, which is a significant difference ($p = 0.0205$). The quality is increased by 19.78 percent (from 61.27 to 81.05). Thus, engineer revision significantly increased the

ISBN: DL:

quality of translator postediting (p = 0.0164, Table 66), indicating that the subject-matter knowledge of the reviser is more important in determining the end-product quality than is that of the posteditor.

Table 66. Translator Postediting and Engineer Revision - Quality Compared - Recurring Errors Penalized

Translator Postediting			
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**
Tr 1	55	88.59	Pass
Tr 2	180	62.66	Fail
Tr 3	258	46.47	Fail
Tr 4	80	83.40	Pass
Tr 5	241	50.00	Fail
Tr 6	284	41.08	Fail
Tr 7	197	59.13	Fail
Tr 8	178	63.07	Fail
Tr 9	182	62.24	Fail
Tr 10	212	56.02	Fail
Mean	**187**	**61.27**	
Engineer Revision of Translator Postediting			
Subjects	**Total Error Points**	**Revision Quality %**	**Result**
En 1	24	95.34	Pass
En 2	171	66.54	Fail
En 3	140	72.11	Fail
En 4	6	98.78	Pass
En 5	64	86.21	Pass
En 6	256	47.65	Fail
En 7	32	93.59	Pass
En 8	21	96.09	Pass
En 9	200	59.35	Fail
En 10	25	94.80	Pass
Mean	**94**	**81.05**	
p-value	**0.0205**	**0.0164**	

We also compared the quality of engineer postediting with its revision by translators. An overview of the data shows that - except for Translator 2 – translator revision increased the quality of engineer postediting. With translator revision, the mean total error points in engineer postediting decreased from 118 to 101. There is a total of 17 points decrease, which does not correspond to a significant difference (p = 0.6626, Table 67). The quality percentage is increased by only 4.10 percent (from 75.58 to 79.68). As a result, although the mean values show that translator revision increased the quality of engineer postediting, it is not possible to talk about a significant increase (p = 0.6100, Table 67). This result affirms our finding above that the subject-matter

ISBN: DL:
knowledge of the reviser is more important in determining the end-product quality than that of the posteditor.

Table 67. Engineer Postediting and Translator Revision - Quality Compared - Recurring Errors Penalized

	Engineer Postediting		
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**
En 1	42	91.29	Pass
En 2	152	68.46	Fail
En 3	225	53.32	Fail
En 4	57	88.17	Pass
En 5	80	83.40	Pass
En 6	235	51.24	Fail
En 7	55	88.59	Pass
En 8	65	86.51	Pass
En 9	228	52.70	Fail
En 10	38	92.12	Pass
Mean	**118**	**75.58**	
Translator Revision of Engineer Postediting			
Subjects	**Total Error Points**	**Revision Quality %**	**Result**
Tr 1	32	93.31	Pass
Tr 2	162	67.73	Fail
Tr 3	219	56.55	Fail
Tr 4	2	99.59	Pass
Tr 5	76	83.80	Fail
Tr 6	210	57.75	Fail
Tr 7	40	92.31	Pass
Tr 8	34	93.54	Pass
Tr 9	210	56.16	Fail
Tr 10	21	96.09	Pass
Mean	**101**	**79.68**	
p-value	**0.6626**	**0.6100**	

Table 68 compares the quality difference between one group's postediting and the other group's revision. Most of the data in Table 68 were presented in the previous tables in this section. However, unlike the previous ones, Table 68 compares the quality differences brought about by revision in the two different processes: translator postediting + engineer revision and engineer postediting + translator revision. Engineer revision decreased the total error points of translator postediting by 93 points and it increased the quality of translator postediting by 19.78 percent. On the other hand, translator revision decreased the total error points of engineer postediting by 17 points and it increased the quality of engineer postediting by 4.10 percent. Consequently, when the recurring errors are penalized, the quality improvement brought about by engineer-

ISBN: DL:

revision of translator-postediting is significantly higher than the quality improvement brought about by translator-revision of engineer-postediting ($p = 0.0073$ for the total error points and $p = 0.0061$ for the quality percentage, Table 68).

Table 68. Quality Difference Between Postediting and Its Revision - Recurring Errors Penalized

	Translator Postediting			Engineer Revision of Translator Postediting					
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**	**Subjects**	**Total Error Points**	**Revision Quality %**	**Result**	**Difference btw. REV and PE Error Pts.**	**Difference btw. REV and PE Quality Percentage %**
Tr 1	55	88.59	Pass	**En 1**	24	95.34	Pass	-31	6.75
Tr 2	180	62.66	Fail	**En 2**	171	66.54	Fail	-9	3.88
Tr 3	258	46.47	Fail	**En 3**	140	72.11	Fail	-118	25.64
Tr 4	80	83.40	Pass	**En 4**	6	98.78	Pass	-74	15.38
Tr 5	241	50.00	Fail	**En 5**	64	86.21	Pass	-177	36.21
Tr 6	284	41.08	Fail	**En 6**	256	47.65	Fail	-28	6.57
Tr 7	197	59.13	Fail	**En 7**	32	93.59	Pass	-165	34.46
Tr 8	178	63.07	Fail	**En 8**	21	96.09	Pass	-157	33.02
Tr 9	182	62.24	Fail	**En 9**	200	59.35	Fail	18	-2.89
Tr 10	212	56.02	Fail	**En 10**	25	94.80	Pass	-187	38.78
Mean	**187**	**61.27**		**Mean**	**94**	**81.05**		**-93**	**19.78**
	Engineer Postediting			**Translator Revision of Engineer Postediting**					
En 1	42	91.29	Pass	**Tr 1**	32	93.31	Pass	-10	2.02
En 2	152	68.46	Fail	**Tr 2**	162	67.73	Fail	10	-0.73
En 3	225	53.32	Fail	**Tr 3**	219	56.55	Fail	-6	3.23
En 4	57	88.17	Pass	**Tr 4**	2	99.59	Pass	-55	11.42
En 5	80	83.40	Pass	**Tr 5**	76	83.80	Fail	-4	0.40
En 6	235	51.24	Fail	**Tr 6**	210	57.75	Fail	-25	6.51
En 7	55	88.59	Pass	**Tr 7**	40	92.31	Pass	-15	3.72
En 8	65	86.51	Pass	**Tr 8**	34	93.54	Pass	-31	7.03
En 9	228	52.70	Fail	**Tr 9**	210	56.16	Fail	-18	3.46
En 10	38	92.12	Pass	**Tr 10**	21	96.09	Pass	-17	3.97
Mean	**118**	**75.58**		**Mean**	**101**	**79.68**		**-17**	**4.10**
p-value	**0.0636**	**0.0636**		**p-value**	**0.8689**	**0.8685**		**0.0073**	**0.0061**

ISBN: DL:

In order to find out whether the participants' perception of the processes corresponds to what really happens in revision, we asked two questions in the post-assignment questionnaire that the participants filled out after the postediting task. (Since our questions on revision were not task-specific, we preferred to ask them in the post-assignment questionnaire given to the participants after the postediting task rather than giving them a separate questionnaire after the revision task.) The first question was: "Which process do you think yields better results in terms of the quality of the final output: translator translation/postediting + engineer revision or engineer translation/postediting + translator revision?" A total of seven out of ten translators (Translators 1, 2, 3, 4, 7, 9 and 10) chose the former, while three out of ten (Translators 5, 6 and 8) chose the latter. Second, we asked the translators to assess the same two combinations in terms of the speed of the postediting process. A total of nine out of ten (except for Translator 5) stated that the postediting process would be faster in translator translation/postediting + engineer revision. This means that only Translator 5 thinks that engineer postediting + translator revision produce higher-quality texts with higher speed. Translators 6 and 8 chose this combination for producing higher-quality products but thought that this process is not as fast as translator postediting + engineer revision.

As for the engineers, a total of seven out of ten (Engineers 1, 2, 3, 4, 6, 7 and 8) stated that translator translation/postediting + engineer revision would result in higher-quality product, while two out of ten (Engineers 9 and 10) stated that engineer translation/postediting + translator revision would bring higher-quality texts. Engineer 9 added that "I have to admit that translation is very tough work. The translator has to be very proficient in the mother tongue. If the translators hold the expertise in the field they translate, their translations will involve fewer errors." Engineer 5 proposed that if translators and engineers work together as a team, the process might be slower, but the translation would be more accurate.

When we asked the engineers to assess the same two combinations in terms of the speed of the postediting process, the same subjects (Engineers 1, 2, 3, 4, 6, 7 and 8) stated that the postediting process would be faster with translator translation/postediting + engineer revision. Engineer 5 kept on saying if translators and engineers worked together as a team, the process might be slower, but the translation would be more accurate. Engineer 9 replied that he did not know the result and added: "It is hard to decide. One group is stranger to the terminology and the other is to the translation profession. The translator must be an expert in his/her field." Engineer 10 stated that

ISBN: DL:
engineer translation/postediting + translator revision would be faster. As is the case in translators, only one of the engineers (Engineer 10) think that engineer translation/postediting + translator revision would produce higher-quality texts with higher speed.

All in all, a total of 70 percent of the participants in each group agreed that translator translation/postediting + engineer revision would yield higher-quality output than vice-versa. The quantitative data do not affirm this opinion. The mean quality percentage of the engineers' revision (81.05) is higher than that of the translators' (79.68), yet the difference is not significant ($p = 0.8685$, Table 65). However, when we look at the added value brought about by revision with respect to quality, the present study indicates that - for technical texts - the quality increase brought about by engineer-revision of translator-postediting is significantly higher than the quality increase brought about by translator-revision of engineer-postediting. ($p = 0.0061$, Table 68).

As for speed, 90 percent of the translators and 70 percent of the engineers think that translator postediting + engineer revision is faster than vice-versa. However, our data do not indicate any superiority of translator postediting + engineer revision over engineer postediting + translator revision with regard to speed ($p = 0.1611$, Table 61).

9.2.2.2. Quality in Revision - Recurring Errors Unpenalized

We repeated the analysis of revision quality by unpenalizing the recurring errors to see if the results change under this condition. We first measured whether the same participant performed differently when postediting and revising the same text. Table 69 presents the quality percentages of each engineer's postediting and revision when the recurrent errors are not penalized. When they did the revision, all of the engineers, except 3 and 6, produced higher-quality texts than when they postedited. The mean quality percentage of engineer postediting is 90.73, while the mean quality percentage of engineer revision is 93.55. The mean difference between the quality percentage of engineer postediting and engineer revision is 2.82. Although the mean numbers indicate an increase in the quality of engineers' revision with respect to their postediting, there is no significant difference between the two ($p = 0.1487$, Table 69).

ISBN: DL:

Table 69. Engineer Postediting and Engineer Revision - Quality Compared - Recurring Errors Unpenalized

Subjects	Postediting Quality %	Revision Quality %	Difference btw REV. and PE Quality %
En 1	92.95	96.70	3.75
En 2	91.70	96.09	4.39
En 3	88.80	86.45	-2.35
En 4	93.36	98.78	5.42
En 5	86.31	91.59	5.28
En 6	85.27	84.25	-1.02
En 7	93.15	96.39	3.24
En 8	93.98	96.46	2.48
En 9	87.97	91.87	3.90
En 10	93.78	96.88	3.10
Mean	**90.73**	**93.55**	**2.82**
P-value of the difference between Engineer PE and REV			**0.1487**

Table 70 presents the quality percentages of each translator's postediting and revision under the recurrent errors unpenalized condition. When they did revision, all of the translators produced higher-quality texts than when they postedited. When we compare the mean quality percentages of translator postediting (89.30) and translator revision (93.76), we find that there is a significant difference (p = 0.0390) between the end-product quality of the two processes.

Table 70. Translator Postediting and Translator Revision - Quality Compared - Recurring Errors Unpenalized

Subjects	Postediting Quality %	Revision Quality %	Difference btw REV. and PE Quality %
Tr 1	95.23	95.61	0.38
Tr 2	90.66	93.63	2.97
Tr 3	82.37	90.87	8.50
Tr 4	93.36	99.59	6.23
Tr 5	83.20	89.77	6.57
Tr 6	81.12	89.34	8.22
Tr 7	91.91	97.31	5.40
Tr 8	92.74	95.82	3.08
Tr 9	89.00	89.56	0.56
Tr 10	93.36	96.09	2.73
Mean	**89.30**	**93.76**	**4.46**
P-value of the difference between Translator PE and REV			**0.0390**

The data show that difference between the engineers' postediting and revision is insignificant both when the recurrent errors are penalized (p = 0.5004, Table 62) and

ISBN: DL:
unpenalized (p = 0.1487, Table 69), while the difference between the translators' postediting and revision is significant under the penalized (p = 0.0232, Table 63) and unpenalized (p = 0.0390, Table 70) conditions. So we can conclude that penalizing and unpenalizing the recurrent errors does not change the significance of the difference between each group's postediting and revision qualities. However, when the recurrent errors are not penalized, the statistical significance of the difference between the engineers' postediting and revision increases, whereas that of the difference between the translators' postediting and revision slightly decreases.

As is the case in the penalized condition, the greater difference between the quality of translators' posteditings and revisions (p = 0.0390, Table 70) as compared to the quality difference between the engineers' posteditings and revisions (p = 0.1487, Table 69) under the unpenalized condition can be explained by the effect of the postediting quality on the revision quality. In other words, the significant quality difference between the translators' posteditings (whose quality is lower than engineers' posteditings) and revisions is due to the higher quality of the posteditings the translators revised, whereas the insignificant quality difference between the engineers' posteditings (whose quality is higher than translators' posteditings) and revisions is due to the lower-quality posteditings the engineers revised. Thus, the results under the unpenalized condition affirm our finding under the penalized condition that the higher quality of the engineers' posteditings significantly increased the quality of the translators' revisions, whereas the lower quality of the translators' posteditings had a negative effect on the quality of the engineers' revisions. This affirms our conclusion that revision quality is not only dependent on the qualifications of the reviser but also on the quality of the text to be revised.

Second, we compared the revision processes performed by translators and engineers. Table 71 shows the distribution of errors in the texts revised by translators and engineers when the recurrent errors are unpenalized. Similar to the penalized condition, none of the error categories under the unpenalized condition yielded significant difference between the translators and engineers with respect to revising one another's postediting (Table 71).

ISBN: DL:

Table 71. Distribution of Errors - Revision - Recurring Errors Unpenalized

	Translator Revision of Engineer Postediting																
Subjects	**LISA QA Model 3.1 Error Categories (Minor/Major/Critical)**																
	Mistranslation			**Accuracy**			**Terminology**			**Language**			**Consistency**			**Number of Language and Terminology Errors**	**Total Number of Errors**
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**		
Tr 1	0	1		2	0		2	2		1			1	0		5	9
Tr 2	1	0		2	0		0	3		14			0	0		17	20
Tr 3	2	2		1	0		2	3		16			0	0		21	26
Tr 4	0	0		0	0		0	0		2			0	0		2	2
Tr 5	0	2		1	1		2	2		10			0	2		14	20
Tr 6	1	2		2	0		1	3		24			0	0		28	33
Tr 7	0	1		1	0		0	1		2			1	0		3	6
Tr 8	0	1		1	0		2	2		4			0	0		8	10
Tr 9	0	3		1	0		2	4		6			1	1		12	18
Tr 10	0	1		1	0		1	1		4			0	1		6	9
Mean	**0.4**	**1.3**		**1.2**	**0.1**		**1.2**	**2.1**		**8.3**			**0.3**	**0.4**		**11.6**	**15.3**
	Engineer Revision of Translator Postediting																
En 1	0	1		1	0		3	1		3			0	0		7	9
En 2	0	0		2	0		2	2		6			0	0		10	12
En 3	0	3		3	0		2	4		13			0	3		19	28
En 4	0	0		2	0		0	0		4			0	0		4	6
En 5	0	1		0	0		1	3		13			0	1		17	19
En 6	1	3		2	0		3	4		26			0	2		33	41
En 7	0	0		2	0		1	2		3			2	0		6	10
En 8	1	1		1	0		2	0		10			0	0		12	15
En 9	0	3		1	0		1	2		8			0	1		11	16
En 10	0	0		0	0		0	1		5			0	1		6	7
Mean	**0.2**	**1.2**		**1.4**	**0**		**1.5**	**1.9**		**9.1**			**0.2**	**0.8**		**12.5**	**16.3**
p-value	**0.549**	**0.8477**		**0.5906**	**0.317**		**0.5120**	**0.7404**		**0.8094**			**0.357**	**0.343**		**0.8172**	**0.8306**

ISBN: DL:

In order to gain a clearer view of both groups' revision qualities, we looked at the total error points (calculated by the LISA QA Model based on the errors entered) and the quality percentages of translators' and engineers' revision (Table 72). According to the mean values, the engineers' total error points is 32 and translators' is 31. On average, the quality percentage of engineers' revision is 93.55, while that of translators' is 93.76. All of the translators passed the quality threshold set for revision (85 percent), while a total of nine out of ten engineers passed this threshold. The engineer who failed had a quality percentage of 84.25, which is very close to the threshold (Table 72). Under the penalized condition, however, a total of five out of ten translators and six out of ten engineers passed the quality test (p = 0.8689 for the total error points and p = 0.8685 for the quality percentage under the penalized condition). Nevertheless, as is the case under the penalized condition, the t-test shows no significant difference between the translators' and engineers' revision qualities under the unpenalized condition (p = 0.9162 for the total error points and 0.9134 for the quality percentages, Table 72).

Table 72. Quality in Revision - Translators and Engineers - Recurring Errors Unpenalized

	Recurring Errors Unpenalized		
Subjects	**Total Error Points**	**Revision Quality %**	**Result**
En 1	17	96.70	Pass
En 2	20	96.09	Pass
En 3	68	86.45	Pass
En 4	6	98.78	Pass
En 5	39	91.59	Pass
En 6	77	84.25	Fail
En 7	18	96.39	Pass
En 8	19	96.46	Pass
En 9	40	91.87	Pass
En 10	15	96.88	Pass
Mean	**32**	**93.55**	
Tr 1	21	95.61	Pass
Tr 2	32	93.63	Pass
Tr 3	46	90.87	Pass
Tr 4	2	99.59	Pass
Tr 5	48	89.77	Pass
Tr 6	53	89.34	Pass
Tr 7	14	97.31	Pass
Tr 8	22	95.82	Pass
Tr 9	50	89.56	Pass
Tr 10	21	96.09	Pass
Mean	**31**	**93.76**	
p-value	**0.9162**	**0.9134**	

ISBN: DL:

We also compared the end-products of the two cycles (translator postediting + engineer revision or engineer postediting + translator revision) in terms of the added value under the unpenalized condition. By "added value" we mean the quality increase brought about by revision. By means of this analysis, we aim to determine the difference between postediting and revision qualities under the unpenalized condition, which will then be compared with the penalized condition.

Table 73 compares the total error points and the quality percentages of translator postediting with engineer revision. An overview of the data shows that engineer revision increased the quality of translator postediting. With engineer revision, the mean total error points in translator postediting decreased from 52 to 32, however the difference is not very significant ($p = 0.0880$). The quality percentage is increased by 4.25 percent (from 89.30 to 93.55), which is not very significant either ($p = 0.0751$). Thus, unlike the penalized condition where the p-value for the difference between the quality percentages of translator postediting and engineer revision is 0.0164 (Table 66), engineer revision did not significantly increase the quality of translator postediting under the recurrent errors unpenalized condition ($p = 0.0751$, Table 73). On the other hand, since it is close to the threshold 0.05, a p-value of 0.0751 does not represent a very low level of significance. We can explain the reason of this difference between the penalized and the unpenalized conditions as follows: The engineers make fewer recurrent errors than translators, thus, the increase in the engineers' quality score brought about by unpenalizing the recurrent errors is low. On the other hand, the translators make more recurrent errors than engineers, thus, the increase in the translators' quality score brought about by unpenalizing the recurrent errors is high. Consequently, under the unpenalized condition, the difference between engineer revision and translator postediting diminishes as compared to the penalized condition. All in all, it is important to note that the level of significance of the difference between translator postediting and engineer revision decreases when we do not penalize the recurrent errors.

ISBN: DL:

Table 73. Translator Postediting and Engineer Revision - Quality Compared - Recurring Errors Unpenalized

Translator Postediting			
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**
Tr 1	23	95.23	Pass
Tr 2	45	90.66	Pass
Tr 3	85	82.37	Pass
Tr 4	32	93.36	Pass
Tr 5	81	83.20	Pass
Tr 6	91	81.12	Pass
Tr 7	39	91.91	Pass
Tr 8	35	92.74	Pass
Tr 9	53	89.00	Pass
Tr 10	32	93.36	Pass
Mean	**52**	**89.30**	
Engineer Revision of Translator Postediting			
Subjects	**Total Error Points**	**Revision Quality %**	**Result**
En 1	17	96.70	Pass
En 2	20	96.09	Pass
En 3	68	86.45	Pass
En 4	6	98.78	Pass
En 5	39	91.59	Pass
En 6	77	84.25	Fail
En 7	18	96.39	Pass
En 8	19	96.46	Pass
En 9	40	91.87	Pass
En 10	15	96.88	Pass
Mean	**32**	**93.55**	
p-value	**0.0880**	**0.0751**	

As for the comparison of the quality of engineer postediting with its revision by translators, our data show that, under the unpenalized condition, translator revision increased the quality of engineer postediting, too. With translator revision, the mean total error points in engineer postediting decreased from 45 to 31. There is a total of 14 points decrease, which does not correspond to a significant difference (p = 0.0827, Table 74). The quality percentage is increased by 3.03 percent (from 90.73 to 93.76) which is not very significant, either (p = 0.0685). However, this is a rather significant figure, at least as compared to the penalized condition where the p-value of the difference between the quality percentages of engineer postediting and translator revision is 0.6100 (Table 67). As stated above, we can explain the reason of this difference between the penalized and the unpenalized conditions as follows: The

ISBN: DL:

translators make more recurrent errors than engineers, thus, the increase in the translators' quality score brought about by unpenalizing the recurrent errors is high. On the other hand, the engineers make fewer recurrent errors than translators, thus, the increase in the engineers' quality score brought about by unpenalizing the recurrent errors is low. Consequently, under the unpenalized condition, the difference between translator revision and engineer postediting increases as compared to the penalized condition. All in all, it is important to note that the level of significance of the difference between engineer postediting and translator revision increases when we do not penalize the recurrent errors.

Table 74. Engineer Postediting and Translator Revision - Quality Compared - Recurring Errors Unpenalized

Engineer Postediting			
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**
En 1	34	92.95	Pass
En 2	40	91.70	Pass
En 3	54	88.80	Pass
En 4	32	93.36	Pass
En 5	66	86.31	Pass
En 6	71	85.27	Pass
En 7	33	93.15	Pass
En 8	29	93.98	Pass
En 9	58	87.97	Pass
En 10	30	93.78	Pass
Mean	**45**	**90.73**	
Translator Revision of Engineer Postediting			
Subjects	**Total Error Points**	**Revision Quality %**	**Result**
Tr 1	21	95.61	Pass
Tr 2	32	93.63	Pass
Tr 3	46	90.87	Pass
Tr 4	2	99.59	Pass
Tr 5	48	89.77	Pass
Tr 6	53	89.34	Pass
Tr 7	14	97.31	Pass
Tr 8	22	95.82	Pass
Tr 9	50	89.56	Pass
Tr 10	21	96.09	Pass
Mean	**31**	**93.76**	
p-value	**0.0827**	**0.0685**	

Table 75 compares the quality difference between one group's postediting and

ISBN: DL:

the other group's revision under the recurrent errors unpenalized condition. Most of the data in Table 75 were presented in the previous tables in this section. However, unlike the previous tables, Table 75 compares the quality differences brought about by revision in two different cycles: translator postediting + engineer revision and engineer postediting + translator revision. In translator postediting + engineer revision, on average, engineer revision decreased the total error points of translator postediting by 20 points and it increased the quality of translator postediting by 4.25 percent (Table 75). Under the recurring errors penalized condition, these figures were 93 and 19.78, respectively (Table 68). On the other hand, in engineer postediting + translator revision, translator revision decreased the total error points of engineer postediting by 14 points and it increased the quality of engineer postediting by 3.03 percent (Table 75). Under the recurring errors penalized condition, these figures were 17 and 4.10 respectively (Table 68).

Under the unpenalized condition, there is no significant difference between the quality improvement brought about by engineer-revision of translator-postediting and translator revision of engineer postediting (the p-value for the total error points is 0.1458 and for the quality percentage 0.1219, Table 75). However, under the penalized condition, the quality improvement brought about by engineer-revision of translator-postediting was significantly higher than the quality improvement brought about by translator-revision of engineer-postediting ($p = 0.0073$ for the total error points and $p = 0.0061$ for the quality percentage, Table 68).

ISBN: DL:

Table 75. Quality Difference Between Postediting and Its Revision - Recurring Errors Unpenalized

	Translator Postediting				**Engineer Revision of Translator Postediting**				
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**	**Subjects**	**Total Error Points**	**Revision Quality %**	**Result**	**Difference btw. REV and PE Error Pts.**	**Difference btw. REV and PE Quality Percentage %**
Tr 1	23	95.23	Pass	**En 1**	17	96.70	Pass	-6	1.47
Tr 2	45	90.66	Pass	**En 2**	20	96.09	Pass	-25	5.43
Tr 3	85	82.37	Pass	**En 3**	68	86.45	Pass	-17	4.08
Tr 4	32	93.36	Pass	**En 4**	6	98.78	Pass	-26	5.42
Tr 5	81	83.20	Pass	**En 5**	39	91.59	Pass	-42	8.39
Tr 6	91	81.12	Pass	**En 6**	77	84.25	Fail	-14	3.13
Tr 7	39	91.91	Pass	**En 7**	18	96.39	Pass	-21	4.48
Tr 8	35	92.74	Pass	**En 8**	19	96.46	Pass	-16	3.72
Tr 9	53	89.00	Pass	**En 9**	40	91.87	Pass	-13	2.87
Tr 10	32	93.36	Pass	**En 10**	15	96.88	Pass	-17	3.52
Mean	**52**	**89.30**	Pass	**Mean**	**32**	**93.55**		**-20**	**4.25**
	Engineer Postediting				**Translator Revision of Engineer Postediting**				
En 1	34	92.95	Pass	**Tr 1**	21	95.61	Pass	-13	2.66
En 2	40	91.70	Pass	**Tr 2**	32	93.63	Pass	-8	1.93
En 3	54	88.80	Pass	**Tr 3**	46	90.87	Pass	-8	2.07
En 4	32	93.36	Pass	**Tr 4**	2	99.59	Pass	-30	6.23
En 5	66	86.31	Pass	**Tr 5**	48	89.77	Pass	-18	3.46
En 6	71	85.27	Pass	**Tr 6**	53	89.34	Pass	-18	4.07
En 7	33	93.15	Pass	**Tr 7**	14	97.31	Pass	-19	4.16
En 8	29	93.98	Pass	**Tr 8**	22	95.82	Pass	-7	1.84
En 9	58	87.97	Pass	**Tr 9**	50	89.56	Pass	-8	1.59
En 10	30	93.78	Pass	**Tr 10**	21	96.09	Pass	-9	2.31
Mean	**45**	**90.73**		**Mean**	**31**	**93.76**		**-14**	**3.03**
p-value	**0.4711**	**0.4708**		**p-value**	**0.9162**	**0.9134**		**0.1458**	**0.1219**

ISBN: DL:

Since we aim to compare what happens under the recurrent errors penalized and unpenalized conditions, we made a series of comparison between the two within the groups. First, we compared the number of errors. Table 76 compares the translators' and Table 77 compares the engineers' errors in revision under the penalized and unpenalized conditions.

The majority of the errors are in terminology and language categories. Under the unpenalized condition, on average, the translators made 1.2 minor and 2.1 major terminology errors, unlike the penalized condition where they made, on average, 3.2 minor and 14.4 major terminology errors (Table 76). The unpaired t-test shows that the translators made significantly fewer major terminology errors under the unpenalized condition than under the penalized condition ($p = 0.0236$, Table 76).

Due to their nonrecurrent nature, the number of language errors did not change considerably when the recurrent errors are unpenalized (8.3 under the unpenalized condition as compared to 9.2 under the penalized condition, Table 76). For translators, the mean total number of errors, which was 31.8 under the penalized condition, decreased by almost 50 percent and became 15.3 (Table 76) when the recurrent errors are unpenalized. Although the difference between the penalized and the unpenalized conditions is not very significant ($p = 0.0738$, Table 76), a 50 percent decrease in the mean total number of errors cannot be underestimated.

ISBN: DL:

Table 76. Distribution of Errors - Translator Revision of Engineer Postediting

	Recurring Errors Penalized																
Subjects	**LISA QA Model 3.1 Error Categories (Minor/Major/Critical)**																
	Mistranslation			**Accuracy**			**Terminology**			**Language**			**Consistency**			**Language and Terminology Errors**	**Total Number of Errors**
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**		
Tr 1	0	1		2	0		6	3		1			3	0		10	16
Tr 2	1	0		2	0		0	29		14			0	0		43	46
Tr 3	3	2		1	0		16	34		19			0	0		69	75
Tr 4	0	0		0	0		0	0		2			0	0		2	2
Tr 5	0	5		1	1		2	3		13			0	3		18	28
Tr 6	1	2		2	0		1	34		26			0	0		61	66
Tr 7	0	4		1	0		0	3		3			1	0		6	12
Tr 8	0	1		1	0		4	4		4			0	0		12	14
Tr 9	0	5		1	0		2	33		6			1	2		41	50
Tr 10	0	1		1	0		1	1		4			0	1		6	9
Mean	**0.5**	**2.1**		**1.2**	**0.1**		**3.2**	**14.4**		**9.2**			**0.5**	**0.6**		**26.8**	**31.8**
	Recurring Errors Unpenalized																
Tr 1	0	1		2	0		2	2		1			1	0		5	9
Tr 2	1	0		2	0		0	3		14			0	0		17	20
Tr 3	2	2		1	0		2	3		16			0	0		21	26
Tr 4	0	0		0	0		0	0		2			0	0		2	2
Tr 5	0	2		1	1		2	2		10			0	2		14	20
Tr 6	1	2		2	0		1	3		24			0	0		28	33
Tr 7	0	1		1	0		0	1		2			1	0		3	6
Tr 8	0	1		1	0		2	2		4			0	0		8	10
Tr 9	0	3		1	0		2	4		6			1	1		12	18
Tr 10	0	1		1	0		1	1		4			0	1		6	9
Mean	**0.4**	**1.3**		**1.2**	**0.1**		**1.2**	**2.1**		**8.3**			**0.3**	**0.4**		**11.6**	**15.3**
p-value	**0.963**	**0.2513**		**1**	**1**		**0.2203**	**0.0236**		**0.8044**			**0.888**	**0.852**		**0.0818**	**0.0738**

ISBN: DL:

As for the engineers, the mean number of minor terminology errors decreased from 5.3 to 1.5 and that of the major terminology errors decreased from 11.8 to 1.9 (Table 77) when we do not penalize the recurrent errors. This corresponds to a significant difference (p = 0.0339 and 0.0384 respectively, Table 77).

Due to their nonrecurrent nature, the mean number of engineers' language errors did not change considerably when the recurrent errors are not penalized (9.1 under the unpenalized condition as compared to 9.6 under the penalized condition, Table 77). Similar to what we found with the translators, the mean total number of errors for engineers decreased by almost 50 percent and became 16.3 (Table 77) when the recurrent errors are unpenalized. Although the difference is not very significant (p = 0.0816, Table 77), a 50 percent decrease in the mean total number of errors is worth considering.

Consequently, the number of terminology errors for both groups (for translators only major terminology errors and for engineers both minor and major terminology errors) decreased significantly under the recurrent errors unpenalized condition as compared to the penalized condition (Table 76 and Table 77). For the rest of the error categories, there is no significant difference between the penalized and unpenalized conditions with respect to the translators' and engineers' distribution of errors when revising each other's posteditings.

ISBN: DL:

Table 77. Distribution of Errors - Engineer Revision of Translator Postediting

Recurring Errors Penalized																		
Subjects	**LISA QA Model 3.1 Error Categories (Minor/Major/Critical)**																	
	Mistranslation			**Accuracy**			**Terminology**			**Language**			**Consistency**			**Language and Terminology Errors**	**Total Number of Errors**
	Min.	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**	**Min.**	**Maj.**	**Cri.**		
En 1	0	1		1	0		10	1		3			0	0		14	16
En 2	0	0		2	0		3	32		6			0	0		41	43
En 3	0	4		3	0		6	13		16			0	6		35	48
En 4	0	0		2	0		0	0		4			0	0		4	6
En 5	0	1		0	0		11	5		13			0	2		29	32
En 6	1	5		2	0		15	35		28			0	2		78	88
En 7	0	0		2	0		3	4		3			4	0		10	16
En 8	1	1		1	0		4	0		10			0	0		14	17
En 9	0	4		1	0		1	26		8			0	8		35	48
En 10	0	0		0	0		0	2		5			0	2		7	9
Mean	**0.2**	**1.6**		**1.4**	**0**		**5.3**	**11.8**		**9.6**			**0.4**	**2.2**		**26.7**	**32.3**
Recurring Errors Unpenalized																		
En 1	0	1		1	0		3	1		3			0	0		7	9
En 2	0	0		2	0		2	2		6			0	0		10	12
En 3	0	3		3	0		2	4		13			0	3		19	28
En 4	0	0		2	0		0	0		4			0	0		4	6
En 5	0	1		0	0		1	3		13			0	1		17	19
En 6	1	3		2	0		3	4		26			0	2		33	41
En 7	0	0		2	0		1	2		3			2	0		6	10
En 8	1	1		1	0		2	0		10			0	0		12	15
En 9	0	3		1	0		1	2		8			0	1		11	16
En 10	0	0		0	0		0	1		5			0	1		6	7
Mean	**0.2**	**1.2**		**1.4**	**0**		**1.5**	**1.9**		**9.1**			**0.2**	**0.8**		**12.5**	**16.3**
p-value	**1**	**0.5981**		**1**	**1**		**0.0339**	**0.0384**		**0.8819**			**0.942**	**0.2237**		**0.0771**	**0.0816**

ISBN: DL:

In order to gain a clearer view of how each group's data compare under the penalized and the unpenalized conditions, we compared the total error points and the quality percentages of each group under both conditions. Table 78 compares the data on the translators' revision of engineer postediting under the penalized and unpenalized conditions. It is interesting to note that there is a significant difference between the penalized and the unpenalized conditions with regard to the total error points ($p = 0.0257$) and the quality percentage ($p = 0.0265$, Table 78). However, both conditions do not differ significantly as far as the added value brought about by translator-revision of engineer-postediting is concerned ($p = 0.5909$ for the difference between revision and postediting total error points and 0.3841 for the difference between revision and postediting quality percentages, Table 78).

ISBN: DL:

Table 78. Translators' Revision Quality - Penalized and Unpenalized Conditions

Recurring Errors Penalized									
Engineer Postediting				**Translator Revision of Engineer Postediting**					
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**	**Subjects**	**Total Error Points**	**Revision Quality %**	**Result**	**Difference btw. REV and PE Error Pts.**	**Difference btw. REV and PE Quality Percentage %**
En 1	42	91.29	Pass	**Tr 1**	32	93.31	Pass	-10	2.02
En 2	152	68.46	Fail	**Tr 2**	162	67.73	Fail	10	-0.73
En 3	225	53.32	Fail	**Tr 3**	219	56.55	Fail	-6	3.23
En 4	57	88.17	Pass	**Tr 4**	2	99.59	Pass	-55	11.42
En 5	80	83.40	Pass	**Tr 5**	76	83.80	Fail	-4	0.40
En 6	235	51.24	Fail	**Tr 6**	210	57.75	Fail	-25	6.51
En 7	55	88.59	Pass	**Tr 7**	40	92.31	Pass	-15	3.72
En 8	65	86.51	Pass	**Tr 8**	34	93.54	Pass	-31	7.03
En 9	228	52.70	Fail	**Tr 9**	210	56.16	Fail	-18	3.46
En 10	38	92.12	Pass	**Tr 10**	21	96.09	Pass	-17	3.97
Mean	**118**	**75.58**		**Mean**	**101**	**79.68**		**-17**	**4.10**
Recurring Errors Unpenalized									
Engineer Postediting				**Translator Revision of Engineer Postediting**					
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**	**Subjects**	**Total Error Points**	**Revision Quality %**	**Result**	**Difference btw. REV and PE Error Pts.**	**Difference btw. REV and PE Quality Percentage %**
En 1	34	92.95	Pass	**Tr 1**	21	95.61	Pass	-13	2.66
En 2	40	91.70	Pass	**Tr 2**	32	93.63	Pass	-8	1.93
En 3	54	88.80	Pass	**Tr 3**	46	90.87	Pass	-8	2.07
En 4	32	93.36	Pass	**Tr 4**	2	99.59	Pass	-30	6.23
En 5	66	86.31	Pass	**Tr 5**	48	89.77	Pass	-18	3.46
En 6	71	85.27	Pass	**Tr 6**	53	89.34	Pass	-18	4.07
En 7	33	93.15	Pass	**Tr 7**	14	97.31	Pass	-19	4.16
En 8	29	93.98	Pass	**Tr 8**	22	95.82	Pass	-7	1.84
En 9	58	87.97	Pass	**Tr 9**	50	89.56	Pass	-8	1.59
En 10	30	93.78	Pass	**Tr 10**	21	96.09	Pass	-9	2.31
Mean	**45**	**90.73**		**Mean**	**31**	**93.76**		**-14**	**3.03**
p-value	**0.0140**	**0.0140**			**0.0257**	**0.0265**		**0.5909**	**0.3841**

ISBN: DL:

Table 79 compares the data on the engineers' revision of translator postediting under the penalized and unpenalized conditions. As is the translators' revision quality, engineers' revision quality differs significantly under the penalized and the unpenalized conditions (p = 0.0497 for the total error points and p = 0.0508 for the quality percentage, Table 79). Unlike the translators' revision, engineers' revision quality also differ significantly under the penalized and the unpenalized conditions with regard to the added value brought about by engineer-revision of translator-postediting (p = 0.0082 for the difference between revision and postediting total error points and 0.0058 for the difference between revision and postediting quality percentages, Table 79).

Consequently, our research findings indicate that penalizing and unpenalizing recurrent errors affects the revision quality significantly.

ISBN: DL:

Table 79. Engineers' Revision Quality - Penalized and Unpenalized Conditions

Recurring Errors Penalized									
Translator Postediting				Engineer Revision of Translator Postediting					
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**	**Subjects**	**Total Error Points**	**Revision Quality %**	**Result**	**Difference btw. REV and PE Error Pts.**	**Difference btw. REV and PE Quality Percentage %**
Tr 1	55	88.59	Pass	**En 1**	24	95.34	Pass	-31	6.75
Tr 2	180	62.66	Fail	**En 2**	171	66.54	Fail	-9	3.88
Tr 3	258	46.47	Fail	**En 3**	140	72.11	Fail	-118	25.64
Tr 4	80	83.40	Pass	**En 4**	6	98.78	Pass	-74	15.38
Tr 5	241	50.00	Fail	**En 5**	64	86.21	Pass	-177	36.21
Tr 6	284	41.08	Fail	**En 6**	256	47.65	Fail	-28	6.57
Tr 7	197	59.13	Fail	**En 7**	32	93.59	Pass	-165	34.46
Tr 8	178	63.07	Fail	**En 8**	21	96.09	Pass	-157	33.02
Tr 9	182	62.24	Fail	**En 9**	200	59.35	Fail	18	-2.89
Tr 10	212	56.02	Fail	**En 10**	25	94.80	Pass	-187	38.78
Mean	**187**	**61.27**		**Mean**	**94**	**81.05**		**-93**	**19.78**
Recurring Errors Unpenalized									
Translator Postediting				**Engineer Revision of Translator Postediting**					
Subjects	**Total Error Points**	**Postediting Quality %**	**Result**	**Subjects**	**Total Error Points**	**Revision Quality %**	**Result**	**Difference btw. REV and PE Error Pts.**	**Difference btw. REV and PE Quality Percentage %**
Tr 1	23	95.23	Pass	**En 1**	17	96.70	Pass	-6	1.47
Tr 2	45	90.66	Pass	**En 2**	20	96.09	Pass	-25	5.43
Tr 3	85	82.37	Pass	**En 3**	68	86.45	Pass	-17	4.08
Tr 4	32	93.36	Pass	**En 4**	6	98.78	Pass	-26	5.42
Tr 5	81	83.20	Pass	**En 5**	39	91.59	Pass	-42	8.39
Tr 6	91	81.12	Pass	**En 6**	77	84.25	Fail	-14	3.13
Tr 7	39	91.91	Pass	**En 7**	18	96.39	Pass	-21	4.48
Tr 8	35	92.74	Pass	**En 8**	19	96.46	Pass	-16	3.72
Tr 9	53	89.00	Pass	**En 9**	40	91.87	Pass	-13	2.87
Tr 10	32	93.36	Pass	**En 10**	15	96.88	Pass	-17	3.52
Mean	**52**	**89.30**	Pass	**Mean**	**32**	**93.55**		**-20**	**4.25**
p-value	**0.0001**	**0.0001**			**0.0497**	**0.0508**		**0.0082**	**0.0058**

ISBN: DL:

10. Conclusions of the Main Study

This research is composed of two parts. Below, we present the final conclusion for each part under the relevant headings.

10.1. Conclusions For Part 1 (Postediting)

In Part 1 we compared the ways engineers and professional translators postedit a technical text, looking at processing speed, time spent on documentation and the number of changes. We also compared the texts postedited by engineers and professional translators, using LISA QA Model 3.1. To this end, we tested the following hypotheses:

For technical texts:

1. Subject-matter experts and professional translators postedit MT output differently in terms of processing speed, time spent on documentation and number of changes.
    a. Professional translators complete the postediting task faster than subject-matter experts do.
    b. Professional translators spend more time on documentation than subject-matter experts do.
    c. Professional translators perform more changes than subject-matter experts do.
2. The quality of postediting by subject-matter experts is higher than postediting by professional translators.

Hypotheses 1a and 1c were not affirmed by our data, while hypothesis 1b was affirmed. Hypothesis 2 was affirmed when we removed the outliers from our data. The following conclusions have thus been derived for postediting technical texts in Part 1:

1. The quantitative data did not reveal a significant difference between translators and engineers with regard to total task time, words processed per minute, number of changes and the time spent on self-revision (Table 31 and Table 32).
2. When evaluating the quality of a postediting (with LISA QA Model 3.1), there is a significant difference between the cases when the recurring errors are

ISBN: DL:

penalized and unpenalized (p = 0.0140 for engineers, Table 40 and 0.0001 for translators, Table 41).

3. When the recurrent errors are penalized and the outlier participants were removed, the engineers' postediting of technical texts is of significantly higher quality than the translators' (p = 0.0339) (Table 47).
4. When the recurrent errors are unpenalized, there is no significant difference between the quality of the translators' and engineers' postediting of a technical text, both when the outliers are included and when they are excluded (p = 0.4708 and 0.7331 respectively) (Table 46).
5. Expertise and experience in the subject-matter is more important for the quality of the postediting than is training in Translation (Table 80).
6. There is a strong correlation between postediting quality and experience in the translation profession (r = 0.50, Table 48). Moreover, for translators, professional experience in the subject-matter (engineering/technical) and professional experience in translation are more important for the quality of the postediting than is training in Translation (Table 48).
7. There is a correlation between the translators' perception of the understandability of the MT output and the quality of the postedited text (Table 48).
8. The correlation between the engineers' level of professional experience in engineering and their postediting quality is very strong (r = 0.87, Table 49). For engineers, experience in their profession as well as the experience with the translating of technical texts as part of their work are important factors affecting postediting quality (Table 49).
9. For engineers, there is an inverse relationship between the quality and the perceived effort spent on the postediting task as compared to translation (Table 49).
10. When postediting technical texts, translators make more terminology errors than engineers (Table 44 and Table 45).
11. There is no significant difference between the translators and engineers with regard to postediting speed, no matter whether we include or exclude the outliers. However, as compared to the translators, the engineers produce higher-quality posteditings with higher speed (Table 55 and Table 56).
12. There is no correlation between the translators' postediting quality and the

ISBN: DL:

number of the words they processed per minute (r = - 0.07, Table 55). However, our data show that the slowest and the fastest translators had lower quality-scores, while those with the mean values of speed produced higher-quality outputs (Table 55). Thus, for translators, spending more time on the task does not always guarantee higher-quality output.

13. There is a strong negative correlation between the engineers' postediting quality and the number of the words they processed per minute (r = - 0.80, Table 56). In other words, when they work slower, engineers produce higher-quality posteditings.
14. As compared to the engineers, the translators perform almost twice as much self-revision, although the difference between the groups is not statistically significant (p = 0.2127 when the outliers are excluded and 0.3498 when the outliers are included) (Table 52).
15. For engineers, there is no correlation between the time spent on self-revision and the postediting quality (r = 0.19, Table 59).
16. For translators, there is a moderate correlation between the time spent on self-revision and the postediting quality (r = 0.37, Table 60).
17. Translators spend more time on documentation than engineers while postediting technical texts, no matter whether we include the outliers (p = 0.0078, Table 33) or exclude them (p = 0.0086, Table 51). In addition, they use on-line dictionaries more and search for more items for longer periods of time. However, both groups use a very limited number of sources on the Internet while performing documentation (Table 33 and Table 51).
18. Both for the engineers and the translators, the time spent on documentation for postediting technical texts is not an important factor affecting the level of postediting quality. For engineers, the correlation between the time spent on documentation and postediting quality is weak (r = 0.30, Table 57). For translators, no such correlation exists (r = - 0.02, Table 58).
19. There is a strong correlation between the translators' postediting quality and the number of changes they performed (r = 0.63, Table 53).
20. There is a very strong correlation between the engineers' postediting quality and the number of changes they performed (r = 0.81, Table 54).

The following hypotheses have thus been formulated after completing Part 1 of the present study:

ISBN: DL:

1. Penalizing or unpenalizing the recurring errors in the target text significantly affects the postediting quality score.
2. Engineers who are experienced in engineering and in the translation of technical texts produce higher-quality postediting of technical texts than a) experienced translators (not particularly in technical translation) and b) translators holding a degree in Translation.
3. Experienced translators holding both a degree in Translation and having experience in technical translation produce similar quality postediting to engineers who are experienced in engineering and in the translation of technical texts.
4. For translators, experience in the subject-matter (technical texts) is more important for postediting quality than is general translation experience.
5. A degree in Translation does not directly correlate with postediting quality unless it is combined with subject-matter knowledge and professional experience in translation.
6. Spending more time on the postediting task does not always bring about higher quality.
7. There is no strong correlation between the time spent on documentation and the postediting quality.
8. When postediting technical texts, translators perform more documentation than engineers.
9. There is no strong correlation between the time spent on self-revision and the postediting quality.

Consequently, expertise and experience in the subject-matter stand out as the main factors determining postediting quality. Table 80 brings together the data from the post-assignment questionnaires that constitute the basis of this conclusion. Further, penalizing and unpenalizing the recurrent errors affects the quality score of the postediting significantly (Table 40 and Table 41).

ISBN: DL:

Table 80. Conclusion in Postediting - Translators and Engineers

Subjects	**Postediting Quality %**	**Quality Result**	**Years of experience in profession (engineering or translating)**	**Experience in the translation of technical texts**	**Education**
Tr 6	41.08	Fail	3-5	No	BA in English Language Teaching
Tr 3	46.47	Fail	3-5	No	BA in English Language and Culture
Tr 5	50.00	Fail	3-5	No	BA in Translation, PhD in Political Sciences
En 6	51.24	Fail	3-5	No	BSc and MSc in Metallurgy and Materials Engineering
En 9	52.70	Fail	3-5	No	BSc in Metallurgy and Materials Engineering
En 3	53.32	Fail	3-5	No	BSc and MSc in Electrical Engineering
Tr 10	56.02	Fail	9-10	No	BA in Translation, MA and PhD in Translation
Tr 7	59.13	Fail	More than 10	No	BA in English Language Teaching
Tr 9	62.24	Fail	6-8	No	BA in International Relations, MA and PhD in Translation
Tr 2	62.66	Fail	More than 10	From time to time	BA in Translation, MA in Communication Sciences
Tr 8	63.07	Fail	9-10	From time to time	BA in Translation
En 2	68.46	Fail	6-8	From time to time	BSc in Mechanical Engineering, MSc in Mechanical Engineering
En 5	83.40	Pass	More than 10	Yes	BSc and MSc in Mechanical Engineering
Tr 4	83.40	Pass	9-10	Yes	MA in Translation, PhD in Translation
En 8	86.51	Pass	9-10	Yes	BSc and MSc in Electrical Electronics Engineering
En 4	88.17	Pass	9-10	Yes	BSc in Mechanical Engineering, MSc in Automotive Engineering
En 7	88.59	Pass	9-10	Yes	BSc in Mechanical Engineering
Tr 1	88.59	Pass	9-10	Yes	BA, MA and PhD in Translation
En 1	91.29	Pass	9-10	Yes	BSc in Mechanical Engineering, MBA
En 10	92.12	Pass	More than 10	Yes	BSc in Mechanical Engineering

ISBN: DL:

10.2. Conclusions for Part 2 (Revision)

In Part 2 we compared the following workflows in terms of speed and end-product quality:

1. Postediting the MT output by subject-matter experts and subsequent revision by professional translators, and
2. Postediting the MT output by professional translators and subsequent revision by subject-matter experts.

We measured speed as words processed per minute in both the postediting and revision tasks. Quality was measured by evaluating the revised end-products using the LISA QA Model 3.1.

We tested the hypothesis that "when professional translators revise postediting by subject-matter experts and when subject-matter experts revise postediting by professional translators, the speed and end-product quality differ". Although this hypothesis was not affirmed, Part 2 yielded interesting results:

1. There is no significant difference between translators and engineers with regard to speed in revising one another's postediting (the p-value for revision duration is 0.4640 and 0.2727 for their w/pm, Table 61).
2. A total of 90 percent of the translators and 70 percent of the engineers think that translator postediting + engineer revision is faster than vice-versa. However, the quantitative data do not confirm this impression (p = 0.1611, Table 61).
3. There is no significant difference between engineer postediting + translator revision and translator postediting + engineer revision with regard to speed (the p-value for postediting + revision task time is 0.3021 and it is 0.1611 for w/pm, Table 61).

When the recurrent errors are penalized:

4. The difference between the quality of the engineers' postediting and engineers' revision of translators' postediting is insignificant (p = 0.5004, Table 62), while the difference between the quality of the translators' postediting and translators' revision of engineers' postediting is significant (p = 0.0232, Table 63).
5. A total of 70 percent of the participants in each group thought that the end-product of translator postediting + engineer revision would be of higher quality than vice-versa. The quantitative data do not affirm this. The mean quality percentage of the engineers' revision is 81.05 and the translators' is 79.68, yet

ISBN: DL:

the difference is not significant (p = 0.8685, Table 65).

6. Engineer revision significantly increased the quality of translator postediting (p = 0.0164, Table 66), indicating that the subject-matter knowledge of the reviser is more important in determining the end-product quality than is that of the posteditor.
7. Translator revision increased the quality of engineer postediting, yet it is not possible to talk about a significant increase (p = 0.6100, Table 67). This result affirms our finding above that the subject-matter knowledge of the reviser is more important in determining the end-product quality than is that of the posteditor.
8. Consequently, for technical texts, when the recurring errors are penalized, the quality improvement brought about by engineer-revision of translator-postediting is significantly higher than the quality improvement brought about by translator-revision of engineer-postediting (the p-value for the total error points is 0.0073 and for the quality percentage 0.0061, Table 68).

When the recurrent errors are unpenalized:

9. As under the penalized condition, the difference between the engineers' postediting and engineers' revision of translators' postediting is insignificant (p = 0.1487, Table 69), while the difference between the translators' postediting and translators' revision of engineers' postediting is significant (p = 0.0390, Table 70).
10. However, when the recurrent errors are not penalized, the statistical significance of the difference between the engineers' postediting and revision increases (Table 62 and Table 69), whereas that of the difference between the translators' postediting and revision slightly decreases (Table 63 and Table 70).
11. As under the penalized condition (Table 65), there is no significant difference between the translators' and engineers' revision qualities under the unpenalized condition (p = 0.9162 for the total error points and 0.9134 for the quality percentages, Table 72).
12. Engineer revision increased the quality of translator postediting, yet unlike the penalized condition (p = 0.0164, Table 66), the difference is not significant (p = 0.0751, Table 73).
13. Translator revision increased the quality of engineer postediting. Although the difference is not quite significant (p = 0.0685, Table 74), it is close to the

ISBN: DL:

threshold 0.05 as compared to the penalized condition, where the p-value for the difference between engineer postediting and translator revision is 0.6100 (Table 67).

14. Unlike the penalized condition (the p-value for the total error points is 0.0073 and for the quality percentage 0.0061, Table 68), the quality improvement brought about by engineer-revision of translator-postediting does not significantly differ from the quality improvement brought about by translator-revision of engineer-postediting (the p-value for the total error points is 0.1458 and for the quality percentage 0.1219, Table 75).
15. The number of terminology errors for both groups (for translators only major terminology errors and for engineers minor and major terminology errors) decreased significantly under the recurrent errors unpenalized condition as compared to the penalized condition (Table 76 and Table 77).
16. The translators' revision quality significantly differs under the penalized and unpenalized conditions ($p = 0.0265$, Table 78). However, as far as the added value brought about by translator-revision of engineer-postediting is concerned, the conditions do not differ significantly ($p = 0.3841$, Table 78).
17. The engineers' revision quality significantly differs under the penalized and unpenalized conditions ($p = 0.0508$, Table 79). Unlike translator revision of engineer postediting, the conditions also differ significantly as far as the added value brought about by engineer-revision of translator-postediting is concerned ($p = 0.0058$, Table 79).

Consequently, the mean differences between postediting and revision indicate that revision brings about a quality improvement in both cycles (translator postediting + engineer revision and engineer postediting + translator revision) under both conditions (penalized and unpenalized) (Table 68 and Table 75). However, under the recurrent errors penalized condition, the added value brought about by engineer-revision of translator-postediting is significant ($p = 0.0164$, Table 66), while it is not very significant under the unpenalized condition ($p = 0.0751$, Table 73). As for the added value brought about by translator-revision of engineer-postediting, we did not find a significant quality difference between the posteditings and revisions either under the penalized ($p = 0.6100$, Table 67) or the unpenalized conditions ($p = 0.0685$, Table 74). However, it is important to note that the significance of the difference between

ISBN: DL:
translator postediting and engineer revision decreased under the unpenalized condition (p = 0.0164, Table 66 compared to p = 0.0751, Table 73), while the significance of the difference between engineer postediting and translator revision increased (p = 0.6100, Table 67 compared to p = 0.0685, Table 74). Moreover, under the penalized condition, the quality improvement brought about by engineer-revision of translator-postediting is significantly higher than vice-versa (p = 0.0061, Table 68). Under the unpenalized condition, however, the difference between the quality improvement brought about by engineer revision and translator revision becomes insignificant (p = 0.1219, Table 75). Finally, for both groups, the quality of revision significantly differs under the recurring errors penalized and unpenalized conditions (Table 78 and Table 79).

10.3. Final Conclusions and Recommendations for the Translation Industry

All in all, the present study shows that, for technical texts, expertise and experience in the subject-matter stand out as the main factors determining postediting quality. In order to get the optimum quality results, the translation companies should give equal priority to the professional translators and subject-matter experts both of whom are experienced in their own profession and in the translation of the subject-matter in question. However, we should also note that — despite the involvement of other actors (i.e. volunteers, subject-matter experts) in the translation/postediting process — the translation profession is principally carried out by professional translators. Therefore, the findings of the present study encourage us to also address translator trainers and trainee translators and recommend that they pay more attention to specialization and work placements during translator training. Further, the stakeholders in the translation process (i.e. translation companies, translator trainers, trainee translators, professional translators) should bear in mind that general experience in the translation profession or a degree in Translation does not yield high-quality outputs unless they are combined with a sound knowledge and experience in the subject-matter.

According to our research results, for technical texts, there is no significant difference between the translators and engineers with regard to postediting and revision speed. However, as compared to the translators, engineers produce higher-quality posteditings with higher speed. Further, spending more time on the task does not always guarantee higher-quality output. The slowest and the fastest translators had lower quality-scores, while those with the mean values of speed produced higher-quality

ISBN: DL:
outputs.

Another finding of this research suggests that the higher quality of engineer postediting has a positive impact on the quality of translator revision and the relatively lower quality of translator postediting has a negative impact on the quality of engineer revision. In other words, the final quality of the end-product is not only the result of the reviser's abilities and expertise/experience in the subject-matter but it is also affected by the quality of the postediting to be revised. Thus, when trying to decide which work cycle (translator postediting + subject-matter expert revision or subject-matter expert postediting + translator revision) produces higher-quality outputs, we need to consider this effect of the quality of the text to be revised on the final output quality. So translation companies should bear in mind that recruiting subject-matter experts for the final revision as a method of quality assurance may not always give the expected results. Rather, they should seek experience and expertise in the subject-matter both for the translator/posteditor and for the reviser if they want to ensure higher-quality outputs.

Our data also show that, for technical texts, the quality improvement (added value) brought about by engineer-revision of translator-postediting is higher than the quality improvement brought about by translator-revision of engineer-postediting. However, this added value is significant only when the recurrent errors are penalized and it is not very significant when the recurrent errors are unpenalized.

Penalizing and unpenalizing recurrent errors are the methods employed in the translation industry for determining the output quality mostly without being questioned. This research shows that both the quality of the postedited texts and their revised versions (either performed by professional translators or subject-matter experts) change significantly as a result of penalizing and unpenalizing recurrent errors. Thus, in order to obtain sound and consistent quality results, the translation industry should take this significant but rather ignored point into consideration when evaluating the quality of the texts postedited/translated or revised.

10.4. Limitations of the Research

Finding financial resources has always been a challenging step during the course of academic research and the present study had its share of this problem. When the researcher works with professionals as in our case, this becomes more crucial, because professionals work for money and, frankly, their time is too limited to be spent on

ISBN: DL:

research experiments.

The experiments conducted for our research required the participants to read task instructions, postedit a 482-word technical text using screen-recording software, and fill out a post-assignment questionnaire. In addition, within a month or so after they completed these tasks, they were asked to revise the text postedited by another participant.

When we consider the fact that even having people fill out questionnaires might be troublesome in some cases, one can imagine how challenging it might be to get professionals to devote their time to our rather long experiments without offering them any financial reward. Moreover, some participants - especially the engineers - due to their tough work schedule, are often not willing to help with experiments even if they are paid. For these reasons, the present research was carried out with a total of twenty participants - ten professional translators and ten engineers using only one text. Obviously, a larger sample would yield more reliable and generalizable results. However, the present study carries its own weight in providing answers to our research questions on postediting MT output and its revision by professional translators versus subject-matter experts.

10.5. Suggestions for Further Research

The present study opens up various lines for further research. First of all, studies with larger samples (with more participants, using more texts) should contribute to more generalizable results.

Another avenue for future research might include the replication of the present study with other subject-matter such as law, medicine or even wine to see if the results change.

An extension could be to use this same methodology with postediting translation memory output or only with translation. Similarly, the method could be used with other language combinations to see if the results are similar, or to what extent they are language dependent.

Another study could be set up with a modification in the methodology by selecting professional translators from only those who have experience and expertise in the subject-matter in question.

It would also be interesting to see what happens when we use a quality evaluation

ISBN: DL:

system and/or software other than LISA QA Model 3.1.

Another perspective could be to examine the views of the end-users or people holding expertise in the subject-matter.

Finally, we hope that this research generates motivation for further research on MT postediting and the involvement of actors other than professional translators in different phases of the postediting/translation process, both of which seem to have a key role in the future of translation.

ISBN: DL:

ISBN: DL:

ISBN: DL:

References

Abekawa, Takeshi, et al. 2010. "Community-based construction of draft and final translation corpus through a translation hosting site Minna no Hon'yaku (MNH)". In *Proceedings of the Seventh international conference on Language Resources and Evaluation,* Nicoletta Calzolari et al. (eds). (LREC). CD. Valletta: European Language Resources Association. 3662-3669.

Aikawa, Takako, et al. 2007. "Impact of Controlled Language on Translation Quality and Postediting in a Statistical Machine Translation Environment". In *Proceedings of the MT Summit* XI. Copenhagen. 1-7.

Allen, Jeffrey. 2000. "The Value of Internet Translation Portals". *International Journal for Language and Documentation IJLD* 6: 45-46.

Allen, Jeffrey. 2001a. "Postediting or no postediting". *International Journal for Language and Documentation* 8: 41-42.

Allen, Jeffrey. 2001b. "Postediting: an integrated part of a translation software program". *Language International* 13 (2): 26-29.

Allen, Jeffrey. 2003. "Postediting". In *Computers and Translation: A Translator's Guide*, H. Somers (ed). Amsterdam: John Benjamins. 297-318.

Allen, Jeffrey. 2004a. "Case study: implementing MT for the translation of pre-sales marketing and post-sales software deployment documentation at Mycom International". In *Proceedings of the 6th Conference of the Association for MT in the Americas.* Washington DC.1-6

Allen, Jeffrey. 2004b. "Inbound versus outbound translation". In Presentation in panel: Localization in customer support. Localization World Bonn Conference.1-20

Allen, Jeffrey. 2004c. "Thinking about Machine Translation". In *Multilingual Computing and Technology* 6: 3.

Allen, Jeffrey. 2005a. "An introduction to using MT software". *The Guide from Multilingual Computing & Technology* 69: 8-12.

Allen, Jeffrey. 2005b. "What is postediting?" *Translation Automation* 4: 1-5.

Arnold, Doug, et al. 1994. *Machine Translation: An Introductory Guide.* London: NCC Blackwell.

Austermühl, Frank. 2001. *Electronic Tools for Translators*. Manchester: St. Jerome

Publishing.

Babel Fish. Accessed March 2012. Free Online Translator. http://www.babelfish.com.

Bass, Scott. 1999. "Machine vs. Human Translation". Advanced Language Translation Inc. http://www.advancedlanguage.com. Accessed January 2010.

BB FlashBack. Accessed January 2011. Screen Recorder Software, Blueberry Software Ltd. http://www.bbsoftware.co.uk/bbflashback/home.aspx.

Bowker, Lynne, and Melissa Ehgoetz. 2007. "Exploring User Acceptance of Machine Translation Output: A Recipient Evaluation". In *Across Boundaries: International Perspectives on Translation Studies*, Dorothy Kenny and Kyongjoo Ryou (eds). Newcastle: Cambridge Scholars Publishing. 209-224.

Bruckner, Christine, and Mirko Plitt. 2001. *Evaluating the Operational Benefit of Using Machine Translation Output as Translation Memory Input*. MT Summit VIII, Santiago de Compostela, Spain. http://www.mt-archive.info/MTS-2001-Bruckner.pdf. Accessed July 2010.

Callison-Burch, Chris, et al. 2006. "Re-evaluating the role of BLEU in Machine Translation Research". In *Proceedings of EACL 2006 (11th Conference of the European Chapter of the Association of Computational Linguistics)*. Trento, Italy, 249-246. http://www.iccs.inf.ed.ac.uk/~miles/papers/eacl06.pdf. Accessed July 2010.

Campbell, S. 2000. "Critical Structures in the Evaluation of Translations from Arabic into English as a Second Language". *The Translator* 6: 211–229.

Carroll, John B. "An Experiment in Evaluating the Quality of Translations". *Mechanical Translation and Computational Linguistics* 9 (3-4): 55-66.

Champollion, Yves. 2001. "Machine translation (MT), and the future of the translation industry" *Translation Journal* 5 (1). http://accurapid.com/journal/15mt.htm. Accessed May 2011.

Chang-Meadows, Shin. 2008. "MT errors in CH-to-EN MT systems: user feedback". In *AMTA-2008. MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*, Waikiki, Hawai.

Chesterman, Andrew, and Emma Wagner. 2002. Can Theory Help Translators? A Dialogue Between the Ivory Tower and the Wordface. Manchester: St. Jerome Publishing.

Christensen, Tina Paulsen. 2003. Translation memory-systemer som værktøj til juridisk oversættelse. Kritisk vurdering af anvendeligheden af translation memory-systemer til oversættelse af selskabsretlig dokumentation. Doctoral Thesis. Syddansk Universitet.

Christensen, Tina Paulsen, and Anne Schjoldager. 2010. "Translation-Memory (TM) Research: What Do We Know and How Do We Know It?" *Hermes, Journal of Language and Communication Studies* (44): 64-89.

Colominas, Carme. 2008. "Towards chunk-based translation memories". *Babel* 54 (4): 343-354.

Dikte. Accessed February 2011. "Home page of Dikte Speech Recognition Software by Yöndata and CTD Systems". http://www.dikte.com.tr.

Dillon, Sarah, and Janet Fraser. 2006. "Translators and TM: An investigation of translators' perceptions of translation memory adoption". *Machine Translation* 20 (2): 67-79.

Dragsted, Barbara. 2004. Segmentation in Translation and Translation Memory Systems-An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM System into the Translation Process. Doctoral thesis. Copenhagen: Samfundslitteratur.

Dragsted, Barbara. 2006. Computer-aided translation as a distributed cognitive task". *Pragmatics & Cognition* 14(2): 443-464.

EAMT. Accessed June 2010. "Home page of the European Association for Machine Translation". www.eamt.org.

eCoLoRe. Accessed December 2010. "Home page of the eContent Localization Resources for Translator Training". http://ecolore.leeds.ac.uk.

ENLASO. Accessed September 2011. "Home page of ENLASO Translation and Localization Solutions Provider". http://www.enlaso.com.

European Standards for Translation Services - prEN15038 (Draft). 2004. http://web.letras.up.pt/egalvao/prEN-15038.pdf. Accessed May 2011.

Facebook. Accessed May 2013. "Facebook Developers". http://developers.facebook.com/docs/internationalization

Fiederer, Rebecca, and Sharon O'Brien. 2009. "Quality and machine translation: a realistic objective?" *Journal of Specialised Translation* 11: 52-74.

Flanagan, Mary. 1997. "MT Today: Emerging Roles, New Successes". *Machine Translation* 12: 25-27.

Fulford, Heather, and Joaquin Granell-Zafra. 2005. "Translation and Technology: a Study of UK Freelance Translators". *Journal of Specialized Translation* 4: 2-

17.

García, Ignacio. 2008. "Translating and Revising for Localisation: What do We Know? What do We Need to Know?". *Perspectives: Studies in Translatology* 16 (1-2): 49-60.

García, Ignacio. 2009. "Beyond Translation Memory: Computers and the Professional Translator". *Journal of Specialised Translation* 12. http://www.jostrans.org/issue12/art_garcia.php. Accessed May 2010.

García, Ignacio. 2010. "Is Machine Translation Ready Yet?" *Target* 22 (1): 7-21.

García, Ignacio. 2012. "A Brief History of Postediting and of Research in Postediting". In *Anglo-saxónica, Special issue on Translation Studies*, Alexandra Rosa Assis and Anthony Pym (eds). 266-291.

Giménez Linares, Jésus Ángel. 2008. *Empirical Machine Translation and its Evaluation*. Doctoral Dissertation, Departament de Llenguatges i Sistemes Informatics, Universitat Politecnica de Catalunya.

Google Translator Toolkit. Accessed November 2010. "Home page of Google Translator Toolkit"**.** http://translate.google.com/toolkit.

GraphPad t-test calculator. Accessed April 2012. http://graphpad.com/quickcalcs/ttest1.cfm

Guerberof, Ana. 2008. Productivity and quality in the postediting of outputs from translation memories and machine translation: A pilot study. Minor Dissertation. Tarragona: Universitat Rovira i Virgili.

Guerberof, Ana. 2012. Productivity and quality in the postediting of outputs from translation memories and machine translation. Doctoral thesis. Tarragona: Universitat Rovira i Virgili.

Guerra Martínez, L. 2003. Human Translation versus Machine Translation and Full Postediting of Raw Machine Translation Output. Masters Dissertation. Dublin: Dublin City University.

Guzmán, Rafael. 2007a. "Manual MT Postediting if it's not broken, don't fix it". *Translation Journal* 11 (4): 1-7

Guzmán, Rafael. 2007b. "Automating MT postediting using regular expressions". *Multilingual 90* 18 (6): 49-52.

Guzmán, Rafael. 2008. "Advanced automatic MT postediting" *Multilingual 95* 19 (3): 52-57.

HISPAWORDS. Accessed September 2011. "Home page of HISPAWORDS Language

ISBN: DL:
Service Provider”. http://hispawords.com/life_sciences.html

Hogan, C. Allen, J. 2000. “Toward the development of a postediting module for raw machine translation output: A controlled language perspective”. Presented at the *Third International Controlled Language Applications Workshop (CLAW2000).* Seattle, Washington.

Holz-Mänttäri, Justa. 1984. *Translatorisches Handeln. Theorie und Methode*. Helsinki: Academia Scientiarum Fennica.

IDIS. Accessed June 2011. “Home page of the International Dismantling Information System”. http://www.idis2.com.

InterTran. Accessed February 2011. “Home page of Translation Experts”. http://www.tranexp.com:2000/Translate/result.shtml.

Izwaini, Sattar. 2006. “Problems of Arabic machine translation: evaluation of three systems”. In *The Challenge of Arabic for NLP/MT*. International conference at the British Computer Society, London. 118-148. http://www.mt-archive.info/BCS-2006-Izwaini.pdf. Accessed December 2010.

Jakobsen, Arnt Lykke. 2003. “Effects of think aloud on translation speed, revision and segmentation”. In *Triangulating Translation: Perspectives in process oriented research*, Fabio Alves (ed). Amsterdam and Philadelphia: John Benjamins. 69–95.

Kaiser-Cooke, Michèle. 1994. “Translatorial expertise - a cross-cultural phenomenon from an inter-disciplinary perspective”. In *Translation Studies. An Interdiscipline*, Mary Snell-Hornby et al. (eds). Amsterdam and Philadelphia: John Benjamins. 135-139.

Kay, Martin. 1982. “Machine Translation”. *American Journal of Computational Linguistics* 8 (2): 74-78.

Kay, Martin. 1997. “The Proper Place of Men and Machines in Language Translation”. *Machine Translation* 12 (1-2): 3-23.

Khalzanova, Serafima. 2008. *Revision and Time Constraints in Translation. Pilot Study.* Tarragona: Universitat Rovira i Virgili. Available from: http://isg.urv.es/publicity/doctorate/research/index.html

Kolmogorov-Smirnov Test. Accessed April 2012. http://www.physics.csbsju.edu/stats/KS-test.n.plot_form.html.

Krings, H. 2001. Repairing Texts: Empirical Investigations of Machine Translation Postediting Processes, G. S. Koby (ed). Ohio: Kent State University Press.

Krollman, Friedrich. 1981. “Computer Aids to Translation”. *Meta* 26 (1): 85-94.

Künzli, Alexander. 2007. “Translation Revision: A study on the Performance of Ten Professional Translators Revising a Legal Text”. In *Doubts and Directions in Translation Studies,* Y. Gambier, M. Shlesinger and R. Stolze (eds). Amsterdam: Benjamins. 115–126.

Lagoudaki, Elina. 2006. *Translation Memories Survey 2006*. London: Imperial College London. http://www.atril.com/docs/tmsurvey.pdf. Accessed August 2010.

Lange, Carmen A. and Winfield S. Bennett. 2000. “Combining machine translation with translation memory at Baan”. In *Translating into success*, Robert C. Sprung (ed). Amsterdam and Philadelphia: John Benjamins. 203-218.

Language Scientific. Accessed September 2011. “Home page of Language Scientific, Translation and Localization Company”. http://www.languagescientific.com/translation-quality-control-process.html

Laurian, Anne Marie. 1984. “Machine Translation: What Type of Postediting on What Type of Documents for What Type of Users”. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting on Association for Computational Linguistics*, Stanford, California. 236-238.

Le Néal, Jocelyne. 2001. “Preparing Multi-Volume Illustrated Terminological Dictionaries”. In *Handbook of Terminology Management: Volume II: Application-Oriented Terminology Management,* S.E. Wright and G. Budin (eds). Amsterdam and Philadelphia: John Benjamins. 645-665.

LISA. 2004. “LISA Best Practice Guides: Implementing Machine Translation”. http://www.lisa.org. Accessed May 2010.

LISA. Accessed July 2010. “Home page of the Localization Industry Standards Association”. http://www.lisa.org/LISA-QA-Model-3-1.124.0.html.

Lopez, Adam. 2008. “Statistical Machine Translation”. *ACM Computing Surveys* 40 (3): 1-49.

Maia, Belinda. 2008. “Machine translation and human translation: Using machine translation engines and corpora for teaching and research”. In *Current Trends in Contrastive Linguistics*, María de los Ángeles Gómez González, J. Lachlan Mackenzie and Elsa M. González Álvarez (eds). Amsterdam: John Benjamins. 123–145.

McElhaney, T. and M. Vasconcellos. 1988. “The Translator and the Postediting

ISBN: DL:
Experience”. In *Technology as Translation Strategy*, Muriel Vasconcellos (ed). American Translators Association Scholarly Monograph Series, Vol. II, State University of New York at Binghamton (SUNY). 140-148.

McElroy Translation. Accessed September 2011. “Home page of McElroy Translation”. http://www.mcelroytranslation.com/processes/qualityassurance/workflowprocess/

Melby, Alan K. 1981. “Translators and Machines – Can They Cooperate?” *Meta* 26 (1): 23-34.

Melby, Alan K. 1992. “The Translator Workstation”. In *Computers in Translation: A Practical Appraisal*, J. Newton (ed). London: Routledge. 147–165.

Melby, Alan K. 1997. “Some Notes on The Proper Place of Men and Machines in Language Translation”. *Machine Translation* 12: 29-34.

Microsoft Bing Translator. Accessed February 2011. “Home page of Microsoft Translator”. http://www.microsofttranslator.com.

Moore Gary C., Benbasat Izak. 2001. “Development of an instrument to measure the perceptions of adopting an information technology innovation”. *Information Systems Research* 2: 192-221.

Mossop, Brian. 2000. “The workplace procedures of professional translators”. In *Translation in Context,* A. Chesterman et al. (eds). Amsterdam and Philadelphia: John Benjamins. 39-48.

Muegge, Uwe. 2011. “Ten good reasons why you should validate your translated terminology”. http://www.tcworld.info/tcworld/translation-and-localization/article/ten-good-reasons-why-you-should-validate-your-translated-terminology. Accessed September 2011.

O’Brien, Sharon. 2002. “Teaching postediting: A proposal for course content”. In *Proceedings of the European Association of Machine Translation*. http://www.mt-archive.info/EAMT-2002-OBrien.pdf. Accessed October 2010.

O’Brien, Sharon. 2005. “Methodologies for Measuring Correlations between Postediting Effort and Machine Translatability”. *Machine Translation* 19 (1): 37-58.

O’Brien, Sharon. 2006a. “Eye-tracking and Translation Memory Matches”. *Perspectives: Studies in Translatology* 14 (3): 185-205

O’Brien, Sharon. 2006b. “Pauses as indicators of cognitive effort in postediting machine translation output”. *Across Languages and Cultures* 7 (1): 1-21.

O'Brien, Sharon. 2008. "Processing fuzzy matches in Translation Memory tools: an eye-tracking analysis". In *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing,* Göpferich S., Jakobsen A.L. and Mees I. M. (eds). Copenhagen: Samfundslitteratur. 79-102.

O'Brien, Sharon. 2012. "Towards a Dynamic Quality Evaluation Model for Translation". *Journal of Specialized Translation* 17: 1-24.

Oflazer, Kemal, and İlknur Durgar El-Kahlout. 2007. "Exploring different representational units in English-to-Turkish statistical machine translation". In *Proceedings of the Second Workshop on Statistical Machine Translation*. Prague, Czech Republic: Association for Computational Linguistics. 25-32.

OPTIMALE Synthesis Report. 2012. "The OPTIMALE Employer Survey and Consultation", LLP Project on Optimizing Professional Translator Training in a Multilingual Europe.

Orr, David B., and Victor H. Small. 1967. "Comprehensibility of machine-aided translations of Russian scientific documents". *Mechanical Translation and Computational Linguistics* 10 (1-2): 1-10.

Papineni et al. 2002. "BLEU: A method for automatic evaluation of machine translation". In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL),* Philadelphia. 311-318.

Park, William M. 1988. "The Ideal Workstation: Perspective of the Independent Translator". In *Technology as Translation Strategy*, Muriel Vasconcellos (ed). American Translators Association Scholarly Monograph Series, Vol. II, State University of New York at Binghamton (SUNY). 29-35.

Pym, Anthony. 1996. "Ideologies of the Expert in Discourses on Translator Training". In *Problems and Trends in the Teaching of Interpreting and Translation.* Mary Snell-Hornby & Yves Gambier (eds). Koinè Anno IV, Misano: Istituto San Pellegrino. 139-149.

Pym, Anthony. 2009. "Using process studies in translator training. Self-discovery through lousy experiments". In *Methodology, Technology and Innovation in Translation Process Research.* S. Göpferich, F. Alves & I. M. Mees (eds). Copenhagen: Samfundslitteratur. 135-156.

Pym, Anthony. 2011a. "Democratizing translation technologies – the role of humanistic research". Presented at *the Luspio Translation Automation Conference*, Rome.

Pym, Anthony. 2011b. "Translation research terms. A tentative glossary for moments of perplexity and dispute". In *Translation Research Projects 3*. Anthony Pym (ed). Tarragona: Intercultural Studies Group. 75-110.

Ramos, Luciana Cecilia. 2010. "Postediting free machine translation: from a language vendor's perspective". In *AMTA 2010: the Ninth conference of the Association for Machine Translation in the Americas*, Denver, Colorado. 1- 5.

Ribas, Carlota. 2007. *Translation Memories as vehicles for error propagation. A pilot study.* Minor Dissertation. Tarragona: Universitat Rovira i Virgili.

Roberts, Martin. 2007. "Optimizing Postediting". Language Weaver 3 (5). http://www.imakenews.com/lweaver/e_article000828087.cfm?x=b11,0,w Accessed May 2011.

Schäfer, Falko. 2003. "MT postediting: How to shed light on the 'unknown task': Experiences at SAP". http://www.mt-archive.info/CLT-2003-Schaefer.pdf. Accessed June 2010.

Science Lab. Accessed April 2010. "Home page of Science Lab: Chemicals&Laboratory Equipment". http://www.sciencelab.com/msdsList.php.

SDL. Accessed September 2011. "Home page of SDL Language Services". http://www.sdl.com/en/language-services/translation-solutions/translation-services.asp

Senez, Dorothy. 1998. "The Machine Translation Help Desk and the Postediting Service". api.ning.com/files/...p/1998_01_tt_senez.pdf. Accessed February 2010.

SimulTrans. Accessed September 2011. "Home page of SimulTrans Language Service Provider". http://www.simultrans.com/quality/quality-process

Snell-Hornby, Mary. 1992. "The professional translator of tomorrow: language specialist or all-round expert?". In *Teaching Translation and Interpreting*, Cay Dollerup and Anne Loddegaard (eds). Amsterdam and Philadelphia: John Benjamins. 9-22.

Soloduk, Judith. Accessed September 2011. "Best Practices: Taking Your Certification Examination Process to a Global Audience". http://www.enlaso.com/Language_Tech_Center/Articles/Taking_Your_Certification_Examination_Process_to_a_Global_Audience.aspx.

Sullivan, Jeremiah J. 1990. "Experts, Expert Systems, and Organizations". In *Organization, Management, and Expert Systems. Models of Abstract Reasoning*, Michael Masuch (ed). Berlin and New York: de Gruyter. 13-33.

TAUS Report. 2010. "A TAUS Report, Postediting in Practice". Netherlands: De Rijp.

TAUS. Accessed April 2013. "A Translation Automation Timeline". http://www.translationautomation.com/timeline/a-translation-automation-timeline.

Tcworld. Accessed September 2011. "Home page of Tcworld Magazine for International Information Management". http://www.tcworld.info.

Tobii. Accessed January 2011. "Home page of Tobii, manufacturer of eye-tracking equipment". http://www.tobii.com.

Trados. Accessed October 2010. "Home page of the SDL Trados". http://www.trados.com/en/.

Translog. "Tool for analyzing text production processes". www.translog.dk.

Twitter. Accessed May 2013. "Twitter Translation Center". http://translate.twttr.com/welcome

Vaezian, Helia. 2010. "Google Translator Toolkit in Translation Classroom". Unpublished Paper. School of Languages, Literacies and translation. Universiti Sains Malaysia.

Van Slype, Georges. 1979. "SYSTRAN: Evaluation of the 1978 Version of the Systran English-French Automatic System of the Commission of the European Communities". *The Incorporated Linguist* 18 (3): 1-4.

Vasconcellos, Muriel and Marjorie Léon. 1985. "SPANAM and ENGSPAN: machine translation at the Pan American Health Organization." *Computational Linguistics* 11 (2-3): 122-136.

Vasconcellos, M. 1986. "Postediting on Screen: Machine Translation from Spanish into English". In *Translating and the Computer 8: A Profession on the Move*, Catriona Picken (ed). London: Aslib. 133-146.

Vasconcellos, Muriel. 1987. "A comparison of MT Postediting and Traditional Revision". In *Proceedings of the 28th Annual Conference of the American Translators Association*, K. Kummer (ed). Medford, NJ: Learned Information. 409-415.

Vaumoron, Jean A. 1988. "The Ideal Workstation: Perspective of the In-House Translator". In *Technology as Translation Strategy*, Muriel Vasconcellos (ed). American Translators Association Scholarly Monograph Series, Vol. II, State University of New York at Binghamton (SUNY). 36-41.

Wagner, Elizabeth. 1985. "Rapid Postediting of SYSTRAN". In *Tools for the Trade:*

Translating and the Computer 5, V. Lawson (ed). London: Aslib. 199-213.

Way, Catherine. 2004. "Making theory reality. An example of interdisciplinary cooperation". In *Translating in the 21st Century. Trends and Prospects. Proceedings*. Georges Androulakis (ed). Thessaloniki: Aristotle University. 584-592.

Wessa, Patrick. Accessed June 2013. Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7. http://www.wessa.net/

WorldLingo. Accessed February 2011. "Home page of WorldLingo Translations". http://www.worldlingo.com/products_services/worldlingo_translator.html.

Yamada, Masaru. 2011. "The effect of translation memory databases on productivity". In *Translation Research Projects 3*, Anthony Pym (ed). Tarragona: Intercultural Studies Group. 63-73.

Yngve, Victor H. 1954. "The Machine and the Man". *Mechanical Translation* 1 (2): 20-22.

Zuo, Li. 2010. "Machine translation from English to Chinese: a study of Google's performance with the UN documents [Abstract]". In *AMTA 2010: the Ninth conference of the Association for Machine Translation in the Americas*, Denver, Colorado.

ISBN: DL:

ISBN: DL:

Appendices

Appendix A – Research Participant Release Form

I voluntarily agree to participate in a series of postediting/revision tests for research conducted for the Intercultural Studies Group at the Rovira i Virgili University in Tarragona, Spain.

I understand that this evaluation is being conducted by Özlem Temizöz and will be part of the subsequent doctoral dissertation supervised by Prof. Anthony Pym.

I understand that the evaluation methods which will involve me are:

1. my postediting and revision of the test text

2. screen recordings of my postediting/revision processes

3. my completion of a post-assignment questionnaire

I grant permission for the postediting/revision protocols to be screen recorded and transcribed together with the questionnaires and to be used only by Özlem Temizöz for analysis of postediting/revision and questionnaire data. I grant permission for the evaluation of data generated from the above methods to be published in the dissertation and future publications.

I understand that the reports and publications will contain no identifiable information with regard to my name.

____

Research Participant's Name / Signature

____

Date

ISBN: DL:

Appendix B – Instructions for Part 1

Preliminary Instructions and Postediting Brief

1. You will postedit a 482-word text in Turkish translated from English with Google Translate. The text is about the dismantling of end-of-life vehicles.
2. The aim is to produce a publishable target text for the information leaflet to be distributed to the technical staff (with Turkish L1) in the automotive technical services.
3. The MT output may contain mistranslations. While postediting, you can refer to the original source text in English to check any missing or incorrect part in the MT output. You are expected to postedit the MT output. Please do not translate the source text from scratch.
4. You can use the Internet and on-line dictionaries. No external Translation Memories (TM) or other tools will be used.
5. You have fifty minutes to postedit the MT output. However, if you do not finish the task in fifty minutes, you can take any extra time.
6. You should perform the postediting on the MT output (do not create a separate target text) at one go without pausing.
7. Your screen activities will be recorded with the screen recorder BB Flashback.

Instructions to initiate BB Flashback

1. Click on the *BB Flashback Pro 3* file provided by the researcher.
2. Set up the program using the user name and the password provided by the researcher.

Instructions for Postediting

1. Click on the *BB Flashback Pro3 Recorder* shortcut icon on the desktop.
2. Click on Record a new movie.
3. Press the red *Record* button of the BB Flashback to initiate the recording.
4. Open “TR-MT Output” in the experiment file and start postediting the text. Save your changes.
5. You can refer to “EN-Source Text” in the experiment file whenever you need.
6. When you complete postediting, press the *Stop* button.
7. Save your BB Flashback recording.

ISBN: DL:

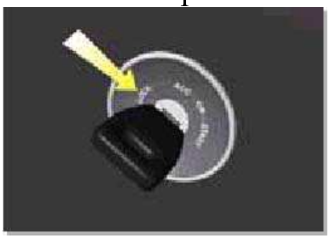
Appendix C – The Source Text in English

2 Method of onboard deployment of electrical pyrotechnic devices

Before performing onboard deployment, be sure to conduct the following preparation.

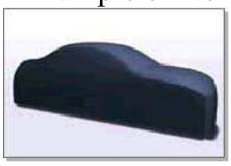
• Turn the ignition switch to OFF (Lock) and disconnect the battery cable terminal from the battery.

• Disconnect the negative, minus terminal and wait for the pyrotechnic system to become electrically discharged. Please refer to the vehicle overview table to obtain the make/model specific discharge time.



• Consider appropriate measures for the prevention of glass scattering and, if necessary, to reduce deployment noise and emission.

• Example of Noise Reduction Measure: Cover up the vehicle.



3 Common procedure after deployment

After onboard deployment, be sure to complete the following steps.

• Ventilate the vehicle passenger air inside the compartment, confirm that all airbags have been deployed and disconnect the deployment harness.

• Since the airbags are hot after deployment, do not touch them immediately after neutralization.

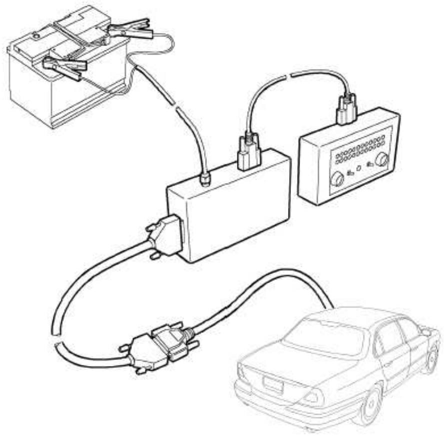
ISBN: DL:

4 Simultaneous deployment tool

4.1 Tool Type

For detailed operational instructions and specifications of the tool, refer to the users manual enclosed with the Tool (**www.seda.at**)

SEDA - Airbag master



4.2 Connection method

Type 1: Directly connect the tool deployment harness to the airbag and/or pretensioner harnesses.

1. Locate and disconnect pyrotechnic device multiplug.

2. Cut off the pyrotechnic device multiplug from the harness.

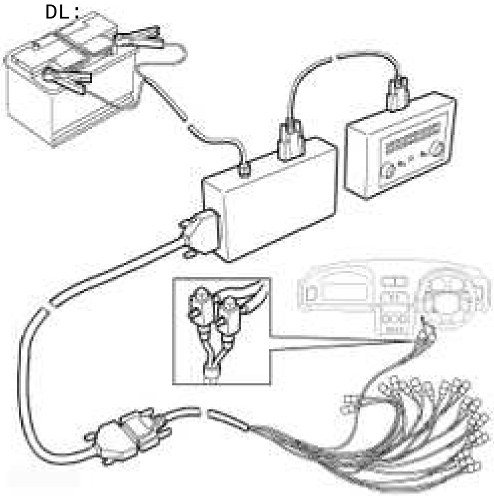
3. Strip off approx. 5mm if insulation from the ends of the cut wires.

4. Connect the stripped wires into the deployment flylead terminals. If connecting to a dual stage device, use a separate pair of flylead terminals for each gas generator.

NOTE: Any pair of flylead terminals can be used for any generator.

5. Connect the deployment tool and carry out the deployment process.

ISBN: DL:

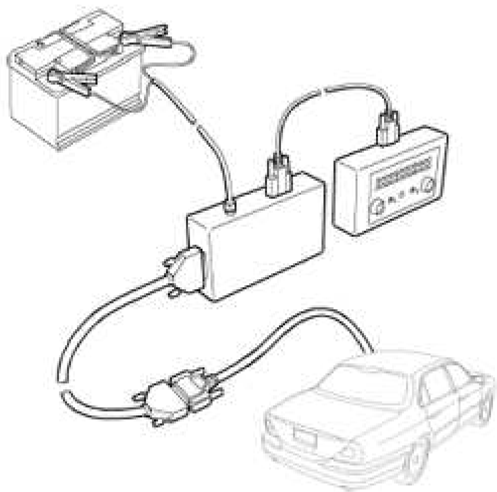


Type 2: Connect the tool model specific deployment adapter to the pyrotechnic system control unit multiplug.

1. Locate the pyrotechnic control unit and disconnect the multiplug.

2. Connect the deployment tool to the pyrotechnic control unit using the appropriate adaptor.

3. Connect the airbag deployment tool and carry out deployment process.



4.3 Deployment procedure

The deployment procedure can be performed manually or in an automatic sequence (See tool manual)

(1) Power up the tool by connecting the battery.

Note: If sufficient voltage is supplied to the tool, the BATT lamp illuminates green. If the light blinks red, the battery voltage is too low. If the battery leads are connected with incorrect polarity, the BATT lamp will not illuminate.

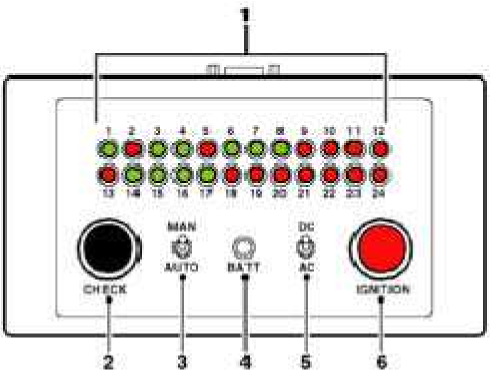
(2) Press the check button, to view the status of the pyrotechnics fitted to the vehicle.

The number of devices fitted to a vehicle determines how many status lamps are lit (green). Two stage devices will be indicated by 2 green lights.

(3) Switch the DC/AC switch to select the correct type of deployment current.

(4) Switch the MAN/AUTO switch to select manual or automatic deployment.

(5) Retire to a safe distance (min. 6 metres) before commencing the deployment sequence.



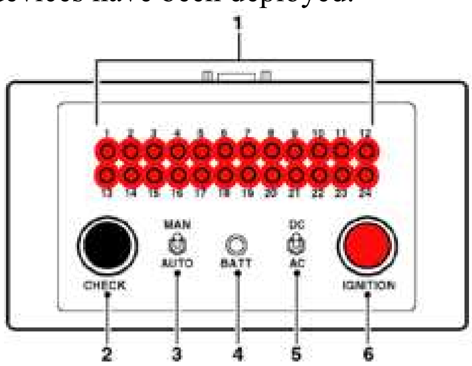
(6) Press the CHECK and the IGNITION button together to deploy the devices.

- Manual (MAN) deployment: Keep pressing the CHECK button and press the IGNITION button repeatedly until all devices are deployed.

- Automatic (AUTO) deployment: Keep pressing the CHECK button and the IGNITION button together to start the activation. The devices will automatically deploy at 3 seconds intervals.

(7) The end of the process is indicated by flashing red lights. Release the buttons after the deployment of all devices.

(8) Press the CHECK button again to ensure all status lamps are red, indicating all devices have been deployed.



Note: All status lamps will flash red when all devices have been activated.

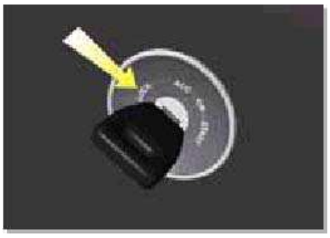
Appendix D – The MT Output in Turkish

2 Elektrik piroteknik cihazların yerleşik dağıtım yöntemi

Onboard dağıtım işlemini gerçekleştirmeden önce, aşağıdaki hazırlık yapmak için emin olun.

• OFF (Lock) Kontak anahtarını açın ve pil pil kablo terminal bağlantısını kesin .

• negatif, eksi terminali bağlantısını kesin ve piroteknik sistemin elektriksel olarak discharge bekleyin. Marka / modele özel deşarj süresi elde etmek için aracın genel tabloya bakınız.



• cam saçılma önlenmesi için uygun önlemleri dikkate alın ve gerekirse, dağıtım gürültü ve emisyon azaltmak için.

• Gürültü Azaltma Önlemi Örnek: araç örtün.

Dağıtımdan sonra 3 Ortak prosedür

Onboard Dağıtımdan sonra, aşağıdaki adımları tamamlamak için emin olun.

• bölmesinin içinde aracın yolcu bölümünü havalandırın, tüm hava yastıklarının dağıtıldığını onaylamak ve dağıtım donanımının bağlantı kesmek.

•, dağıtımdan sonra hava yastıkları sıcak olduğundan, hemen nötralizasyon sonra onlara

ISBN: DL:

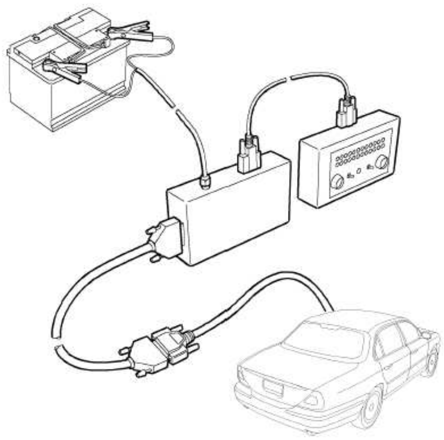
dokunmak yok .

4 Eşzamanlı dağıtım aracı

4.1 Aracı Türü

Aracın detaylı operasyonel talimatları ve teknik özellikleri için, bkz.
Aracı ile kapalı kullanım kılavuzu (www.seda.at)

SEDA – Hava yastığı master



4.2 Bağlantı yöntemi

Tip 1: Doğrudan araç dağıtım harness hava yastığına ve / veya gerdirici harnesses bağlayın

1. Piroteknik cihazı, çoklu bulun ve çekin.

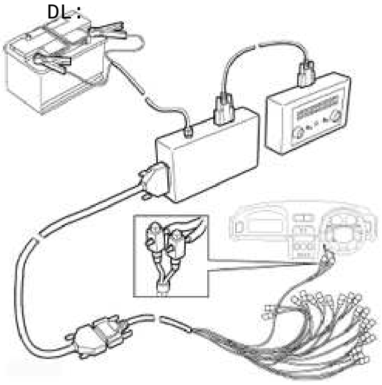
2. Piroteknik cihazı çoklu harness kesin.

3. Yaklaşık 5mm soyun, kesilen teller ucundan yalıtım ise.

4. Dağıtım flylead terminalleri içine soyulan teller bağlayın. Bir çift stage cihazına bağlıyorsanız, her bir gaz jeneratörü için ayrı bir çift flylead terminalleri kullanın.

NOT: Herhangi bir çift flylead terminalleri herhangi bir jeneratör için kullanılabilir.

5. Dağıtım aracı bağlayın ve dağıtım işlemini yürütmek.

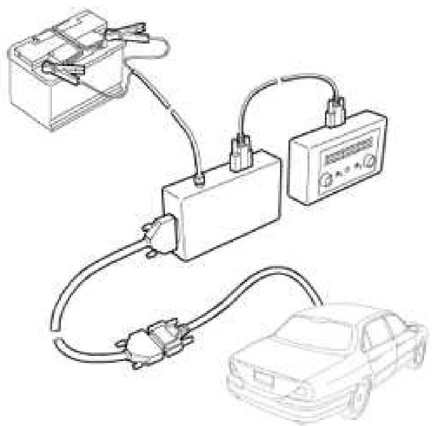


Tip 2: Araç modele özel dağıtım adaptörü piroteknik sistem kontrol ünitesi çoklu bağlayın.

1. Piroteknik kontrol ünitesi bulun ve çoklu bağlantısı kesin.

2. Dağıtım aracı, uygun adaptörü kullanarak piroteknik kontrol ünitesine bağlayın

3. Hava yastığı dağıtım aracı bağlayın ve dağıtım işlemini yürütmek.



4.3 Dağıtım prosedürü

Dağıtım işlemi elle veya otomatik bir sıra yapılabilir (aracı kılavuzuna bakın)

(1) Güç verin araca pil bağlayarak.

Not: Araca yeterli voltajı verilirse, BATT lambası yeşil yanar. Işık kırmızı olarak yanıp söner, pil voltajı çok düşük. Pil kablolarında yanlış polarite ile bağlı iseniz, BATT lamba yanmaz.

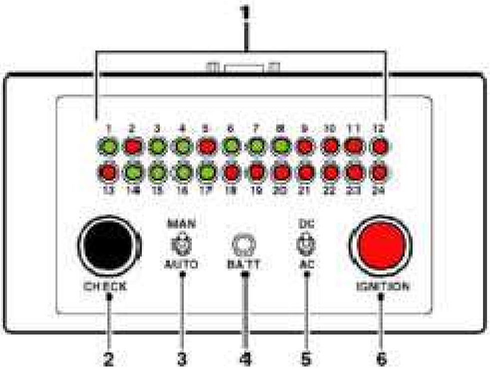
(2), araçta bulunan piroteknik durumunu görmek için, onay tuşuna basın. Bir araca takılan cihazların sayısı (yeşil) kaç durum lambaları yanar belirler. İki aşamalı cihazı, 2 adet yeşil ışıklar gösterilir.

(3) dağıtım akımın doğru türünü seçmek için DC / AC switch.

(4), manuel veya otomatik dağıtım seçmek için MAN / OTO switch.

ISBN:
DL:

(5), dağıtım dizisi başlamadan önce güvenli bir mesafede (en az 6 metre) çekilin.



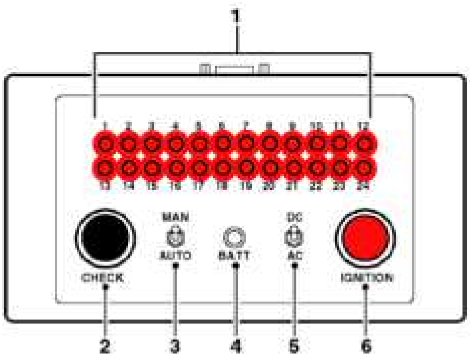
(6) cihazlar dağıtmak için CHECK ve ATESLEME düğmesine birlikte basın.

- Manuel (MAN) dağıtım: CHECK tuşuna basılı tutun ve tüm cihazlar dağıtılan kadar art arda ATESLEME düğmesine basın .

- Otomatik (AUTO) dağıtım: aktivasyon başlatmak için CHECK düğmesi ve ATESLEME düğmesini birlikte basılı tutun .Cihazlar, 3 saniye aralıklarla otomatik olarak dağıtılacak.

(7) sürecinin sonunda yanıp sönen kırmızı ışıklar ile gösterilir. Tüm aygıtların dağıtımdan sonra düğmeleri serbest bırakın.

(8) tüm cihazlar dağıtılmış gösteren, tüm durum lambaları kırmızı olması için tekrar CHECK düğmesine basın.



Not: tüm cihazlar aktif olduğunda tüm durum lambaları kırmızı yanıp söner.

ISBN: DL:

Appendix E – Post-Assignment Questionnaire for Translators

Research Participant: ____________ Date: …./…./20….

Thank you for taking time to complete this questionnaire, which should take no more than 30 minutes.

All data will be completely anonymous and treated confidentially.

Please answer the following questions. (You can add whatever comments you might have, either next to the questions or at the end of the questionnaire.)

About the Participant

Sex:

☐ Male
☐ Female

Age:

☐ Less than 25 or 25
☐ 26-30
☐ 31-35
☐ 36-40
☐ More than 40

1. What is your education level? Please, at each level, state the subject area you have completed or been studying (e.g. BA: Translation, MA: Politics)

☐ BA (major)______________________________
☐ BSc (major)______________________________
☐ MA (subject)______________________________
☐ MBA______________________________
☐ MSc (subject)______________________________
☐ PhD (subject)______________________________
☐ Other:______________________________

2. How long have you been working as a professional translator?

☐ Less than 2 years or 2 years
☐ From 3 to 5 years
☐ From 6 to 8 years
☐ From 9 to 10 years
☐ More than 10 years

3. Have you ever received formal training in engineering?

☐ Yes (please specify where and how long): ______________________________
____

ISBN: DL:

☐ No

4. Do you have any professional experience in engineering?

☐ Yes
☐ No
☐ Other (please specify):____
____
____

5. If you answered "Yes" to question 4, please state how long you have worked as an engineer.

☐ Less than 2 years or 2 years
☐ From 3 to 5 years
☐ From 6 to 8 years
☐ From 9 to 10 years
☐ More than 10 years

6. What type of texts do you usually translate?

☐ Technical texts
☐ Legal texts
☐ Literary texts
☐ Web-sites
☐ User manuals
☐ Other (please specify): ____
____
____

7. Do you use a translation memory (TM) or machine translation (MT)?

☐ Yes (please state the name of the system(s) and the purpose of use)_______________
____
____
____
____

☐ No

8. Please estimate your average daily throughput when you translate <u>without</u> any translation aid.

☐ Less than 2000 words per day.
☐ Between 2100 and 3000 words per day.
☐ Between 3100 and 5000 words per day.
☐ More than 5100 words per day.
Comment:____
____
____

ISBN: DL:

9. Have you ever used Google Translate?

☐ Yes
☐ No

10. If you answered "Yes" to question 9, please state the purpose of using Google Translate.

☐ To translate electronic mails
☐ To translate pages I come across while surfing the Internet
☐ To make drafts of the texts I will translate
☐ To translate technical documents
☐ To get the gist of any type of texts
☐ To see what a text is like
☐ Other (please specify): ____

____
____
____
____

11. If you answered "Yes" to question 9, please state what you think about Google Translate.

____
____
____
____

12. Have you ever used Google Translator Toolkit (GTT)?

☐ Yes
☐ No

13. If you answered "Yes" to question 12, please state the purpose of using GTT and what you think about it.

____
____
____
____

14. Do you have experience postediting translation memory or machine translation output?

Translation Memory Output ☐ Yes ☐ No

Machine Translation Output ☐ Yes ☐ No

15. If you answered "Yes" to question 14, how long have you been postediting?

Translation Memory Output
☐ Less than 1 year or 1 year

Translation Memory Output
☐ Less than 1 year or 1 year

ISBN: DL:

☐ From 2 to 3 years
☐ From 4 to 5 years
☐ From 6 to 8 years
☐ More than 8 years

☐ From 2 to 3 years
☐ From 4 to 5 years
☐ From 6 to 8 years
☐ More than 8 years

16. If you answered "Yes" to question 14, please estimate the percentage, on average, that postediting represents in a month of your work.

☐ 0 % to 25 %
☐ 26 % to 49 %
☐ 50 % to 74 %
☐ 75 % to 90 %
☐ 91 % to 100 %

17. If you answered "Yes" to question 14, what type of texts do you usually postedit?

____
____
____
____

18. Besides translating, what tasks does your work involve?

☐ Postediting
☐ Revising translations (using both the source and the target texts)
☐ Proofreading (only target texts)
☐ Terminology work
☐ Other (please specify): ____________________________
____
____
____

19. Have you received any training in postediting?

☐ Yes (please state where and how long) ____________________
____
____
____

☐ No

For 20-23, please state your opinion on the given statement.

20. Translators translate/postedit technical texts faster than engineers.

☐ Strongly disagree
☐ Disagree
☐ Neutral

ISBN: DL:

☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

____

____

____

21. While translating/postediting technical texts, translators spend more time on documentation than engineers.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

____

____

____

22. When translators translate/postedit technical texts, the result is more readable (of higher <u>language quality)</u> than when engineers translate/postedit.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

____

____

____

23. When engineers translate/postedit technical texts, the result is of higher <u>terminological quality</u> than when translators translate/postedit.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

____

ISBN: DL:

24. Which process do you think yields better results in terms of the overall <u>quality</u> of the final output?

☐ Translation/postediting by a professional translator + revision of it by a subject-matter expert (i.e. engineer, doctor, lawyer)

☐ Translation/postediting by a subject-matter expert (i.e. engineer, doctor, lawyer) + revision of it by a professional translator.

☐ I do not know.

Comments: ____________________

25. Which process do you think yields better results in terms of the <u>speed with which the process is completed</u>?

☐ Translation/postediting by a professional translator + revision of it by a subject-matter expert (i.e. engineer, doctor, lawyer)

☐ Translation/postediting by a subject-matter expert (i.e. engineer, doctor, lawyer) + revision of it by a professional translator.

☐ I do not know.

Comments: ____________________

About the Postediting Task

1. How easily could you understand the MT output?

☐ Not understandable
☐ Slightly understandable
☐ Understandable
☐ Quite understandable
☐ Fully understandable
☐ I do not know

2. Please estimate how often the following statements describe your postediting procedure in this experiment:

a. As I postedit, I go back and forth and I recheck my postediting before going to the

ISBN: DL:

next sentence.

☐ Never
☐ Rarely
☐ Sometimes
☐ Frequently
☐ Always

b. Immediately after I finish postediting the whole text, I go back to all sentences and review them one by one again.

☐ Never
☐ Rarely
☐ Sometimes
☐ Frequently
☐ Always

c. ☐ Other (please explain): ____
____
____
____

3. Based on your postediting experience (either previous or from this experiment), do you prefer MT postediting to translating?

☐ Never
☐ Rarely
☐ Sometimes
☐ Frequently
☐ Always

Comments: ____
____
____
____

4. Based on your postediting experience (either previous or from this experiment), which of the options below best represents your experience?

☐ Postediting, for me, requires the *same* effort as translating
☐ Postediting, for me, requires *more* effort than translating
☐ Postediting, for me, requires *less* effort than translating
☐ I do not know

Comments: ____
____
____
____

5. Please briefly state the advantages and disadvantages of MT postediting.

ISBN: DL:

Advantages:

*a.*____

*b.*____

*c.*____

Disadvantages:

*a.*____

*b.*____

*c.*____

6. Please briefly mention the problems you came across during the postediting task.

____

7. Is there anything else that you would like to add that we might not have asked you? Please, feel free to comment.

____

ISBN: DL:

Appendix F – Post-Assignment Questionnaire for Engineers

Research Participant: ____________ Date: …./…../20….

Thank you for taking time to complete this questionnaire, which should take no more than 30 minutes.

All data will be completely anonymous and treated confidentially.

Please answer the following questions. (You can add whatever comments you might have, either next to the questions or at the end of the questionnaire.)

About the Participant

Sex:

☐ Male
☐ Female

Age:

☐ Less than 25 or 25
☐ 26-30
☐ 31-35
☐ 36-40
☐ More than 40

1. What is your major in engineering?

☐ Mechanical Engineering
☐ Electrical-Electronics Engineering
☐ Computer Engineering
☐ Industrial Engineering
☐ Metallurgy and Materials Engineering
☐ Other (Please specify): ____

2. What is your education level?

☐ BSc
☐ MBA
☐ MSc (subject)
☐ PhD (subject)

3. How long have you been working as an engineer?

☐ Less than 2 years or 2 years
☐ From 3 to 5 years
☐ From 6 to 8 years
☐ From 9 to 10 years
☐ More than 10 years

ISBN: DL:

4. Have you received any formal training in translation?

☐ Yes (please specify where and how long): ______________________________

____

☐ No

5. Do you have any professional experience in translation?

☐ Yes
☐ No
☐ Other (please specify): ______________________________

____

____

6. If you answered "Yes" to question 4, please state how long you have translated/been translating.

☐ Less than 2 years or 2 years
☐ From 3 to 5 years
☐ From 6 to 8 years
☐ From 9 to 10 years
☐ More than 10 years

7. If you answered "Yes" to question 4, please state the type of texts you have translated/been translating.

☐ Automotive related texts
☐ Computer texts
☐ Technical manuals
☐ Other (please specify): ______________________________

____

____

8. Do you need to do translation as part of your work as an engineer in your company?

☐ Yes
☐ No
☐ Other (please specify): ______________________________

____

____

9. If you answered "Yes" to question 8, please state the type of texts you have been translating.

☐ Electronic mails
☐ Minutes of meetings
☐ Reports
☐ Technical documents
☐ Other (please specify): ______________________________

____

ISBN: DL:

10. Have you ever used machine translation (MT)?

☐ Yes (please specify which MT system you used and what you think about it) ______
____
____

☐ No

11. Have you ever used Google Translate?

☐ Yes (please state what you think about it) ____________________
____
____

☐ No

12. If you answered "Yes" to question 10 or 11, please state what purposes you use MT for.

☐ To translate electronic mails
☐ To translate texts I come across surfing the Internet
☐ To provide draft translations
☐ To translate technical documents
☐ To get the gist of any type of texts
☐ To see what a text is like
☐ Other (please specify): ____________________
____
____

13. Have you ever used Google Translator Toolkit (GTT)?

☐ Yes
☐ No

14. If you answered "Yes" to question 13, please state the purpose of using GTT and what you think about it.

____
____
____

15. If you use MT, how do you use it?

☐ I use raw MT output
☐ I postedit MT output myself
☐ I have the MT output postedited by a professional translator
☐ Other (please specify): ____________________

For 16-19, please state your opinion on the given statement.

16. Translators translate/postedit technical texts faster than engineers.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

17. While translating/postediting technical texts, translators spend more time on documentation than engineers.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

18. When translators translate/postedit technical texts, the result is more readable (of higher <u>language quality)</u> than when engineers translate/postedit.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____

19. When engineers translate/postedit technical texts, the result is of higher <u>terminological quality</u> than when translators translate/postedit.

☐ Strongly disagree
☐ Disagree
☐ Neutral
☐ Agree
☐ Strongly agree
☐ I do not know

Comments: ____
____
____
____

20. Which process do you think yields better results in terms of the overall quality of the final output?

☐ Translation/postediting by a professional translator + revision of it by a subject-matter expert (i.e. engineer, doctor, lawyer)

☐ Translation/postediting by a subject-matter expert (i.e. engineer, doctor, lawyer) + revision of it by a professional translator.

☐ I do not know

Comments: ____
____
____
____

21. Which process do you think yields better results in terms of the speed with which the process is completed?

☐ Translation/postediting by a professional translator + revision of it by a subject-matter expert (i.e. engineer, doctor, lawyer)

☐ Translation/postediting by a subject-matter expert (i.e. engineer, doctor, lawyer) + revision of it by a professional translator.

☐ I do not know

Comments: ____
____
____
____

About the Postediting Task

1. How easily could you understand the MT output?

☐ Not understandable
☐ Slightly understandable

ISBN: DL:

☐ Understandable
☐ Quite understandable
☐ Fully understandable
☐ I do not know

2. Please estimate how often the following statements describe your postediting procedure in this experiment:

a. As I postedit, I go back and forth and I recheck my postediting before going to the next sentence.

☐ Never
☐ Rarely
☐ Sometimes
☐ Often
☐ Always

b. Immediately after I finish postediting the whole text, I go back to all sentences and review them one by one again

☐ Never
☐ Rarely
☐ Sometimes
☐ Often
☐ Always

c. ☐ Other (please explain): ____
____
____
____

3. Based on your postediting experience (either previous or from this experiment), do you prefer MT postediting to translating?

☐ Never
☐ Rarely
☐ Sometimes
☐ Often
☐ Always

4. Based on your postediting experience (either previous or from this experiment), which of the options below best represents your experience?

☐ Postediting, for me, requires the *same* effort as translating
☐ Postediting, for me, requires *more* effort than translating
☐ Postediting, for me, requires *less* effort than translating
☐ I do not know

Comments: ____
____

5. Please briefly state the advantages and disadvantages of MT postediting.

Advantages:

a.

b.

c.

Disadvantages:

a.

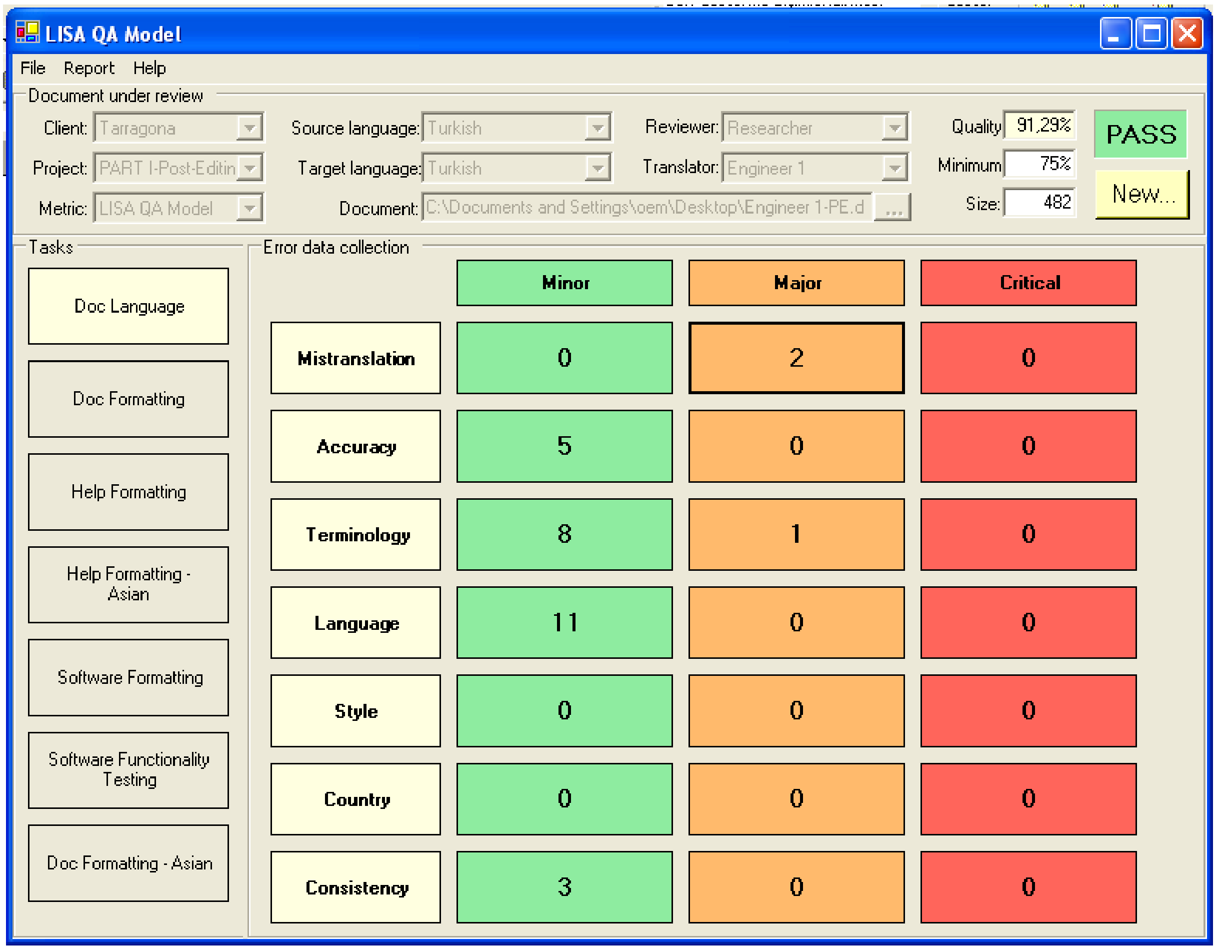
b.

c.

6. Please briefly mention the problems you came across during the postediting task.

7. Is there anything else that you would like to add that we might not have asked you? Please, feel free to comment.

Appendix G – LISA QA 3.1 Interface



ISBN: DL:

Appendix H – LISA QA 3.1 - Sample Project Review Report

Project Review Report

Client: Tarragona	**Source Language:** Turkish	**Reviewer:** Researcher	
Project: PART I-Post-Editing	**Target Language:** Turkish	**Translator:** Engineer 1	
Document: ...ents and Settings\oem\Desktop\Engineer 1-PE.doc		**Minimum:** 75%	
Metric: LISA QA Model		**Size:** 482	

Task	Error Category	Severity Level	Error Count
Doc Language	Mistranslation	Major	2
Doc Language	Accuracy	Minor	5
Doc Language	Terminology	Minor	8
Doc Language	Language	Minor	11
Doc Language	Terminology	Major	1
Doc Language	Consistency	Minor	3

Total error points: 42
Translation quality: 91.29%

PASS