The need for practice in the acquisition of the post-editing skill-set

Lessons learned from the industry

Clara Ginovart Cid

TESI DOCTORAL UPF / year 2021

THESIS SUPERVISORS
Dra. Carme Colominas
Dr. Antoni Oliver
Department Departament de Traducció i Ciències del Llenguatge

A muns iaios, Paca i Pere, i Teresina i Juanito, que tant me van cuidar.

A muns pares, que m'han ensenyat a viure per ser feliç.

Agraïments

To all the persons that contributed in any way to this research or my well-being in the past years, named below or not, I would like to thank you all.

En el ámbito académico, me gustaría agradecer la ayuda y oportunidades que me han brindado, desde la Faculté de traduction et d'interprétation de l'Université de Genève, la profesora Marta Villacampa y la Dra. Pierrette Bouillon. Un reconocimiento especial a los estudiantes cuya implicación en el experimento ha permitido llevar a cabo la investigación.

A los expertos, por serlo y por vuestra paciencia y ayuda: Anabel Blasco, del Servei d'Estadística Aplicada de la UAB: I don't think I hate statistics anymore; Dr. LianTze Lim and the team at Overleaf; y Motoblouz y Adrià Morgó, por su aportación, decisiva, desde el sector industrial.

Por descontado, la tesis no hubiera sido posible sin la participación, crucial, de los 66 representantes de empresas, los 142 traductoras y traductores, y los 54 profesores universitarios. Agradezco especialmente el interés y disponibilidad a los 48 entrevistadas y entrevistados.

Depuis Datawords, j'ai reçu plein d'encouragement et de la participation active dans la partie expérimentale. J'aimerais remercier les équipes anglaise et italienne pour leur contribution dans les articles qui ont marqué le début de mes années de recherche. Je suis également reconnaissante à Marie-Santhia et à l'ancienne référence ISO, Irina, pour leur support. Un merci tout particulier à mes +1 au boulot : Marina Frattino et Silvia Giampiccoli, grazie mille pour votre accompagnement.

Debo mi más sincero agradecimiento al equipo de traducción español, tanto de Levallois-Perret, como de Barcelona. Todas y cada una de las personas de este gran equipo, antiguos y actuales, junto con los traductores autónomos, constituyen una parte esencial en el estudio experimental. Gracias por vuestra pacienca con el piloto, el warm-up, el pretest, la formación y el postest. Gracias por vuestra dedicación y esfuerzo. Lo mismo les digo a mis querídismes amigues ginebrines, que de ginebrinos no tienen nada, más que la ciudad de residencia y la elegancia de sus personalidades. Me tomo los mates a vuestra salud.

Finalmente, a mis queridos supervisores, Dra. Carme Colominas y Dr. Antoni Oliver, que han pasado tantas horas leyendo mis correos y hablándole a la cámara de Skype, aconsejándome y ayudándome a aprender, a progresar. Gracias por dudar y decidir conmigo en los momentos clave.

A la meva estimadíssima familia, que m'acompanya sempre, vagi on vagi. Merci à toi, Alexis Fouquet, je t'ai trouvé plus tard au long du chemin, mais je t'aime de tout mon coeur. I gràcies als meus pares, Montse Cid i Gabriel Ginovart, no seria jo sense vatros.

Abstract

While (translation) technology and the language industry continue to evolve, varieties of professional profiles and nuances to the relevant skill-sets emerge. The field of machine translation post-editing, and, in particular, research on the learnability of its skills, has become one important focus of attention.

With an exploratory sequential mixed-methods design, this dissertation seeks to attain two goals. First, in the survey-research, the identification of errors, the decision-making, and the respect of guidelines are confirmed as core post-editing skills from an efficient editing standpoint. Second, in the experimental study based on a pretest-posttest method, we evaluate the treatment, that is, a proposal for a French-Spanish split-training course on the three mentioned skills. The data analyses indicate positive results from the quality perspective and suggest that longer practice is required to reach proceduralisation.

Resumen

Mientras las tecnologías (de la traducción) y la industria lingüística siguen evolucionando, emergen variedades de perfiles profesionales cuyas habilidades se ven matizadas. El campo de la posedición de traducción automática y, en particular, la investigación en lo que respecta a la adquisición de las habilidades adecuadas, se han convertido en focos de atención primordiales.

Con una metodología secuencial exploratoria mixta, en esta tesis se plantean dos objetivos. El primero, confirmar las principales habilidades de posedición desde una perspectiva de edición eficiente gracias a tres encuestas. Se concluye que son la identificación de errores, la toma de decisiones, y la aplicación de directrices. El segundo, evaluar una propuesta de formación fragmentada (*split-training*) en el par de lenguas francés-castellano sobre las tres habilidades mencionadas mediante un estudio experimental con método pretest-postest. Los resultados son positivos en cuanto a calidad traductológica, y sugieren que con más práctica y mayor experiencia se puede llegar a un grado de proceduralización adecuado en el ámbito profesional.

Resum

Mentre les tecnologies (de la traducció) i la indústria lingüística segueixen evolucionant, emergeixen varietats de perfils professionals i les habilitats que se'ls vinculen es van matisant. El camp de la post-edició de traducció automàtica i, en particular, la recerca sobre l'adquisició de les habilitats pertinents, s'han convertit en focus d'atenció cabdals.

Amb una metodologia seqüencial exploratòria mixta, aquesta tesi té dos objectius. En primer lloc, la recerca mitjançant enquestes permet confirmar les principals habilitats de post-edició des d'una perspectiva d'edició efficient: la identificació d'errors, la presa de decisions, i l'aplicació de directrius. En segon lloc, l'estudi experimental basat en el mètode pretest-postest avalua el tractament, és a dir, la proposta d'una formació fragmentada (*split-training*) aplicada al parell d'idiomes francès-castellà sobre les tres habilitats esmentades. L'anàlisi conclou resultats positius quant a qualitat i suggereix que, per assolir un grau de proceduralització més adient, cal incrementar la pràctica.

Introduction

Motivation

In my very first year of professional experience, as a translation project manager in a start-up, the way Neural Machine Translation (NMT) was "radically alter[ing] the language industry landscape" (Angelone et al. 2019) was patent. The volumes of words were higher than one could have ever imagined during the Master studies, and the deadline, always "yesterday".

I quickly learned the acronym "asap" and I appreciated the necessity for a translation Project Manager (PM), to have cross-functional skills and switch quickly from one role to another. In that sense, some years later, a Slator[1] article would appear confirming that language industry jobs could add up to around 600 titles (Bond 2018). While juggling with sequences of varying tasks, that is, with my role of PM, terminologist, translator, post-editor, reviewer, proofreader, vendor manager... I realised that, despite my postgraduate training in Translation Technologies, I lacked tools to explain to the stakeholders of my projects (director or chief executive officer, client, translators, and others), how post-editing of machine translation could be performed. Post-editing of Machine Translation (PEMT), or Machine Translation Post-editing (MTPE)[2], has now evolved since then thanks to major contributions to the field that will be presented in part I.

In the past decade, there has been an exponential growth of the Post-Editing (PE) service. Indeed, more Language Service Companies (LSCs) are using machine translation in their professional workflow (Blagodarna 2019, p. 3). This can be explained by the higher quality of the outputs, which is well reflected in recent reports on the value of the machine translation market reaching ranges from USD130 million to USD400 million (Nimdzi 2019). To understand how and why PE has become one of the main demanded services and to establish basic terminology[3] for the present

[1] https://slator.com/

[2] Syntactically more adequate in English.

[3] It should be noted that the industry-based discourse is not "clear-cut" and is "far from uniform", as commented by Mazur (2009). The main aim here is not to theorise on strict definitions, this is why, in the present research, it is possible that some terms are used inter-

dissertation, in the following paragraphs one recent publication is quoted on three occasions to summarise and reflect the state of the art in the PE market before moving to the Theoretical Framework in part I.

> The introduction of MT technology has caused some disruption in the translation community [...] mainly due to the quality of the output to post-edit, on occasions too low to be of benefit, [or to] the time allowed for the PE task, with sometimes overly optimistic deadlines, and the price paid for the PE assignment. (Guerberof Arenas 2019, p. 333)

The technological progress[4] and its impact on the translator job has been observed within academia (e.g. Kornacki 2018), but also within the industry. Indeed, as Guerberof Arenas (2019, pp. 343–344) reminds the reader:

> Many companies have presented results from the work done internally using MT in combination with PE (or without PE) at conferences such as Localization World[5], GALA[6] and the TAUS forums[7], or at more specialized conferences such as AMTA[8], EAMT[9] or the MT Summit[10]".[11]

Despite the numerous publications on MT and PE, it is worth noting that the research on these fields is still considered to be in an emergent trend:

> Empirical research on the impact of machine translation and artificial intelligence on translation and interpreting is still in its infancy. This is even more the case when it comes to research on its impact in domains such as terminology management [...], pre- and post-editing [...] and project management. (Angelone et al. 2019, p. 4)

It is certain that the inclusion of post-editing in *European Master's in Translation Competence Framework* (EMT Network 2017, p. 7) as part of Translation Competence (TC) marks the patent recognition by scholars of the evolution of the language industry.

---

changeably (or not) depending on the context.

[4]The Nimdzi Language Technology Atlas (the old versions and recently the edition for 2020) is a real-life example of the variety and evolution in the field of translation technologies (Nimdzi 2020).

[5]https://locworld.com

[6]https://www.gala-global.org

[7]https://www.taus.net

[8]https://amtaweb.org

[9]http://eamt.org

[10]https://www.mtsummit2019.com

[11]The footnotes to the websites have been added.

Datawords and the Industrial Doctorate

The possibility of carrying out the research motivated in the previous section in an industrial setting can be seen as a cross-sector partnership. The Industrial Doctorate[12] allows for this possibility. It enables innovation and competitiveness within one of the company's strategic research projects while bringing technology and knowledge transfer into the thesis at the University. It is advantageous from the perspective of Datawords[13], as innovative methods are applied to improve the translation workflows, and from the perspective of research institutions, since the knowledge and technology transfer is enriched thanks to real data and resources. In fact, the main experimental study is set up and conducted within Datawords.

The Industrial Doctorate grant allocated to the development of this thesis has the reference 2017 DI010. In order to provide context to the present thesis, in the following paragraphs a brief introduction to Datawords Group is provided.

Datasia was founded in 2000 in Saint-Ouen, France, by "a multicultural group of acquantainces" as one of the founders, Alexandre Crazover, relates in his Wordpress blog in French[14]:

> Une équipe d'associés aux origines diverses (un Français, une Chinoise, une Japonaise, un Suisse Italien et un Indien) pour fonder DATASIA (Direct Access To Asia), un groupe spécialisé en communication en langues asiatiques. (Crazover 2010)

Datawords appears in 2002 to fulfill the expressed requests of their clients of specializing in translation of European languages too. Datawords' headquarters are today in Levallois-Perret (France) and the services provided include multilingual technology, international marketing, web integration, linguistic services, and multicultural consulting, among others. Recently, it became Datawords Group[15] with the acquisition of agencies that enforce or add to the cited portfolio (such as Vanksen, Whatsquare, and 87seconds), and the development of a proprietary localization solution: Wezen.

The progress made by Datawords has escalated these last years with three CEO at its head: Stanislas de Nervo, Alexandre Crazover, and Didier Rosenberg. Datawords has over 500 employees who represent over 40 nationalities and more than 50 languages and cultures. The company is established in France, Italy, Spain, Germany, the United Kingdom, South Korea, Hong Kong, Philippines, Japan, the United States, Canada, and Mexico.

[12] http://doctoratsindustrials.gencat.cat/en

[13] https://www.datawordsgroup.com

[14] http://xl8.link/Crazover2010

[15] Hereinafter "Datawords".

The department New Language Technologies, where I have been working as a Computer-Aided Translation (CAT) and MT Tool Consultant since September 2016[16], was created in 2013 by Marina Frattino and is part of the Multilingual Department[17]. The firm has been renewing the ISO 9001 certification for quality since 2008[18], which is especially accounted for in the offering of MT and PE since 2014. One of the first clients to become interested in the MTPE service is Motoblouz. Motoblouz is a commercial distributor and producer of motorbike accessories and gear for motorbikes[19]. This customer has been trusting Datawords for the localisation of their website since then.

Research Objectives and Research Design Framework

The research objectives of this thesis are twofold. Firstly, three questionnaires aim at setting the landscape in regards of the PE activity and the corresponding professional profile(s) in the current language industry. To do so, the research questions (RQs) are established in chapter 5. Secondly, our objective is to determine the extent to which a split-training course[20] may or may not enhance the PE performance of semi-professional translators. The design is tailored to the Motoblouz project at Datawords and the results set the path to continue the work in the field of PE pedagogy, especially in market-oriented settings.

In terms of design framework, following J. W. Creswell and J. D. Creswell (2017), the overall method is sequential exploratory. It contains embedded two mixed-methods designs: one, presented in part II, is similar to a sequential explanatory design (Two-Phase); and the second one, presented in part III, resembles an exploratory sequential (Three-Phase) design. To sum up, we could name the present research design framework as "Exploratory Sequential (Five-Phase Design)" and represent it as in figure 1.

In words of J. W. Creswell and J. D. Creswell (ibid., p. 216):

> At a practical level, mixed methods provides a sophisticated, complex approach to research that appeals to those on the forefront of new research procedures. It also can be an ideal approach if the researcher has access to both quantitative and qualitative data. At a a procedural level, it is a useful strategy to

[16] And since 2019 as Machine Translation Solutions Manager.
[17] Internally abbreviated "ML" and often referred to as "linguistic production".
[18] *ISO 9001:2008*.
[19] https://corporate.motoblouz.com/en
[20] A concept that will be defined in section 4.1.

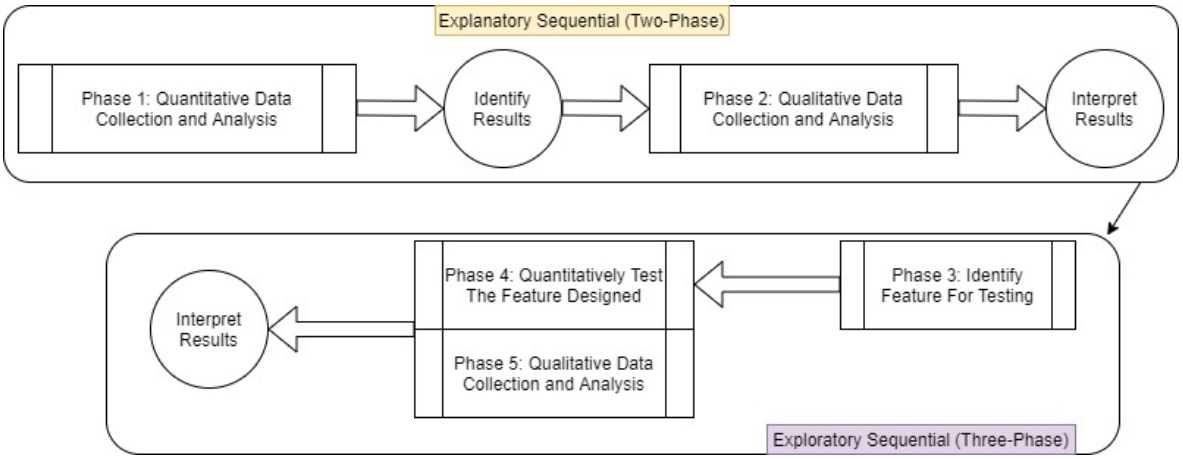

Figure 1: Research Design Framework: Exploratory Sequential (Five-Phase)

> have a more complete understanding of research problems and questions [...].

Thus, the quantitative results from the survey-based research (see chapter 6) along with the qualitative results from the syllabi analysis and the interviews (see sections 6.2 and 6.3, respectively), help at identifying the features for testing (see the Final thoughts of part II). With such features identified, the instrument is designed as described in chapter 9, and quantitatively tested.

Qualitative data is also collected to further interpret the usefulness of the tested instrument, also known as "treatment" in a pretest-posttest setting. In our case, the treatment is the PE split-training course[21] and the discussion of outcomes is provided in chapter 10 and in section Final thoughts of part III.

Limitations of this Research

In terms of scope, the present research is limited to PE as a service, that is, an offer made and bought by a customer. While MT-aided translation and even revision are activities very close to one another and to PE, the focus of this dissertation does not fall on them.

One limitation of this thesis affects part II. As a matter of fact, in our survey-based research carried out through three questionnaires, the number of submissions received could be considered a medium-sized sample. We

[21] Named Post-editing Practice Application (PEPA).

received 66 responses from industry stakeholders and 142 submissions from linguists. When contrasted to the 286 responses from language companies and 946 from individuals at the 2019 edition of the Language Industry Survey (EUATC et al. 2019), it can be considered a limitation. However, the cited survey also states that "less than 20% of the translation companies report that they are frequently using machine translation". Moreover, the 54 studied syllabi in section 7.2 along with the 49 interviews[22] in section 7.3, contribute with qualitative data and, thus, counter the mentioned limitation.

In terms of methodology design, the lack of previous expert knowledge by the main researcher in some specific fields constitutes another limitation. For example, such limitation could affect the definition of criteria for recruiting professionally, as discussed in part II, especially in regards of table 6.2. As argued in part III, a third limitation can be found in the methodology of the pretest-posttest study. As a matter of fact, there is no control group within the experimental setting of our split-training provision. There are two main reasons why the possibility of a control group was disregarded. Firstly, the availability of resources. Without financial motivation, the recruitment of volunteers to an experiment in a language pair without English can become challenging. Secondly, the logical reasoning behind certain peer-reviews upon conference paper submissions. Fellow researchers argued that a certain treatment (split-training) of any kind would undoubtedly have a positive impact in the performance and results of the posttest, leading to conclude that a control group was not necessary. Finally, a limitation that affects every research in the field is linked to the variability between individuals. For instance, in our experiment in part III, the selected participants do not undergo a proficiency test regarding ICT skills (Kornacki 2018, pp. 71–72).

Structure of this Dissertation

The present thesis is structured in three parts. Part I presents an overview of the theoretical framework. Its four chapters review the state of the art on NMT, PE, PE skills, and PE training for translators (see chapters 1, 2, 3, and 4, respectively). It ends with some conclusive remarks (see Final thoughts).

Part II constitutes the survey-based research. It sets out the research RQs in chapter 5. In chapter 6 we describe the methodology for the questionnaires. The results are discussed in chapter 7, before closing the part with a short reflection on the impact of the survey outcomes (see Final thoughts).

In part III, three chapters deal with the experimental study. First, chapter 8 sets out the relevant RQs for this part. Second, chapter 9 establishes the

[22] With 48 interviewees, as one educator taught two PE courses.

methods applied to design and perform the experimental study. Finally, in chapter 10, the results of the experiment are examined. The part is closed with some concluding remarks in section Final thoughts.

The present thesis finishes with the Conclusions, which covers in retrospective the main hypotheses of this research. In the Conclusions chapter, the future lines of investigation are also highlighted.

Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence

AER Actual Edit Rate

ASR Automatic Speech Recognition

ATA American Translators Association

CAT Computer-Aided Translation

CDI Calibration of Dichotomous Items

CL Controlled Language

CNL Controlled Natural Language

CPD Continuing Professional Development

DQF Dynamic Quality Framework

ECTS European Credit Transfer and Accumulation System

ED Edit Distance

EES Efficient Editing Skill

EMT European Master's in Translation

EPH Edits per Hour

FAHQT Fully Automatic High Quality Machine Translation

HAMT Human-Aided Machine Translation

HT Human Translation

IIS Instrumental Input Skill

ICT Information and Communications Technology

LSCs Language Service Companies

MT Machine Translation

MAHT Machine-Aided Human Translation

MQM Multidimensional Quality Metrics

MTPE Machine Translation Post-editing

MTQE Machine Translation Quality Estimation

NMT Neural Machine Translation

NTI Negative Translatability Indicators

PACTE Process of Acquisition of Translation Competence and Evaluation

PE Post-Editing

PEA Post-editing Action

PEPA Post-editing Practice Application

PIE Preselected Item Evaluation

PM Project Manager

QA Quality Assurance

RegEx Regular Expressions

RBMT Rule-based Machine Translation

SaaS Software as a Service

SEA Servei d'Estadística Aplicada

SMT Statistical Machine Translation

ST Source Text

TAP Think-Aloud Protocol

TAUS Translation Automation User Society

TES TAUS Efficiency Score

TB Termbank or Termbase

TC Translation Competence

TCC Translator-relevant Computer Competence

TEP Translation-Editing-Proofreading

TER Translation Error/Edit Rate

TES TAUS Efficiency Score

TM Translation Memory

TQA Translation Quality Assessment

TS Translation Studies

TT Target Text

WPH Words per Hour

WPS Words per Second

Part I

Theoretical Framework

Despite the existence of a variety of ontologies in the field of Translation Studies (TS) and also in the subfield of translation technologies, there are benefits to having clear taxonomies (Kenny 2020, p. 17). Hence, even if we must acknowledge their constructed nature and the need to constantly adapt to changes over time, we explore below how this thesis is framed within the research done until date.

If we consider the maps generally accepted[23] by TS scholars, such as the one proposed by Holmes (1988) and the more well-known diagram drawn by Toury (2012)[24], it could be stated that the present thesis is framed within process-oriented studies, part of Descriptive TS, but more relevant to the Applied TS (which include Translator Training and Translation Aids). Indeed, we investigate Translator-relevant Computer Competence (TCC) (Kornacki 2018, p. 42) and, according to the more recent alternative proposed by Vandepitte (2008), our work finds its place within the studies of translation profession research.

In particular, we study a recent branch of research closely related to the two above-mentioned fields (TCC and translation profession). This branch concerns the service that comes after pretranslating a file with Machine Translation (MT): Post-Editing (PE). In fact, we do not investigate the Machine Translation Post-editing (MTPE) process, but rather which translator/translation skills are core to the PE activity, if there are any new skills, and how they can be acquired. Thus, it is also relevant to the branch of studies within translation competence development research. Indeed, Massey et al. (2019, p. 211), citing Piotrowska and Tyupa (2014), claimed that translation pedagogy is a sub-discipline of its own. All things considered, the present thesis is framed at the intersection of the above-mentioned fields: PE pedagogy.

In chapter 1, an overview of the history of translation technology is presented to conclude how the evolution of MT and its intertwining with Translation Memory (TM) and Computer-Aided Translation (CAT) tools has affected and keeps affecting the translation profession. Chapter 2 reviews the general characteristics of PE and the state of the art. First, from a professional perspective (job descriptions, tasks and profiles, difference with revisers and translators, etc.). Second, from an empirical perspective (research in PE effort, PE operations, productivity, etc.). And finally, from an "instrumental" perspective; probably at the crossroads between the em-

[23] As argued in Gambier and Van Doorslaer (2009), these maps are often criticised by the simplicity of their relationships and they are unavoidably incomplete. The following lines are not intended as a reference or definition of the discipline or subfield in question, but rather as a panorama view to readers who could regard it as a helpful prop.

[24] The combination of both is sometimes referred to as the "Holmes/Toury map".

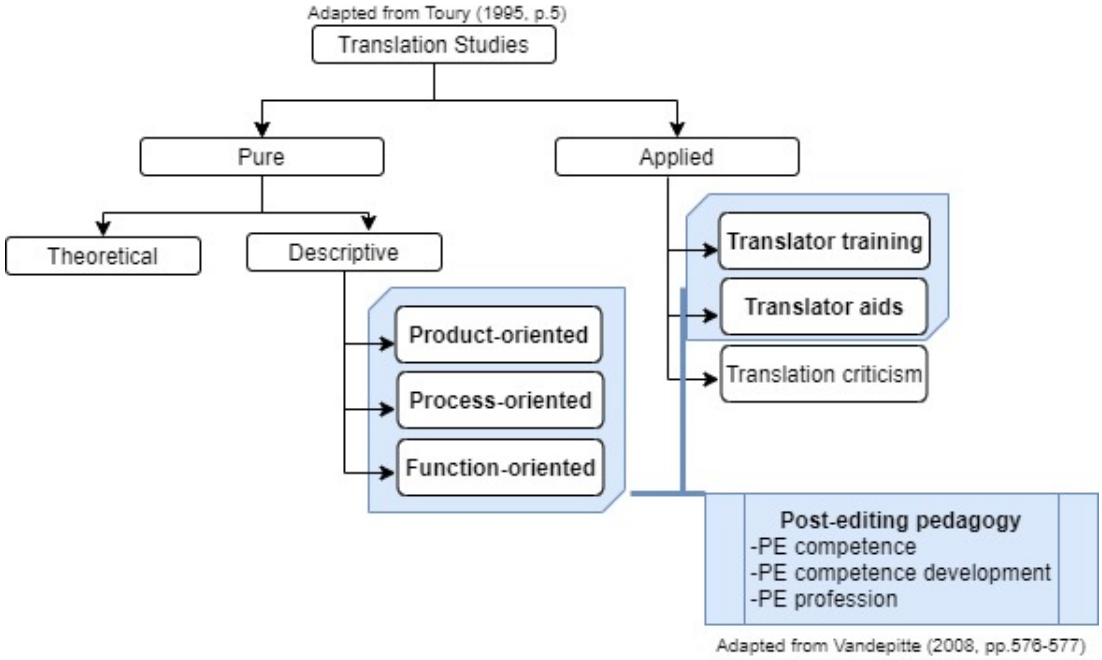

Figure 2: Post-editing pedagogy studies

pirical and the industry areas: PE assignment[25]. In chapter 3, we delve into the post-editing competence or skill-set, by summarizing previous work in this field, at the same time that we acknowledge how such a discussion cannot possibly provide definitive and comprehensive answers to the varied challenges currently being researched by the community (competence or skill, to what extent it is or not language-dependent, to what extent it is or not similar to translation[26] or revision, when it is best introduced during graduate studies...). In chapter 4, the work published until present about practical training on PE for translators is reviewed. A difference is made between training courses in universities, and professional modules within the industry. To conclude, in the last chapter of the present part (Final thoughts) we provide a summary of the literature review presented in its four preceding chapters. Such summary will set the path to the survey-based research of part II.

[25] Also named "brief" or "commission", as explained in Fraser (2000, p. 53).

[26] Now often called Human Translation (HT).

Chapter 1

MACHINE TRANSLATION

1.1. Brief Review of MT History

Research on MT will soon be one century old. Even so, translation technology research and specifically the field of CAT or its main technology, TM, was truly developed in the 1990s. Authors such as J. Hutchins (2007), Kornacki (2018, p. 97-106), and Oliver (2016) have provided summaries of the history of one or both fields. Considering that the main research topic of this dissertation does not fall on MT alone, the evolution of the area is summarised below with the list of the main milestones:

- 1930s - Beginning of modern MT history (with George Artsrouni's device and Petr Petrobich Troyanskii's formulation).
- 1949 - Warren Weaver's memorandum (with the subsequent work by Reifler [1950] and the statistical approach by Abraham Kaplan).
- 1950s - First MT conference (at the Massachusetts Institute of Technology, reported in Dodd [1952] and first demonstration, in 1954, when Georgetown and IBM present an MT working system live).
- 1966 - Automatic Language Processing Advisory Committee (ALPAC) report, which brought up the scepticism around MT and research in the field slowed down (J. Hutchins 2003).
- 1976 - Météo in Canada, Systran in the European Commission.
- 1980s - Rule-based Machine Translation (RBMT) (such as American Weidner, METAL, or EUROTRA) and first proto-CAT tool (Translation Support System by Automated Language Processing Systems).
- 1990s - Multi-purpose personal computers become popular and affordable (Kornacki 2018, p. 103); Machine-Aided Translation (MAT) and Moses[1] (Koehn, Federico, et al. 2007) appear.
- 2006 - MT becomes consumer goods, for which a good example is Google Translator with its Statistical Machine Translation (SMT) technology. However, as noted by Oliver (2006, p. 9) and Christensen

[1] http://www.statmt.org/moses

and Schjoldager (2016, p. 92), towards the end of the 2000s MT was not integrated (or rarely) into TM suites.

- 2013 - LeBlanc (2013), cited in Christensen and Schjoldager (2016, p. 92), report a "widespread use of TM technology, regardless of genre or subject-matter". Full TM/MT integration is reported by authors like Sanz Villa (2015).
- 2016 - First Neural Machine Translation (NMT) systems for which Google can again be cited as a well-known example. There is a refinement of TM features and their integration with MT (e.g. adaptive MT[2]) and the Software as a Service (SaaS) model becomes mainstream.
- 2017 to present - DeepL Translator[3] is presented by the company that had founded, in 2009, Linguee[4]. Several open-source NMT "toolkits" appear[5]. Research towards seamless integration and development of language technologies leaps into a new era of optimism, the so-called movement from Convergence into Singularity, commented in the concluding chapter of the present part (Final thoughts).

A visual overview of the history of MT has also been published by Translation Automation User Society (TAUS)[6]. TAUS publishes regularly a Translation Technology Landscape Report where the trends in the translation industry regarding technology are extensively presented, for instance the first edition: "From Luxury to Utility" (TAUS 2013). In the following paragraphs, the discussion about the evolution of MT since the 1960s allows us to introduce the different types of MT technologies. As presented above, the first MT systems started operating in the 1960s (Kenny 2020, p. 5), just before the ALPAC report was released.

After the pessimist outlook, and thanks to the research that nevertheless took place in countries other than the USA (e.g. Canada, France or Germany), in the 1980s, with the first good-enough results of RBMT output[7] (at least good enough to be editable by a professional translator), PE flourished as a new research field. Some examples of the first successful and well-known PE studies are those of Vasconcellos (1986), Vasconcellos (1987), and Vasconcellos (1988) and the seminal work by Krings (2001). In the former, the author, Chief of Terminology and Machine Translation Program at the time at the Pan American Health Organization (PAHO), describes the

[2]Daems and Macken (2019).

[3]https://www.deepl.com

[4]https://www.linguee.es

[5]Such as Junczys-Dowmunt et al. (2018)or Klein et al. (2017).

[6]https://www.taus.net/academy/timelines/translation-automation-timeline

[7]In the context of PE, the MT output can also be called "seed translations" (Van Ess-Dykema et al. 2010).

strategies used at the international organisation to facilitate the translators interaction with MT. The author emphasises that a post-editor can reach a sensible level of maturity in their PE skills after post-editing 100,000 words, and observed how some professionals at the PAHO produced 10,000 words in a working day (p. 145). In the latter, it is concluded that PE may be more complex than translating, given the third text[8].

In the late 1970s and until the late 1980s a new discipline called "computational linguistics" (Vasconcellos 1988, p. 3) appeared, which has the main objective of managing linguistic data with technological aids, for instance, for purposes of natural language processing. MT systems at the time were RBMT systems. Such engines contain grammatical and lexical rules for the source language, grammatical and lexical rules for the Target Language (TL), and a transfer set of grammatical and lexical rules (bilingual dictionary) for the relevant language pair. However, the development of such rules and RBMT systems, and their refinement, was soon too expensive for the results obtained (Kornacki 2018, p. 100).

At the post-computer era (Mossop 2006a), with the appearance of large mono- and bilingual corpora, easily accessible on the Internet, and thanks to the increase in computing power, we saw a rapid development of SMT. An SMT system requires large aligned corpora to calculate the probability of a target block of words (phrase) appearing as a translation of a source phrase, along with large monolingual corpora in the target language to calculate the probability of a target phrase appearing next or near another phrase. Kornacki (2018, p. 101) presents a table comparing the main drawbacks and positive points of each system: RBMT is consistent and predictable, robust, and knows grammatical rules, but it lacks fluency and does not handle well exceptions to rules. Despite its higher fluent outputs and its cost-effectiveness, SMT shows more inconsistencies, it is very corpora-dependent, it does not "know" grammar, and its quality is unpredictable. Yet, with the improvements made in the output of SMT systems, semi-automatic MT (another term for the already mentioned "post-editing") began to gain importance both within the industry and academia. Finally, Hybrid systems combine the best features of RBMT with SMT, leading to even more qualitative outputs and another reason to keep developing the PE activity.

W. J. Hutchins and H. L. Somers (1992) classified technologies depending on the degree of mechanisation they offered, leading to three main groups:

[8]In regular translation, we had Source Text (ST) and Target Text (TT), with MTPE we have the ST (probably fragmented or constantly updated), the MT output, and the post-edited text we deliver.

- Fully Automatic High Quality Machine Translation (FAHQT) a type of Fully Automatic Machine Translation (FAMT)
- Human-Aided Machine Translation (HAMT)
- Machine-Aided Human Translation (MAHT)

Even with hybrid systems, FAHQT was still unattainable. In fact, some "prominent individuals" (Kornacki 2018, p. 96), such as Bar-Hillel, had already considered that the very foundation of MT was, in principle, impossible. Thus, even if research on MT did not stop, the acknowledgment of the inherent limits of the quality of MT outputs led to further develop CAT tools and TM (ibid. pp. 99–100).

Since the 1980s, research on MT and CAT has been ongoing in parallel. Kenny (2020, p. 8) highlights the publication of seminal textbooks about CAT: *Electronic Tools for Translators*, by Austermühl (2001); and Bowker's *Computer-Aided Translation Technology. A Practical Introduction* (2002). Kenny notes how these sources complemented existing textbooks on MT such as the ones published by W. J. Hutchins and H. L. Somers (1992), Arnold et al. (1994), and Trujillo (1999).

Traditionally, studies on CAT tools (TM translation) and MT-aided translation[9] have been separated (Teixeira 2014, p. 45) since the core technology and the required competences by the user differ. Nonetheless, the combination and interaction that nowadays TM databases and MT engines allow for have led to new research trends that investigate their intertwining from varying perspectives (technical, user experience, quality, etc.). A more deep review of recent advances from the PE perspective is provided in chapter 2.

1.2. TM/MT Interaction

With the increased research in TM technology, CAT tools become widely commercially used with robust features (TM, alignment, term management, filters...), and thanks to the European Commission's tender bids in 1996-97 Trados[10] becomes the default industry standard (Garcia 2015). Kenny (2020, p. 17) observed that the term "Computer-Aided Translation" or "Computer-Assisted Translation", sometimes seen as a synonym for, or as an advanced form of, 'Machine-Aided Human Translation' (Kornacki 2018, p.107-126), found favour with most authors in translation pedagogy (e.g. Bowker 2002; Quah 2006). Whereas CAT as synonym of MAT is a process, within the industry, the collocation "CAT tools" is more often em-

[9] Often treated as synonym of PE, even if it should not be, given the crucial nuances one term or the other (should) hold in industrial settings.

[10] https://www.sdl.com/es/products-and-solutions/translation/software/sdl-trados-studio/

ployed, which is another term for Translation Environment Tool Translation Environment Tool (TEnT). While it is out of the scope of the present dissertation to review the nature and evolution of TM and CAT, in the following lines it is briefly discussed how TM and MT have been combined in the past and until now.

In the mid 2010s, the interaction between both technologies started to occur in the translation market. First, their integration was not mainstream as CAT tools did not provide a seamless connection between TM results, the editor, and the MT provider. For instance, in a survey carried out in Denmark, in 2013, out of 22 translation service providers using CAT tools "five respondents said they combine TM and MT technology, but they only use this combination rarely" (Christensen and Schjoldager 2016, p. 97). However, there are not yet numerous studies about scenarios in which MT and TM are combined in the same workflow[11]. (Teixeira 2014, p. 46).

Nowadays, the possibility of plugging an MT engine via an API to virtually any CAT environment has led to a "spreading use of MT in combination with translation memory databases", which "is moving the translator into a space where texts are not linear and the binary distinction between source and target text gets blurred" (Celia Rico 2017, p. 79). In the same vein, Mossop (2006a, abstract) had previously noted this behaviour with the term "collage" translations.

The classification of MT systems cited in the previous section (W. J. Hutchins and H. L. Somers 1992) is later reproduced and rethought by other researchers, such as L'Homme (2008, pp. 10–11). Likewise, Christensen and Schjoldager (2016, p. 90) compared MAHT to TM translation and HAMT to PE. However, the boundaries between HAMT and MAHT are not completely clear, and keep becoming more blurred as the technology and the industry evolve. This distinction has already been questioned and it will probably fade completely as seamless integration of varied features concur in CAT environments. Indeed, the authors pointed to that direction:

> Relatively recently, TSPs[12] have started to implement (statistical) MT software that is integrated into a TM suite, meaning that an MT match is provided when no match can be found in the TM database [...]. In this new kind of translation, which we shall refer to as MT-assisted TM translation, the translation process is more automated than in regular TM translation, but not

[11]The existing studies often use a purpose-built PE environment, instead of a widely-used CAT tool.

[12]Footnote added to this dissertation: Translation Service Providers include individual freelance translators. We use Language Service Companies (LSCs) to refer to firms and agencies with bigger structures.

> so automated as in MT, which places it in the middle ground of Hutchins and Somers' (1992) technology continuum. (Christensen and Schjoldager 2016, p. 90)

The same observation is made by researchers such as Kornacki (2018, p. 127-128). The truth is that, currently, almost all translation projects are carried out using various computer software (Christensen and Schjoldager 2016, p. 89) and TM and MT are completely integrated in CAT tools: "blending the difference between editing MT and translating". If not already in the present, in the near future, "most translations will be assisted with a smart blend of translation memory and MT" (Nimdzi 2019).

To conclude, we can establish four types of translation workflows regarding translation technologies:

- With no technology (text-processing software, such as Ms Word, and probably the Internet for searches).
- With CAT (with TM, termbases and all the old and new features, from filtering to sub-segment recall or "repaired" segments).
- With MT only (within an MT editor or a CAT environment; it can be "interactive" or what is called "downstream" by Muzii (2016).
- Merged scenarios (within a CAT tool, the TM provides results above a certain threshold and the MT below it; moreover, the Termbank or Termbase (TB) or other resources can contribute to "repairing" segments and adaptive or interactive PE may be enabled too).

The last two scenarios, if interactive instead of downstream, can allow for features such as adaptation of MT on-the-fly (adaptive MT) or predictive writing powered by MT[13]. Other tools that implement auto-completion or predictive features are CasMaCat[14] (Alabau et al. 2014) and HandyCAT[15] (Hokamp and Liu 2015; as commented by do Carmo 2020b, pp. 422–423).

The last scenario (merged TM/MT), as already mentioned, can be referred to with the mouthful of "MT-assisted TM translation", and it has undergone disruptive changes in the past five years due to the development of NMT. As observed by Ovchinnikova and Morozova (2019), the origin and typology of errors the translators edit or correct from TM results compared to NMT errors may show the direction for translator and post-editor training. Hence, the quality and errors of NMT is concisely presented in the next section.

[13] Lilt may be one of the only commercially available tools with this feature at the time of writing, according toTeixeira (2019, p. 222).

[14] https://github.com/casmacat

[15] http://handycat.github.io/

1.3. NMT and its better quality

Both "post-editing" and "MT-assisted TM translation" are activities/services that have thrived in the past five years. Indeed, recent research in neural networks and deep learning have led to a very good quality of MT outputs. The first neural networks were applied to memory retrieval (used in connection with TM and fuzzy matching). Even though neural networks exist since the mid 1980s, they have been more successfully applied to MT since the appearance of deep learning and Artificial Intelligence (AI). NMT became well-known within the translation industry between 2016 and 2017, as a field of knowledge in AI and deep learning.

Forcada (2017) presents a good introduction of the principles of NMT. Pouliquen and Junczys-Dowmunt (2016) explains, in brief and clear words, the history of NMT in a University lecture[16]; in it, one can see how commercial providers such as Systran, Google and Microsoft released their new NMT product and service at the end of 2016. One controverted publication soon reported "human parity" regarding the quality of the NMT outputs: Hassan et al. (2018)[17]. As a consequence, and despite the controversy and disapproval (e.g. Toral et al. 2018), the increased quality and fluency of the NMT outputs (Castilho, Moorkens, et al. 2017) has since then brought major interest in MTPE within academia, and a higher demand in the industry.

Moreover, Amazon[18], Google[19] and Microsoft[20] launched services to set up customised engines through their cloud service platforms (Muzii 2016), but, as the author puts it, "NMT engines are still quite pricey and challenging as to technical requirements and operational complexity". The author also notes the typical issues in NMT outputs that have been reported by other researchers: missing words (omissions), additions and mistranslations (e.g. Guerberof Arenas 2019, p. 349). Such problems may be caused by the so-called "out-of-vocabulary" words, by spelling and grammatical errors in the ST[21] or by very long or very short sentences (Muzii 2016).

Having explained the main flaws of NMT regarding accuracy, the aspect of fluency may now be commented. According to Muzii (ibid.), fluency in NMT outputs may be misleading, as the NMT errors can be harder to spot. The author highlights how the "cultural appeal" may be underesti-

[16] Within a Global Initiative of Academic Networks (GIAN) course, delivered at the Indian Institute of Technology of Banaras Hindu University (IIT-BHU) University, Varanasi, 19-22 December 2016.

[17] https://www.microsoft.com/en-us/translator/blog/2018/11/14/nextgennmt/

[18] https://aws.amazon.com/translate/

[19] https://cloud.google.com/automl/

[20] https://portal.customtranslator.azure.ai/

[21] The abbreviation ST can be read as "source text" or "start text" as advocated by Pym (2013), cited in Angelone et al. (2019, p. 3).

mated and, "while NMT may reduce the effort spent on repetitive errors that might escape an automatic script, more cognitive effort might be necessary for messaging refinement". Hence, still citing Muzii:

> Especially with NMT outputs, post-editors should know not to be swayed by fluent sentences and to be extra vigilant against any missed word that might change the meaning entirely [especially because] error patterns are not consistent from segment to segment.

In figure 1.1, the purple line represents all these issues that separate NMT outputs from FAHQT.

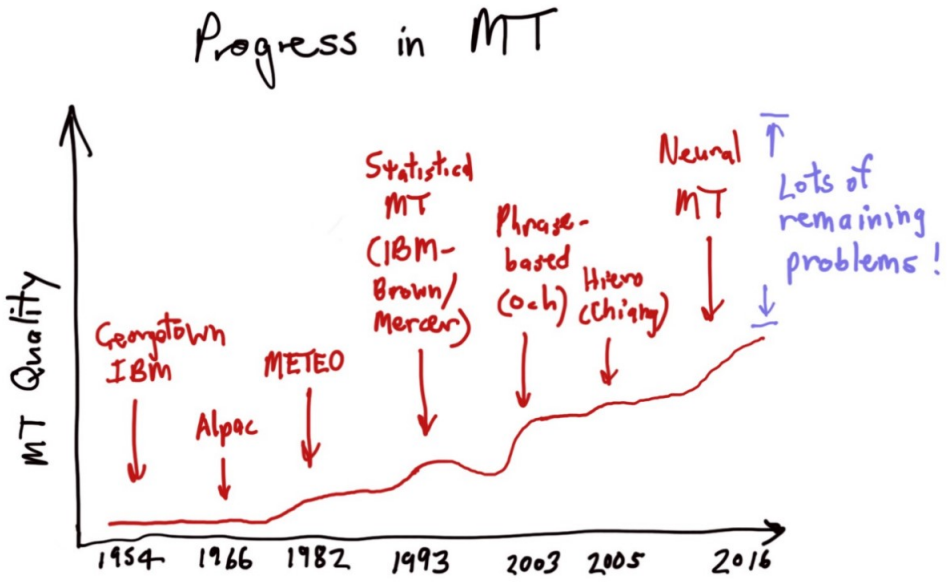

Figure 1.1: Luong et al.'s (2016) evolution of MT quality

It should be precised that the Y axis of figure 1.1[22] reads "MT Quality" and not just "Quality", "Translation quality" or "Human parity"[23].

Moving on now to consider the concept of quality. Since 2017, all the dimensions of the notion of quality in translation have been further researched to continuously improve and develop reliable measures. While MT evaluation is a whole field of research (Escartín and Arcedillo 2015, p. 132), the classification presented by Specia (2017, October 11) in a TAUS webinar

[22] The author thanks Christopher D. Manning for authorising the reproduction.

[23] Actually, at the TAUS Global Content Summit in 2019, Chris Wendt, from Microsoft, presented a version adapted from figure 1.1 where human parity would be accounted for in 2018 (Wendt 2019). The concept of human parity is not discussed in the present dissertation.

in October 2017 is summarised below to provide a concise overview of the said field:

1. Automatic Evaluation Metric (AEM)
   - Reference-based (BLEU, Meteor, NIST, TER, WER, PER, CDER, BEER, CiDER, Cobalt, RATATOUILLE, RED, AMBER, PARMESAN,...)
   - Machine Translation Quality Estimation (MTQE)[24]
2. Manual evaluation ("human evaluation")
   - Direct assessment
     - Scoring
     - Ranking
     - Error annotation
   - Task-based
     - Post-editing (key-logging and time)
     - Reading comprehension
     - Eye-tracking

The automatic metrics are useful in the upstream evaluation for MT engine development. Especially BLEU (Papineni et al. 2002) has a long tradition and it is still the most cited metric and used reference in the industry for MT quality evaluation. Even though, its reliability has been questioned again with the appearance of NMT (Forcada et al. 2017). One challenging aspect regarding MT output quality is MTQE (de Gibert Bonet 2018, p. 22). MTQE should help the translator quickly decide which MT segments are worth (1) to ignore and accept as one would sometimes do with a perfect match; (2) to edit; or (3) discard and retranslate. The research community is working to make this measure a reliable counterpart of the traditional fuzzy matches.

The human evaluation techniques (such as error annotation) have been reported to be time-consuming (hence, expensive) by authors like Koponen (2016a, p. 23). Luckily, the Edit Distance (ED) metrics offer a range of possibilities between automatic and manual. This is why they represent the most widely-used method to account for technical effort in PE. They are undoubtedly the most relevant when assessing post-editor performance and PE productivity, as we will argue in section 2.1.1. Before that, it is necessary here to clarify exactly what are the differences between the main ED-based metrics:

1. Automatic with golden reference: Translation Error/Edit Rate (TER)
2. Automatic with human-targeted reference, that is, the post-edited version of the output: Human-targeted Translation Error/Edit Rate (HTER)

[24] Sometimes abbreviated just QE.

3. Manual light: PE distance (e.g. with Qualitivity[25])
4. Manual full: Actual Edit Rate (AER)

The first two can be computed with TERCOM[26]. The last two are better defined in next chapter.

All things considered, the imminent arrival of MT discussed by Mossop (2006a, p. 788) is no longer imminent, it has been a reality for a decade. Such reality has brought new lines of research, such as "PE pedagogy". As observed by Blagodarna (2019, p. 4) in her dissertation:

> PE competency is a relatively new area of research and the ways in which post-editors can acquire it have been given insufficient academic attention to date.

Even when it has received attention, it mainly concerned RBMT or SMT systems. NMT has represented considerable changes in the translation industry and academia landscapes, for instance, for the PE task, it demands increased attention to fluency errors (as commented in work by Castilho, Moorkens, et al.), but less keystrokes may be necessary according to Muzii (2016). In fact, NMT has been disruptive for the overall workflow, for it has technologised the profession and it is paving the path towards Fully Automatic Usable Translation (FAUT) (Massey 2018, p. 3; O'Brien and Rossetti 2021, p. 95). In this regard, it introduces many unknowns into the life cycle of translation projects, mainly related to productivity, pricing, processes and quality (Vieira and Alonso 2020, p. 11). In next chapter, the industry landscape regarding the PE activity is explored to provide the state of the art regarding such unknowns.

[25]The name of this plug-in corresponds with the concept of "qualitivity" introduced in section 2.1.3 and used throughout the dissertation. The advantage of this plug-in is that it accounts for the so-called "revisited" segments.

[26]https://github.com/jhclark/tercom (Bandyopadhyay 2012, p. 97).

Chapter 2

POST-EDITING OR MT-AIDED TRANSLATION

As briefly introduced at the end of chapter 1, different terms have been used for the "activity of editing and correcting MT output" (definition in 3.1.4 of *ISO 18587:2017*). It is at the already mentioned Conference on Mechanical Translation at the Massachusetts Institute of Technology that the terms "pre-editing" and "post-editing" are used for the first time (Dodd 1952). The first mention for the role, "post-editor", is made in 1950 by Erwin Reifler, according to W. J. Hutchins (1986, p. 31) and it can later be found in the phrase "MT with a Post-Editor" (Reifler 1952). After seventy years, the term retained within academia and the industry is still "post-editing". The above-quoted definition by *ISO 18587:2017* seems larger than others found in literature, for instance: "comparing a source text with the machine translation and making changes to it to make it acceptable for its intended purpose" in Geoffrey S Koby (2001).

While the different definitions may include or not post-editing MT monolingually, that is, when the post-editor edits the TT without referring to the ST at all, the present dissertation puts the focus on bilingual editing. The fact is that, when listing the tasks of a post-editor and the requirements of a professional translation with MTPE, the reference to the ST can hardly ever be avoided. For instance, in *ISO 18587:2017* (p. 7), "using the source language content as reference in order to understand and, if necessary, correct the target language content" is one of the three listed tasks for a post-editor. What is common to every definition of PE is its primary goal: to increase productivity (i.e. improvement of turn-around times and cost reduction).

Some of the main concepts that have attracted attention in research are:

- The comparison of the PE activity to tasks with a longer research tradition (translation or revision)
- PE Productivity
- PE Quality

These two last dimensions are often studied combined and according to other factors, such as: type of errors in the MT output, language pair, text type, profile of the post-editor (expert or linguist), mother-tongue principle, purpose or intended audience, etc. Other research lines fall within the professionalism approach, for instance: resistance and agency of the linguist, project management, pricing, employability... Or within the technology approach: software or tool development, use of certain features like concordance search, integration between tools, new features such as adaptive or interactive environments, etc.

In conclusion, if the complexity of each notion individually is substantial, for the PE intersection it is exacerbated by the nature of the "scope" in the so-called "Project Management Iron Triangle", also known as "Triple Constraint" following Iizuka (2018), and its relation to the other three factors, as depicted in figure 2.1[1].

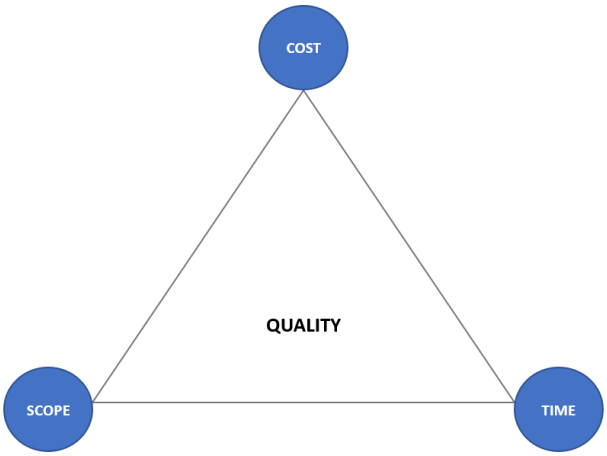

Figure 2.1: Triple constraint triangle

The chronological progress in research tackled first the comparison of PE to translation and revision (term "scope" in figure 2.1), to gradually move to productivity (ratio, as we will come to conclude, between not only the two bottom extremes of the triangle but its centre too), and quality (at the centre of the triangle). However, the opposite direction of thoughts is applied in the present dissertation.

In section 2.1, a brief overview of related work in PE productivity (2.1.1) and quality (2.1.2) is presented, to conclude that one encompasses the other in section 2.1.3. In section 2.2, the PE assignment is presented as the

[1]As a matter of fact, the triangle is adapted to each situation, this is why the scope and quality notions often swap places.

tool that determines the scope for PE in a professional situation and that must guide towards the aimed quality and productivity. Finally, in section 2.3, we explore the impact that PE productivity, PE quality and PE assignment have had (or keep having) on the professional profile of the linguist (reviser/translator/post-editor/etc.).

2.1. The post-editing process and product

As noted by do Carmo (2017, p. 148), the concept of PE lies between Translation Studies and Machine Translation (topic already addressed in chapter 1), and it is a complex one. As argued in the Introduction, how one refers to it and its subfields may vary depending on the approach or perspective taken. From the perspective of the overall translation workflow, the PE process occurs at the centre, simultaneously with the translation process and editing of fuzzy matches. The PE process would, thus, replace translation. It is represented in the figure by Mason and Rinsche (1995, p. 33). As

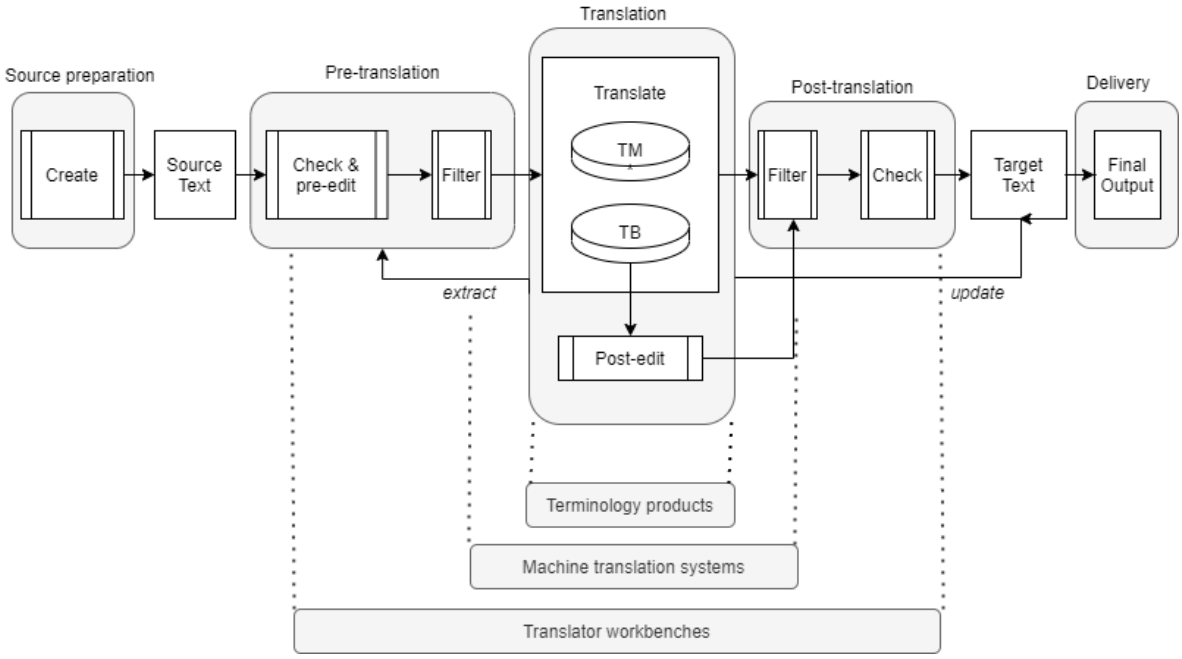

Figure 2.2: OVUM Translation process

observed by Mason and Rinsche (ibid., p. 94), the translation mode can be in batch (fully automatic, called “downstream” by Muzii, as seen in chapter

1) or interactive[2]. The interactive mode and its most recent lines of research are out of the scope of the present dissertation.

The driver for replacing the translation step with MTPE, as commented earlier, is to increase productivity. This notion is explored in the following section, where the concept PE effort is also introduced, since it is recognised as an influential factor from the business perspective to appreciate the cost-saving potential of MTPE (Blain et al. 2011).

2.1.1. Post-editing productivity

"Productivity", in translation, has traditionally been measured as the rate between volume of translated words and time spent. Indeed, it is defined by the Oxford Advanced Learner's Dictionary of English as

> the rate at which a worker, a company or a country produces goods, and the amount produced, compared with how much time, work and money is needed to produce them.

Hence, it would be quasi-synonymous of the term "throughput". The definition of this last term from the same dictionary is:

> the amount of work that is done, or the number of people that are dealt with, in a particular period of time.

In PE research, there had been references to absolute numbers for some time to justify increases in productivity. To give some examples:

- 10,000 words per day were reported at the Pan American Health Organization (PAHO) with their customised MT systems (Vasconcellos 1986, p. 145), as mentioned in the previous chapter;
- Joscelyne (2006) reports a turnaround of approximately 5,000 words per day and it is recognised as more realistic by O'Brien (2006a); and
- A.-M. Robert (2013, p. 32) compares the throughput by a translator (on average, 2,000 words/day) to that of a post-editor (3,500 words/day).

In terms of relative numbers, Plitt and Masselot (2010, p. 10) found an average increase of 74% in the number of words post-edited per hour with a Moses engine at Autodesk. Oliver (2016, p. 205) highlights that the productivity gains may depend on the language pair and can vary from 15% to 40%. This same maximum percentage (40%) is reported as the average increase by Gene (2019). It is claimed by do Carmo (2017, p. 153) that productivity gains would go from 2 to 4 times the translation productivity rates.

[2] A recent reference about the radical change that interactive workflows could bring to the profession is Carl and Planas (2020).

Finally, Läubli et al. (2013, p. 90) warned that, in a realistic environment (as opposed to isolated, which may be useful to study specific aspects), gains in throughput can be around 15-20%.

Indeed, since NMT appeared, such claims had to be reviewed. For example, Jia et al. (2019) find that the time-saving effect with PE varies according to text types, and is particularly true for domain-specific texts (which corresponds to previous findings O'Brien 2007, Plitt and Masselot 2010 and Guerberof Arenas 2009). All things considered, it is noteworthy to remind that several conditioning factors underpin any productivity increases and their extent. In that sense, researchers like Guerberof Arenas (2010, p. 3), Garcia (2011, p. 228) and Zhechev (2014) highlight the following aspects: language pair, ST type and if pre-edited or not, profile of the linguist and their experience, domain-specific MT engine...

In his dissertation, do Carmo (2017, p. 260) analyses PE productivity from the perspective of temporal effort (higher speed). The author also highlights the importance of this notion within the industry and for the definition of the PE activity (p.153-154) and, once again, how it depends on a large range of factors:

- Language pair;
- Quality of MT output (by extension: the bilingual data used for its training, the type of MT technology, the domain and style of the ST, etc.);
- Post-editor profile (by extension: their experience and learning curve); and
- The purpose and intended audience of the PE product; quality expectations, called by Muzii (2016) "Acceptable Quality Limits".

Despite all these increases in productivity reported in past literature, the norm *ISO 18587:2017* only mentions this notion as one of the objectives for MTPE in the Introduction, without defining the term or mentioning its components anywhere else in the standard (speed, rate, throughput...). It seems that the general belief is that

> [p]roductivity is an average measure of the efficiency of production, and, in translation, it is measured as the throughput or speed expressed in the number of words per hour. (ibid.)

To gain a deeper understanding of the concept of productivity, we should refer back to the above-mentioned triple constraint: speed, cost and scope. These three notions, when applied to PE, are referred to with the term "PE effort", introduced by the above-cited work by Krings (2001).

PE effort

In the last two decades of PE research, the work by Krings (2001) has become a reference (O'Brien 2004, p. 4). Krings (2001, p. 531) used Think-Aloud Protocol (TAP) to learn more about the PE process and the author found there are three types for PE effort(p. 178): temporal, technical, and cognitive. As summarised by Lacruz and Shreve (2014, p. 247), the temporal effort refers to the time the post-editor spends; the technical effort, to the keyboard and mouse activity; and the cognitive effort refers to mental processing. These have been thoroughly reviewed by researchers in doctoral dissertations such as Koponen (2016b) and do Carmo (2017) and some disadvantages of the TAP method are observed by Koponen (2016a, p. 22), since TAP "can only capture the conscious part of cognitive processing, and has the effect of slowing down the process and potentially even changing the cognitive processing involved", especially studied in O'Brien (2005).

The three types of PE effort (temporal, technical and cognitive) seem to have been retained as valid measures, ideally combined, within the research and (maybe less often) industrial communities. Their interrelation is well applied to the above-mentioned triangle by de Gibert Bonet (2018) reproduced in figure 2.3[3].

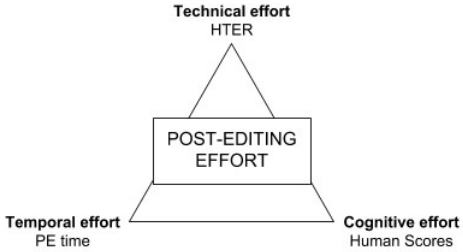

Figure 2.3: PE Effort

Regarding the temporal effort, the speed at which a post-editor corrects an MT output can be recorded with many tools and it can be computed as Words per Hour (WPH) or as Words per Second (WPS). Regarding the cognitive effort, unlike temporal and technical efforts, it is difficult to analyse objectively (Lacruz and Shreve 2014, p. 247). Indeed, cognitive effort poses a challenge to PE research. Several authors have used "user activity data", such as key logging, with tools like Translog II[4] (O'Brien 2006a), and eye-tracking software to record every part of the PE process, thus analysing

[3]The author thanks the authors for authorising the reproduction.

[4]https://sites.google.com/site/centretranslationinnovation/translog-ii?authuser=0

gaze data, pauses (Koehn and Haddow 2009), etc. to come up with different indicators that could possibly reflect cognitive effort:

- The length of a pause (Couto Vale 2017) or the pause density (Lacruz and Shreve 2014, p. 263);
- Gaze behaviour (Doherty, O'Brien, et al. 2010; Mesa-Lao 2014); such as duration of fixations (Vieira 2014) or fixation count (Carl, Dragsted, et al. 2011, p. 138), or their combination; or
- Keyboard activity (ibid.).

As can be observed, they are all an extension or deduction from temporal effort (length, duration) or from technical effort (number, density, count...). Along with the more or less intrusive methodologies summarised above, Likert scales (four-point or five-point scales, most often) have also been used to evaluate the cognitive load by human raters (Koponen 2016a, p. 22-23), citing Specia, Cancedda, et al. (2010, p. 3376) and Callison-Burch et al. (2010). In that sense, authors like Moorkens, O'brien, et al. (2015) performed a study to try to find a relation between the metrics that would indicate cognitive effort and the actual perceptions of effort expressed by the translators. The complexity in identifying and characterizing cognitive effort has attracted a considerable amount of research whose review is out of the scope of the present dissertation.

Finally, to analyse the PE technical effort one can find several studies with a wide range of proposals. To give just one example, Barrachina et al. (2009) studied interactive-predictive performance measures:

1. Keystroke ratio (KSR): n of keystrokes divided by the total n of reference characters
2. Mouse-action ratio (MAR): n of pointer movements plus one more count per sentence divided by the total n of reference characters
3. Keystroke and mouse-action ration (KSMR): KSR plus MAR

The most widely-used measure, however, in commercial settings, is the ED[5], which is another word for the Levenshtein Distance (LD) (Levenshtein 1966), according to Black (2008):

> Definition:
>
> (1) The smallest number of insertions, deletions, and substitutions required to change one string or tree into another.
>
> (2)
>
> $$\Theta(m \times n)$$

[5]Defined by Snover et al. (2006) as "the number of insertions, deletions, and substitutions that are required in order to make a system translation equivalent in meaning to that of a reference translation".

> algorithm to compute the distance between strings, where m and n are the lengths of the strings. Also known as edit distance.

Many tools can be connected as plug-ins or add-ons to the existent CAT environments and record the edit operations a linguist performs when correcting an MT output. As commented above, they are often based on the Damerau-Levenshtein distance or the more recent variants of the metric TER. The TER formula is depicted in figure 2.4 and it "measures the amount of editing that a human would have to perform to change a system output so it exactly matches a reference translation" (Snover et al. 2006). The authors conclude that "if humans are to be used to judge the quality of MT output, this should be done by creating a new reference and counting errors, rather than by making subjective judgments".

$$\text{TER} = \frac{\text{\# of edits}}{\text{average \# of reference words}}$$

Figure 2.4: Translation Edit Rate formula by Snover et al. (2006)

Using a new reference, post-edited from the particular MT output, instead of a gold reference, is also in line with the work presented carried out by Ive (2017, p. 28). Other authors, such as Specia and Farzindar (2010), explain how

> [HTER] consists in measuring the ED between the translation produced by the MT system and its minimally post-edited version produced by a human translator.

With this comparison, HTER computes the minimum number of word-level changes between the two versions of the segment (Koponen 2016a, p. 23).

More recently, Marg (2018) highlighted how ED is often used to measure how many character-edits are necessary to transform a given MT output segment to its post-edited version "that is compliant with the translation quality requirements specified by the buyer". Thus, we obtain a segment-level score. In the present dissertation, especially in part III, the character-based ED score at the document-level is retained as basis for our research. As we will explain in the said part, it is done in a manner as to analyse also the so-called "revisited" segments.

As Sanchez-Torron and Koehn (2016) observe:

> HTER is concerned about the PE product, not the process. It therefore does not measure translators' actual edit operations, which may involve going back and applying corrections to previously post-edited parts of the text.

Other researchers like Maučec and Donaj (2019, p. 16) agree that the measure which best accounts for the technical effort during the PE process is the AER that Sanchez-Torron and Koehn (2016) proposed, and Ive (2017) also uses, as commented above. In order terms, AER has been referred to "human-edit" as "the average count of actions computed from human keystrokes" in Góis, Cho, et al. (2020).

Concerning how the ED (technical effort) relates to the WPH rates (temporal effort), several claims have been made about discrepancies between PE time and HTER (e.g. M. Aziz W. K. and Specia 2014, p. 172 and Muzii 2016). However, the comparison of these two indicators of PE effort (or more) remain one of the most cost-efficient methods to date to follow up on any productivity increases when using MTPE. In that sense, in 2015, a composite indicator was developed by Nikos Argyropoulos called the "TAUS Efficiency Score" (Attila Görög 2015a; Ruopp 2015; Attila Görög 2015b) that is, however, not widely used either by industry stakeholders or researchers, for the only mention found in Google Scholar is by Valli (2015, p. 131). This scored will be applied to the dataset in part III.

Not only the MT errors in the raw output affect PE effort. Muzii (2016) lists six elements upon which the PE effort depends: (1) The user requirements; (2) The quality expectations; (3) The perishability of the content and its scope; (4) The overall volume of content to process; (5) The function for which the content is intended; (6) The overall turn-around time for the PE task. These necessarily will remind the reader of the conditions we had already listed as circumstances that can guarantee increased PE productivity rates. It could seem that the connections between such factors may be too vaguely defined in the overview of literature provided above, hence, they call for further precision. To sum up:

- Language pair
- MT engine type (RBMT, SMT, NMT)
- MT engine maturity (in-domain customised, which, in turns, depends on the quality of the training data and other factors)
- ST type (domain and genre)
- ST linguistic characteristics (indicators of translatability[6] that may be improved through pre-editing)

all have an impact on **MT output quality**. Then:

- MT output quality (and if it is previously well estimated and managed)
- Profile of the professional (expert or linguist, experience, training and

[6]Translatability Indicators (TI) and Negative Translatability Indicators (NTI) are concepts scarcely researched in languages other than English. Some well-known research on the topic that could be cited is the work by Bernth and Gdaniec (2001) and O'Brien (2005) and O'Brien (2006b).

educational background, etc.)
- *Skopos* (quality expectations, perishability of the content, audience or end-users, etc.)
- Assignment context and ergonomics (CAT environment, volume of words and turn-around time, chair, screen and input devices, etc.)

all have an impact on **productivity** (a concept built up, at least, by the three types of PE effort), **quality**, and the balance between both concepts (the real "PE productivity").

Given the high complexity of the concept and the number of factors that come into play, PE researchers have triangulated technical and temporal effort analysis with the PE product quality. An important contribution to the field is made by Guerberof Arenas (2008), Guerberof Arenas (2009), and Guerberof Arenas (2014), who systematically analyses translation quality through error classification. In our opinion, productivity should not only encompass the rate between volume and speed[7], but it should also include the quality of the post-edited product. This is in line with the work by Guerberof Arenas (2019, p. 344-346), who, after noting the gains in productivity within the industry (p.344) without compromising quality, and within academia (p.345), highlights that such increases in productivity are often relevant to particular environments with customised engines, or to certain type of content or language pairs, etc. and concludes that "an analysis of the quality of the product is needed" (p.346).

2.1.2. Quality of the post-edited product

If the productivity notion is already complex and a challenge to PE research (especially the component of the cognitive effort, as seen in section 2.1.1), the concept of quality is a whole other research field in Translation Studies that comes with its own challenges:

> [a]ssessing translation quality is known to be an extremely complicated and subjective task, and low agreement between raters was also found in previous studies. (Jia et al. 2019)

Thus, the field of translation assessment is still considered to be in its infancy (Akbari and Segers 2017, p. 4). Even with such constraints, a higher qualitative product with PE (versus HT) has nevertheless been acknowledged in research. For instance, Fiederer and O'Brien (2009) reported that three post-edited TTs were judged to be of higher clarity and accuracy than the versions translated from scratch. Other studies, such as the ones published by Läubli et al. (2013, p. 89) and Fiederer and O'Brien (2009, p. 68),

[7] That is WPH or WPS, for instance.

confirmed that MTPE translations' quality is or can be at least equivalent to conventionally produced translations. While a comprehensive review of the concept of quality in TS is out of the scope of the present dissertation[8], a brief discussion regarding MT and PE is presented in the following paragraphs.

In 2014, the Tradumàtica journal published a three-part series[9] of articles co-authored by four researchers. After having discussed the definition of translation in the first article, in the second one, Fields et al. (2014, p. 406-407) offer a review of the different approaches to translation quality devised by Garvin (1984) and how several authors align more with one or another:

1. Transcendent approach (quality as innate characteristic, absolute and universally recognizable);
2. Product-based approach (quality as a quantifiable feature, that can be ranked, according to certain attributes);
3. User-based approach (quality as a degree to which a product or a service satisfies the end user's needs or preferences);
4. Production-based approach (as renamed by Fields et al., quality as a degree to which a product or service complies with predefined requirements or specifications); and
5. Value-based approach (quality as the ratio between costs and benefits).

It would seem that maximizing accuracy and fluency is a basic notion both for transcendent and product-based approaches. However, the authors highlight that in MTPE scenarios an inaccurate or "less-than-fluent" translation may be useful (Fields et al. 2014, p. 408). In that sense, it is acknowledged that a production-based approach to translation quality may be relevant for the translation industry, which is why it is the approach adopted by the norms *ISO 17100:2015*; *ISO 18587:2017* (ibid. p. 409). All in all, the authors express their disagreement on a number of factors around translation quality:

1. the degree to which the end-user's perspective is helpful to translation-quality assessment;
2. the relevance of the value approach to translation quality; and
3. the role and nature of translation specifications.

Finally, the authors conclude that functionality has an influence on quality measurement, and that quality management is a good tool to address downward price pressure (ibid., p. 411).

In the third article of the above-mentioned three-part series, Geoffrey S.

[8]We refer the reader to this two monographs for a more comprehensive review of the topic: *Quality aspects in institutional translation* (Biel et al. 2017) and *Translation Quality Assessment: From Principles to Practice* (Moorkens, Castilho, et al. 2018).

[9]The specific number on quality is summarised in an English editorial in Görög (2014).

Koby et al. (2014) make a proposal of two definitions for translation quality while expressing their lack of consensus on several points. They present one broad definition of translation quality, and one narrow. The first one reads:

> A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs. (Geoffrey S. Koby et al. 2014, p. 416)

And the narrow one reads as follows:

> A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture. (ibid., p. 416-417)

It is particularly relevant to the present dissertation how the authors (p. 419) argue that the post-editor's role (called "operator" by the authors) and the PE task require great skill and understanding of varied products and services. While they claim not to have reached consensus concerning how quality management can be applied to the translation industry, they highlight the points of agreement, one of which is the necessity to define the translation project according to the presence or absence of explicit specifications (see section 2.2).

When it comes to PE research, certain studies (e.g. Koponen 2016b, p. 5; O'Brien 2011, p. 2) lacked the evaluation of the quality of the post-edited product. Others, however, took the quality factor into consideration when assessing productivity gains in PE (e.g. Guerberof Arenas 2008; Guerberof Arenas 2009; Guerberof Arenas 2014).

Guerberof Arenas (2019, p. 346) notes that studies in the past have demonstrated little difference between the quality of fully post-edited segments and "human-translated" (HT) segments. To evaluate or assess translation quality, several error taxonomies have been created and used in the past within the translation industry and academia. The most recent and relevant for the present dissertation are explored in the following section.

Error types and their assessment/evaluation

To examine quality, what defines an "error" must be previously established. The error taxonomies in translation have a long research tradition.

However, as one could expect after noticing the lack of consensus regarding the notion of quality in the previous paragraphs, there is not one single reference for error classification upon which all stakeholders agree either.

Koponen (2016b, p. 27) reviewed several proposals for error categorisations, for instance: Temnikova's (2010) 10-error-type classification (specific to classifying errors in terms of cognitive effort), and Lacruz, Denkowski, et al.'s (p. 77) 5-error-type classification: mistranslation, omission or addition, syntax, word form, and punctuation. Similarly, Läubli et al. (2013, p. 87) used ZHAW's internal evaluation scheme consisting of five ordinal scales for:

- target language expression
- target language grammar
- target language syntax
- semantic accuracy
- translation strategy

As observed in Moorkens, Castilho, et al. (2018, p. 17), the most recent effort and the proposal that seems to have gained considerable consensus (Lommel and Melby 2018, p. 34) for a range of purposes and workflows (MT output, human translations, and PE) is the combination of the TAUS Dynamic Quality Framework (DQF) (Görög 2014) with the Multidimensional Quality Metrics (MQM) (Burchardt et al. 2014; Lommel and Burchardt 2015). Indeed, in Lommel (2018)[10], the author presents the DQF/MQM Error Typology, an approach that appeared in 2014 by integrating the two above-mentioned typologies. Since 2018, the merged DQF/MQM framework has been in the ASTM International standardisation process to provide a unified systematic framework for the translation industry.

The specifics of how the two models were integrated into a single one are described by Lommel (ibid., p. 125) with a figure that represents the final resulting model. Guerberof Arenas (2019, p. 338) summarises their harmonisation with these words:

> errors [are] classified, firstly, according to a broader DQF error typology and, subsequently, by the subcategories as defined in MQM.

MQM has become a widely used model, to the extent that Mariana et al. (2015) provide a comparison with the American Translators Association (ATA) error categories, traditionally used for the ATA Certification.

[10] A chapter in Moorkens, Castilho, et al. (2018), to which we refer the reader in terms of state of the art for what concerns Translation Quality Assessment (TQA), especially to the chapter written by Castilho, Doherty, et al. (2018), where the authors provide a critical overview to TQA approaches and highlight some remaining fundamental issues.

The correspondences can be consulted in the authors' table "ATA and MQM Mapped Error Categories", while the divergences concern the scoring direction (0 for absence of errors or 100 for top quality) and the pass/fail threshold (Mariana et al. 2015, pp. 142–143).

During the MTPE process there are four stages where quality may be evaluated. These stages, along with the method and its outcomes are well represented in Popović (2018, p. 132) and reproduced in figure 2.5[11], where the manual method is a circle and the automatic method is a square.

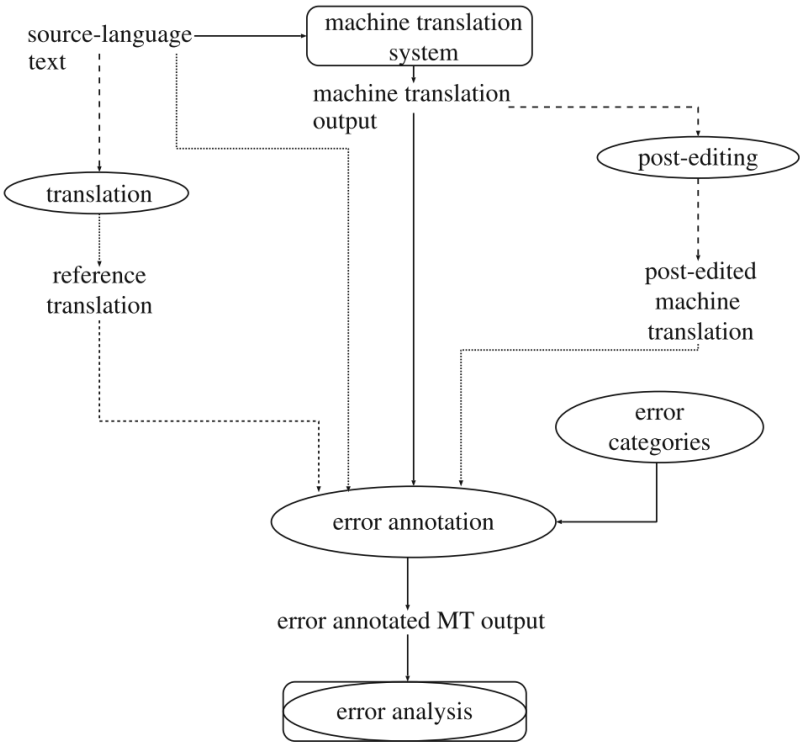

Figure 2.5: Manual error annotation by Popović (2018)

As observed in Castilho, Doherty, et al. (2018, p. 11) and Thelen (2019, p. 14), the reason why we implement a TQA method brings with it certain variability (for instance, if it is in a commercial setting or within a research project). Later in the same publication Doherty, Moorkens, et al. (2018, p. 100) claim that "evaluation needs in industry and academia necessarily differ based on the pragmatic requirements of each scenario".

The advantage of the DQF-MQM model, in that sense, is that it is tunable. The person responsible for evaluating translation quality (e.g. PM in a

[11]The author thanks Maja Popović for authorising the reproduction.

commercial setting, teacher at the university, researcher of a study, etc.) may select between a more holistic or, on the contrary, analytic approach (Lommel 2018, p. 122). The person in charge also decides which "branches", "rubrics" or "dimensions" are relevant to the project (accuracy, design, fluency, terminology, verity, style, and locale convention) and to which extent the evaluation/assessment must be fine-grained.

Besides the bright future for DQF-MQM (ibid., p. 126), another model for translation quality evaluation/assessment that seems to be understudied is the Preselected Item Evaluation (PIE) devised by Kockaert and Segers (2014), Kockaert and Segers (2017), and Segers and Kockaert (2016). If we follow Akbari and Segers (2017, p. 20), who claimed that holistic and analytic assessments can be too subjective, the PIE method shows greater validity and reliability (abstract and p.20), especially when applied to students' tests. Likewise, Van Egdom, Verplaetse, et al. (2019), after reviewing again the 5 approaches by Garvin (1984)[12], acknowledge the success of the analytical methods such as DQF and MQM as well as their flaws (p.32-33). The authors go on (p.34-37) to describe the PIE method and how it has been used in several case studies[13], and their own case study.

In short, Segers and Kockaert claim to have brought together the major strengths of criterion-based and norm-referenced evaluation with the PIE method, which presents the following features:

- As a pragmatic evolution of the Calibration of Dichotomous Items (CDI), it is a **calibration method**, meaning that "the accuracy of the measuring instrument is checked and adjusted" (Kockaert and Segers 2017, p. 150).
- It is a **dichotomous method**: it makes "distinction between correct and wrong solutions, [it does] not distinguish between levels of error".
- It is characterised by a **preselection of items in the ST** on the basis of a number of criteria, such as translation brief, domain or test-specific factors (ibid., p. 152).

To calibrate it, one must decide how many items will bring enough validity and reliability to the test, that is selecting the "minimum number of items needed for a desired level of score reliability or measurement accuracy" (Lei and Wu 2007, p. 527; cited in Kockaert and Segers 2017, p. 151). To do so, the concepts of "item difficulty" (p-value) and "item discrimination" (D-index) are used. The first one is simply the percentage of candidates who answer the item correctly (Kockaert and Segers 2017, p. 151). The second one is used to represent how much "discriminating power" a given

[12]Briefly commented before when discussing the work by Fields et al. (2014) and Geoffrey S. Koby et al. (2014).

[13]For instance, by the Process of Acquisition of Translation Competence and Evaluation (PACTE) group, who "entertained the wish to even relate items to translation competences".

preselected item holds, that is, to what extent it truly reflects the difference between those who do well and those who do poorly as maintained by Matlock-Hetzel (1997, p. 5), quoted in Kockaert and Segers (2017, p. 151). To obtain it, and following Jurs and Wiersma (1990), Kockaert and Segers (2017, p. 152) subtract the number of candidates in the bottom group (those who answered the item incorrectly) to the number of candidates in the top group (those who answered correctly).

> Preselected items that do not respond to docimological standards (too high or too low p values, weak discriminating power) may be removed from the translation test and replaced by other items. (ibid., p. 152)

Thus, PIE can be adopted to ensure increased test validity. The authors also claim that different evaluators using the PIE method will have the same value judgment on the TT. Finally, it must be acknowledged that even the PIE method has its flaws. For instance: "What do evaluators do with a candidate who proposes an incorrect solution for an item that was not preselected?" (p.153). The authors suggest that the evaluator calculate the item difficulty and its discriminating power to then decide if it is worth including such item in the translation test. As acknowledged by Melby, Fields and Housely, and reported by the editors of Balling and Carl (2014):

> studies of translation processes will lead to inconsistent results if researchers do not define and measure the quality of the output translation in explicit and similar ways.

As argued in the previous paragraph, the research community has so far provided methods to fulfil this requirement (e.g.: DQF-MQM or PIE). To conclude, one more challenging question that comes with the inherent subjectivity in translation (quality), and, thus, may be raised for every single evaluation model, even the PIE method, is: Which solutions are correct and which are not?

2.1.3. Between quality and productivity

As seen in the last two sections, the notion of productivity has often lacked the quality analysis of the product, which makes any claims of higher productivity void. We follow O'Curran (2014) to consider that linguistic quality assessment must be performed to guarantee that the "productivity gains are valid, not occurring at the expense of quality". From the production and value-based perspectives discussed in the previous section, a qualitative PE project would be one that shows productivity gains when

compared to "traditional" translation workflows (with CAT tools, but without MT, in other words, what nowadays can be considered HT). Hence, in PE, not only quality would encompass productivity, but productivity must include quality (Angelone et al. 2019, p. 5). To express that notion, the already cited plug-in for SDL Trados Studio comes to mind: "Qualitivity"[14]. Paradoxically this tool does not allow for evaluation of linguistic quality, but it provides key-logging[15] and time recording, as well as the Damerau-Levenshtein ED. The term "qualitivity" is used in the present dissertation to refer to the concept of balance between "productivity", understood as the rate of volume (words or characters) per time unit (hours or seconds), and "quality", understood as a combination of the production- and product-based approaches[16].

To find out what makes a professional linguist attain (successful levels of) qualitivity (again: the right balance between the misused term "productivity" and quality of the PE product), one must necessarily study the process one more time. As anticipated in section 2.1.1, the concept of "edit operations" has often been used to study PE effort and PE productivity. What the post-editor does when post-editing (setting aside terminological research and other neighbouring tasks and, thus, focusing on the proper "edit operations") and how it is done are the core elements that must lead to a productive process with a qualitative output (qualitivity). For this reason, such edit operations or Post-editing Action (PEA) are concisely reviewed below.

The analysis of PEA can have different applications: for example, Góis and Martins (2019), by grouping sequences of actions (text-editing actions: inserting, deleting, replacing a single word, and inserting or deleting a block of words; and non-editing actions: jump forward or backward, sentence jump, mouse clicks and selections, and pauses between actions) use it to identify the editor's identity and behaviour, and to predict PE time.

According to M. Aziz W. K. and Specia (2014, p. 173), the concept of PEA is introduced by Blain et al. (2011) as a "logical edit" counterpart of the mechanical edits that had been used before. Blain et al. (ibid.) uses this new unit, PEA, to analyse PE effort from a qualitative perspective, claiming that previous work on describing error taxonomies does not provide a methodology for fixing them. The authors defend that their approach is "less interested in understanding the errors than defining the correct action

[14] http://xl8.link/wikiSDL

[15] In part III, the key-logging complement is not used due to practical reasons.

[16] This notion is expressed in *ISO 18587:2017* with the characterisation of the product of full PE as "indistinguishable from human translation output [but] it is recommended that post-editors use as much of the MT output as possible". Please refer to do Carmo (2020a) to read more about how *ISO 18587:2017* characterises PE mainly through the factors of time and money.

to obtain a good translation".

Blain et al. (2011) define a typology of PEA by reviewing existing classifications such as the ones published in Dugast et al. (2007), Llitjós et al. (2005), and Vilar et al. (2006), to move from the mechanical edits (addition, deletion, insertion, shift) to a proposal of 8 linguistically logically motivated edits:

1. Lexical changes, with 8 subdivisions (determiner choice, meaning, stylistic, number, case, adjective, multi-word, structure)
2. Grammatical changes, with 4 subdivisions (verb agreement, verb phrase structure, meaning, stylistic)
3. Preposition change
4. Co-reference change
5. Reordering
6. Pseudo-editing
7. Misc style (described as unnecessary stylistic change, thus, over-editing)
8. Misc (PEAs "that we cannot classify")

Other authors (e.g. Koponen and Salmi 2017; Vieira 2014; Guerberof Arenas 2019, p. 347) had noted that certain types of edit operations[17] demanded higher temporal or technical effort. Especially word-order changes, correcting mistranslations and incorrect syntax are the edits that took the longest time in PE (Popović, Avramidis, et al. 2014, p. 197).

Herbig et al. (2019, pp. 2–3) propose a comprehensive classification of the mechanical edits described in previous paragraphs by citing Koponen (2012), Popović, Lommel, et al. (2014), and Temnikova (2010). Respectively, Popović, Lommel, et al. (2014) set 5 PE operations: correcting word form, correcting word order, adding omission, deleting addition, and correcting lexical choice; Koponen (2012) contributes with the difference between moving single words or groups, and calculating the distance of the movement; and Temnikova (2010) also distinguishes between the addition or replacement of punctuation, correction of mistranslated idiomatic expressions, and replacing a word with a different lexical item vs. with a different style synonym. Herbig et al. (2019) create 7 referents from the cited operations and perform an elicitation study to investigate the cognitive process of PE with the aim of developing the modality. The authors findings confirm that reordering is the most complex operation for post-editors, and that other modes of input, such as speech, touch or digital pen can be an alternative solution to "move away from mouse and keyboard-only approaches".

Some difficulties in the analysis of PEA remain: e.g. when the output of the MT makes no sense (called "word salad" by Blain et al. 2011); when the

[17] Considered as a synonym of PEA.

structure of a segment is radically changed, "making the decomposition into PEAs impossible; or when a post-editor introduces a mistake" (Blain et al. 2011). To evaluate or asses the PEAs of the PE process, four categories have been used in previous literature that were probably introduced in the 1980s (Loffler-Laurian 1984, p. 237):

1. fit-for-purpose (OK), called "necessary PE" by Loffler-Laurian (ibid.);
2. over-editing (KO, as it vulnerates the definition of qualitivity in PE at least as per the value-based approach to quality, and obviously in the misused sense of productivity), it happens when the post-editor applies "preferential" or "unnecessary" changes to the MT output, as will be discussed in the next section, and it is called "superfluous PE" by Loffler-Laurian (ibid.);
3. under-editing (KO, as it vulnerates the qualitivity in the quality side of the notion, in most of its approaches), it happens when the linguist fails to apply "necessary" changes; and
4. pseudo-editing (super KO, it vulnerates qualitivity as a whole, both sides of the notion), it happens when the linguist introduces a mistake that was not originally present in the MT output.

These four concepts appear again in the next section and chapters, as they are essential in PE assessment[18]. For now, they allow us to conclude that, to find qualitivity, at least two prerequisites must be met: (1) clear instructions and guidelines (in that sense, the MTPE assignment is reviewed in the next section), and (2) adequate training and practice. Only with a relevant assignment and satisfactory training will the linguist be in a position to avoid under- over- and pseudo-editing as much as possible (see chapters 3 and 4). Figure 2.6 should summarise the discussions of the section we close here.

[18]For the purpose of the experiment presented in part III, only necessary editing actions are selected for analysis and the PEAs of the 34 participants are not classified into the four listed categories on the ground of practicality.

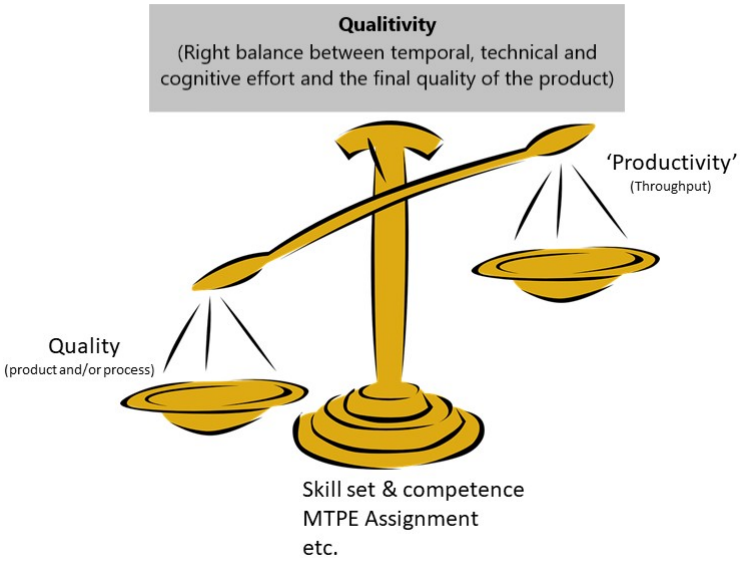

Figure 2.6: The concept of "qualitivity"

2.2. The post-editing assignment/brief

In 2.1.1 it was discussed how one of the conditioning factors to analysing any productivity gains is the *Skopos* (i.e. purpose and intended audience, among others) of the PE product. Moreover, as seen in section 2.1.2, the criteria upon which we base our method for evaluation of translation quality should be related to the translation brief, which is also in line with the "false picture of the quality situation" evoked by Mossop (2019, p. 232). Indeed, as concluded in section 2.1.3, one of the main instruments (along with skill/competence, topic discussed in chapter 3) that should help the post-editor act one way or another (that is, perform one PEA or another, and do it in a certain way or with a different method) is the MTPE assignment.

Mossop (ibid., p. 124) reminds the reader that "[t]he brief is a set of specifications including such matters as who will be reading the translation, whether it is a publication, and preferred terminology" and that it may or not include instructions. The author highlights that the content of the brief is shaped with information coming from three sources: explicitly stated by the client (orally or in writing), assumed from previous similar jobs from the same client, and elicited by the translation services. Moreover, Mossop underlines that the brief must be known for the linguist to choose the adequate strategy (ibid., p. 124). Two concepts that appeared soon in PE research, and

that should be part of any MTPE commission/brief/assignment, are the PE levels (see the next section: 2.2.1) and the PE guidelines (see the subsequent section: 2.2.2).

2.2.1. Post-editing levels

Teixeira (2019, p. 216) noted that "[t]he demand for different levels of quality is the result of a business strategy that has long been adopted in the translation industry as well as in other industries: cost-benefit analysis". For some time, two levels of PE have been accepted (Joscelyne and Brace 2010, p. 7):

- Full PE: publishable document (comparable to high quality HT), "but still different from human parity?" some still wonder, for example in a webinar organised by GALA on 21 May 2020 reported in Gene (2020).
- Light PE: understandable document (correct terminology and names, unambiguous but not necessarily elegant style).

To come up to the two accepted PE levels previous research in PE was considered. Loffler-Laurian (1986) and Elizabeth Wagner (1985) are the first authors to discuss such types of PE. The same dichotomy is presented by Krings (2001) and by Allen (2003). The relativity and subjectivity inherent to the notion of "translation quality expectation" is exacerbated in PE, given the notion of "good enough" introduced with the above-mentioned light PE level, frequent source of disagreement and discomfort amongst all stakeholders. This relativity is well captured by Absolon (2019)[19].

Between the two extremes there are a lot of "grey areas", as acknowledged by Koponen (2016a, p. 21), in turn citing Green (1982). These grey areas are described as cases where an MT output is reasonably good, but contains "doubtful translations and near misses" (ibid., p. 102), and they are source of trouble for post-editors. In that situation, the post-editor must make subjective decisions about whether corrections are needed, and if so, how extensive they should be. This issue is further explored by Krings (2001, p. 539), who found that such medium-quality sentences (grey areas) involved more effort than poor ones.

In terms of PE assessment, we could use the term introduced by Loffler-Laurian (1984, p. 237): "possible post-editing" (that should therefore be added to the list seen at the end of section 2.1.3: necessary or fit-for-purpose, over-, under-, pseudo- and possible editing). To decide which PE level, if any, is appropriate in a given translation project, one considerably comprehensive decision tree has been created by Nitzke, Hansen-Schirra, et al.

[19] Figure 4: *Illustrating the relativity of the term "Good Enough Quality".*

(2019, p. 246) and is reproduced in figure 2.7[20].

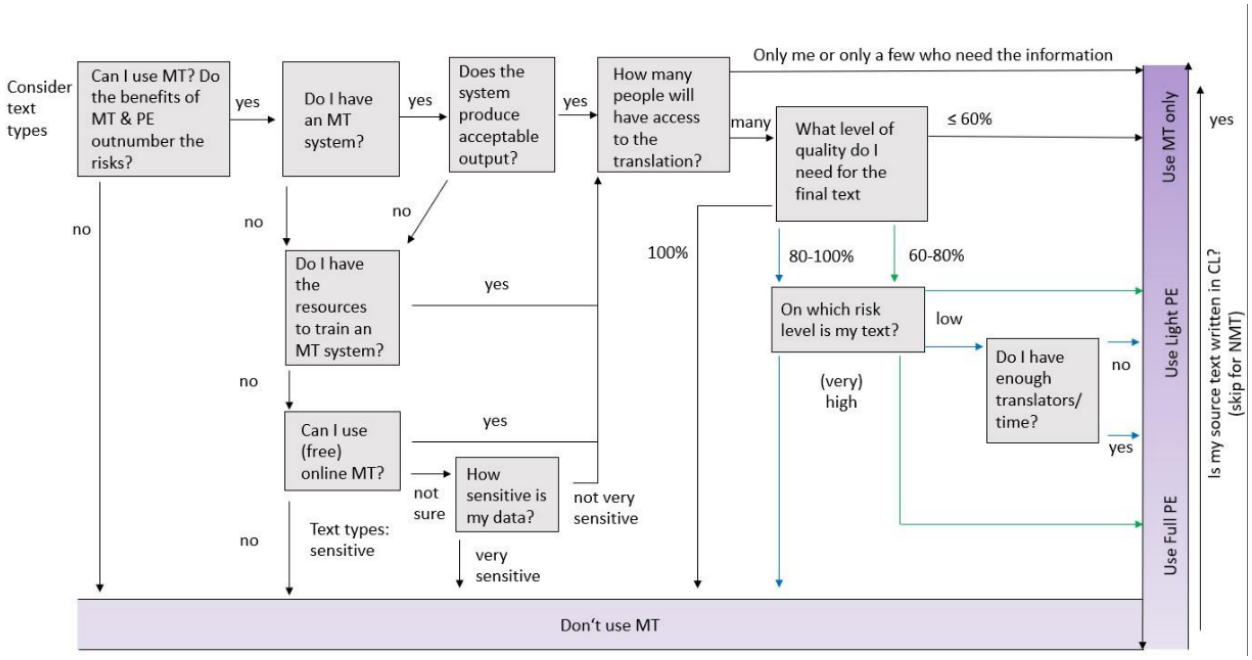

Figure 2.7: PE decision tree (Nitzke, Hansen-Schirra, et al. 2019)

The terms "light" and "full" are to be understood from the quality expectations perspective, rather than from the perspective of PE effort. They do not make the link between the MT output quality and the final desired outcome. Especially with the good quality that NMT outputs are showing nowadays, we may find useful to add nuances with adjectival quantifiers ("very light") or a two-factor terminology such as:

- light-light PE
- light-full PE
- full-full PE

In such scenario, the full-full PE would be one from a bad MT output quality up to a publishable quality expectation, with high PE effort, low (if none) productivity gains, let alone qualitivity. The links that could be established between three hypothetical levels of MT output quality and the quality expectations are depicted in figure 2.8: An explicit solution such as the one proposed in figure 2.8 (whereby the vertical connections are under the circle identifying the quality of the MT output and diagonal lines point towards the connection between raw quality to the expectation threshold) would solve the problem that "people misunderstand that these [two PE levels] describe

[20]The author thanks Jean Nitzke for authorising the reproduction.

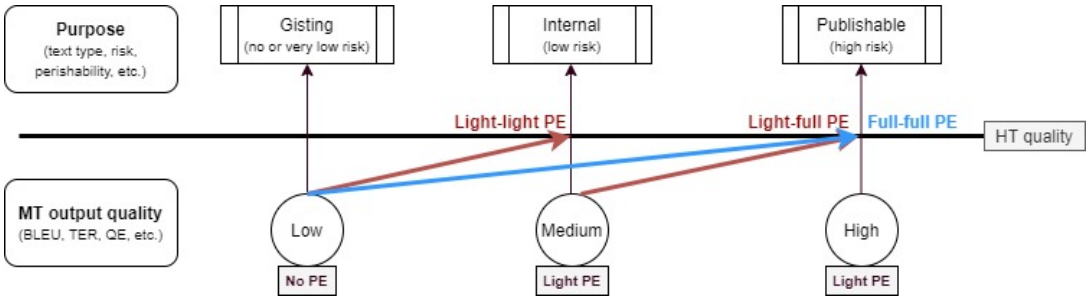

Figure 2.8: Connecting the quality expectations to the departure point

how much editing needs to be done [...] rather than what the final translation quality should be" (Nunziatini and Marg 2020, p. 3). However, it must be noted that moving from a higher degree of MT output quality to a lower level in the scale of purpose may still require checking, considering the notary-like role a post-editor could hold (Pym 2019). As a matter of fact, the number of quality levels in translation or PE has never found consensus, it can go from 2 to 7 depending on the source (Teixeira 2019, p. 215). Allen (2003, p. 301), like Nitzke, Hansen-Schirra, et al. (2019) and other authors, observed that the choice of PE level for a given project depends on the same factors upon which qualitivity depends (see 2.1.1).

Likewise, Allen drew the reader's attention to the fact that the subjectivity around the concept of quality called for explicit instructions: "Post-editing guidelines and criteria" (Allen 2003, p. 306). While the model reproduced in figure 2.7 accounts for the quantity of human resources (there is a question about how many linguists are available), the model fails to consider if the linguists have the relevant PE expertise. In the same way, the notion of brief or assignment is not considered as a key factor in the decision tree. One good question to add would be: "Do I have clear and appropriate PE guidelines for my linguists?".

2.2.2. Post-editing guidelines

Allen (ibid., p. 307-315) reviewed some of the industry PE guidelines available at the time. More recently, a comprehensive study of PE guidelines was published by Hu and Cadwell (2016). The authors provide a summary of previous approaches to PE guidelines by reviewing the work by Elizabeth Wagner (1985), who wrote some of the PE guidelines for the translation service of the European Commission. They also comment on the "do's and don'ts" proposed by Belam (2003, p. 10), and on the guidelines by Mesa-Lao (2013, p. 14-16), who provides a general description of PE guidelines and enters into detail of what they should cover if they are for fast PE or for

full PE. Mesa-Lao also presents the "5-10 second evaluation" and the "high 5 and low 5" rules used at Microsoft to decide when to edit or discard an MT output. Massardo et al. (2016) finally published what would become a widespread reference in terms of PE guidelines: a short set for "good-enough" quality (p.17) and a larger set to achieve "quality similar or equal to human translation" (p.18; note the variability introduced by the use of both terms "similar or equal"):

- Aim for grammatically, syntactically and semantically correct translation.
- Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of Do Not Translate terms.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling, punctuation and hyphenation apply.
- Ensure that formatting is correct.

Flanagan and Christensen (2014) tested the TAUS 2010 guidelines with translation trainees, which resulted in a tailored set of PE guidelines for their use in class.

Within the industry, as observed by Hu and Cadwell (2016), there have not been many publications on tailored PE guidelines by LSCs. With the information they had to date, the authors published two very clear tables. The first table (ibid., p. 349) compares the light PE guidelines by Densmer (2014), Massardo et al. (2016), Mesa-Lao (2013), and O'Brien (2010). The second table compares full PE guidelines by the same four sources (Hu and Cadwell 2016, p. 350) and it includes two categories that were absent from the first table dedicated to light PE: punctuation and formatting. The elements that are common to both tables are: accuracy, terminology, grammar, semantics, spelling, syntax, style, restructure, culture, and information. Both tables include a field "Others" which contains miscellaneous instructions that did not fall in the other categories.

The authors conclude that many guidelines overlapped, especially for light PE, and that the main differences affect stylistic improvements which, according to the authors, depend on the use and type of text. In conclusion, they insist on the lack of standard PE guidelines and sum up other sources that published language-dependent or aim-specific PE guidelines. Consistent with their work is the article by Nunziatini and Marg (2020), in which Appendix A and B compare more recent sets of PE guidelines on the basis of DQF-MQM Framework.

Around the time when NMT appeared, the industry standard containing a set of full PE guidelines, ISO Central Secretary (2017, p. 8), was published. In it, the seven guidelines by Massardo et al. (2016) are reordered,

two are added ("reestructuring sentences in the case of incorrect or unclear meaning" and "ensuring that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed"), and one is deleted ("use as much of the raw MT output as possible"). Thus, the community disposes now of eight general PE guidelines.

2.3. Post-editing: a translator's job?

At the beginning of chapter 2, it was noted how, to come up with the PE levels and guidelines discussed in the previous section, the first researchers had explored the PE activity by comparing it to translation and revision, both having a longer tradition. Undoubtedly, the recent evolution regarding the PE process and product (2.1) and regarding the figure of the MTPE assignment or commission (2.2) have had an impact on the profile of the professional in charge of the PE activity. Who is this professional?

Since PE research was born, we wondered "who are the post-editors?" (Allen 2003, p. 298) as it was seen as a new role. Traditionally, the PE work has been attributed to professional translators, for example in Krollmann et al. (1974), Lehmann et al. (1981), and Vasconcellos (1986), and many authors later, such as Rico Pérez and Enrique Torrejón (2012, p. 167). Besides translation, the task of PE has traditionally been compared to revision (Krings 2001; Vasconcellos 1987). Soon it was highlighted that "[p]ost-editing is not revision, nor correction, nor rewriting" (Loffler-Laurian 1984, p. 237).

The main difference between translation and PE, happens to be the main similarity between PE and revision, that is: whereas translation involves two texts, a ST and the TT produced entirely by the translator, revision and PE involve three texts, a ST, a draft TT (either translated by a fellow professional or pretranslated by an engine), and the final TT that the post-editor delivers. In that sense, Mossop (2006b)'s work on revision is interesting insofar it can be applied to the tasks performed by a post-editor. For instance, one of the central principles (ibid., p. 5), defined as "guides to action" are, for PE, the instruction of using as much MT output as possible, more specifically, of "handling the trade-off between necessary changes and over-editing" (Nitzke, Hansen-Schirra, et al. 2019, p. 249). In particular, such guides to action can also be easily compared to PE guidelines (an aspect reviewed in section 2.2.2).

The section on revision of MT output (Mossop 2006b, p. 115-116) was brief in this second edition of the book, since PE as a research field was still in its infancy. In it, Mossop highlighted the main challenge for PE, which is still valid today: "how, with the fewest possible keyboard operations, can I

achieve at least the low end of the readability/clarity scale?". In the fourth edition of the book (Mossop 2019), a chapter is dedicated to the topic, and the quoted question is posed again (Teixeira 2019, p. 218). Therefore, it has been established (e.g. Muzii 2016) that PE and revision are distinct tasks that require different skills.

In the book published by Angelone et al. (2019), the chapter by Guerberof Arenas (2019) is a good summary of the state of the art regarding the profile of the post-editor. Two notable observations made by the author are that the PE activity is still perceived as a tiring task (p.347) and that further research and collaboration is needed (p.349) to continue the effort in building knowledge regarding PE, especially because the (MT) technology keeps evolving at a fast pace.

For a long time, MT was perceived negatively among professional translators, since their new role as post-editors was less central and carried possibly negative influence on their remuneration and status. To give just some examples of such opinions one can read the work by Christensen and Schjoldager (2016, p. 90), Doherty (2016, p. 962), Kornacki (2018, p. 128), Pérez Macías (2020, p. 13-16), and Moorkens (2017). In regard of the post-editor prestige and status, Pym et al. (2013) noted that the National Research Council of the United States, in 2001, did not include translators in the status of "profession". Furthermore Neubauer (2015, p. 32) critically commented the "de-professionalisation" that a PE practitioner may undergo unless certain standards by all stakeholders are met.

Elizabeth Wagner (1985) indicated that the PE activity should not be undertaken by "inexperienced staff" and that the post-editor needs "a high level of linguistic and technical knowledge [to] post-edit the raw output to a reasonable standard in the recommended time". Similarly, Wheeler (1995) recognised the importance of attributing the PE task to a professional translator who also is an expert in the matter. Since achieving both a high expertise in translation and a deep subject-matter knowledge certainly requires years of working experience. As De Almeida and O'Brien (2010) observed "more experienced translators are also faster and more accurate post-editors. However, experience as a translator might also lead to a propensity to implement a higher number of preferential (or stylistic) changes, which is often contrary to PE guidelines". To cope with the difficulty in finding experienced translators who have enough expertise in a given subject matter and who can (or want to) post-edit MT output, authors like Temizöz (2013) have researched the difference in speed and quality between translators and experts. In her dissertation, Temizöz (ibid.) also concluded that any professional translator has to combine his or her experience with solid knowledge in the subject-matter.

The industry has similarly wondered in the past who the right post-editor

is. For instance, Gene (2019) asked "Is the post-editor a translator or a reviewer?", whereas Marheinecke (2016b) imagined other terms, such as "language expert", "geek", "translator" and "half techie", as possible combinations into a more hybrid profile[21]. One under-researched profile seems to be the paralinguist, which is at the centre of the study published by Van Ess-Dykema et al. (2010), who thank the Canadian [National] Translation Bureau for raising awareness on this term, which can be described as "analogous to the term and function of a paralegal professional" (ibid., p. 2).

The reality in the market reflects some confusion, to the extent that an "amateur crowd" is sometimes in charge of professional translation services (Absolon 2017, p. 4). For this reason, instead of defining the role and profile with a single restrictive term (translator, reviser, proofreader, reviewer, digital linguist, paralinguist, post-editor, editor, etc.), and to acknowledge its transdisciplinarity[22], one must obtain deeper insights about the core tasks of the professional whose one of the main activities is PE. Multi- and transdisciplinarity have been defined in the following terms by Gambier (2006):

> [L]'interdisciplinarité [...] [c]omme mode de proximité ou juxtaposition, jonction de disciplines: c'est la pluridisciplinarité qui fait de la traductologie une discipline-carrefour.[23]
>
> [...]
>
> [L]'interdisciplinarité [...] [c]omme mode de passage ou synergie entre deux ou plusieurs disciplines qui peuvent aller jusqu'à fusionner, c'est-à-dire à modifier leurs frontières respectives, pour un bénefice mutuel. C'est la transdisciplinarité[24].

To build up on the proposal provided by Sánchez-Gijón (2016), a list of PE tasks is designed in part II for the survey-based research. Thanks to it, one obtains insightful ideas about the tasks that may be combined with PE. In fact, authors such as A.-M. Robert (2013, p. 38) observe that PE can constitute up to 25% of a linguist/translator's professional activity, leaving 75% of their working day available for other tasks.

The profile of the post-editor, from the tasks perspective, has mainly been studied by Rico Pérez and Enrique Torrejón (2012) and Sánchez-Gijón

[21] This should remind the reader of the list by Bond (2018) mentioned in the Introduction.

[22] Interdisciplinarity is an extremely complex concept that encompasses multi-, cross-, pluri-, inter- and transdisciplinarity (Van den Besselaar, Heimeriks, et al. 2001, p. 706), especially in application-oriented research. Here, the definitions by Gambier (2006) are adopted as reference.

[23] The word "multidisciplinary" is used in this dissertation, defined by the Merriam-Webster's dictionary as "combining or involving more than one discipline or field of study".

[24] The term "transdisciplinarity" is used in the present dissertation.

(2016). The former defines several PE tasks based on Krings (2001, p. 321-522) and groups them in seven domains (Rico Pérez and Enrique Torrejón 2012, p. 168-169):

1. Source text-related processes
2. Machine translation-related processes
3. Target text production processes
4. Target text evaluation processes
5. Reference work-related processes
6. Physical writing processes
7. Global task-related processes

Inside the seventh domain, the authors see the post-editor as a stakeholder of the controlled translation scenario they investigated in Torrejón and Rico (2002). On the other hand, Sánchez-Gijón (2016) proposes two post-editor profiles. One with the role of "segments validator" (i.e. the task of editing the MT output); and one extended, more comprehensive, profile (called "provider of translation services"). The latter would be in charge of editing MT output but would also engage in tasks within other three, more managerial, groups of tasks: (1) Preparing materials for PE, (2) Preparing MT engines and (3) Managing MT systems. Similarly, Geoffrey S Koby (2001) highlighted how rarely would in reality a translator spend all the working day post-editing exclusively.

In a new role such as the one of the post-editor, where the limits of its scope are blurred, the definition of the professional profile is considered a prerequisite to determine skills and competences (Plaza Lara 2014, p. 81-82). Having preliminarily addressed it in the present section[25], the following chapter delves into the skills and competences.

[25]The professional profile of the post-editor is further researched in part II.

Chapter 3

THE POST-EDITING SKILL-SET

In the following section (3.1), we provide a summary of recent and relevant Translation Competence (TC) models. For a deep review of the definition of "competence" and reflections on the acquisition of TC, we refer the reader to the publications of Hurtado Albir (2007) and Hurtado Albir (2015). To read more about the difference between translation and translator competence, we refer the reader to the work of Kiraly (2012b), Kiraly (2014), Kiraly (2015), and Kiraly and Massey (2016), who considers translation education, the profession and translation assessment from a constructivist perspective.

As observed by Koponen and Salmi (2017), previous studies often assumed that the edits by post-editors were both correct and necessary. However, De Almeida (2013) and Koponen and Salmi (2015) highlight that errors may remain even after PE, or that post-editors can introduce errors (pseudo-editing). Indeed, De Almeida (2013, pp. 189, 192) states that up to 25% of the edits were preferential in the studied PE sessions in the framework of her doctoral dissertation. This is the reason why the focus of the present chapter falls not on TC but on specific PE skills, discussed in section 3.2.

3.1. General considerations

The complexity, nuances and controversy of the concept of "competence" have been object of discussion and research for a long time (Celia Rico 2017, p. 84). The varied definitions and the changing nature of the concept depending on the perspective where we view it from were recently tackled by Esfandiari et al. (2019, p. 2). The same observation is made by Piotrowska and Tyupa (2014) as reported in Kornacki (2018, p. 63):

> The main problem with translation competence is that it is

> not something constant. It changes, evolves, constantly adapts in order to accommodate market demands and new resources available to translators.

To provide a general definition, within the field of project management, we may look at "competence" as "the skill and capacity required to complete assigned activities within the project constraints" (PMI 2017, p. 319). In TS, thorough reviews of TC models have been published in the past, for instance the works by several authors cited in Calvo (2011, p. abstract). Whilst it is out of the scope of the present dissertation to comment extensively on the evolution of TC research[1], it must be highlighted that such variations may come from the distinction made between "translator" competence and "translation" competence. The former would be seen from the "vocational" and "transferable" approach, whereas the latter would correspond to the "academic" and "vocational specialised rationale" (ibid., p.13).

The three more consensual TC models are the PACTE model (Albir et al. 2020; PACTE Group 2005), the TransComp model (Göpferich 2009), and the EMT[2] model (EMT Expert Group and others 2009). They are reviewed and compared by Kornacki (2018, pp. 30–42), and presented again in figure 3.1: The PACTE Group revised previous and parallel research in the field of TC and defined it as: (a) expert knowledge; (b) predominantly procedural knowledge (i.e. non-declarative); (c) comprising different interrelated subcompetences; and (d) including a strategic component which is of particular importance" (Beeby, Fernández Rodríguez, et al. 2011). Furthermore, the PACTE model allows for variations in the scope and role of competencies, in particular:

> in each translation specialty greater importance will be given to different psychological abilities (logical reasoning in technical translation, creativity in literary translation) and the translation context (translation brief, time, etc.) may require a certain subcompetence to be activated (instrumental/professional, psychophysiological, etc.). (Kornacki 2018, p. 31-32)

As Kornacki (ibid., p. 35-36) comments by quoting Klimkowski et al. (2015), the EMT TC model "revolves around the notion of Translation Service Provision competence" and the author concludes that the TransComp model can be considered a combination of the previous two (Kornacki 2018, p. 41-42).

[1] From the "supercompetence" concept (Wilss 1982), going through the associative and macro-strategy competences put forward by Hönig (1991), to the minimalist views by Lörscher (1991, p. 2) and Pym (2003).

[2] European Master's in Translation (EMT).

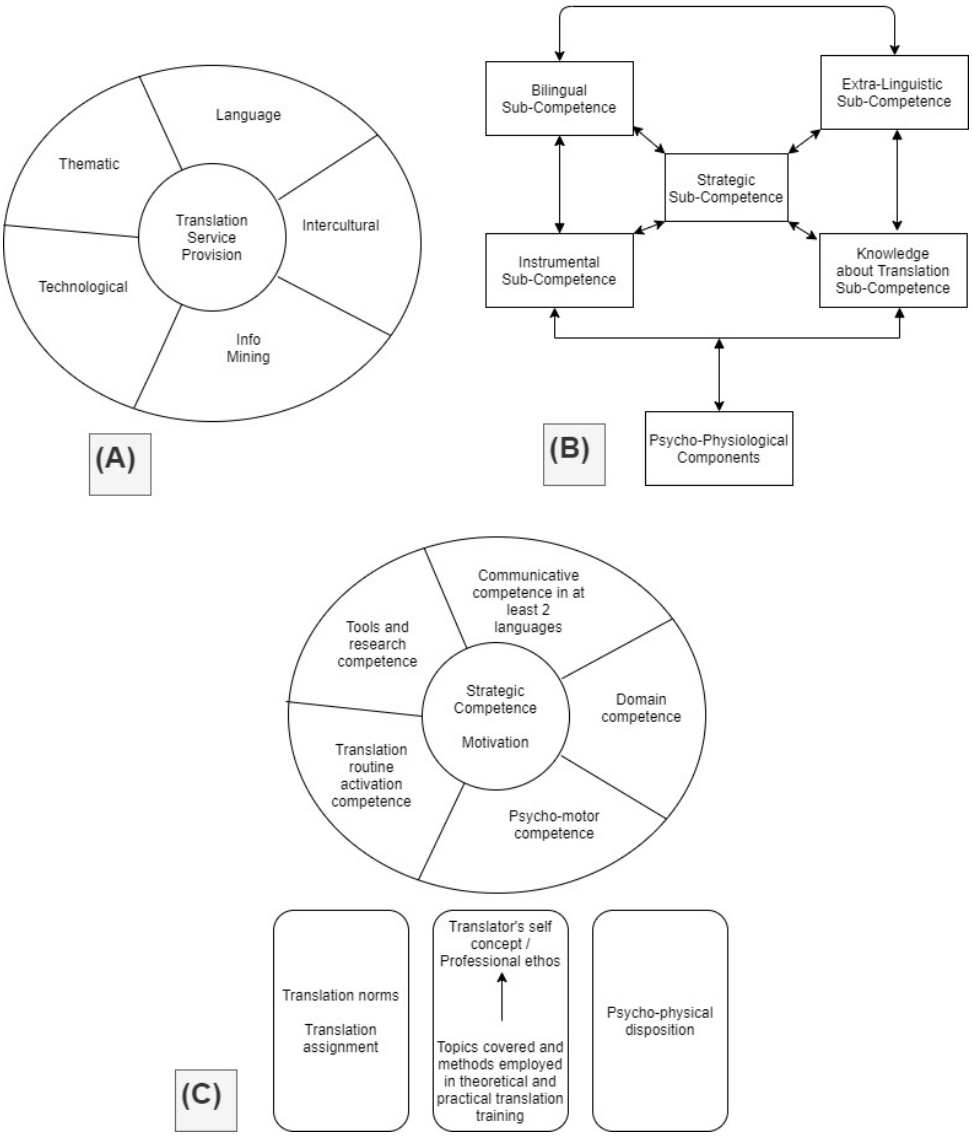

Figure 3.1: EMT (A), PACTE (B) and TransComp (C) models

All things considered, and in line with the observations commented above by Esfandiari et al. (2019) and Piotrowska and Tyupa (2014), Kornacki (2018, p. 36-37) also claims that due to how the profession evolves, "any attempt at setting a fixed framework is doomed to failure in the long run". The author highlights that the translation profession has become (or keeps becoming) a "hybrid profession"; one which combines competencies of different roles, jobs or activities (ibid., p.37). Moreover, the three illus-

trated TC models still have some detractors: Kornacki (2018, p. 63) cites Piotrowska and Tyupa (2014):

> the existing models of translation competence (PACTE 2003, EMT 2009) are inadequate in terms of translation teaching [...] translation pedagogy remains peripheral and enjoys less prestige in comparison to more theoretically-oriented sub-branches of the discipline (p.20) [and] most translators educators are not trained to teach translation as a profession. (Kornacki 2018, p. 63)

Kornacki (ibid., p. 63) himself, based on interviews with 2nd year Bachelor students of English at the inception of their translation programme, claims that "it is the practical approach that attracts students to translation courses". Moreover, according to Kornacki (ibid., p. 64), the emergentist view by Kiraly (2014) and Kiraly and Piotrowska (2014)[3] shows that, although such TC models (PACTE 2003, EMT 2009 and TransComp 2009) can be regarded as milestones of TC discussion, "a new learner-oriented approach is required". Likewise, Pym (2003, abstract) claims that such multicomponential models are "conceptually flawed in that they will always be one or two steps behind market demands".

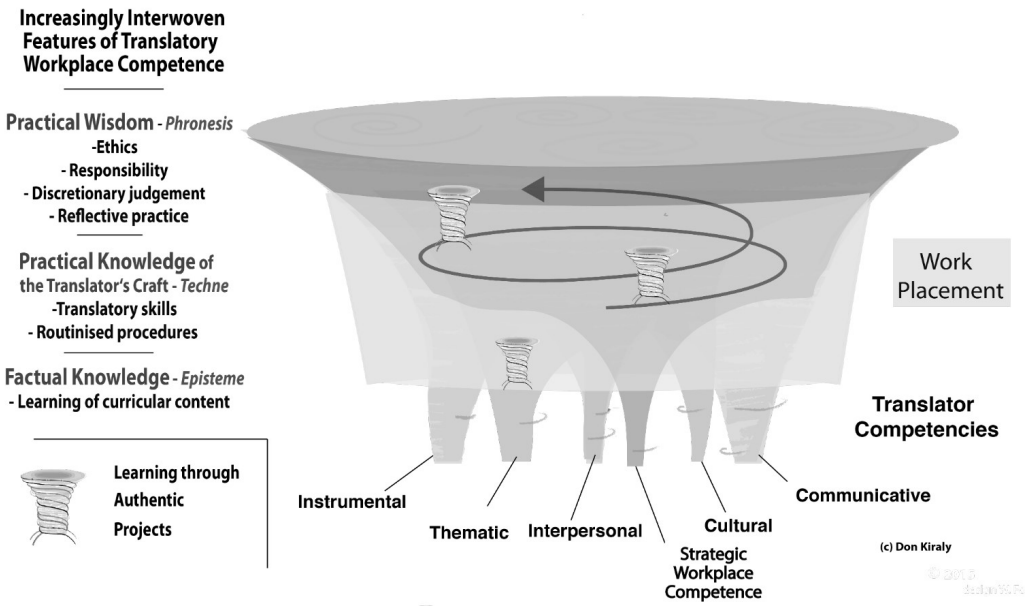

Figure 3.2: The interwoven features of TC by Kiraly (2016, p. 7)

[3]The author thanks Donald Kiraly for authorising the reproduction of figure 3.2.

What is certain is that, as Hurtado Albir (2008, p. 20) puts it, professional competence is a complex know-how combined of skills, abilities, values, knowledge, etc. that guarantees the effectiveness of a craft. The author notes how such professional competences should be the departure point for syllabi design, which is discussed in chapter 4. As observed by Calvo (2011, p. 5):

> [t]he existing plethora of definitions and competence models can make the concept difficult to pin down, particularly when trying to apply translators' and/or translation skills formulae to specific contexts and operational needs.

For instance, to the extent that revision differs from translation, even if studies are scarce, it has its own Translation Revision Competence model, which is represented in figure 4 of I. S. Robert et al. (2017, p. 13). The author also provides a list of 5 criteria that allow for classification of existing models (p.6). Considering how PE is a particular process in localisation and translation, and bearing in mind the increasing demand for this service in the language market (as argued in chapters 1 and 2), the need emerges to define a specific PE skill-set.

3.2. What is specific to PE?

PE is certainly a specific context and presents operational needs that other linguistic services do not require. However, there is still not a consensual PE competence model, or rather, a widely-agreed PE skill-set within the TC model. In the following paragraphs, the literature on PE skills is reviewed. The first mention found in Google Scholar to the "Skill-Set for Post-Editing" is by Krings (2001). Indeed, as Kenny (2020, p. 13) reminds us, journals that started providing major visibility to the pedagogical research on translation technology appeared in the 2000s: "Revista Tradumàtica"[4] (founded in 2001), "The Journal of Specialised Translation"[5] (2004), and "The Interpreter and Translator Trainer"[6] (2007). Understandably, the road to defining a skill-set for PE departed from the already researched neighbouring activities: translating and revising.

Recently, Guerberof Arenas (2019, p. 348) provided an overview of previous work on PE skills citing O'Brien (2002), Rico Pérez and Enrique Torrejón (2012) and Pym (2013). Several researchers soon agreed that PE differs from conventional HT and that it requires specific skills. Indeed,

[4]https://revistes.uab.cat/tradumatica

[5]https://www.jostrans.org/about.php

[6]https://www.tandfonline.com/loi/ritt20

O'Brien (2002, p. 100) claimed that "post-editing skills are different from translation skills and we cannot assume that a qualified translator will be a successful post-editor", which was further approved by authors such as Şahin (2011, p. 2). Hence, the new skills that are added to the TC model, particularly applicable for the PE activity have been researched for the past two decades to identify in what ways or to what extent the accepted translation skills differ or are exacerbated for PE, and if any need to be removed or added to the model. O'Brien (2002, p. 102-103) builds up on fellow-researchers' work to sum up and propose a skill-set for PE:

1. The expertise in the subject area, the target language, the text-type and contrastive knowledge (Johnson and Whitelock 1987)
2. Excellent knowledge of the source language, perfect command of the target language, specialised subject knowledge, word-processing experience and tolerance (Emma Wagner 1987, p. 76)
3. Full key proficiency, cursor positioning, search and replace, use of macros, functional treatment of linguistic constructions, etc. (Vasconcellos 1986)
4. Positive disposition towards MT (Vasconcellos 1986; Emma Wagner 1987).
5. Knowledge of MT technology, current limitations and how this technology might improve in the future.
6. Terminology Management Skills with knowledge of several term management tools and terminology exchange formats, as well as how to code dictionary for their use in MT systems.
7. Pre-editing/Controlled Language (CL) Skills to be able to apply CL rules to a text before it is translated with the MT system.
8. Programming skills to be able to write macros or scripts for automatically correcting repetitive errors of the MT system.
9. Text linguistics skills useful both for PE as well as for programming macros and automatic PE modules.

Moreover, Doherty and Gaspari (2013) also enumerate a set of eight skills specific for PE:

1. excellent word-processing and editing skills
2. ability to work and make corrections directly on screen
3. general knowledge of the problems and challenges faced by MT
4. specific knowledge of the weaknesses of the particular MT system
5. knowledge of source and target languages
6. quick in making decisions as to what and how to correct or ignore errors
7. ability to always balance PE speed and cost with respect to required quality
8. ability to adapt to different specifications required for each job

Ellements 1 and 2 in Doherty and Gaspari (ibid.)'s list would correspond to 3 in the list by O'Brien (2002); 3 and 4 can be compared to 5 in O'Brien (ibid.); 5 is less specific for the 1 and 2 mentioned by O'Brien (ibid.); while the last three skills suggested by Doherty and Gaspari (2013) were not yet present in the proposal by O'Brien (2002). On the other hand, Doherty and Gaspari (2013) did not mention attitude or positive disposition towards MT, CL or pre-editing, higher programming skills, or TB management skills.

According to EDIT-TA project findings (Celia Rico and Díez 2013; Celia Rico and Orzas 2013), the methodology for PE is designed in three steps: (1) preliminary analysis (performed by the PE team coordinator); (2) post-editing of MT (done by post-editors); and (3) error reporting and quality control (done by both roles in collaboration). PE coordinators and post-editors collaborate in some tasks before and after the PE takes place, which leads to conclude that they share some competences. These competences are classified in the three groups presented in figure 3.3.

- Core competences (attitudinal or psycho-physiological competence and strategic competence).
- Linguistic skills (excellent knowledge of source and target language, PE guidelines, communicative and textual competence, cultural and intercultural competence, subject area competence).
- Instrumental competence (knowledge of MT systems, term management, MT dictionary maintenance, corpus quality assessment skills and some programming skills).

The step (2) "post-editing of MT" called for greater definition and delimitation both within the industry and academia. Guerberof Arenas, Depraetere, et al. (2012) reports an interview to three specialists in PE, and comments PE skills in one question. From their answers, some challenges that remain regarding the skills needed to perform the PE activity are highlighted:

1. The knowledge a post-editor has or should have of different MT technologies and how to interact with this technology;
2. How a post-editor measures their own PE effort compared to the expected quality to set a fair price for their work;
3. How a post-editor can minimise the technical (keyboard and mouse) effort;
4. How much practice should a post-editor have on the task of reviewing texts translated by humans;
5. What type and how much knowledge of error typologies should a post-editor have;
6. How a post-editor handles different aspects related to quality, as definitions or levels of quality and the customer expectations; and
7. How a post-editor uses, adapts or writes PE guidelines.

Saint-André (2015, p. 62-65) listed 34 competences for a post-editor: three belong to the group of writing skills; six, to translation skills; eight, to "correction skills"; eight, to "specialised" skills; three, to Information and Communications Technology (ICT) skills; and six competences belong to the group "professional aspects". The author concludes that the six most valued competences for PE are (p.105)[7]:

1. Abide to target language orthography, grammar, syntax, punctuation and typography. (CR3).
2. Master the target language (CT2)
3. Guarantee fidelity to the source message (CT3)
4. Quickly decide if an MT output should be kept, edited or deleted and translated from scratch (CC7)
5. Spot what needs editing (CC6)
6. Be able to identify the differences between the ST and the TT (CC5)

One can acknowledge how, from these six top PE competences, the first three correspond exactly to the TC competence model. CC7, CC6 and CC5, three competences belonging to the group of "correction skills", seem to be exclusive to PE. In particular, CC6 intrinsically depends on CC5.

While the subskills for PE have not yet found consensus, there are already some attempts to link preliminary subskills to higher categories. One example is the work by Nitzke, Hansen-Schirra, et al. (2019, p. 247-250), where the authors propose a set of four core competences, listed below.

1. Risk Assessment Competence
2. Strategic Competence
3. Consulting Competence
4. Service Competence

And a set of eight subsidiary competences:

1. Bilingual Competence
2. Extralinguistic Competence
3. Instrumental Competence
4. Research Competence
5. Revision Competence
6. Translation Competence
7. MT Competence
8. PE Competence

However, item 8 (PE competence), which refers to "spotting exactly [...] more fine-grained problems [of NMT]" (CC5 and CC6 in Saint-André (2015)'s work commented above), still calls for further definition and delimitation. Indeed, the fact of spotting an error in the MT output does not imply that the post-editor would choose the right solution. Hence, meta-

[7]Translated from French by the author of the present dissertation.

editing techniques for PE must be researched. Considering the previous literature discussed above, some questions that may be raised are depicted in schema 3.3, adapted from Rico Pérez and Enrique Torrejón (2012)[8].

The purple questions are inspired from the work on the post-editor profile(s) by Sánchez-Gijón (2016). The white questions (the minimum level of source language competence and the minimum level of TC), and the green questions are mainly inspired by the publications of O'Brien (2002) and Doherty and Gaspari (2013). The red questions result from what seems a balanced suggestion of PE processing speed (productivity) by Temizöz (2013, p. 15), and also from the distinction between necessary and non-necessary edits commented in section 2.1.3.

Finally, the yellow bit does not constitute lacunae or unknowns, but speculation or theories. While the skills reviewed in the present chapter have often dealt with competencies linked to language, technology or attitude (blue, green, and red, respectively, in figure 3.3), as argued in section 2.1.3, there must be "something" that must lead to qualitivity in PE. That "something" is obviously dependent or conditioned by, at least, (1) the characteristics of the project and how it is prepared (MTPE brief seen in section 2.2), and (2) the PE skill-set reviewed in the present chapter. Our conjecture, in yellow in the cited figure, is that the combination of attitudinal or core competencies with linguistic skills results in what once could call the Efficient Editing Skill (EES), which entails the capacity to decide how to edit a word or expression, or to replace an error from the MT output with a given word (shorter, for instance) instead of another (longer). In other words, from a linguistic perspective, one should make relevant and fast editing decisions, and choose one PEA rather than another.

In addition, the combination of the core competences with the instrumental could result in what we have named Instrumental Input Skill (IIS), which is defined as the ability to prioritise the most appropriate action and tool according to the PEA the linguist is performing. In other words, a professional post-editor should not only be comfortable, but make relevant decisions regarding the use of the mouse, keyboard (shortcuts, etc.), voice recognition, and other input methods taking into consideration the PEA at hand. Having EES and IIS interwoven could hypothetically lead to an empowerment of translation students and novice translators, allowing them to "conduct effective editing behaviors" (Yang and Wang 2020, p. 10). In their study, Yang and Wang (ibid.) present and discuss self-reported data of 109 novice Chinese student translators without MTPE experience which points towards positive effects of self-regulation on PE performance.

In an exploratory study, Part II aims to elucidate how PE is understood

[8]The author thanks the authors for granting permission to modify and publish the figure.

by LSCs, linguists, and trainers; that is, to find out to which extent there is (dis)agreement around very specific matters discussed in the previous chapters. With this goal in mind, a list of 11 skills is drafted in section 6.1.2 thanks to the review of previous work commented in the present chapter. To conclude, we may establish that the PE skill-set is not entirely defined to date and additional research is required. Nonetheless, the proactive response of training (PMI 2017, p. 319) has already emerged both within the industry and academic settings. We delve into it in the next chapter.

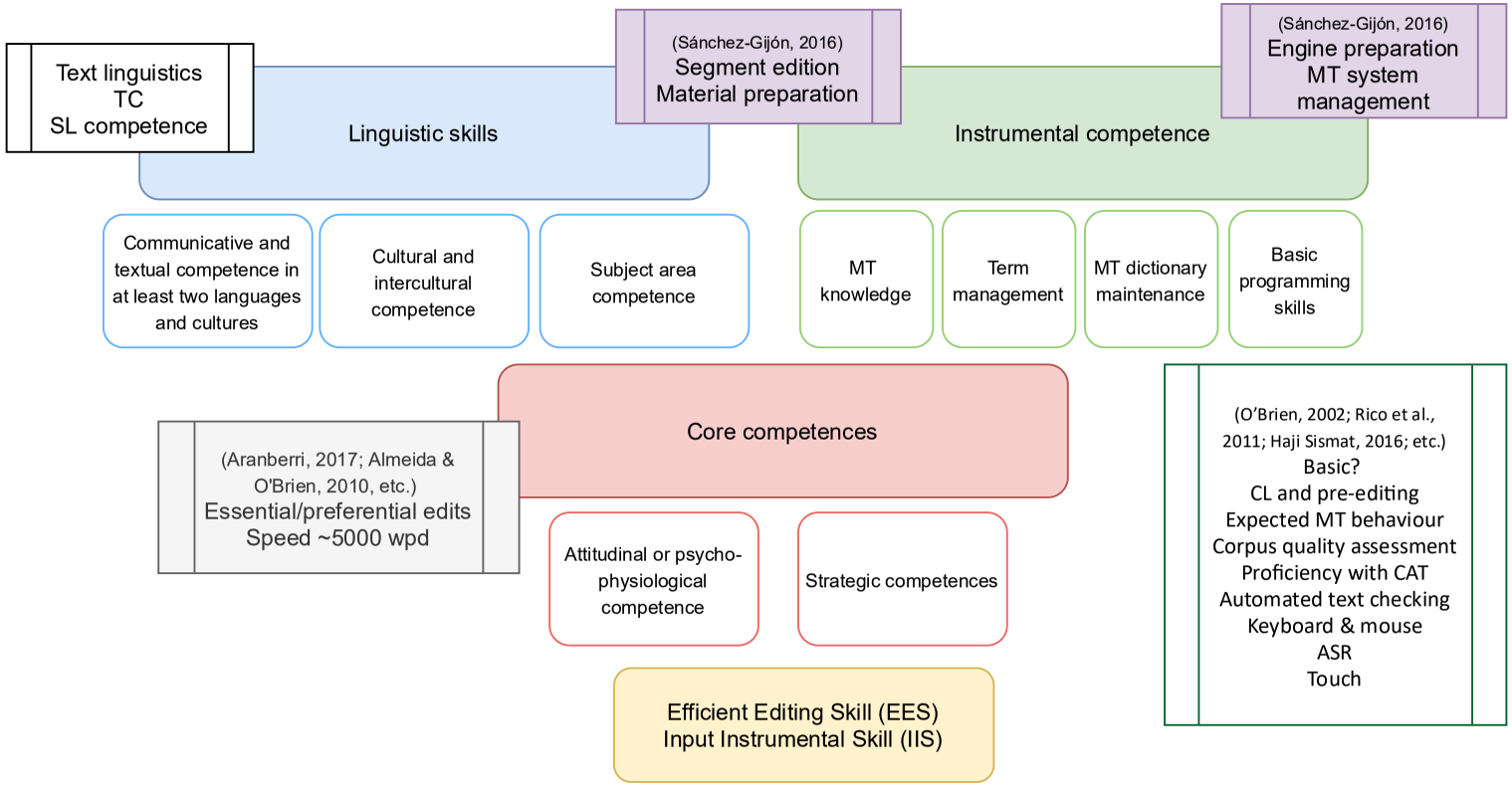

Figure 3.3: Adapted from Rico Pérez and Enrique Torrejón (2012)

Chapter 4

POST-EDITING TRAINING FOR TRANSLATORS

The field of TCC pedagogy places the tool (CAT or MT) as the object of learning, either in stand-alone modules or in a more crosswise manner of delivery, whereby translation technology would be taught "when the need arises across the entire translator-training curriculum" (Kenny 2020, p. 21). In the emerging field of PE pedagogy, however, the object of study is the learning of PE skills. Until recently, it represented a gap between the translation industry and the training programmes, which lacked the PE training component (e.g. Gaspari et al. 2015; Gene 2019).

> By the mid-2000s, many major software developers such as Microsoft, IBM, Autodesk, and SAP had implemented MT and were requesting trained post-editors [...]. (Guerberof Arenas and Moorkens 2019, p. 218)

Despite some criticism within the academic community about introducing MT in the classroom (Pokrivcakova 2019, p. 140), since the 1980s, the focus was more often put on MT (the technical side of it) rather than the service associated to it, that is, PE. Cadwell et al. (2018, p. 317) highlight that:

> [v]arious universities, such as University College London, offer courses in the use of MT and other translation technologies to professional translators. These courses are typically research-based and practical, and are conducted within the context of relevant research findings and with the collaboration of a commercial partner.

The research and work reviewed in chapter 3, along with the observation that the lack of experienced and trained post-editors was a challenge for training and recruitment (Stevens and Fuentes Corradi 2016), has favoured

the emergence of two capital statements: post-editors should be trained in advance (e.g. Massardo et al. 2016, p. 15; Hazbavi 2015) or established professionals should follow a Continuing Professional Development (CPD) to become competent in PE (Secretariat 2018); and training for trainers should be provided (Nitzke, Tardel, et al. 2019, p. 299).

In this chapter, the focus falls on the training of PE as an activity (and professional service), rather than other uses and applications of MT. In section 4.1, some PE courses for established professionals are briefly described. In section 4.2, we review current efforts to provide PE training within translation curricula in higher-education programmes. Finally, in section 4.3, a short presentation is made of Train-the-trainer courses in PE.

4.1. Continuous Professional Development

There have been discussions about matching translator training with the demands of the industry for a long time. To read more about this subject and practice-oriented syllabi, the reader is referred to the chapters by Fraser (2000) and Anderman and M. Rogers (2000).

In the following paragraphs, we provide an outline of some employability-focused PE training courses as opposed to purely academic approaches (Calvo 2011). Within the industry, translation technology providers such as SDL[1], Systran[2], KantanMT[3], Welocalize[4], among others, provide some short PE courses that may serve as introduction to a senior linguist (Aranberri 2014, p. 473). Furthermore, professional associations in some countries have also developed more comprehensive PE courses in the framework of CPD. In Spain, for example, one can find the English into Spanish online training provided by Trágora[5], which includes 6 theoretical chapters and 11 practical exercises. In particular, during the 80h course, the student will see topics such as:

1. Introduction to PE
2. The MTPE market
3. The translation process and PE tools
4. MT quality evaluation and evaluation of post-edited texts
5. PE project management

A well-known PE course set up within the industry is the one by TAUS[6], which may be less comprehensive, but includes some hands-on exercises.

[1] https://www.sdl.com/
[2] https://www.systransoft.com/
[3] https://kantanmt.com/
[4] https://www.welocalize.com/
[5] http://xl8.link/Tragora
[6] https://elearning.taus.net/course/index.php#pe-outline

The six theoretical modules are in English:

1. Introduction: MT history
2. Types of Machine Translation: MT systems
3. Evaluation of Machine Translation: MT performance (assessment and metrics)
4. Pre-Editing and Controlled Natural Language (CNL)
5. Post-Editing
6. Organisation and Setting up of MT Projects: Project management for MT

It contains some examples of MT raw output (in 13 languages), two language-specific exercises (you can choose among 31 languages), and some practice in DQF (error typology evaluation and productivity measurement). Excluding the hands-on part, it can be completed in less than 2 hours. One aspect that should be highlighted is that the figure of 100,000 words (or one month) of PE hands-on practice to reach a certain level of comfort with the activity, presented in section 1.1, is cited in TAUS PE course again as a reference.

It is noteworthy that most, if not all, professional PE courses are introductory. In that sense, they tackle MT history and some basic technological background since the range of audience profiles is usually wide. This is illustrated by the weight given to the PE module: it represents only one of the six parts in TAUS PE course. The length and detail of any course is therefore an important factor to consider. Recently, Absolon (2017), Absolon (2018), and Absolon (2019) has researched the "trainability of PE skills and competences" and recommends a method called "split techniques" for PE training. In Absolon (2017, pp. 7–9) the author explains the origin of the so-called "split-training". According to the author, having one exercise for each specific error type (if we think of language competency) could yield "more results in a shorted period of time". The author devised a research project that involved 39 first-year Master's students at Constantine the Philosopher University in Nitra. The study had the following phases:

1. Pilot project
2. Observation and analysis of post-editors work
3. Post-editor job profile creation
4. Design of a training plan
5. Set up of tests
6. Comparison of performance between HT and MTPE
7. Verification

While the translation experience and MTPE awareness of the 39 students was, in general, modest, a majority of them used MT for unknown words or for whole sentences and almost 56% of them claimed that "MT quality varies but it is a big help". In the pilot project, Absolon observes that the

increase of productivity is higher compared to the loss in quality when comparing post-edited texts to HT. He suggests three possible explanations: (1) the use of MT could be "second-nature for young people", (2) the ST was appropriate for MT, and (3) the students benefit from MTPE as they are not experienced translators. Without commenting on items 2 and 3 of the previous enumeration, we turn now to the design of a training plan. Absolon (2017, p. 23) prepared 12 exercises based on the split-technique:

1. Decision-making
2. PE segments with morphological errors
3. PE segments with typographical errors
4. PE segments with syntax errors
5. PE segments with semantic errors (omissions, mistranslations)
6. PE segments with terminological errors
7. PE segments with factual errors
8. Identification of errors
9. Comprehension of the topic

> [W]e train individual techniques separately and intensively and only later begin training all techniques together. It is similar to when an athlete practices particular individual techniques of his performance and only later the entire activity as a whole. (Absolon 2019, p. 20)

To conclude, the author created a course in cooperation with the Department of Translation Studies at Faculty of Arts, Constantine the Philosopher University in Nitra to help students, novice translators, and professionals "to deal with challenges of the future translation market" (Absolon n.d.).

4.2. PE within translation and translator curricula

Within academia, the observation is often made that, despite several translation technology handbooks (Austermühl 2001; Bowker 2002; Quah 2006; H. Somers 2003) being available, PE is still relatively new and training has not been implemented homogeneously in University translation programmes (e.g. Celia Rico 2017, p. 80-81 and Oliver 2016, p. 205). Indeed, one first outline for a PE course module was proposed by O'Brien (2002, p. 103), but at least a decade would pass until it became a reality in a yet small number of universities.

It has been reported by Moorkens, Castilho, et al. (2018, p. 87) and Guerberof Arenas (2019, p. 348) that syllabi which include PE have been explained and described since 2012. However, many of the examples given

focus again on MT or translation technologies in general, and may eventually include a short PE session (theoretical or practical) during the course. For this reason, they can be seen as part of the MT pedagogy field, which clearly shares boundaries with the PE pedagogy field. Doherty, Kenny, et al. (2012), for instance, studied how to introduce SMT in the classroom and concluded that "improvements were found for general MT knowledge". Doherty and Moorkens (2013) put the emphasis on the benefits of hands-on experience in translation technologies, not specifically in PE, and conclude that the updating process of such labs must be continuous.

Few studies have their focal point of research on PE pedagogy. One of them is Koponen (2015), where the author presents a PE course at the University of Helsinki in Autumn 2014. Koponen describes the content of the seven two-hour lectures, five homework assignments, in-class exercises, and a final reflective essay. The Moodle platform was used for this PE course, which was completed by 13 students and was structured as follows:

1. Theory and history of MT and PE
2. Practical use of MT and PE
3. Controlled language and pre-editing for MT
4. Post-editing without source text
5. Post-editing process research
6. Post-editing quality levels and guidelines
7. MT quality evaluation and PE effort
8. PE competences

On the other hand, the five assignments in Koponen (ibid.) are the (1) comparison of MT versions, (2) Pre-editing, (3) PE without source text, (4) Quality levels, and (5) MT quality evaluation. In terms of MT providers, being limited by the availability of Finnish language, Koponen (ibid.) used the system of the European Commission, a rule-based system by Sunda Systems[7], and Google Translate. To post-edit, the students used Ms Word, and the Appraise evaluation tool[8] (Federmann 2012) for one exercise.

One of the main conclusions of Koponen (2015)'s work is that, through training, the negative views of students towards MT change to more positive attitudes. Moreover, thanks to the reflective essays and collection of feedback, pre-editing is found to be a less interesting/useful topic for the students, an opinion that some also expressed about PE without ST. Moreover, a big challenge according to the author "relates to the fact that the course is offered to students in all language and translation subjects, which makes it difficult to provide materials for all the language pairs".

Another substantial contribution to the field is the work by Guerberof

[7] http://www.sunda.fi/kaantaja.html

[8] Despite being a tool thought for MT evaluation.

Arenas and Moorkens (2019), where the authors present the introduction of a PE module at Universitat Autònoma de Barcelona. The 8-hour PE module, greatly inspired by O'Brien (2002), includes since 2009 the following topics:

1. Basic definitions of PE (PE versus revision, post-editor profile, and CNL and pre-editing)
2. Quality (metrics for raw MT output, expected quality of the assignment, and metrics to assess a post-edited text)
3. PE levels (light and full PE; which seems to partially overlap with the "expected quality of the assignment" in the previous item)
4. General rules for PE (guidelines for light and full PE)
5. Common MT errors
6. PE effort and productivity
7. PE and pricing

The module is complemented with six practical exercises: (1) translating a technical text from scratch; (2) comparing their translation from exercise 1 to different MT outputs; (3) CL and pre-editing; (4) Monolingual PE (as seen before in Koponen, 2015); (5) PE using an web-based tool (MateCat) following PE guidelines; and (6) PE using SDL Trados following PE guidelines. In line with the discussion above about Koponen (2015)'s work, Guerberof Arenas and Moorkens (2019, p. 225) also note that the varied and unpredictable language combinations represent the main challenge for a PE module and suggests that assessing through closed questions using a questionnaire may be the optimal route. Another case study with two cohorts of students is presented by Moorkens (2018), where trainees assessed NMT quality (adequacy), performed error annotation, and experienced the task of PE. In particular, they learned about PE temporal effort.

Recent work in PE pedagogy, without any case study but with alternative and innovating ideas, has been carried out by Mellinger (2017). Unlike previous work presented above, the author argues that MT should be a part of translation courses across the curriculum, rather than a single module. In that sense, regular translation courses could tackle topics such as terminology management, controlled authoring, post-editing, and engine tuning. Mellinger (ibid., p. 287-288) observes the need of incorporating PE in language-specific translation courses and highlights how "post-editing tasks may yet be one more approach to translation". However, one question that still comes up regularly when discussing PE in translation/translator curricula is when exactly should be introduced. O'Brien (2002) argued it should be optional at the postgraduate level.

Despite the logical and justified reasons (mastery of source language, mature TC and transfer skills, etc.), the reality of students using MT since the first year of undergraduate programmes is opening the question again.

In the study of Schumacher (2020, p. 260) with 28 students, the author, by citing Garcia (2011) and Killman (2018), highlights that there can be a "leveling effect" when exposing translation students to MTPE, which consists in an increase of the quality of the post-edited texts delivered by mediocre students, while lowering the scores for the best students, whose creativity is undercut when post-editing.

To conclude, we should be aware that several challenges may arise during PE training, and that they may vary depending on the type of PE course, as a stand-alone or as a curriculum-wide implementation. For example, A.-M. Robert (2013, p. 37-38) notes that, unless a collaboration agreement is signed with an industry agent (LSCs, MT provider, etc.), the professional PE as a service cannot be taught in good conditions, if only due to the MT output quality (as seen in chapter 1, customised proprietary systems are used to provide better quality than free online MT systems, which also pose confidentiality issues). On the contrary, one single agreement with a given LSC may provide a very deterministic view of the service that should be contrasted with practices in other firms.

All in all, we agree with Guerberof Arenas and Moorkens (2019, p. 231) that NMT is probably not changing PE training completely. However, there are a number of factors that motivate studying PE training further:

1. A consensus on some core concepts underlying the PE activity has not yet been found (PE pricing and rates, quality levels, the benefits CL and pre-editing may still have with NMT outputs with a view to lessen the PE effort, etc.).
2. The competence models and training modules should undergo continuous updating as the world evolves and progresses (not only translation technology keeps evolving, but new industry standards appear, and research is always ongoing).
3. PE strategies and techniques are yet to be defined (even if Blain et al. (2011)'s PEA represent a first step, to our best knowledge, the link between translation techniques and MT errors and PEAs has not yet been made (the meta-editing techniques mentioned in section 3.2).
4. As observed by Allen (2003, p.298-299), there is not a formal PE qualification or certificate. Even if there were, the author acknowledges how it would never "convey the professional pedigree of a full university degree program" (Neubauer 2015, p. 32).
5. Some translators may still be sceptical regarding MTPE and attitude seem to affect PE effort and performance (Stasimioti and Sosoni 2019, p. abstract).

At the time of writing, some PE modules and courses are being developed and implemented in translation/translator curricula, for instance, the Eras-

mus+ DigiLing project[9] presented by Nitzke, Tardel, et al. (2019) and acknowledged as a good initiative in EMT Network (2019). This course is under construction and it should represent 120 hours and 5 ECTS[10] credits. The topics it is set out to cover are distributed in the ten units listed below.

1. MT history
2. MT approaches
3. General introduction to PE
4. PE and text types
5. Light vs. Full PE vs. Proofreading
6. PE and Translation Memory Systems
7. Controlled languages and pre-editing
8. ST defects
9. PE in research
10. PE in practice

It is announced for a level "introductory to intermediate", and it uses presentations, recordings, exercises, quizzes, and readings. It also clearly states its five learning objectives: (1) learn how MT works, (2) understand PE, (3) distinguish HT, PE and proofreading, (4) evaluate factors that affect PE, and (5) learn about PE "in practice and research".

In part II, section 6.2 constitutes a comparison of 49 syllabi in postgraduate or master programmes which either are a single PE stand-alone course or include, at least, one PE hands-on task or module.

4.3. PE training for translators (for trainers)

Kornacki (2018, p. 84-85) argues that MT has to be considered in translator training programmes because it has "a profound influence on the way translators do their job" and follows Pym (2009) to see MT not as a "tool" but as a factor affecting the nature of the translation process. The author purports that one cannot provide probing training in state-of-the-art translation without a deep understanding of translation technology; to the extent to which it affects the very nature of Translation-Editing-Proofreading (TEP).

Several programmes and projects seek to continuously update the knowledge and competences of TS scholars. For example, the Optimale (Optimising Professional Translator Training in a Multilingual Europe) project was launched in 2011 by the EU Commission, based on a questionnaire survey of LSCs, and two of its goals were to "strengthen the relevance of translator training programmes by identifying and disseminating good

[9] http://www.digiling.eu/ and https://learn.digiling.eu/

[10] European Credit Transfer and Accumulation System (ECTS); https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en

practice in a certain number of key areas" and to "promote the quality of the training of teachers in setting up a platform for exchange of resources and organizing workshops on teaching practices"[11]. Some other projects, such as eCoLoTrain[12] (now collaborating with Translation Commons[13]), aim to improving the translation trainers' ICT skills. eCoLoTrain provides materials (curriculum, vocational training, etc.) in topics such as CAT and project management.

Such train-the-trainer courses would help regular translation teachers design realistic PE assignments (Marheinecke 2016b, p. 8), for example, recommends no less than 10,000 words per student. As reported in EMT Network (2019, p. 10), an agreement between the EU and the UN Office in Nairobi in support of the Pan-African Consortium of Master programmes in Conference Interpreting and Translation provides the possibility of training-for-trainers seminars in translation. For instance, from 9-13 July 2018, Pr. Andrew Rothwell (University of Swansea) and Pr. David Orrego-Carmona (Aston University), delivered train-the-trainers seminars on CAT tools. Such on-site or webinar workshops can happen regularly to cover different universities. Indeed, another example is the session "Neural Machine Translation and Post-editing: required skills and competences (a hands-on workshop and a hands-off presentation)" (ibid., p. 11) that took place in Hungary on 28 September 2018.

[11] https://sites-formations.univ-rennes2.fr/lea-cfttr/networks/?lang=en
[12] http://xl8.link/eCoLoTrain
[13] https://translationcommons.org/

Final thoughts

The present part has explored, compared and critically examined the state-of-the-art of Post-Editing (PE) from four distinct perspectives. In chapter 1, from the perspective of Machine Translation (MT) evolution, we have provided a review of the history of MT. The substantial improvements in Neural Machine Translation (NMT), leading to increased translation quality, have contributed towards the establishment of a new activity and service: PE. In fact, Artificial Intelligence (AI) has brought disruptive changes recently to move from "Convergence" in year 2020 towards a period named "Singularity" towards 2030, as depicted in figure 4.1[14].

In chapter 2, an overview of the profile of the "post-editor" has been provided. To read more about the similarities between the profile of a translator and the profile of a post-editor, an interesting summary was published by Saint-André (2011, p. 5). For varied reasons (monotony of work, pertinence of Machine Translation Post-editing (MTPE) for a given assignment, etc.), a digital linguist is most likely to have a hybrid profile (multi- and transdisciplinary), and this is why it is investigated with a thorough survey-based research project in part II. In words of Lee and Liao (2011):

> Future translators are faced with the need to develop new work skills conforming to the multitasked translator profile, which is increasingly becoming a must-have in the translation market.

Moreover, the variability from one professional to another is high. For instance, students and less experienced translators may over-rely on digital resources (the so-called "blind faith"), which would result in more or less serious translation errors, according to Kornacki (2018, p. 14), himself citing Bowker (2005) and Doherty (2016). Just like Kornacki (2018, p. 129) noted that CAT tools can reduce the capacity for learning in some individuals (it is no longer necessary to learn or memorise thanks to the software), one can understand that working solely on MTPE would affect the translator's perception of language (its fluency, its natural or unnatural wordings,

[14]The author thanks Milica Panić and TAUS for authorising the reproduction.

etc. as rejected first by Daems, De Clercq, et al. (2017) and later demonstrated by Toral 2019, when researching the phenomenon of "post-editese"). If the profile of the post-editor is hard to pin down and some translators still feel frustrated when asked to produce less-than-maximum-quality (Teixeira 2019, p. 216), the PE skill-set is yet another challenging question tackled by many scholars and researchers.

In chapter 3, we have presented the state-of-the-art about the PE skill-set. We agree with Kornacki (2018, p. 38) that PE is not just a technological/instrumental skill, neither is it a practical market-oriented skill, it is becoming, along and combined with transcreation and other neighbouring skills, an essential transdisciplinary element to Translation Competence (TC). Models for TC are thus becoming more diffuse, as they must now encompass varied subsets of skills. In particular, considering the PE competence models discussed in chapter 3, one could deduce that the activity of editing MT output relies on a competence formed by the intertwining of two groups of skills[15]:

1. one from the linguistic perspective, which would encompass all the traditional TC models' elements: thematic, linguistic, etc. that we can call Efficient Editing Skill (EES).
2. one from the purely technical perspective, which would encompass the already researched skills for PE regarding the instrumental abilities: use of shortcuts, knowledge of MT, knowledge of Regular Expressions (RegEx), editing mode, etc. that we shall name Instrumental Input Skill (IIS).

The EES are the combination of all the traditional translation competences with qualitivity. That is, if a synonym is a valid word in a PE project, and it requires no editing or less editing than using another synonym, the post-editor would pick the linguistic solution that is more efficient. Similarly, the IIS would be, thanks to future research, the proceduralisation and application of the more relevant input mode (a given shortcut, use of the keyboard or mouse, speech or touch, etc.) to the corresponding Post-editing Action (PEA). For example, if a deletion in the middle of a sentence, double click and key `SUPR`; but if an addition at the beginning of the sentence, rather key `Home` plus input by speech if the addition is long and understandable by the speech-recognition software of the post-editor (e.g. not an homonym).

The combination of EES with IIS would result in the ideal (maximised) level of qualitivity. In the savvy words of Engeström and Sannino (2010, p. 3):

[15] And always nuanced by the service offering: translation (e.g. possibly MT-aided TM translation, who knows) versus PE.

> Nobody knows exactly what needs to be learned. The design of the new activity and the acquisition of the knowledge and skills it requires are increasingly intertwined. In expansive learning activity, they merge.

For the purpose of this dissertation and with the aim of assuring feasibility and practicality, it is considered that participants to the experimental study in part III have homogeneous IIS. Hence, the study focuses largely on EES. Finally, in chapter 4, the increasing (necessity of the) presence of MT in translator training has been observed. We also discussed the need for formal and homogeneous programmes addressing PE transversely in regular translation courses and the popular impression that the university training does not serve the needs of the market (Kornacki 2018, p. 81). This perceived gap has been documented both at professional fora (e.g. European Language Industry Association, ELIA and Directorate-General for Translation, DGT) and academic networks (Optimale and EMT[16]) and it calls for further research (De Cespedes 2018, p. 18; Van Egdom, Verplaetse, et al. 2019, p. 28).

In addition, as observed in Paradowska (2015) and Celia Rico, Oliver, et al. (2020), MT is used by students of translation and we can no longer disregard the applications it has in professional translation nowadays. We would like to conclude that, whereas the following statement may hold true for some particular sectors, for some particular Language Service Companies (LSCs) or for some particular translators, the PE meta-techniques have not been explicitly defined yet.

> The differentiation between human and machine translation is further called into question in interactive, adaptive machine translation, [...] [a]gainst the backdrop of such technological developments, Castilho et al. have gone so far as to claim that the traditional separation of human and machine is no longer valid [...];

Including meta-techniques in PE in translation courses, we believe, would allow for proceduralisation at later stages, when a linguist would further practise editing of MT output (in postgraduate or Continuing Professional Development (CPD) courses, work placements or later in their career). As Mossop (2006b) said and Göpferich (2013) would further develop, procedures, eventually "become second-nature, but the point of studying [post-editing] is to formulate them". According to Absolon (2019, p. 20), it requires practice, in other words, time is of essence to let learned techniques turn into automatic actions. The speed at which they may become

[16]European Master's in Translation (EMT).

automatic varies, but the author argues that one can speed it up by applying the so-called "split principle" to training design and provision.

Furthermore, as it probably happens with any type of change, in any other sphere, each professional translator adapts differently to technological change, as indicated by E. M. Rogers (1962), in turn cited in Mihalache (2010, p. 166). Hence, there are "innovators and early adopters"; an "early majority" willing to become post-editors, as anticipated by Pym (2013); the "late majority"; and some "laggards". We hope that research in PE pedagogy will allow soon for the explicitation of PE strategies/procedures/techniques/etc.[17] and that both the early adopters and the laggards groups can follow their professional career and continuous development of competencies with a sound basis. To provide as much support as possible, the use of pedagogic presentations and detailed follow-up both for post-editors and clients as suggested in Nunziatini and Marg (2020) may be a beneficial solution.

One assumption from the literature review presented in this part is that "regular translation"[18] syllabi suffer from a slow development in the sense that the ones that are still considered "regular" translation courses, not only do not include any MTPE, but the use of CAT tools is also rarely promoted (Kornacki 2018, p. 154). Departing from this discussion, in part II four draft lists are defined to launch a large-scale survey-based study. The listed elements of the core questions will be regarding:

1. PE neighbouring tasks (to grasp better the multi- and transdisciplinarity of the post-editor profile as per chapters 1 and 2);
2. Hiring criteria for post-editors (to obtain insights from the industry regarding employability in this branch of the sector, also in line with chapter 2);
3. PE-related skills (as per chapter 3).
4. PE elements included in MTPE training courses (as per chapter 4).

It is hoped that the present research will be helpful in making informed choices in the translation classroom (also in CPD or in the industry), at the same time that the obstacles mentioned by Pokrivcakova (2019, p. 145) can be overcome: (1) lack of information and ICT[19] skills, (2) lack of experience, (3) lack of motivation, (4) struggle to integrate ICT with teacher's learning style, (5) feeling of being out of one's comfortable zone, (6 and 7) losing a dominant position or weakening control, and (8) losing students' respect.

[17]The theory of translation is not the main object of study to the present dissertation. The Handbook of Translation Studies and some entries in it, such as Gambier (2010), are recommended readings in this matter.

[18]That is translation without the help of Computer-Aided Translation (CAT) tools.

[19]Information and Communications Technology (ICT).

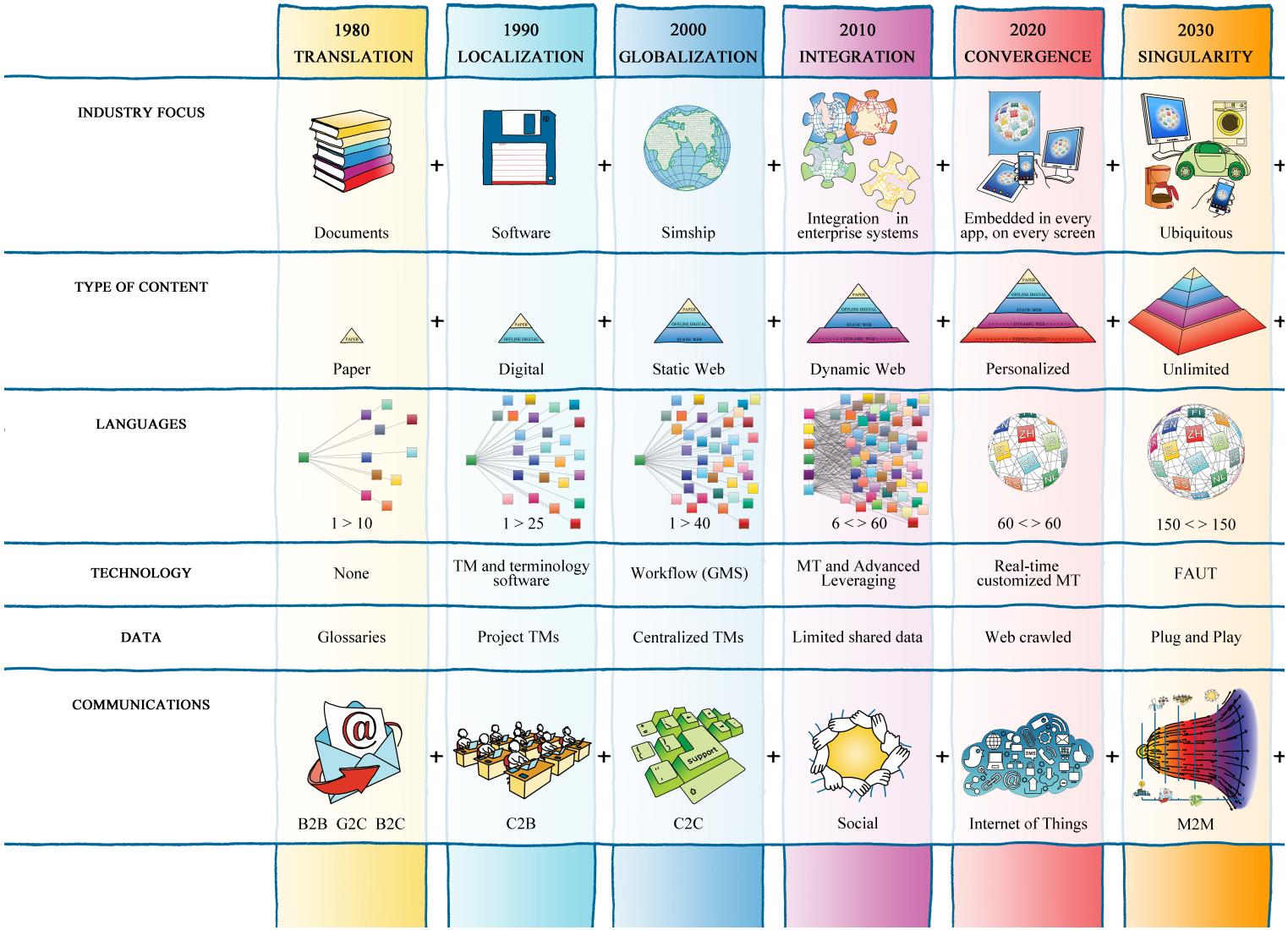

Figure 4.1: 2020 Convergence 2030 Singularity (Van der Meer 2013)

Part II

Survey-based Research

Many studies have resorted to survey research to confirm assumptions about the translation industry and the profiles of the professionals. For instance, in her doctoral dissertation, Blagodarna (2019, p. 67) departed from 124 submissions from active translators who work on Machine Translation Post-editing (MTPE) or related tasks either combined or not with other professional activities. The definition of the target population and the goal of the survey are essential criteria to put into perspective the results of any survey-based research.

For example, in their article «A survey of machine translation competences: Insights for translation technology educators and practitioners», Gaspari et al. (2015) addresses a wider target population: freelance translators, Language Service Companies (LSCs), translator trainers, and academics, and received 438 valid submissions, of which 285 answered positively at having some experience with Post-Editing (PE). Just another example is the doctoral dissertation of Saint-André (2015) whose survey has a much more restricted target population and goal. The audience are translators above the age of 18 who post-edit texts in the English-French language pair. And the goal was to identify the most useful skills for English-French post-editors, and, of those, to identify the ones that are less often acquired at University. Saint-André (ibid., p. 94) received 34 complete submissions.

Surveys are the most widely-spread method to quickly obtain up-to-date data, and the industry stakeholders and other organisations use it frequently (e.g. TAUS, SDL, Common Sense Advisory, and the Language Industry Survey[20]). Moreover, amongst researchers, consulting professionals to find out their opinions and carrying out situational analysis is also considered a valid methodology at a preliminary stage (Calvo 2011, p. 16-17; Kearns 2006, p. 286). As Mihalache (2010, p. 176) suggests, an online survey could bring insight into the training methods and strategies around translation technology; thus, we consider that survey-based research is the most appropriate method in our sequential exploratory design to shed some light on the current PE landscape within academia and within the industry.

Our goal, at this preliminary stage, is to obtain a general perception about what PE means or should entail according to three distinct stakeholders: LSCs and firms as professional entities, professional post-editors as individuals, and PE trainers[21]. Since the profiles of respondents can become

[20] Submitted regularly by the group formed by Elia, EMT, EUATC, FIT Europe, GALA and LIND.

[21] Despite the terminological choice defended in Massey et al. (2019, pp. 212–213), the terms “instructor” and “educator” are also used in this thesis to refer largely to any professional that would guide trainees in a split-training course, either in Continuing Professional Development (CPD) or higher-education programmes.

mixed in such large-scale surveys[22], one distinct questionnaire is designed for each audience.

In the following chapters, the initial survey-based study is presented. Basing ourselves on the literature review in part I, we devise the research questions to get insights from the current PE market and the post-editor profile. In particular, the RQs are introduced in chapter 5. In chapter 6, the methods applied to design and disseminate the questionnaires are presented, along with two other tools that bring further insight about the current PE training practices in European universities. These two tools concern the third audience (PE educators) and consist of an analysis of PE syllabi and interviews, respectively. The methodology to collect and analyse the data is described in the same chapter. Finally, the results are explored in chapter 7, with the discussion of their impact and implications. In the chapter Final thoughts of the present part, we draw the main conclusions that represent the basis for the experimental work carried out in part III.

[22] As seen before, some surveys have wider circles as recipients: professional translators, students, Machine Translation (MT) or Computer-Aided Translation (CAT) vendors, end-customers, etc.

Chapter 5

RESEARCH QUESTIONS

According to the Project Management Institute[1], the term "project stakeholder" refers to:

> an individual, group, or organisation, who may affect, be affected by, or perceive itself to be affected by a decision, activity, or outcome of a project.

ISO 21500:2012 uses a similar definition. In the PE market, logically, the main stakeholders are the vendors and buyers of the product or service. Hence, since the perspective of an individual freelance (or in-house) worker may be different from the perspective of a business or corporation, it was decided that two distinct questionnaires were needed to vary the wording according to the audience. Furthermore, the future post-editors and the MTPE trainers are also affected and can themselves affect how the PE activity will evolve and be perceived in the PE market. Thus, a third questionnaire would be directed at the training institutions. To tie together the data and provide arbitrary delimitation to the study, the European continent was primarily aimed.

The first major question that we wish to provide an answer to is the following: **RQ1** To what extent do the PE stakeholders (LSCs, individual linguists, and trainers) agree on the definition of PE as an activity (in its large sense) and of the post-editor profile?

In particular:

1. RQ1.1 LSCs' questionnaire: How is PE done and viewed within the industry, more specifically, by LSCs in Europe?
2. RQ1.2 Linguists' questionnaire: How is PE done and viewed within the industry, more specifically, by individual linguists, either freelance or in-house professionals with PE experience, especially in Europe?

[1] https://www.pmi.org/

3. RQ1.3 Trainers' questionnaire: How is PE taught and viewed by European postgraduate instructors?

Contrasting and comparing the results of the three questionnaires should allow us to answer our **RQ1**. Mainly, through testing the following hypotheses:

> Hypothesis 1
>
> H0 = There is no significant difference among the three groups regarding their views about PE-related tasks.
>
> H1 = There is a significant difference among the three groups regarding their views about PE neighbouring tasks.
>
> Hypothesis 2
>
> H0 = There is no significant difference among the three groups regarding their views about PE-related skills.
>
> H1 = There is a significant difference among the three groups regarding their views about PE-related skills.
>
> Hypothesis 3
>
> H0 = There is no significant difference among the three groups regarding their views about the design of PE training courses.
>
> H1 = There is a significant difference among the three groups regarding their views about the design of PE training courses.
>
> Hypothesis 4
>
> H0 = There is no significant difference regarding the views on hiring criteria for post-editors between linguists and LSCs.
>
> H1 = There is a significant difference regarding the views on hiring criteria for post-editors between linguists and LSCs.

The second major question we pose is: **RQ2** To what extent is the profile of an MT post-editor multi- and transdisciplinary? (Gambier 2006). The nuclear questions for each questionnaire stem from the main topics explored in part I and each one of them contributes to presume or theorise an answer for **RQ2**.

1. RQ2.1 What are the current PE practices in the industry?[2]
2. RQ2.2 What are the skills most valued by the PE stakeholders?[3]
3. RQ2.3 What do current MTPE training courses tackle according to each PE stakeholder?[4]

Four publications resulting from this survey-based research summarise the LSCs' view (Ginovart-Cid, Colominas, and Oliver 2020); compare the results from LSCs to those obtained from linguists (Ginovart-Cid 2020b); explore the MTPE trainers' view (Ginovart-Cid and Colominas 2020); and

[2]Core topic (PE-related tasks and hiring criteria) seen in chapter 2.

[3]Core topic (PE-related skills) discussed in chapter 3.

[4]Core topic (PE training elements) from chapter 4.

cross-examine the overall results for the three questionnaires (Ginovart-Cid and Oliver 2020).

With the aim of defining clear boundaries for the study, the population of PE trainers was limited to those with PE courses at the postgraduate level. This decision was made considering, first, that authors like Lara (2019) and O'Brien (2002) recommend that PE is introduced either at the second cycle (undergraduate) or, if possible, at the postgraduate level, and, second, that PE pedagogy being a young field, one can expect that the most advanced signs of progress could be found within this segment of academic settings.

Chapter 6

METHODOLOGY

The methodology adopted to answer the RQs posed in the previous chapter is of a mixed-methods nature, exploratory sequential (J. W. Creswell and J. D. Creswell 2017, p. 15), in particular: a **survey** followed by the analysis of a set of selected University **syllabi** and **interviews**. At this preliminary stage, we seek to explore the general trends and practices in the European language industry regarding the PE activity and how they meet the expectations set for a post-editor profile at University.

First, the survey-based research is designed for the three above-mentioned audiences, through the creation, dissemination, and data collection of three online questionnaires that are discussed in further detail in next section. Second, the questionnaire addressed at MTPE trainers is complemented with an analysis of the syllabi outlines and with one-on-one interviews with the educators, in sections 6.2 and 6.3, respectively.

As discussed above, the **populations** for our survey-based research are LSCs who sell and buy PE as a service, linguists who post-edit professionally, and trainers who teach PE at postgraduate programmes. While there is not a specific closed set of targeted LSCs and linguists (probabilistic and convenience dissemination are methods used for these two populations), the syllabi to be studied and, thus, the PE educators to be interviewed, are carefully selected thanks to the online resources, such as the list of members of the European Master's in Translation (EMT)[1]. To make sure non-EMT schools are not excluded of the study, other lists and databases[2] of TI schools are used to contact the faculties. To reduce the risk of failing to include a European faculty with a PE course, the websites of European faculties with Master or Postgraduate studies in Translation, Linguistics or related studies are carefully browsed, and the academic directors of the relevant degrees are contacted by e-mail to inquire about the existence of a course where MTPE is *practised* by the students. Even though the first contact to potential PE

[1] http://xl8.link/EMT2019-2024

[2] The author thanks Jeanette Steward from Translation Commons for easy access to the learning centre and resources.

educators referred to the PE content of the course in more general terms, further interaction with the contact person would confirm if there was any hands-on PE at all for the students.

Following J. W. Creswell and J. D. Creswell (2017, p. 149), the Internet is selected as **form of data collection** due to the ease of access to the identified populations. Even if the sample size cannot be accurately defined (ibid., p. 151) due to the lack of a collegiate body or official register of European LSCs or translators, a minimum of 50 participants may be considered sufficient at the light of the previous works commented at the beginning of part II. Indeed, statisticians have found that it takes a sample size of about 30 participants to fulfill the assumption that the sample represents a population (Salkind 2017, p. 387). In terms of creation, the **design** of the questions and answers and the interconnection between them (that is, the logical conditioning, e.g. hiding or displaying certain questions or redirecting to relevant parts of the survey according to previous answers) was performed in a spreadsheet. It can be consulted in text format in Ginovart-Cid (2020c)[3]. In the publication, the introductory paragraph for each questionnaire is also reproduced. In it, the respondent would find the informed consent statement, among other practical information about the survey-based research.

The selected **tool** to distribute the questionnaires and collect the submissions is Jotform[4]. Solutions like SurveyMonkey[5] or Google Forms[6] were considered, but they did not fulfill the requirements for the present study. In particular, some types of redirections in the conditional logic between questions and/or sections were not possible, and certain question types (e.g. matrix) were either not available or returned some display and technical bugs. Jotform, on the contrary, allowed for much more flexibility regarding conditional logic[7] and had successfully been used in the past by fellow-researchers (Vintar et al. 2019). The possibility of exporting the collected data to CSV and Excel formats[8] is also considered a positive aspect of the tool. It is thus decided that the use of Jotform could facilitate "data collection into organised spreadsheets for data analysis, reducing data entry errors and accelerating hypothesis testing" (J. W. Creswell and J. D. Creswell 2017, p. 153).

To ensure **construct validity and reliability**, the researcher contacted the Servei d'Estadística Aplicada (SEA) of the Universitat Autònoma de Barcelona to request a formal analysis of the design of the questionnaire

[3]This report functions as an appendix to the present dissertation.
[4]https://www.jotform.com
[5]https://www.surveymonkey.com
[6]https://www.google.com/forms/about
[7]https://www.jotform.com/features/conditional-logic/
[8]https://www.jotform.com/help/44-how-to-export-form-data-to-excel

to LSCs. The advice of the SEA would be applied to increment construct validity and reliability in the whole survey-based study.

Firstly, the content of the survey is formed mainly by closed-ended questions, with some open-ended ones at the end, or free text fields. Among the closed-ended questions, most are multiple-choice questions that allow for more than one answer (check-box), either with a limited number of two or three selections, or with no limit (this last case is represented by one of our four core topics: the PE training elements). Several questions are multiple-choice but with a single answer to be chosen (radio-buttons). Finally, three out of the four core topics (PE-related tasks, PE hiring criteria, and PE-related skills) are matrix-type questions (a multiple-choice question/answer formed by a set of columns and rows). The typology for each question and their description or limitation in terms of number of answers can be consulted in Ginovart-Cid (2020c).

Secondly, and in an effort to reduce the dropout rate, it was specified in bubbles that leaving an answer blank or clicking on the "N/A" option was equivalent, which would reduce the total number of clicks (Callegaro et al. 2015). Similarly, the fact that certain questions are mandatory and some are optional, along with the progress bar at the top, should help the participants foresee the time they invest in filling out the questionnaire and avoid dropout rate. Furthermore, when asking about the importance of a certain item in a Likert scale, an even number of values was used to avoid temptation from the respondents to pick too often the mean value, which would undermine the reliability of survey results as has been observed by Blagodarna (2019, p. 64) and suggested in the SEA report mentioned above. Finally, the phrasing of the questions and answers was proofread by a native professional linguist[9].

Finally, in terms of data collection and analysis, the feature pivot tables of Ms Excel is the main tool used. However, for the multiple-choice questions allowing for more than one answer, before creating the pivot table a plug-in named Kutools[10] helps at dividing the line breaks of a single cell into several cells in the spreadsheet. The statistical significance results explored in chapter 7 are calculated with the following formula (Blanco Claraco 2019):

$$\sigma = \sqrt{p_1(100 - p_1)/n_1 + p_2(100 - p_2)/n_2}$$

The following section (6.1.2) presents the data analysis methods and

[9]In spite of the linguistic revision, one poor choice of words remained in the questionnaires: "eventual" was used instead of the adjective "occasional". It is hoped, however, that next to "main task" and "secondary task", the respondents could deduce the right meaning for "eventual task".

[10]https://www.extendoffice.com/download/kutools-for-excel.html

the four draft lists that are designed after the review of literature in part I and used in the questionnaires. For further detail on the phrasing of peripheral questions and answers and logical conditioning of the questionnaires, we refer the reader to the above-cited report: «Report about a survey-based research on machine translation post-editing: common ground and gaps between LSCs, linguists and trainers» (Ginovart-Cid 2020c).

6.1. Questionnaires

6.1.1. Data Analysis

To test the hypotheses established in the previous chapter, the one-way ANOVA (also called "single-factor") is the experimental design to be applied. The treatment variable in this case is rather a "grouping factor" (Salkind 2017, p. 265), that is, the data is compared among the three aimed groups: LSCs, linguists, and trainers. The *p*-value for significance applied is the usual .05

Because we are examining differences between groups, the groups are not tested more than once, we are dealing with more than two groups, and we consider only one factor (the grouping factor mentioned above), the appropriate test is the simple analysis of variance (ibid., p. 264). In particular, the test statistic needed is the computation of the *F*-value (ibid., p. 266). Regarding the hiring criteria the applied data analysis tool is the *t*-test for independent samples, for only two groups are compared: LSCs and linguists (ibid., p. 264). Both tests are computed using the Analysis ToolPak[11].

6.1.2. Core topics

The core topics have been announced in chapter 5 when phrasing the RQs for this part of the thesis. They are the same throughout the three questionnaires, excluding the hiring criteria of the trainers' questionnaire, as explained below when introducing the sources for the design of table 6.2.

Division of labour is both a factor to help define a job description in a narrow manner as to be highly specialised and also a reason to become multi-talented and multi-skilled (Mossop 2006a, p. 791). Thus, it is interesting to identify which neighbouring tasks are more often combined with PE. In that sense, the works commented in part I were helpful in designing a list of 14 PE-related tasks in table 6.1. **Post-editing machine translation output**, T7, has been described in detail in Chapter 2 and is kept in the list to provide focus.

[11] http://xl8.link/Ms-toolpak

Table 6.1: List of 14 PE-related tasks

Code	Name of task
T1	Customisation/Tuning of MT engines
T2	Feedback collection on MT output quality for solution engineers
T3	Management of MTPE projects (e.g. outsourcing)
T4	Material preparation for MT engine training (building corpora, alignment, cleaning TM...)
T5	MT output quality evaluation (error categorisation...)
T6	PE guidelines design
T7	Post-editing machine translation output
T8	Pre-editing the source text
T9	Proofreading of post-edited output (monolingual)
T10	Quality control and text checking
T11	Revision of post-edited MT output (bilingual)
T12	Support users with CAT/MT tools
T13	Terminology extraction and TB management
T14	Tracking PE productivity

To include T1, **Customisation/Tuning of MT engines**, we considered the previous work of Van Ess-Dykema et al. (2010, p. 3) who state that a paralinguist should have specific knowledge of the performance of baseline and customised MT systems. It seems that other researchers have wondered if this activity is to be inscribed alongside translation and PE training modules, since it appeared twice in the POST-IT questionnaire (Tradumàtica Research Group n.d.)[12]. It appeared, first, when asking the instructors which restricted access resources they enable to their students, among the answers there was "Access to corpora to train MT systems". Second, it appeared when asking the instructors which topics they would like to cover in a course covering PE, as the answer "MT systems training" was offered. Finally, T1 is especially relevant if we consider that Sánchez-Gijón (2016, p. 156) explicitly includes the task of preparing MT systems in one of the two post-editor profiles (in the more broad, instrumental one), and so is advised by Vieira and Alonso (2020, p. 16). To give just one more example, this time from within the language industry, it is one of the main 5 tasks listed by Muzii (2016): "Engine preparation".

T2, **Feedback collection on MT output quality for solution engineers**, was particularly inspired by work coming from the industry. For in-

[12]The quotes on the phrasing of questions and answers of this survey are translated by the author of the present dissertation.

stance, Tinsley (2014), mentions the feedback and being capable of "bringing the translator into the loop" as one of the "four pillars of happiness". Tinsley (ibid.) also showed examples of direct feedback from end-users, such as talking to the translators/post-editors, but he also offers more structured ideas. Likewise, Muzii (2016) states that "[f]eedback forms should be prepared to allow post-editors to submit information that help improve the engine performances". Particularly structured is the template proposed by Guerrero (2017). The form she uses at CPSL with post-editors allows the LSC to collect useful feedback for the further improvement of their MT systems and MTPE workflow. The importance of this activity has also been acknowledged by researchers like Oliver (2016, p. 207) and Rico Pérez and Enrique Torrejón (2012, p. 169); for the latter within the "Global task-related processes" discussed in chapter 2.

Considering the largest profile suggested by Sánchez-Gijón (2016), "Post-editor / Provider of translation services", the author underlines the importance of the soft skills needed by any translation Project Manager (PM). This is why T3, **Management of MTPE projects (e.g. outsourcing)**, was included in the draft list. Moreover, it is also considered by other authors, such as Cid-Leal et al. (2019) and Marín Juarros (2017). The latter stresses the need to handle the client's expectations through the strategic competence. In addition, it can be considered a possible topic to cover in a higher-education translation course: under the group "Professional aspects" in the already mentioned POST-IT questionnaire by Tradumàtica Research Group.

T4, **Material preparation for MT engine training (building corpora, alignment, cleaning TM...)**, stems from tasks mentioned mainly by Sánchez-Gijón (2014) "Corpus management and cleanup". Again, the authors of the questionnaire (Tradumàtica Research Group n.d.) provide a place for this task under the group "Preparing resources for MT/PE", where the surveyed participants could choose elements such as "Compilation of a corpus". Similarly, sentence alignment and analysis of large parallel corpora have also been highlighted as challenging tasks within the industry (e.g. Muzii 2016).

As established in section 2.1, PE is at the intersection of at least two fields: MT and Translation Studies (TS). As such, T5, **MT output quality evaluation**, is potentially a task that could be carried out by the same person who edits the MT output. In Geoffrey S Koby (2001, p. 21), it is noted that:

> [t]he quality of the post-edited output depends very much on the quality of the raw machine translation output, and the only person qualified to judge that quality is a trained, experienced professional translator.

Within the industry, MT output evaluation is also listed by Muzii (2016)

as one of the main tasks to be done when setting up the MTPE workflow. In that sense, several authors (for instance Guerrero and Soloviev 2019; Marheinecke 2016a) have used scales or annotation trees such as the Multidimensional Quality Metrics (MQM) (Lommel and Burchardt 2015), introduced in section 2.1.2.

T6, **PE guidelines design**, was highly inspired by Hu and Cadwell (2016). If we assume that the post-editor is a hybrid profile of a professional who does not post-edit MT output the whole working day but who also translates, revises, manages translation projects, etc., one ancillary or neighbouring task could be the creation of PE guidelines that are tailored to a client or to a specific assignment (see section 2.2.2), or evaluating which guidelines are relevant to be activated or not in a certain scenario. Otherwise, facing an absence of explicit instructions, a post-editor could deduce his/her own implicit PE guideline(s). The inclusion of T6 in the list is in line with the questions raised about the use, adaptation or writing of PE guidelines we commented in section 3.2 (Guerberof Arenas, Depraetere, et al. 2012).

T8, **Pre-editing the source text**, stems from research on Controlled Language (CL). As commented in section 2.1.1, authors like O'Brien (2002) researched the so-called "translatability indicators" that predict if a ST can be more or less easily translated by an MT system (Koponen 2016a, p. 28) and, therefore, to what extent it could benefit from pre-editing[13]. Recently, the research community has expressed some doubt about the extent of the usefulness of CL in certain scenarios related to Neural Machine Translation (NMT) and PE. For example, Nitzke, Hansen-Schirra, et al. (2019), citing the work of Marzouk and Hansen-Schirra (2019), claim that "the use of controlled language has no influence on the quality of the NMT output". For these reasons, it can be interesting to gather the current opinions of the three stakeholders.

T9, **Proofreading of post-edited output (monolingual)**, T10, **Quality control and text checking**, and T11, **Revision of post-edited MT output (bilingual)**, were a must. They are intrinsically related to PE, as has been discussed in chapter 2 and also at the beginning of the present section when commenting publications such as Krings (2001), Mossop (2006b), or Mossop (2019). We do not use the terms "revision", "quality control",

[13]It is of particular interest the concept of "negative translatability indicators", researched by O'Brien (2005, p. 38), Bernth and Gdaniec (2001), and Yamagata (2013). Most of the research involving Negative Translatability Indicators (NTI) has focused on English; e.g. a tool called "Translatability Checker" that calculates the translatability index of an English ST is described in Underwood and Jongejan (2001). Other languages, such as French or German, have been scarcely researched in terms of CL and pre-editing (e.g. Winkler et al. 2014).

and "proofreading" as synonymous[14]. Quality control can also be referred as "text checking", which implies nowadays making use of Quality Assurance (QA) tools, often integrated in the CAT tool. Revision is defined as contrasting the Target Text (TT) to the Source Text (ST), whereas proofreading consists in reading the TT only, no transfer (accuracy) errors can possibly be corrected. Certainly, Translation Quality Assessment (TQA), as discussed in section 2.1.2, is also regarded as a topic relevant for PE courses by Tradumàtica Research Group (n.d.), under the group "Aspects about target language": "QA in translation and PE".

On her preliminary findings, Gene (2019) claims that the "CAT tools interface is more challenging than the post-editing process itself". Also, the personal experience of the author of the present dissertation showed that a big part of the working day of a PM, at least in a start-up, was spent helping colleagues with CAT tools or solving issues for oneself. For this reason, it is decided to include T12, **Support users with CAT/MT tools**, in our list. There is indeed evidence of a multiplication of translation technology tools[15], this is why the improvement of their features (Bundgaard and Christensen 2019; Moorkens and O'Brien 2017), along with the mastery of different features of a number of tools, seems a good asset to suggest for the profile of the professional post-editor.

T13, **Terminology extraction and TB management**, is an ancillary task to many translators and revisers. Indeed, Sánchez-Gijón (2014) and Tradumàtica Research Group (n.d.) cited it under "Preparing resources for MT/PE" the possibility: "Creating glossaries/lists of forbidden words".

As discussed in chapter 2, specifically in section 2.1.1, the introduction of MT and PE is expected to bring improved productivity (Blatt et al. 1985), which has as underlying cause a desire to reduce costs according to Krings (2001, p. 21), citing Vasconcellos and Bostad (1992) among others. For this reason, T14, **Tracking PE productivity**, is included in the list. New tools do not cease to appear in the market to provide a precise follow-up of the cost reduction and productivity gains (e.g. DQF tool by TAUS presented by Attila Görög 2014), or the Qualitivity plug-in to be added to SDL Trados Studio). Other CAT tools, such as Memsource[16] or memoQ[17], had already integrated the editing distance as a reference in a post-analysis before. Hence, either when requested by their employer or customer, or be-

[14] As noted by Mossop (2019, p. 20) these terms are often used interchangeably by translators to refer to the activity of copyediting, to monolingual re-reading, or to revision.

[15] The multiplication of tools is noticed with a glance to the European Association for Machine Translation (EAMT) compendium (J. Hutchins et al. 2009) or the Nimdzi Language Technology Atlas mentioned in the Introduction.

[16] https://www.memsource.com/

[17] https://www.memoq.com/

cause they are willing to keep track themselves, professional post-editors may spend some time setting up such tools and analysing the results of the subsequent reports after they complete a PE project.

All things considered, some tasks in table 6.1 belong to the same phase in the PE process. For instance, T1, T4, T8, T13 all belong to the preparatory phase (Martín-Mor et al. 2016). Not only the tasks are interrelated, but they are necessary related to recruitment criteria (6.2), to skills (6.3), and to PE training elements (6.4), as we attempt to show in table 6.5.

To conclude the discussion about PE-related tasks, it must be highlighted that the publication of *ISO 18587:2017* served to stabilise the role of the professionals within the industry. In particular, it insists on the fact that a post-editor should have a deeper skill-set in translation technologies (p.9). The listed competences for PE in the cited standard have been reviewed in chapter 3 in detail, but they are considered again in the following paragraphs insofar they impact the decision factors by employers who would like to hire a post-editor.

In the next section, we draw up a list of factors a company could consider during the selection process of a linguist whose one of the main activities would be to post-edit MT output as a service. Recruitment criteria for post-editors are mentioned as a challenge in the translation industry by LSCs (for example, see the webinar by Intertranslations Gene 2019). However, one limitation of the suggested list (see table 6.2) is the lack of specialisation in the Human Resources (HR) discipline. Hence, the potential knowledge gaps concerning specific methods in the HR field would explain any bias in our list of recruitment or selection criteria, which are solely based on the translation industry research discussed in Part I.

Table 6.2: List of 17 hiring criteria for post-editors

Code	**Criterion**
C1	Capacity to post-edit into both directions
C2	CAT tool(s) knowledge
C3	Certification in PE by a professional association (ProZ, TAUS, etc.)
C4	Experience in project management
C5	MT system knowledge
C6	Pre-editing or CL skills
C7	Previous experience in post-editing MT output
C8	Productivity (processing speed)
C9	Quality assurance (QA) checking skills
C10	Revision & proofreading skills
Continued on next page	

Table 6.2 – continued from previous page

Code	Criterion
C11	Skills using automatic speech recognition (ASR) or touch-screen technology
C12	Specific locale (variant, sublanguage)
C13	Subject field knowledge or specialisation
C14	Technical skills: macros, xliff, tmx, Java, RegEx...
C15	Terminology management & information mining skills
C16	Translation Quality Assessment (TQA) skills (scores, metrics, evaluation, etc.)
C17	University degree in Translation or related studies

C1, **Capacity to post-edit into both directions**, means a linguist in charge of a PE project should be able to correct and edit output not only in their native language but also for languages B or C. This criterion stems from the questioning of the mother tongue principle (e.g. Newmark 1988). Indeed, bidirectionality in professional translation has become more accepted (e.g. Sánchez-Gijón and Torres Hostench 2014; Emma Wagner et al. 2014; Kornacki 2018, p. 141), so it is only natural to wonder about bidirectionality in PE too.

C2, **CAT tool(s) knowledge**, is especially related to T12. Often, translators show in their CV or other public profiles (TranslatorsCafe[18], ProZ, etc.) which tools they can use. It is also customary for LSCs to ask a translator to use their chosen tool.

Since C17 (holding a **University degree in Translation or related studies**) has become a frequent prerequisite to select a translator according to *ISO 17100:2015*, it seemed interesting to find out if the PE job also comes with needs or requirements regarding an extra certification[19], this is the reason why C3, **Certification in PE by a professional association (ProZ, TAUS, etc.)**, is included. Likewise, work by Absolon (2018) and Van Egdom, Vieira, et al. (2018) have been a contribution towards designing a PE test, which motivates the presence of C3 in the list.

C4, **Experience in project management**, is particularly related to T3 and assumes that there are managing tasks that are handled by the post-editor, according to the second profile proposed by Sánchez-Gijón (2016): "Post-editor / Provider of translation services".

C5, **MT system knowledge**, is lightly related to T1 and it assumes that, to a greater (actual engine training) or lesser (being aware of the history

[18] https://www.translatorscafe.com/cafe/

[19] In *ISO 18587:2017*, section 5.2 Qualifications, it is stated that the LSC should make sure the post-editor fulfills one of three criteria (the same as for translators).

and different technologies in MT) extent, the recruiters would like a candidate post-editor to have previous knowledge of MT systems. The degree is deliberately not precised[20].

C6, **Pre-editing or CL skills**, is mainly related to T8. It is included to see if, and to what extent, pre-editing and CLs are still considered nowadays useful with a view to improve the MT output and reduce the PE effort. Translators in Cadwell et al. (2018, p. 317) valued positively improvements in the ST, but we have already stated for T8 that NMT has thrown some doubts on the usefulness of applying CL on the ST for the purpose of reducing PE effort.

About C7, **Previous experience in post-editing MT output**, it must be highlighted how experience is frequently a factor for selecting one translator/post-editor or another. For instance, experience seems to have more importance in *ISO 17100:2015* than in *ISO 18587:2017*. Thus, it entered the list for the survey.

C8, **Productivity (processing speed)**, deliberately in the misused sense discussed in section 2.1.3, is related to T14. For instance, De Almeida and O'Brien (2010, p. 2) said that a good post-editor must be able to "carry out the post-editing task with reasonable speed, so as to meet the expectations of daily productivity for this type of activity (approximately 5,000 words post-edited per day, on average)".

Regarding C9, **QA checking skills**, it is similar to T10, and it assumes that, if phases are cut out of the translation project (as discussed above, it would replace TE in Translation-Editing-Proofreading (TEP), and potentially there is not even a revision or proofreading step), a QA step before delivery will probably become crucial. Even when revision and proofreading phases are kept, due to the fact that more errors can go unseen with NMT output and with the bigger volumes of fragmented content that we are handling today, QA presents itself as the next star of the show. A good management of the tools and resources to perform the QA should certainly fall within the scope of the post-editor.

C10, **Revision and proofreading skills** , is directly related to T9 and T11, without entering into so much detail for the HR. It also summarises the view expressed by Kornacki (2018, p. 125) that "[c]ertain projects may require translators to spend as much (or more) time proofreading automated translation than translating", where "automated" includes not only MT but also Translation Memory (TM).

As for C11, **Skills using ASR or touch-screen technology**, authors like Teixeira and Moorkens (2017) have started researching new input methods

[20]Some tasks, criteria, skills, or elements are deliberately vague to allow the contrast between topics in the large sense amongst the three stakeholders.

for translating and PE. Research in the field is still ongoing, but, if we consider the pace of the progress in translation technology, we can certainly be curious about how such new features are (or not yet) implemented within the industry, and this is why it was incorporated in the list. Mossop (2006a, p. 791) also mentions Automatic Speech Recognition (ASR) as "worth monitoring", and other authors have more recently explored its utility, especially for the revision task (Ciobanu et al. 2019).

Localizing for a specific region is often a requirement of a translation project. This is why C12, **Specific locale (variant, sublanguage)**, inspired by Mossop (2006b, p. 6), is included in our list. Within the industry and specifically for PE, it is also mentioned as a requirement by Muzii (2016).

C13, **Subject field knowledge or specialisation**, comes directly from the Process of Acquisition of Translation Competence and Evaluation (PACTE) model; as cited in Rico Pérez and Enrique Torrejón (2012) and reviewed in chapter 3. Furthermore, as discussed at the beginning of section 2.3, experts have been considered as possible post-editors, instead of professional linguists.

While macros themselves may not be as frequently used as when PE was done in text-processing software, a basic programming knowledge (O'Brien 2002) may still be needed to use Regular Expressions (RegEx) efficiently, or at least to understand how filtering and other features in CAT tools work to make the work of post-editors more efficient. This is the motivation for including C14, **Technical skills: macros, xliff, tmx, Java, RegEx...**, in the list.

C15, **Terminology management and information mining skills**, is partly related to T13, therefore inspired from the cited work by Sánchez-Gijón (2014) and Tradumàtica Research Group (n.d.), but also from three well-known Translation Competence (TC) models highlighting the importance of the information mining competence (discussed in chapter 3).

Metrics are mentioned as a challenge for PE within the industry (for instance, in the already cited webinar by Intertranslations, conducted by Gene). Moreover, TQA is a field on its own (Moorkens, Castilho, et al. 2018; Van Egdom, Verplaetse, et al. 2019, p. 27), discussed in section 2.1.2. This field holds a strong relation to translation quality and is therefore of potential interest to employers when hiring a post-editor. For this reason we include C16, **Translation Quality Assessment TQA skills (scores, metrics, evaluation, etc.)**[21], in the list of hiring criteria.

Finally, as anticipated in the discussion about C3, and considering how industry standards such as *ISO 17100:2015 ISO 18587:2017* point to higher degrees as a condition to ensure translation quality when hiring an em-

[21] It deliberately mixes automatic metrics with human evaluation and assessment.

ployee, C17, **University degree in Translation or related studies**, is added to the possible hiring criteria and closes our list.

The inclusion of the four core topics along with contextual questions resulted in long questionnaires (from 15 up to 25 minutes, as mentioned before). With the aim of shortening the time that the group of educators would invest in filling out the online survey and allow for the interview, several questions that had been asked to linguists and LSCs were absent in the survey sent to educators. This is the case for one of the four core topics: the hiring criteria.

Besides the actual factors for hiring, a professional in the translation industry could see one skill more useful than another for the PE activity, even when it is not easily measured or accounted for during the selection process. Hence, we include as core topic the PE skill-set. This topic has been discussed in detail in chapter 3.

Table 6.3: List of 11 PE-related skills

Code	Name of skill
S1	Capacity to decide when to edit or discard (translating from scratch) an MT result
S2	Capacity to post-edit according to PE guidelines
S3	Capacity to post-edit up to human quality (full PE)
S4	Capacity to post-edit to a good enough quality (light PE)
S5	Capacity to pre-edit a source language according to a controlled language
S6	Capacity to train and tune an MT engine
S7	Capacity to identify MT output errors
S8	Capacity to apply the right correction strategy
S9	Capacity to advise when MTPE is appropriate for a text or project
S10	Capacity to provide feedback for the MT solution engineers
S11	Capacity to learn about new technologies

S1, **Capacity to decide when to edit or discard (translating from scratch) an MT result**, is a skill that results from the reading of Mesa-Lao (2013)'s work on the 5-10 seconds rule to evaluate an MT output. As we have commented in chapter 3 while citing the work by Doherty and Gaspari (2013), not only is it important to make the right decision, but for a good post-editor this good decision must be made in a timely manner (see section 2.1.1). Teixeira (2019, p. 215) notes that "[w]ith both levels of post-editing, before starting to make changes you should decide whether it will require

more effort to reuse the MT output or to delete the suggestion and translate the entire segment from scratch. As a rule of thumb, this decision should not take longer than two seconds [... t]he ability to quickly assess the quality of the output is an important skill to be acquired by post-editors" (p.221). It corresponds perfectly well with CC7, "Be able to quickly decide to keep, modify ore replace a textual element"[22] in Saint-André (2015).

It has been commented in section 2.2 how the brief must be guides to action. In that sense, S2, **Capacity to post-edit according to PE guidelines**, is included in the list insofar PE guidelines are one piece of the MTPE assignment. PE rules and instructions are mentioned as a challenge within the translation industry by Gene (2019) and section 2.2.2 has been dedicated to this topic, which, of course, found a place within the list in table 6.3. It also matches the AP3 "Be able to respect a commission that changes from one project to the other"[23] in Saint-André (2015).

About S3, **Capacity to post-edit up to human quality (full PE)**, it can be said the same that we commented for T7 (table 6.1). It constitutes the very focus of the present research, and it is kept in the list to give focus. Further, it is also included as the other extreme of the dichotomy of PE levels, that is, to see the relative importance it has compared to S4.

A brief overview of the variety of error typologies and the corresponding correction techniques or edit operations, also known as Post-editing Action (PEA), has been provided in section 2.1. For instance, Guerberof Arenas, Depraetere, et al. (2012)) insist that the results of such research must be put into perspective according to the concept of "fitness for purpose", which is embodied in S4, **Capacity to post-edit to a good enough quality (light PE)**. The two PE levels have been explored in section 2.2.1.

S5, **Capacity to pre-edit a source language according to a controlled language**, just like T8 (table 6.1), is inspired by the close relationship between the linguistic features of a ST and the MT output quality (as argued in section 2.1.1), which in turn has a direct impact on the work of the post-editor. Authors like Bernth and Gdaniec (2001) and O'Brien (2006b) observe the general rules to improve the chances at having an MT output suitable for PE, for instance, reducing length and ambiguity. Nonetheless, it must be noted that most of them focus on the English language, as commented before. Similarly, Muzii (2016) lists the preparation of the ST with plain and consistent language as one of the four recommended steps for any MTPE project.

S6, **Capacity to train and tune an MT engine**, is T1's counterpart (table 6.1) in the skills list, and represents the most technical side of the

[22]Translated from French by the author of this dissertation.

[23]Translated from French by the author of the present dissertation.

post-editor profile which has been recognised as a possible new activity for the linguist, for instance in the round table and presentations done on 3 February 2020 in Celia Rico, Oliver, et al. (2020).

Researchers like De Almeida and O'Brien (2010, p. 2) and Guerberof Arenas, Depraetere, et al. (2012) have worked on S7, **Capacity to identify MT output errors**, which seems a *sine qua non* condition to become a professional post-editor. Likewise, Kornacki (2018, p. 121) observes[24], that the match of a TM (and therefore even more an MT output) "may help the translator or may hamper him/her instead". S7 is also in line with CC5 and CC6 commented in section 3.2. Any errors not identified, or even if spotted but not correctly edited, will be considered the translator's errors by the client, as it is further argued in the following paragraph.

S8, **Capacity to apply the right correction strategy**, stems from previous work by De Almeida and O'Brien (2010, p. 2) and Marín Juarros (2017), and it represents the ability of refraining from over-editing (preferential unnecessary changes are made to the MT output), but obviously also avoiding both under-editing (necessary changes are lacking) and, more importantly, pseudo-editing (new errors that were not in the MT output are introduced by the post-editor). Moreover, it also represents the only "strategies" or "techniques" that have been identified so far in PE, either mechanical (additions, deletions, modifications, and shifting) or linguistically-motivated, as discussed in section 2.1.3).

S9, **Capacity to advise when MTPE is appropriate for a text or project**, represents the ability to assess the usefulness of a certain MT output (seed translation) for PE and is, therefore, closely related to the Risk Assessment and the Service Competences considered in section 3.2. Furthermore, considering the second profile of a post-editor by Sánchez-Gijón (2016), we may see S9 as related to T3. According to Van Ess-Dykema (2011), a paralinguist should know the elements that are strong predictors of PE suitability. For example, Moorkens (2017, p. 470), citing A. Way (2013, p. 2), highlights the importance of a post-editor correlating his/her effort in editing to the content lifespan or "perishability of content". The same skill is seen as bringing more agency to the translator/post-editor in the MT age by Vieira and Alonso (2020, p. 16-17). It is also referred to by Muzii (2016) under the section "Content evaluation".

S10, **Capacity to provide feedback for the MT solution engineers**, is the skill counterpart for T2 (see table 6.1). As such, it is inspired by publications like the paper by Marheinecke (2016a, p. 74), who concludes that "qualified expert feedback that serves as a basis for further optimisation" is needed by MT researchers. In regards of this skill Marheinecke (2016b, p. 8)

[24] Himself citing Doherty, 2016.

comments that a post-editor should "learn how to give valuable feedback".

Finally, S11, **Capacity to learn about new technologies**, is inspired from research by Pym (2013, p. 494), where the ability of learning to learn is linked to the attitudinal competence. In the same vein, Marín Juarros (2017, p. 35) and Absolon (2019, p. 21) also apply these requirements to the post-editor profile; in terms of the latter author: "characteristics such as diligence, perseverance, responsibility, accuracy, or resilience are important to us". In her work, A.-M. Robert (2013) underpins the importance for a post-editor of being "agile d'esprit" as she puts it in French.

Table 6.4: List of 15 PE training elements

Code	Name of element
E1	Description of content profiles and text types according to MT system
E2	Integration between CAT tool and MT system
E3	MT evaluation: automatic & reference-based (metrics)
E4	MT evaluation: human (scoring, ranking, error categorisation)
E5	MT systems: rule-based, example-based, statistical, hybrid, or neural
E6	PE attitude: decide when to discard and translate from scratch
E7	PE guidelines: exhaustive list and examples in the relevant language pair
E8	PE levels: light and full post-editing
E9	Practical PE exercises in the relevant language pair
E10	PE risks: under-editing, over-editing & pseudo-editing
E11	PE techniques and strategies (shift, replacement, addition, deletion)
E12	PE technology: PE tool, ASR, touch-screen, etc.
E13	Pre-editing and controlled languages
E14	Productivity tracking tools
E15	Quality estimation (predictive, without post-edited gold reference)

The training element E1, **Description of content profiles and text types according to MT system**, comes from the wide-spread knowledge in the industry that some type of content (technical, repetitive) may be more apt to MT than others (marketing, for instance) (Stefaniak 2020, p. 263). It seems interesting to find out if such theories are contemplated and justified

either at LSCs or at universities.

E2, **Integration between CAT tool and MT system**, is something that not only the industry but also academia has been struggling about for some time, as discussed in chapter 1. Students often learned about CAT tools in a course and MT and PE in another one, without combining both, and LSCs have been exporting `TMX` and `xliff` files to translate them in Rule-based Machine Translation (RBMT) and Statistical Machine Translation (SMT) for decades. Hence, one may wonder if the "seamless" integration (nowadays possible) is taught at all to novice post-editors.

Regarding E3, **MT evaluation: automatic & reference-based (metrics)**, we have considered it is something becoming more and more present in PE courses at universities and sometimes argued by scholars as one of the main task of digital linguists, thus, post-editors, as we commented in section 4.2.

Similarly, E4, **MT evaluation: human (scoring, ranking, error categorisation)**, is included in the list as an essential task for a post-editor after numerous publications and authors have argued it would become a primary activity for digital linguists. The topic has been discussed in sections 1.3 and 2.1.2.

The reason why E5, **MT systems: rule-based, example-based, statistical, hybrid, or neural**, has been added to the list of 16 topics in MTPE training courses is the same why T1, C5, and S6 were considered relevant. It should help us determine to which extent the "hybridness" of the post-editor weights towards the technical side of its profile.

About E6, **PE attitude: decide when to discard and translate from scratch**, we shall remind how authors such as Doherty and Gaspari (2013) noted that it is important for a post-editor to be able to make quick decisions as to what are errors and what are not, and how to correct them. It is the counterpart of S1, seen in table 6.3.

Regarding E7, **PE guidelines: exhaustive list and examples in the relevant language pair**, it is an element related vaguely to T6 (table 6.1), and mostly to S2 (table 6.3). If we expect a professional post-editor to be able to make quick and right decisions at the same time a set of instructions are respected, it seems only appropriate that the student or novice employee is familiarised with such constraints and how to find the right balance. Or otherwise that s/he is able to note the non-feasibility of an MTPE project or the inconsistencies between the real *Skopos* and the explicit guidelines or instructions.

E8, **PE levels: light and full post-editing**, is the counterpart for S3 and S4 in the skills question (table 6.3).

E9, **Practical PE exercises in the relevant language pair**, is the result of some anecdotal observations by the author of the present dissertation that

most of the PE courses available online for translators did not include much hands-on practice on PE. Moreover, most publications analyse the typical MT errors and PE strategies for language pairs with English as source or target language (Bié et al. 2020, p. 477).

The elements E10, **PE risks: under-editing, over-editing & pseudo-editing**, and E11, **PE techniques and strategies (shift, replacement, addition, deletion)**, have been introduced in section 2.1.3 of part I and they are arguably essential concepts in the PE field still being researched.

E12, **PE technology: PE tool, ASR, touch-screen, etc.**, like C11 (table 6.2), is included because new input technologies for translation-related activities seem worth monitoring to us and to other researchers, such as Mossop (2006a, p. 791).

E13, **Pre-editing and controlled languages**, is the counterpart for T8 (table 6.1), C6 (table 6.2) and S5 (table 6.3).

E14, **Productivity tracking tools**, is the counterpart for T14 and C8 (tables 6.1 and 6.2, respectively), since tight deadlines are more and more often a reality within the PE market as acknowledged by Bowker (2016, p. 23), who was, in turn, quoting other authors.

The last element that we suggest, E15[25], **Quality estimation (predictive, without post-edited gold reference)**, refers to the confidence estimates that some CAT tools may display about MT suggestions, such as PET (W. Aziz et al. 2012) and CasMaCat (Alabau et al. 2014) as commented by Teixeira (2014, p. 45), and Memsource (Tamchyna 2020). Indeed, it has been observed how hard it is to estimate MT output quality *a priori*, especially in real scenarios (Muzii 2016), and how it is still "not [...] reliable enough for some domains/language combinations", in words of Tamchyna (2020, p. 287). The explanations and references provided in the previous paragraphs are summarised in a table that tentatively reflects the interrelations between items (table 6.5), without following any particular order.

Table 6.5: Connections between tasks, criteria, skills, and elements

Concept or notion from Part I	**Related Task (T), Criterion (C), Skill (S) or Training element (E)**
Assertive attitude + Learn to learn	S11, E6, E12
Traditional translation/translator competence model	T9, T11, C1, C3, C10, C12, C13, C17
Proofreading and quality checking (more or less automated)	T10, T4, T13, C9, C14
Continued on next page	

[25] A free-text field was added, E16, to allow the respondent to share any topic covered in his/her MTPE course that was not present in the list.

Table 6.5 – continued from previous page

Concept or notion from Part I	**Related Task (T), Criterion (C), Skill (S) or Training element (E)**
Corpus and information mining	T4, C15
Terminology management	T13, C15
Advanced knowledge in CAT tools	T12, C2, E2
Quality estimation (predictive, confidence scores)	T3, T14, E15
Quality evaluation	T5, C16, E4
MT system knowledge (deep and technical, or user perspective)	T1, C5, S6, E5
Abstraction for PE feedback	T2, C16, S7, S10
Efficient Editing Skills	S1, S2, S7, S8, E6, T14, C8
Input Instrumental Skills	C8, C11, E2, E12, E14, T14, C8
PE techniques, strategies and procedures	S7, S8, E10, E11
CL and pre-editing	T8, C6, S5, E13
Professionalism, PE service and standards	T3, T6, T14, S2, S3, S9, E1, E7, E8, C4, C8

As commented at the end of part I, the intertwining of Efficient Editing Skill (EES) and Instrumental Input Skill (IIS), along with the capacity to focus and refocus fast, are hypothesised to be the core of qualitivity. However, the limited time and resources, along with the fact that input modes are presently being developed and researched, lead the present dissertation to focus on EES only.

The following three sections present the methods used to design each questionnaire: section 6.1.3 presents the questionnaire to LSCs; section 6.1.4 presents the questionnaire to post-editors; and section 6.1.5 describes the methodology for the questionnaire to PE trainers.

6.1.3. Questionnaire to LSCs

The title of the questionnaire to LSCs is "Machine Translation and Post-Editing in the Industry"[26] and it is addressed to LSCs or companies with a translation department that sell (or just do) MTPE.

The SEA of the Universitat Autònoma de Barcelona produced an evaluative report. The advice produced by the SEA report was followed, thus improving the design of the questionnaire. In particular, the longest answers

[26] https://form.jotformeu.com/82863740587368, hereinafter "Questionnaire to LSCs".

were randomised to avoid influencing the responses; a progress bar indicator was added at the top of the questionnaire; and some questions were rephrased or their matrix or scale was adapted according to the suggestions in the report. A pilot survey with 15 participants was completed to adjust any remaining wording before launching the survey. The results of the pilot and the report by the SEA also helped delete unclear items (J. W. Creswell and J. D. Creswell 2017, p 153).

Both probabilistic and convenience (i.e. non-probability) sampling methods were used to reach maximum representation. The convenience sampling is performed with the snowball technique, whereby initially sampled respondents refer the survey to other persons that match the required characteristics. In that sense, the questionnaire was disseminated in early December 2018, by e-mail, to over 200 European groups, associations, and firms. In terms of probabilistic dissemination, it was published on leading social media platforms, both in the user's feed of the author and in specific closed or open groups.

The questionnaire to LSCs contains 69 questions which can be consulted in pages 3 to 17 of Ginovart-Cid (2020c). It can be filled out in approximately 15 to 20 minutes, and it is structured in 7 sections, as summarised below.

1. Basic Information. Here we collect demographic data of the respondent.
2. MTPE Projects. Here we define the concept of 'post-editing' according to *ISO 18587:2017*. We also filter here for the right audience. More specifically, if the respondent answers 'No' to the question Does your company handle translation projects with post-editing of machine translation?, the survey comes to an end and s/he is invited to submit.
3. Client-TSP Agreement. This section is only visible to LSCs, not to firms with a translation department.
4. Current Trends in MTPE Projects. This section is subdivided as follows:
    - Your MTPE Background. It inquires about the LSC's background in MTPE; e.g. language pairs, the decision of using MT, PE and revision levels, and controlled authoring.
    - The MT Post-editors' Profile. It covers the three core topics that are commented in the present article, namely: the tasks performed by in-house and freelance post-editors, the recruitment criteria of the firm or LSCs, and the rating of skills according to the respondents' beliefs.
    - The MT Post-Editing Project. In this section we inquire about workflows and procedures of how MTPE projects are handled;

e.g. the PE assignment (if different of the translation assignment), PE guidelines, PE feedback, and TQA.

5. Training in MTPE. Here we inquire if the company provides training on MT, either to in-house staff or to freelance collaborators, and how it is provided. It also researches one of the key topics presented in this article: the MTPE training contents.
6. Technology & Tools. Here we collect information on the CAT tools and MT systems used. It is subdivided in two parts ('TM/MT not integrated' and 'Integrated TM/MT').
7. Feelings & Thoughts. Here we ask the respondents their opinion on the maturity of existing courses on MTPE. We also ask if they would potentially like to have access to an online platform to submit MTPE projects to trainees. Finally, we collect information on the awareness of the industry regarding *ISO 18587:2017*, and the draft under development *ASTM WK46396*.

The above-mentioned publication (ibid.) includes the logical condition and the possible answers of the questionnaire to LSCs. Furthermore, in pages 44 to 58, the collected data is displayed. It will be discussed in detail in chapter 7.

6.1.4. Questionnaire to linguists

The title of the questionnaire to linguists is "Survey for Post-editors of Machine Translation"[27]. It was addressed to individual professional translators who do PE as part of their job. It has 79 questions and it could take between 15 to 25 minutes to fill out, depending on the profile of the respondent.

The seven sections from the questionnaire to LSCs (commented in previous section: 6.1.3) were also used in the questionnaire to linguists. Likewise, probabilistic and convenience sampling methods were used: groups and associations of freelance translators were contacted by e-mail, and the questionnaire was posted in leading social media channels (i.e. Twitter, LinkedIn, Facebook) and online forums such as Translators without Borders[28] or ProZ[29].

The questions and answers, their typology and logical conditioning between them can be consulted in Ginovart-Cid (ibid., pp. 19–33). The corresponding collected data is presented in the same publication, in pages 61 to 82, and it will be commented in section 7.1.3. The questionnaires to LSCs

[27] https://form.jotformeu.com/82855955787379, hereinafter "Questionnaire to linguists".
[28] https://translatorswithoutborders.org/.
[29] https://esl.proz.com/forum/categories/foros_sobre_prozcom-proz.html.

and to linguists were launched in December 2018 and remained open for submissions until April 2019.

6.1.5. Questionnaire to trainers

The questionnaire to trainers is titled "Survey for MTPE training providers"[30] and its audience are PE instructors in European postgraduate studies. It contains 32 questions and takes approximately 10 minutes to fill out.

On this occasion, only convenience sampling was applied as dissemination method. As commented before, only the European translation faculties with masters or postgraduate courses in translation or related studies were contacted[31].

The consent form to be signed (appendix 10.4) was exclusively sent to the 54 relevant educators who accepted to be part of the study. They received the link to the questionnaire by e-mail after signing the consent form, and they could fill it out between May and August 2019. Since some expressed concerns about their participation considering how the PE course they were in charge of would evolve in the coming months (precisely, for the next academic year), Question 3 in the interview (see 6.3) was included to reassure them.

Regarding its content, the questions were distributed in 5 sections:

1. MTPE Training. Demographic information is collected, and the concept of PE is defined. The right audience is filtered in case the convenience sampling was not 100% effective.
2. Standards & Certifications. We inquire which standards and certifications of the language industry are presented or discussed with the translation students.
3. Current trends in MTPE teaching. This section is subdivided as follows:
    - Your profile as MTPE trainer. Three questions aim at drafting the general overview of the approach taken by the instructor in the syllabus.
    - Contents of the MTPE training. Ten questions enter into more detailed matters and techniques of the MTPE training approach.
    - The MT post-editing project. Ten questions inquire on the related matters around a typical PE assignment (i.e. professionalism and ethics, QA, etc.).

[30] https://form.jotformeu.com/82844920241354, hereinafter "Questionnaire to trainers".

[31] As a matter of fact, 200 scholars received the e-mail and their replies helped exclude the degrees where PE was not practised at all, or only at the undergraduate stage.

4. Translation Technology Tools. Here we collect information on the CAT tools and MT systems used, and if and how they are combined.
5. Feelings & Thoughts. Here we ask the respondents their opinion on the maturity of existing courses on MTPE. We also ask if they would potentially like to have access to an online platform where their students could practice on real(istic) MTPE assignments.

The questions and answers can be consulted in Ginovart-Cid (2020c), which also covers their logical conditioning and the answers in pages 35 to 43. The results, which will be discussed in detail in chapter 7, are displayed in pages 83 to 93 of the mentioned publication.

6.2. Syllabi outlines

In addition to the online questionnaires, two other data collection instruments are used for the smaller audience group: PE trainers and educators. The two instruments are the syllabi outlines (presented below) and interviews (in next section).

As argued in Ginovart-Cid and Colominas (2020), "syllabus" is a term that designs for this research the "summary outline of a [...] course of study or of examination requirements." (the written document). On the other hand, "course" is employed to refer to "a number of lectures or other matter dealing with a subject" (following the definitions of the Merriam-Webster's dictionary). It is acknowledged that both during the questionnaire or interviews the meanings might overlap for certain questions.

The syllabi were printed and read before each instructor filled out the online questionnaire, if available on the website of the relevant Faculty. If not available, the PE instructor was asked via e-mail and would provide it either before or during the interview, if a draft version of it could be shared. At the time of the study, out of the 54 studied PE courses, five did not have an official written outline (i.e. a syllabus) that could be shared with the author. The method applied to analyse the 49 available outlines is basic and rudimentary. Through reading, a number of salient elements are compared from one course to another. Namely, the contrasted elements are:

1. Name of the course
2. Compulsory or elective
3. Workload in hours
4. European Credit Transfer and Accumulation System (ECTS)
5. Learning outcomes
6. Content (Elements)
7. Specific tools
8. Assessment method (presence or absence of PE examination)

9. Language for teaching
10. Language pairs
11. Enrolment prerequisites
12. Distance learning

A spreadsheet is used to collect the cited salient elements of the syllabi, with one column per topic and one row per PE course. The results are presented in Ginovart-Cid and Colominas (2020, p 238) and discussed in more detail in chapter 7.

6.3. Interviews

As observed by J. W. Creswell and J. D. Creswell (2017, p. 19), combining open-ended qualitative data with closed-ended quantitative data has proved advantageous in the past. We agree with Saldanha and O'Brien (2014, p. 169) and De Cespedes (2018) in that interviews are a method to access privileged information, such as the opinions or thoughts that a person has about a given topic. Authors like Vieira and Alonso (2020, p. 4) have used interviews in their research, however, in ours, we do not employ any software for the analysis of transcripts. Yet, a recording device is used when the interviewee expresses consent by writing to allow re-listening to the interview at later stages of the study.

In the present research, the author had the possibility to interview 48 PE trainers[32]. An outline with eleven questions was prepared in a spreadsheet beforehand, and it could be completed with questions specific to the trainer's submitted questionnaire or syllabus outline, if available. The eleven questions are expected to make the interview last between 15 and 25 minutes, and they are listed in Ginovart-Cid and Colominas (2020, p 240) and below:

1. how long PE has been included in their course ('Age of syllabus');
2. which tools and software they present to the students, and if they have a hands-on class about MT engine training ('Tools and software');
3. whether they had or knew at the time of the interview of any plans to increase the PE presence in the curriculum ('Plans to increase MTPE');
4. whether their colleagues encourage the use of MT in 'traditional' translation courses ('Use of MT in regular translation courses');
5. whether they use the task-based or project-based approach as a pedagogical method in the course ('Teaching methods');
6. whether or not their students have hands-on practice in error categorisation and, independently, what is their opinion of the error typologies with neural MT outputs ('Error categorisation of the MT output');

[32]There were 49 interviews, as one educator was in charge of two different PE courses.

7. whether they think trainee post-editors should be encouraged to read the source or the target segment first ('Source or target segment first');
8. whether they include pre-editing of the ST in their course, and if they think this is useful to obtain a higher NMT output quality ('Pre-editing of the source text');
9. whether they consider the final quality of the post-edited product in the students' assessment ('Evaluation of the post-edited text');
10. whether deontological issues with MTPE are discussed in class and what their views are on this topic ('Deontological issues with MTPE');
11. whether they know the so-called 'split principle' as a training method for MTPE ('split-training').

Some of the questions are intended to complete the answers provided via the submitted questionnaire or to facilitate the understanding of the syllabus outline. In particular, Question 2 checks if the tools mentioned in the syllabi are the ones effectively used. Question 7 is intended to obtain a more clear answer to the submitted questionnaire, which often was nuanced with the free text field with sentences such as "It depends". Question 8 intends to complete the landscape of pre-editing today in the translation industry regarding PE services by contributing to both the submitted questionnaire and the syllabus outline. Finally, Question 9 mainly seeks confirmation of the type of examination announced in the syllabus.

Some syllabi had missing information and the researcher took the opportunity at the interview to inquire on those topics. For example, when not present in the syllabus outline, the interviewee would be asked if his/her course is compulsory or elective; how many ECTS credits the course represents; or if distance-learning is a possibility. Regarding the environment, even if most interviews were held via Skype[33], some were on the phone, one was face to face, and two were answered in written form (e-mail) due to the unavailability and time constraints of the instructors.

[33] https://www.skype.com/

Chapter 7

RESULTS AND DISCUSSION

The questionnaire to LSCs received 66 valid submissions. The questionnaire to post-editors received 142 valid submissions. And the questionnaire to PE trainers records the 54 aimed submissions.

The data collected through the three questionnaires presented in the previous chapter are published in Ginovart-Cid (2020c). In next section of the present chapter (7.1), it is discussed how the three surveyed stakeholders convene or disagree on the PE core topics introduced in section 6.1.2. In the last two sections (7.2 and 7.3), we comment on the impact of what has been observed and gathered through the study of the 49 PE syllabi and the interviews with the 48 PE educators, respectively.

Similar to the methodology of data analysis used by Saint-André (2015, p. 104), we ponderate the average score for three of the core topics, namely, the ones that are presented in a matrix-type of question: PE tasks (Likert scale 1-3), PE hiring criteria (Likert scale 1-4) and PE skills (Likert scale 1-5). The first part of next section presents an overview of the results obtained regarding the four core topics through the overall survey-based research (7.1.1). The last three parts of next section (7.1.2, 7.1.3, and 7.1.4) discuss briefly the most salient implications of each individual questionnaire, considering the contextual questions.

7.1. Questionnaires

7.1.1. Core topics of the questionnaires

In terms of PE-related tasks, as expected, the proper PE task has been the main reference to the three groups, thus, it occupies the first position in the ranking. In general, it must be highlighted how the industry stakeholders express that, on average, the number of tasks combined with PE (including PE itself) on a typical working day, are 6.17 tasks for LSCs[1]; and 6.65 tasks

[1] As a total of 580 tasks are checked either as main, secondary or occasional task by 94 respondents, if we refer to the results of tables 31 and 32 of Ginovart-Cid (2020c, pp. 51–

according to linguists[2].

It contrasts with the average number of tasks that PE trainers think a professional post-editors does, on average. Indeed, according to the university teachers it would be 13 neighbouring activities[3], which makes clear how, despite being a hybrid profile, it seems more multidisciplinary in the eyes of stakeholders in academic settings, whereas the industry representatives expect a more specialised profile. In table 7.1 the computed score for each PE-related task is expressed by computing 3 points to "Main tasks", 2 points to "Secondary tasks" and 1 point to "Occasional tasks", and the weighted score as a percentage is also displayed, according to the data available in tables 33, 106, and 173 of Ginovart-Cid (2020c).

Table 7.1: Scores of three audiences to PE-related tasks

	LSCs		Post-editors		Trainers	
Task	Score	%	Score	%	Score	%
T1	0.38	13	0.36	12	1.21	40
T2	0.77	26	0.44	15	1.36	45
T3	0.40	13	0.51	17	1.92	64
T4	0.39	13	0.44	15	1.62	54
T5	1.10	37	0.89	30	2.25	75
T6	0.46	15	0.45	15	1.91	64
T7	2.23	74	2.44	81	2.87	96
T8	0.33	11	0.61	20	1.81	60
T9	1.12	37	1.33	44	2.23	74
T10	1.71	57	1.74	58	2.70	90
T11	1.83	61	1.91	64	2.68	89
T12	0.46	15	0.52	17	1.57	52
T13	0.72	24	0.71	24	1.66	55
T14	0.57	19	0.60	20	1.74	58

The data collected shows that trainers and LSCs believe that revision of post-edited MT output (T11) and proofreading of post-edited output (T9) are activities closely related to PE, which a post-editor frequently performs during the working day. LSCs seem to agree with trainers that the task of collecting feedback (T2) is important in the job of a post-editor. Trainers say that post-editors evaluate MT output quality (T5) more often than they proofread a job that a colleague has already post-edited (T9). They also voted the activities of management of MTPE projects (T3), PE guidelines design (T6), and pre-editing the ST (T8) as more recurrent in the working

52).

[2]Please refer to table 106 in Ginovart-Cid (2020c, p. 72).

[3]Please refer to table 173 in Ginovart-Cid (ibid., p. 90).

day of a post-editor than the other two groups of stakeholders. The comparison of the answers of the three surveyed audiences (curves) to the average score (bars) of each task is depicted in figure 7.1[4].

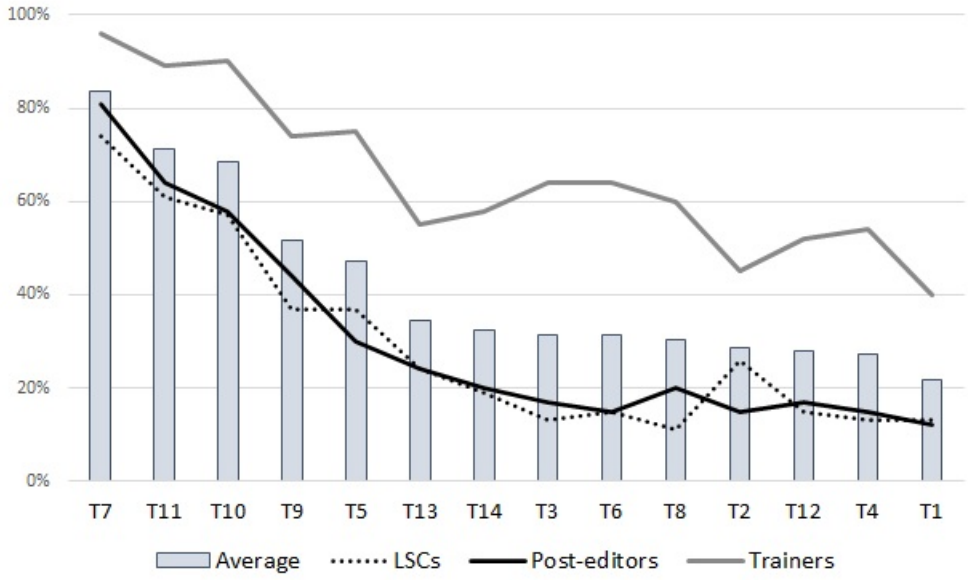

Figure 7.1: Scores for the 14 PE-related tasks by the three audiences

According to the one-way ANOVA test on PE tasks, the null hypothesis is rejected: there is a significant difference amongst populations. The F-test result for the PE-related tasks is summarised below:

$$F_{(2,39)} = 14.41, p < .05$$

The difference is highly significant (p-value 0.00002). It is clearly coming from the view the trainers have about the industry: the totality of tasks, from T1 to T14, are considered much more present in the working day of a post-editor from the point of view of trainers (see figure 7.1).

As a conclusion on the PE-related tasks, one could assume that, from the perspective of university training, it is safer to prepare students and novice linguists to a wider range of activities. Despite the phrasing of the question[5] referring to the reality of PE work in the industry, the educators' answers may reflect on this desire to leave open paths in the careers of their graduates.

[4]Previous publications, such as Ginovart-Cid and Oliver (2020), had only considered for the average the LSCs and post-editors' submissions.

[5]*What workload do you think the following PE-related tasks might bear in the everyday work of a professional post-editor?* published in Ginovart-Cid (2020c).

About hiring criteria for candidate post-editors, as explained in section 6.1.2, only the two industry stakeholders were asked this question (LSCs and linguists). Hence, the comparison of scores is displayed in table 7.2 for these two groups, considering tables 36 and 108 in Ginovart-Cid (2020c).

Table 7.2: Scores of three audiences to hiring criteria

	LSCs		Post-editors	
Criterion	Score	%	Score	%
C1	0.52	17	0.73	24
C2	2.06	69	2.01	67
C3	0.50	17	0.56	19
C4	0.50	17	0.51	17
C5	1.06	35	1.27	42
C6	0.52	17	0.79	26
C7	1.47	49	1.65	55
C8	1.58	53	1.88	63
C9	1.58	53	1.99	66
C10	2.29	76	2.32	77
C11	0.20	7	0.46	15
C12	1.11	37	1.16	39
C13	2.18	73	1.90	63
C14	0.58	19	0.79	26
C15	1.15	38	1.25	42
C16	1.02	34	1.12	37
C17	1.59	53	1.42	47

Linguists agree with LSCs about the first most valued criterion: revision and proofreading skills (C10). Their opinions differ on the weight other features (should) have when selecting a professional post-editor for a PE job. For individual post-editors the CAT tools knowledge (C2) and QA skills (CC9) could be more valued, than LSCs think. It is surprising to note that the University degree (C17) is actually more valued by respondents from LSCs than linguists believe to be. To conclude on the hiring criteria analysis, it must be noted that neither of the mentioned divergences between LSCs and linguists is found to be statistically significant. The contrast is depicted in figure 7.2.

Concerning the hiring criteria, the result of the *t*-test does not allow to reject the null hypothesis. In other words, there is no statistically significant difference in the way LSCs and linguists view hiring criteria (p-value=.60).

In regard of PE skills, the results consisting of the weighted scores abstracted from tables 37, 107, and 171 in Ginovart-Cid (ibid.) can be found in table 7.3.

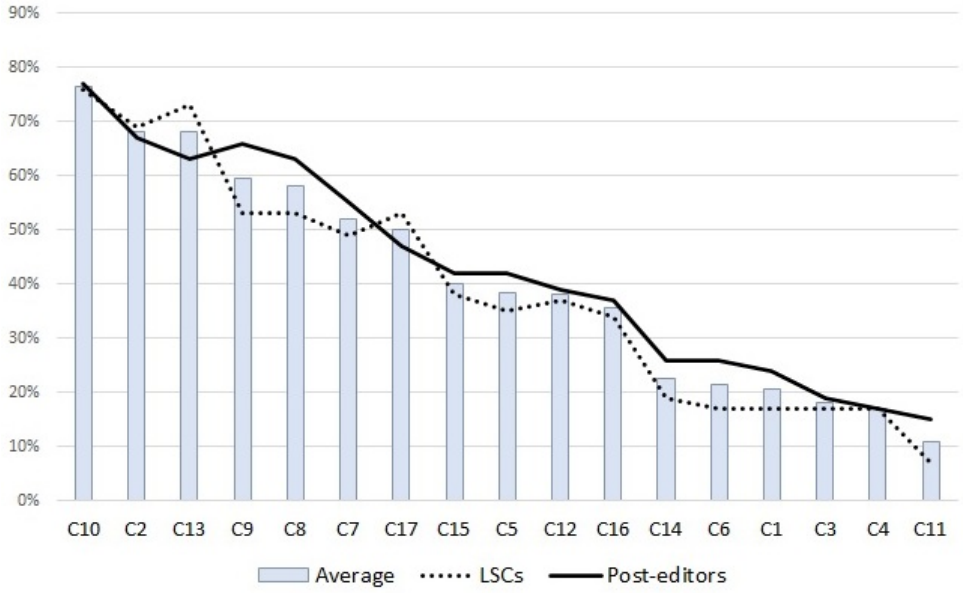

Figure 7.2: Scores to the 17 PE hiring criteria by LSCs and linguists

Table 7.3: Scores of three audiences to PE-related skills

	LSCs		Post-editors		Trainers	
Skill	Score	%	Score	%	Score	%
S1	4.76	95	4.16	83	4.55	91
S2	4.50	90	4.02	80	4.66	93
S3	4.77	95	4.42	88	4.75	95
S4	4.20	84	3.94	79	4.23	85
S5	2.00	40	2.61	52	3.04	61
S6	1.92	38	2.39	48	2.68	54
S7	4.39	88	4.30	86	4.43	89
S8	4.05	81	3.92	78	4.40	88
S9	3.38	68	3.40	68	4.45	89
S10	3.20	64	3.00	60	3.45	69
S11	3.35	67	3.62	72	3.40	68

A higher inter-agreement score is found among the three surveyed audiences (see figure 7.3). Only three skills show statistically significant differences, namely: trainers value more than the other two stakeholders the (S9) capacity to advise when MTPE is appropriate, (3.9σ, 99.9% confidence level). The (S1) capacity to decide when to edit or discard and the (S2) capacity to post-edit according to PE guidelines are significantly better valued by LSCs than by the individual professional respondents (2.9σ and

2σ, respectively).

The one-way ANOVA test does not allow to reject the null hypothesis. In other words, there is no statistically significant difference in the way LSCs, linguists, and trainers view PE skills.

$$F_{(2,30)} = 0.70, p = .50$$

This is a meaningful and positive outcome, since the selection of the top PE skills for future research can be supported by the common ground found in this question amongst the three surveyed groups.

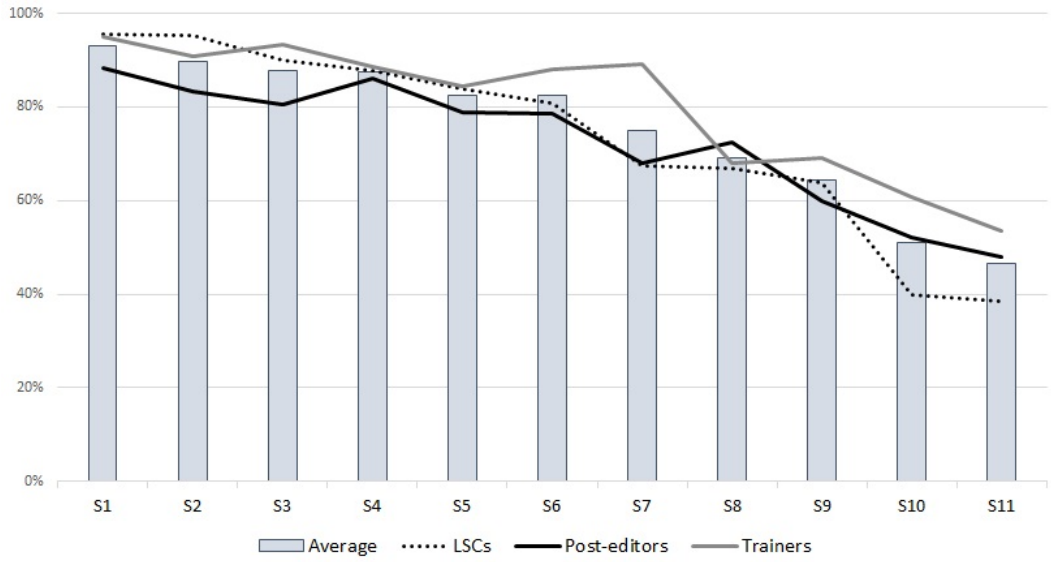

Figure 7.3: Scores to 11 PE-related skills by three audiences

The answers provided by the three stakeholders on PE training elements are commented in the following lines. Keeping in mind that the scope (goals and learning outcomes, time constraints, varying range of audiences, etc.) cannot possibly match between academic programmes and professional or CPD training courses, the comparison is nevertheless made between topics that LSCs include in their proprietary PE courses (according to LSCs and linguists), with the content of PE courses at European master and postgraduate programmes. As a matter of fact, for the post-editors questionnaire, Question 49 (published in Ginovart-Cid (2020c): "Which elements does the training include?") was not replicated under the section of the questionnaire entitled "Training provided by a University", since the submissions by edu-

cators combined with syllabi outlines and interviews are judged as reliable enough in that respect.

LSCs will most often include presentations and training on (1) PE levels: light and full post-editing; (2) PE guidelines: exhaustive list and examples in the relevant language pair; (3) PE attitude: decide when to discard and translate from scratch; (4) MT systems: rule-based, example-based, statistical, hybrid or neural; and (5) PE risks: under-editing, over-editing and pseudo-editing.

Post-editors that attended a PE course provided by a customer express that the elements more often present in such private training settings are: (1) MT evaluation: human (scoring, ranking, error categorisation), (2) PE levels: light and full post-editing; (3) PE guidelines: exhaustive list and examples in the relevant language pair; (4) Practical PE exercises in the relevant language pair; and (5) Integration between CAT tool and MT system.

Finally, for trainers at European faculties the order is (1) MT systems: rule-based, example-based, statistical, hybrid or neural; (2) PE levels: light and full post-editing; (3) Practical PE exercises in the relevant language pair; (4) MT evaluation: human (scoring, ranking, error categorisation), and (5) Integration between CAT and MT system. The comparison of the three groups is depicted in figure 7.4 and the scores are displayed in table 7.4:

Table 7.4: Scores of three audiences to PE training elements

Element	LSCs %	Post-editors %	Trainers %
E1	28	30	33
E2	61	55	80
E3	50	24	50
E4	56	73	83
E5	72	48	93
E6	78	45	72
E7	78	58	54
E8	94	67	87
E9	67	55	85
E10	61	48	69
E11	17	48	56
E12	56	12	13
E13	22	18	65
E14	28	30	28
E15	28	15	19

The scores in table 7.4 are obtained by dividing the total number of times an item is chosen by each population by the number of respondents, that is, its potential maximum score. The raw data gathered via the surveys can be

consulted in tables 54, 129, and 170 in Ginovart-Cid (2020c).

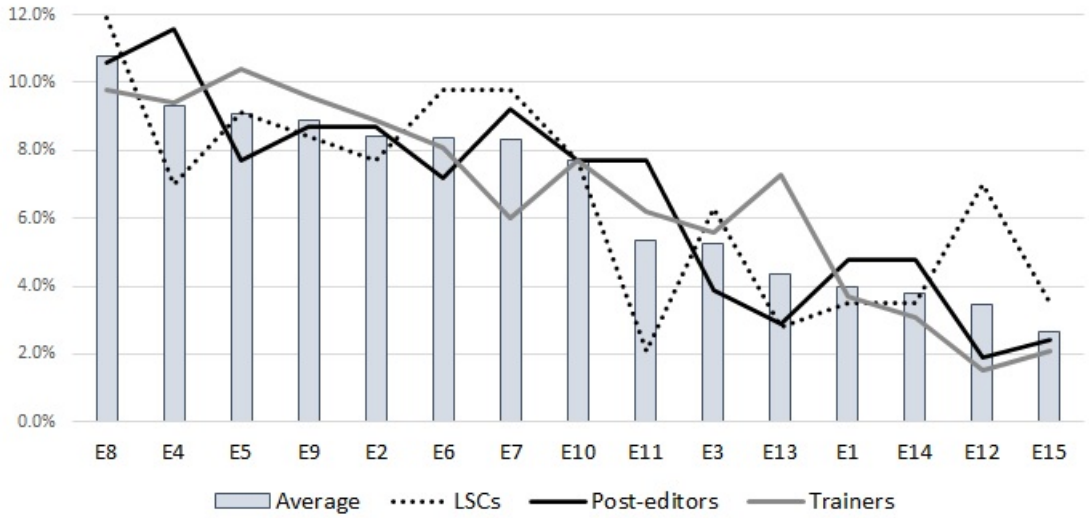

Figure 7.4: Scores to the 15 PE courses' elements by the three audiences

Applying one more time the calculator provided by Blanco Claraco (2019) to easily analyse poll results, some significant differences are observed between pairs of stakeholders, as presented in table 7.5.

Table 7.5: Difference and significance for training elements

Element	**Between**	**Difference**	**Confidence**
E2	Industry/trainers	3.4σ	99.9%
E3	LSCs/linguists	3.0σ	99.7%
E3	Linguists/trainers	3.3σ	99.9%
E4	LSCs/linguists	2.4σ	98.3%
E4	LSCs/trainers	3.4σ	99.9%
E5	LSCs/linguists	3.4σ	99.9%
E5	LSCs/trainers	3.2σ	99.8%
E5	Linguists/trainers	8.2σ	100%
E6	LSCs/linguists	5.0σ	100%
E6	Linguists/trainers	3.6σ	99.9%
E7	LSCs/linguists	3.0σ	99.7%
E7	LSCs/trainers	2.8σ	99.5%
E8	LSCs/linguists	5.5σ	100%
E8	Linguists/trainers	3.3σ	99.8%
E9	Industry/trainers	4.0σ	99.9 %
E11	LSCs/linguists	4.9σ	100%

Continued on next page

Table 7.5 – continued from previous page

Element	Between	Difference	Confidence
E11	LSCs/trainers	4.7σ	99.9%
E12	LSCs/linguists	6.5σ	100%
E12	LSCs/trainers	5.6σ	100%
E13	Industry/trainers	6.4σ	100%
E15	LSCs/linguists	2.0σ	96.3%

Table 7.5 shows how (E5) the more or less technical training on MT systems, (E12) the mastery of the so-called "PE technology" that embodies innovative inputs methods, and (E14) pre-editing are the most significant sources of differences between the stakeholders.

Notwithstanding the results commented above, the one-way ANOVA test does not allow to reject the null hypothesis for training elements. That is, despite the low-level significant differences between pairs of stakeholders on single elements, the overall disagreement is not statistically significant for the topics LSCs, linguists, and trainers see included in their known MTPE courses.

$$F_{(2,42)} = 2.15, p = .12$$

The reasons why we cannot reject H0 for hypotheses 2, 3, and 4 could suggest, to a certain extent, agreement between the three stakeholders. However, one more plausible explanation is that the samples are not large enough or that there is too much variability within groups (Salkind 2017, p. 238).

To provide an overall analysis of the core topics thoroughly commented in the previous paragraphs, one may tentatively contrast the selected topics. In particular, one noticeable statistically significant difference is the comparison between the PE-related skills and the PE-related tasks: our respondents give high scores to the (S5) capacity to pre-edit a source text according to CL, whereas the task (T8) pre-editing the source text is not as important (3.6σ), nor is it considered as important from a hiring perspective (5.8σ for C6).

7.1.2. Questionnaire to LSCs

From the submissions by the 66 respondents to the questionnaire to industry stakeholders, almost 88% are representatives of an LSC (as opposed to an employer at a firm with an in-house translation department), which is why the label "questionnaire to LSCs" is used throughout the thesis.

The topics that have been discussed in section 7.1.1 about PE-related tasks, skills, criteria for hiring and training elements are now completed thanks to contextual data gathered in the rest of the questionnaire. In fact,

to put into perspective the answers described in the previous section, one should keep in mind the profile of the surveyed audience. Most of the submissions came from project managers or executive managers (approximately 65%), and most of the respondents hold a Master's degree (approximately 64%).

We have observed that the in-house teams are not as big as the outsourcing resources in LSCs, according to the data in tables 28 and 29 in Ginovart-Cid (2020, p. 51). Moreover, the awareness of industry standards amongst the surveyed LSCs is very balanced: approximately 60% know either the *ISO 18587:2017* or *ASTM WK46396* standard, for instance, but over 40% do not know any of the listed standards and chose not to use the free-text field to precise another standard. Against this landscape on the profile of participants, the reader may find interesting to see if the practice of PE (selling and buying MTPE projects) has been carried out for a long time or if its workload (versus classical TEP projects) is considerable nowadays.

The results of combining their years of PE experience with PE workload show a prevailing "Intermediate" level, that is, the PE activity seems to be progressing at a steady pace. To obtain this level of expertise, each answer to Question 14 ("How long have you been handling MTPE projects?" as can be seen in Ginovart-Cid 2020) is assigned one value: 1 for "Less than 1 year", 2 for "Between 1 and 5 years", and 3 for "More than 5 years". This value, for each respondent, is multiplied with the value obtained in Question 15 ("What is the percentage of your translation production via post-editing (PE) of machine translation (MT)?" of the same publication), namely: 1 for "25% or less", 2 for "Between 26% and 50%", 3 for "Between 51% and 75%", and 4 for "More than 76%". The expertise scale ranges from 1 to 12. These values are arbitrarily distributed as follows to make the analysis easier and to aim at more or less balanced groups.

- Beginner: Value 1
- Intermediate: Values 2 and 3
- Advanced: Values 4 to 12

Consequently, the distribution of the respondents to the questionnaire in these three expertise levels is presented in table 7.6.

Table 7.6: MTPE expertise level for LSCs

Label	**Percentage of respondents**
Beginner	11%
Intermediate	65%
Advanced	24%

As argued in Ginovart-Cid (2020b), such three levels of expertise may be indicative of the current state-of-the-art in PE by pointing toward the

effective practices, tools and future trends of the profession.

If we consider other contextual topics, for instance, the PE risk that linguists fall most often into according to LSCs, we see that under-editing is the picked answer for over 42% of the respondents. Hence, it could be argued that courses with components of the traditional TC are currently as relevant as anytime before. Indeed, properly acquired traditional translation and revision techniques would probably reduce the risk of not spotting an MT output error. On the contrary, it could also be argued that if current courses centered on regular translation strategies and techniques fail to produce professionals who systematically avoid under-editing, a change in PE pedagogy is necessary.

All in all, it must be noted that there are more factors that come into play. As discussed in section 2.2, the conditions have an obvious impact on the quality of the product. If the expected productivity is unreasonable to the MT raw output and volume of words to post-edit, the best linguist would be incapable of editing all the MT errors. Furthermore, as mentioned by other authors (Aranberri 2017, p. 91; Offersgaard et al. 2008, p. 155), not every good translator is systematically a good post-editor.

The current situation thus seemed to call for specific PE training, tailored to each audience. It is noteworthy that only 13.6% of the surveyed LSCs offer PE training on a regular basis to their providers (in-house or freelance), and over 66% have never organised one. About the current PE courses available and known by the respondents, almost 38% consider they do not fulfill the needs of the professionals who carry out the PE jobs, and over 27% of the participants express their lack of opinion or knowledge about such courses. Consequently, almost 64% would appreciate an access to some sort of "virtual PE internship", where their providers could learn and practice in a controlled environment.

7.1.3. Questionnaire to linguists

From the submissions by the 142 respondents to the questionnaire to linguists who post-edit MT output as a professional activity, the findings analysed in section 7.1.1 about PE-related tasks, skills, hiring criteria and training elements shall be considered from the perspective of contextual data gathered in the rest of the questionnaire.

The large majority of the respondents are freelance translators (84.2%) and do not hold a continuous-development professional certification (68.2%) apart from their University degree. Moreover, most of the respondents (67.6%) are not aware of industry standards such as *ISO 18587:2017*. When looking at the years of experience in MTPE and the workload of PE projects (percentage versus classical TEP projects), as has been done in sec-

tion 7.1.2, it becomes patent that the profile is not yet mature, since the greatest percentage of respondents are found to be "Intermediate" as per the expertise level commented before.

Table 7.7: MTPE expertise level for post-editors

Label	**Percentage of respondents**
Novice	15%
Intermediate	54%
Expert	31%

As argued in Ginovart-Cid (2020b), these results evidence the young age of the profession of PE post-editor[6]. The fact that the dedication to PE is ancillary nowadays by our respondents is supported by some of the comments sent via the free-text field at the end of the questionnaire (Ginovart-Cid 2020c, p. 33). For example, one respondent says

> *I have mostly done PEMT on assignments from Translators Without Borders on the Kató platform.*

Another also expressed the lack of knowledge of the empirical parameters and evolution in the field like this:

> *You've gathered everything that I never knew about what I've already been doing for a while and now I want to actually follow a course. Looks promising. Thanks.*

Another interesting result to comment about the current conditions in which MTPE projects are carried out is the tool. From the 142 submissions one clearly notices that there is an aspect that has not changed: it is the layout of the CAT tool (horizontal or vertical editor), since it is mostly keeping two main fields (source and target). Indeed, almost 86% of the respondents work in a tool that allows for editing with a bilingual display of the text. For instance, one respondent sent the following comment at the end of the submission:

> *I am greatful*[sic] *that I usually only have to use MT as a "support tool addon" to the CAT tm*[sic] *matches, as I would find it difficult to start thinking about translation in "MT mode".*

About PE training, it is worth mentioning that over half of the respondents had never attended a PE course of any kind (almost 52%). As discussed in chapter 4, the implementation of PE courses or PE content in current University courses is under way. However, the reality in the industry is

[6]The classification in three levels may be seen as a simplified application of the Dreyfus model (Dreyfus et al. 2000) for skill acquisition which Kiraly and Sascha (2019, pp. 60, 72) argued to be the centerpiece of Chesterman's *Memes of translation* (1997).

still reflecting a general lack of theoretical and proceduralised knowledge, which can be expressed in the words of one linguist who filled out the free-text field as follows:

> *I think the approach to MT is generally amateurish. I work with international companies, so I guess it's a general issue, although more severe in Italy where the approach to translation is poor and unprofessional. Clients/LSPs dictate the pace and the conditions (tools, productivity, rates, quality...), making it not worthwhile to spend time, money and efforts on formation*[sic].

7.1.4. Questionnaire to trainers

From the submissions by the 53 respondents to the questionnaire to trainers who teach PE at a master or postgraduate programme, the core topics commented in section 7.1.1 are now framed within the contextual data gathered in the rest of the questionnaire.

The large majority of the respondents (over 60%) are female professors (full-position), and only a minority are guest speakers on a more or less regular basis that bring a professional insight as they hold a position in the translation industry. Some of them (around 19%) teach at more than one institution. The countries with a higher number of PE courses studied are Germany, Spain, and the United Kingdom (15.1% each country), and also France (11.3%), and Italy (9.4%), but submissions were received from 17 countries in Europe. Regarding industry standards, 30% do not present any to the students. The most praised standards are: *ISO 17100:2015*, *ISO 18587:2017*, and *ISO 9001:2008*.

Another interesting topic to look at is the weight that PE has in the whole course, and the type of instruments used to teach it. Almost 52% respondents say only a quarter of the course or less is dedicated to PE, while approximately 20% focus on PE for almost half of the course, and approximately 28% of the surveyed PE trainers dedicate more than half of their total number of classes to PE. Regarding the instruments, while a slides presentation and some readings are almost always present in any tertiary education programme, it is clear that PE is a practical activity. As a matter of fact, the hands-on activities have received also quite high percentages in the multiple-choice checkbox question reported in Ginovart-Cid (ibid., p. 86).

If we consider other contextual topics, for instance, the advice a PE teacher may give about reading the source sentence or the target sentence first, half of the respondents do not tackle this matter or nuance their answer. About 31% recommend reading the ST first, while the remaining minority (approximately 18%) are inclined to suggest to their students to

skim the TT in the first place. Finally, to conclude the commentary on peripheral questions of the questionnaire to trainers, it is noteworthy that the PE risk that trainers consider PE students fall most often into is over-editing (approximately 48%), contrary to what LSCs expressed. It must be noted, however, that it is almost a balanced answer, since more than 40% of the surveyed educators think it is under-editing. The already stated instruments that complete this last and smallest questionnaire have been discussed in sections 6.2 and 6.3, and their results are discussed in the next two sections.

7.2. Syllabi outlines

After the educators had been contacted and they had expressed their interest in taking part in this study, and signed the consent form, we requested the syllabus of the course they are in charge of, if it was not available at their institution's website. The 49 syllabi available at the time enabled us to gain insight into the way PE is currently being taught in European universities. As has been introduced in section 6.2, 11 selected information points were checked for the set of syllabi.

It should be firstly highlighted that the written outlines contain highly varied levels of information, as it depends on each faculty and country. While certain contain even the name of the instructor, others lack some basic pieces of information that must be asked at the interview stage. The data collected from the analysis of the syllabi has been described in detail in Ginovart-Cid and Colominas (2020), and the resulting implications of the results are summarised in the following paragraphs. The fact that not even half of the courses include an explicit mention of "post-editing" in their title may be indicative of the fact that, at the time of writing, most translation faculties do not consider PE as a core element to the service-provision competence in the translation market. In fact, the most common titles would include a mention to "computer-assisted translation", "translation tools" or "translation technology" instead of highlighting the development of skills and competences for the translation profession in a more holistic manner.

In relation to the necessity to complete the PE course as a compulsory one in the relevant curriculum, it is noted from the study of the syllabi, that more than a quarter of them are elective. Another scenario found is that the same PE course is mandatory in a certain path or major but optional for another branch of the same curriculum or a related Master programme, which certainly increases the difficulty in terms of assessment and evaluation. In reference to the weight of PE in the syllabi, the outlines hardly ever present a specific distribution of the contact and study hours dedicated to PE. Firstly, without referring to PE content only, most of them announce between 12 to

50 hours of class. Secondly, the majority announce from 8 to 160 hours of study time at home. Certainly, the extent to which the workload focuses on PE needs to be clarified thanks to the interviews that are discussed in section 7.3.

As for ECTS credits, which should be in concordance to the number of hours of a course, it is noted that most syllabi range from 2 to 14 ECTS credits. Like with the contact and study hours, the reasons behind this are clarified through the interviews presented in next section.

Approximately 20 syllabi cover one single language pair (uni- or bi-directional). On the other hand, two of the studied courses can cover up to 14 language pairs, depending, of course, on the variety of native languages the students attending each year have in their profiles. Undoubtedly, the multilingual setting makes the task of designing a PE course more difficult and the activities that should allow for the proceduralisation of the PE competence become either too general ("language-agnostic") or too difficult to assess by a single educator (who probably cannot evaluate as many native language as their students have in a multilingual setting).

As a conclusion, it could be argued that PE courses should be, like translation and revision courses, part of the general branch of translation curricula, instead of a different specialisation or major. Furthermore, the lack of presence of PE in the examination (only four syllabi include PE in the test) is a good reason to support the PE pedagogy research and, in particular, the arduous task that authors like Van Egdom, Vieira, et al. (2018) started, as commented in section 6.1.2.

About the prerequisites for enrolling, it has been gathered from the syllabi that around 70% of the courses do not have any. It may be due to the fact that being enrolled in the Master's programme or having completed the first year should guarantee that the foundation of the required knowledge is laid for all the students. Some courses include recommendations about how the student should be able to use an Ms Office suite (word-processing, spreadsheet and presentation software), be already familiar with CAT tools, and, more generally, have ICT skills.

Because the study of the syllabi and the interviews took place before the global coronavirus pandemic, a big percentage of the studied PE courses did not allow for distance learning, which can probably be explained by the need for a room equipped with licensed software. However, this may have now rapidly changed. It must be noted that despite the technology available (virtual private network -VPN-, etc.), it is a more complex course to teach when the possibility of the hands-on on-site support is completely out of question, let alone the possibility of sporadically combining practice with discussions and reflection about PE choices.

7.3. Interviews

The interviews are the last instrument used to complete the questionnaire to PE educators. The design of the interview has been introduced in section 6.3. They took place between September and November of 2019. Before asking Question 1 (see section 6.3), the interviewee and the interviewer introduced themselves. Most of the time the interviews were held in English, but some were conducted in French, Spanish or Catalan, depending on the interviewee. After the acquaintance, interviewer and interviewee confirmed that the syllabus the respondent had in mind when filling out the online questionnaire was indeed the one previously identified, printed and analysed, since some faculties had slight variations of the same course with different names, or for different branches of the curriculum.

As has been noted in section 6.1.5, a question about the plans to increase the weight of PE in the syllabus or the curriculum is asked during the interview to include the participants whose syllabi were undergoing modifications at the time of the study. From May to August 2019, approximately half of the participants were informed or actively included in the process of editing the PE syllabus. They said that more ECTS and hours would be dedicated to PE in the next academic year, or that the course would be split into two, for example, leading to one stand-alone course in revision and PE, and one stand-alone course in translation technology (or localisation or the content that was originally mixed with PE). Also, two of the interviewees said that PE would become as from 2020 part of the undergraduate programme.

Considering that PE has traditionally been more close to courses focused on CAT tools or localisation, it may not be striking that some syllabi are as old as 2000 and 2005, even if the majority have been created between 2015 and 2020. This traditional link between PE and translation technology may also explain why some interviewees highlight their efforts towards making the syllabus as "language-agnostic" as possible.

The fact that the population of students enrolled can be international either made it impossible to evaluate the quality of the post-edited text (if the educator had not mastered the target language) or led the students to post-edit languages in which they are not native. From the discussions on the age of the syllabus and the language pair at the interview, it is concluded that PE shows a growing curve, since PE courses are now more often included in translation curricula. This probably finds an explanation in the recent inclusion of PE-related skills in the already mentioned EMT, which is regularly updated to reflect the evolution of the language industry.

Concerning the tools used, two observations come as a conclusion of the 48 interviews. Firstly, some syllabi have longer lists of software than the ones that can actually be tested during the course, given the number of

contact/study hours. Secondly, on the contrary, certain outlines do not list any CAT tool or name only one, while the PE trainer has in fact a varied range of activities including hands-on practice on tools such as Memsource, SDL Trados Studio, Matecat, memoQ, etc. Six interviewees recognised the usefulness of some PE exercises in Ms Excel and Ms Word. Finally, in terms of MT systems, the most widely used is Google Translate[7]. It comes as no surprise that the second one is DeepL. The rest of the trainers ask their students to use Microsoft[8] and/or Bing[9], KantanMT, Tilde[10], e-Translation[11] and SDL Language Cloud[12]. Only two PE courses include hands-on on training MT systems with bilingual corpora and preprocessing, postprocessing, and RegEx rules.

About the use of MT in regular translation courses, despite an effort to increase the use of CAT tools in translation classes (Ginovart-Cid and Colominas 2020), around 50% of the interviewed educators say most of their colleagues do not introduce CAT, and even less so MT, in their translation courses. One interviewee mentioned that some "Train the trainer" courses are organised to empower the professors of the translation programme with sufficient knowledge on translation technologies to use them to a certain extent in the other courses.

When asked about the teaching methodology, we considered one of the "cornerstones of competence-based translator training, namely the [...] translation task and project-based approach as a methodological and curriculum design framework" summarised in (Hurtado Albir 2015). In fact, Kiraly (2012a, p. 84) pointed towards the interest in organizing learning activities in real working environments to prepare the trainees to multifaceted situations. Thus, the interviewees were asked whether they had such a project-based approach and a significant number of educators claim that, while the syllabus is not project-based, it can be considered as a whole set of task-based learning activities. Overall, approximately ten of the 49 syllabi follow Kiraly's project-based approach.

In terms of error categorisation, the interviews lead to the finding that almost half of the interviewees do not currently have a structured hands-on activity on classifying the types of errors about an MT output extract, which seems striking considering how the MT errors' analysis is considered to be an important task in the industry, as has been argued in section 7.1.1. However, the educators are nevertheless asked about their opinion on the

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

[8]https://microsoft.github.io/inmt/ and https://www.microsoft.com/en-us/translator/

[9]https://www.bing.com/translator/

[10]https://translate.tilde.com/

[11]https://webgate.ec.europa.eu/etranslation/public/welcome.html

[12]https://languagecloud.sdl.com/

NMT errors, when compared to SMT or RBMT engines. All except one see NMT error types as substantially different. As a conclusion, almost half of the interviewees even added the thought that what is becoming more challenging for novice translators nowadays with NMT engines is *actually spotting* accuracy errors. This observation is in line with the outcome from the questionnaire to LSCs that under-editing is a frequent error by post-editors.

Considering how the PE trainers answered the questionnaire on the topic of reading ST or TT first, they are asked why none was chosen when filling out the survey and how is it so nuanced during the PE course. As an answer, almost half of the interviewees acknowledged that there is lack of empirical evidence supporting one course of action or the other, and therefore the choice is made by each student. In particular, one professor commented that reading the target segment first may lead to:

> [...] a more error-prone state of mind of the post-editor. However, I try not to influence my students, and I try to make it clear to them that both approaches have merits and flaws. And, despite the fact that I do not have the "scientific" data to support my theory, I would say the students that have a stronger background in "traditional" translation tend to focus more on the ST. (Anonymous)

On the topic of ST pre-editing and CL, despite the observations made in section 7.1.1 about how it is not a central task in the everyday work of a post-editor but still a valued skill, mostly by trainers, the interview leads to the finding that hardly ever pre-editing constitutes a hands-on activity in the PE syllabi. Moreover, the educators' opinion about how useful can CL be with NMT outputs was generally not clearly defined. They expressed hesitation and conceded that pre-editing the ST may now be less capital with NMT than it had been with SMT (agreeing with Nitzke, Hansen-Schirra, et al. 2019). Yet, some noted that pre-editing can still be useful for some scenarios and genre of the text.

One of the last questions asked to some interviewees, if time allowed, is whether they debate the ethical implications of using MT professionally with their students. For instance, if they think the customer should be informed when MT is being used. Some educators consider that it is ethically necessary to inform the client, while, on the contrary, and more often, interviewees consider that MT should be an available resource to any professional translator, as long as they are used like TMs, glossaries, or any other tool that leads to a product of quality, in all cases not inferior to the quality that would have been provided with HT.

Lastly, the concept of split-training (presented in chapter 4.3) was introduced as a last question before concluding the interview. Except for two professors, the rest claimed not to know such split-technique concept from the PE training perspective. Hence, the author's understanding of it was shared with the interviewee first, to define the concept. Most professors did not have a straightforward opinion. Some expressed motivation and positive ideas about researching it in more detail, while others considered it unfit to their PE courses.

Final thoughts

The review of the state-of-the-art of PE neighbouring tasks, hiring criteria, PE skills, and training elements for a professional post-editor in part I led to conclude that updated and detailed knowledge could be useful to draw the current landscape of the profession from the Machine Translation Post-editing (MTPE) perspective. In chapter 5 we set the two RQs for this exploratory study. They are reminded below, without the corresponding subquestions.

> **RQ1** To what extent do the Post-Editing (PE) stakeholders agree on the definition of PE as an activity and of the post-editor profile?
>
> **RQ2** To what extent is the profile of an MT post-editor multi- and transdisciplinary?

To answer them, the data obtained via three fine-grained online questionnaires, along with the analysis of University PE syllabi and interviews to the educators, contribute with valuable information from the industry and academia to find out how each stakeholder views the PE activity. The methodology has been described in chapter 6, where the design of the survey-based research (three questionnaires to three populations) has been presented. The chapter also served as an introduction to the methods used for the analysis of the selected 49 syllabi, as well as the one-on-one interviews with the 48 corresponding PE educators. The results and implications of the data gathered through these three instruments have been exposed in chapter 7. Whereas the number of submissions to the questionnaires may be considered to be mediocre for some type of statistical analysis, it is acknowledged that "[the] total number [of the populations] is difficult to determine, though, particularly since the profession is not regulated" (Schäffner 2019; Saint-André 2015, p. 149), and it it suffices to provide a preliminary view of the current PE market.

Most Language Service Companies (LSCs) belong to the Intermediate expertise level (see table 7.6). The perspective by LSCs seems to be that PE is the solution to increase productivity (66.7% of the respondents track

it, as reported in table 25 of Ginovart-Cid (2020c, p. 50), while keeping the same quality: 72.7% of the respondents work with MTPE to improve acceptable raw Machine Translation (MT) output up to high or publishable quality, as reported in table 24 of Ginovart-Cid (ibid., p. 50). However, one can notice some lack of knowledge around pricing methods, for source word rate is still the main method, as reported in table 34 of Ginovart-Cid (ibid., p. 53) and around PE training, for only 9 LSCs provide PE training courses regularly, as reported in table 48 of Ginovart-Cid (ibid., p. 56). Yet, the need for such training is acknowledged, if we consider that over 60% of the respondents would be willing to access a digital solution to provide PE internships, according to the findings in table 69 of Ginovart-Cid (ibid., p. 61).

Like with LSCs, the majority of the surveyed post-editors belong to the expertise group that we called "Intermediate" (see table 7.7). They express generally a loss of control/agency and they have often highlighted the need for empirical data and training in PE. The view from the trainers is that PE courses are just now being created and upgraded (thanks to resources such as "Train the trainer" modules), and PE is starting to be included in undergraduate programmes. They have noted the need for time/resources to go into further detail in their PE courses. Overall, the most complex issue to be solved in the academic setting is how to address the PE courses where more than two or three target (native) languages should be handled (the "language-agnostic" issue). Likewise, more empirical data on PE research is needed to define all the ins and outs of the post-editor profile. For instance: is there empirical evidence to support that post-editors should read first the Source Text (ST), the Target Text (TT), or that it does not have an objective impact on qualitivity?

To answer **RQ1**, it can be established that LSCs, linguists and PE trainers agree to a medium extent about the definition of PE as a professional activity and the definition of the post-editor profile. There is common ground on certain key topics. For instance, a positive outcome is the agreement by the three groups that the top PE skills are S1, S2, and S7 (drafted in 6.3), whose results are presented in table 7.3. As the null hypotheses could not be rejected in chapter 7 for skills, criteria, and elements, any differences observed for these matters may be due to chance alone.

The F-value of the one-way ANOVA performed in chapter 7 made us reject H0 for Hypothesis 1. Hence, there are "mismatched expectations" (Vieira and Alonso 2020, p. 15) for which we are 95% sure that are not due to chance alone. There are indeed statistically significant differences in the views on the distribution and weight of activities professional post-editors perform in their working day. The disagreement is found between industry stakeholders (linguists and LSCs) on one side, and train-

ers/instructors/educators on the other. As the day has 24 hours for everyone, the higher weighted scores for trainers may not be meaningful from a pragmatic perspective: the fact they chose more main tasks than the other two groups could indicate their wish to prepare students for more than one job position, as argued in section 7.1.

While there cannot be a one-size-fits-all answer to account for all the items and unsolved questions addressed in the present part, one can only presume the PE landscape. In hesitant scenarios, one could follow Muzii's recommendation (2016) that translators and LSCs should reject any MTPE project that does not fulfill four features that he listed as requirements for an MTPE project "to prevent being paid for iron while providing gold". The four features, without any special order, are (1) Existence of a reliable glossary, (2) Proper MT engine training and resulting quality, (3) Absence of typographical errors in the ST, and (4) Using plain language in the ST. Notwithstanding, we have seen that Controlled Language (CL) is not widely used within the industry (see tables 7.1 and 7.2), despite the weight trainers tend to award to it (see table 7.4).

While the industry seems to have a very practical way of assessing if a candidate is a good fit for a job —subject field and Computer-Aided Translation (CAT) tools knowledge were the second and third-most valued criteria—, MTPE training courses tend to focus more on core PE skills, such as deciding when to edit or discard a segment. However, we can imagine how the third-most trained element (PE guidelines: exhaustive list and examples in the relevant language pair) may also include 'relevant domain' or 'relevant text-type' examples. Regardless of what degree of specialisation is included in MTPE training plans at LSCs, it is a positive result to see how they succeed at connecting their perception of the PE skill-set to their training programmes.

To provide an answer to **RQ2**, the practices in the industry seem to be less comprehensive than theorised by scholars: 13 average tasks as per educators submissions, 6.65 average tasks as per linguists submissions and 6.17 tasks as per LSCs' submissions, as argued in section 7.1.1. Regardless, the multi- and transdisciplinary character of the craft seems to be exacerbated with PE. The reason for this is the weight that the instrumental component of MT brings to the PE competence model. Depending on the profile of each post-editor, discussed in part I —segment validator or Project Manager (PM), according to Sánchez-Gijón (2016), or notary-like, following Pym (2019), or others—, and depending on their distribution of PE-related tasks, their skills and profile may become more or less hybrid[13], towards

[13]In other words, the plurality of the translation industry certainly increases how hybrid the post-editor profile may be, but the situation of each professional is different (Pym et al.

one discipline or another. What is certain is that additional and deeper skills in matters of Information and Communications Technology (ICT), error spotting and decision-making, Quality Assurance (QA), and terminology management are gaining interest, thus augmenting the multi- and transdisciplinary nature of the PE activity, in terms of Gambier (2006).

If we proceed by decomposing **RQ2** to its three subquestions, one could better grasp the implications of what has been discussed in the previous chapter for the PE service in the current language industry. The overall agreement between LSCs and linguists in terms of tasks (see figure 7.1) and hiring criteria (see figure 7.2) indicates that current practices investigated by RQ2.1 are clearly established. The profile of a professional translator is not "devolving" but rather "evolving" (Pym 2013) not only to that of a post-editor or even a "sophisticated" post-editor (Pym 2019), but to more varied profiles, as commented in chapter 3 citing Kornacki (2018, p. 37), and in Ginovart-Cid, Colominas, and Oliver (2020).

About RQ2.2, the three top PE skills have been identified: S1, S2, and S7, that is (without any particular order):

> S1[14] - Capacity to decide when to edit or discard (translating from scratch) an MT result.
>
> S2[15] - Capacity to post-edit according to PE guidelines
>
> S7[16] - Capacity to identify MT output errors

The coincidence of the three top PE skills in this study with AP3, CC7 and CC6 as being part of the top six skills found by Saint-André (2015) encourages the idea that drill and practice on these activities should reinforce and develop the PE competence of novice and intermediate linguists. Indeed, these three PE-related skills could represent a good foundation for the Efficient Editing Skill (EES) group commented in the chapter Final thoughts of part I. As a result, they are selected as the main tools to organise the split-technique training presented and discussed in part III, where each skill will have a dedicated hands-on exercise. In fact, they correspond perfectly with the skills listed by in section 4.1 by Absolon (2017, p. 23)[17].

Finally, regarding RQ2.3, despite the low quantity of LSCs providing training in PE, the average number of topics covered in such courses remains similar to the average number of elements chosen by the respondents

2013, p. 119).

[14] In line with the findings of Saint-André (2015) regarding CC7.

[15] In line with the findings of Saint-André (ibid.) regarding AP3.

[16] In line with the findings of Saint-André (ibid.) regarding CC5 and CC6.

[17] The 1st one was "Decision-making" and the 8th one was "Identification of errors", while the rest of them ("Comprehension of the topic" and all the split-technique MT errors) build up for what we retain here as "Post-edit according to PE guidelines".

to the questionnaire to trainers. On average, University educators in PE include 8.9 subject matters, while LSCs cover 7.94, and linguists reduce it, for their customer-provided PE courses, to 6.27 topics; as seen in the findings reported in tables 54, 129, and 170 in Ginovart-Cid (2020c). Taking into consideration the implications presented in the previous paragraphs, if we now consider the insights and knowledge gathered via the study of syllabi outlines and one-on-one interviews described in sections 7.2 and 7.3, respectively, we can draw well-informed conclusions about the state-of-the-art of the PE activity and the profile of the professional.

It could be suggested that the top PE skills presented in the previous chapter could be grouped into a more comprehensive competency model[18], and that the top training elements should be rethought in a manner so as to be introduced in translation curricula in parallel or just after the thematic knowledge and the revision courses begin. In fact, the qualitative data obtained in the above-mentioned sections points towards the trend of European universities making more efforts to embrace PE. It is noted how the PE courses are gaining presence: in the undergraduate studies, with more credits and hours, or by intertwining the practice with their translation courses, and not only in CAT or localisation courses.

While the MT providers sometimes have claimed that a good-quality MT output allows you to "retain" your post-editors (Thicke 2013), a professional post-editor should have the right tools and knowledge to decide which job s/he wants to accept and on what conditions. By designing the experiment presented in part III and its split-technique exercises on the three above-mentioned top PE skills, it is expected to provide a contribution in the PE pedagogy field with new insights regarding the French-Spanish language pair, the split-training technique in a real scenario within the language industry.

Unlike the study presented by Blagodarna (2019, p. 124), the training proposal designed in chapter 9 will have a language-dependent nature to allow for analysis and discussion of the EES in chapter 10. In terms of Massey (2018, p. 11), the following split-training course is set up to allow learning "with" translation technologies for PE (and not "about" translation technologies). Advanced linguistic skills, text-processing skills, and Translator-relevant Computer Competence (TCC) competences are not the object of the proposal in part III. As explained in chapter Final thoughts of part I, the Instrumental Input Skill (IIS) are assumed to be homogeneous amongst participants with the notion of practicality in mind (see section 9.3.2). Furthermore, as stated in section 4.1, the length and detail of a PE course is of

[18] As described by EMT Network (2017), for instance, where PE is considered a skill related to translation, and MT abilities are encompassed within the technological competence.

capital importance depending on its purpose. In the present dissertation, the industrial setting that frames the split-training brings with it the requirement of a relatively short plan.

Part III

Experimental Study

In this part, we describe the training proposal design[19], the selection of participants, the training provision, and the results about the Post-Editing (PE) performance of the 34 participants. In particular, chapter 8 enunciates the research question for this empirical study. Chapter 9 sets out the methodology of the pretest/treatment/posttest[20] study through the description of the participants, the materials for the training, and the data collection and data analysis methods applied.

In the framework of the Industrial Doctorate, the experimental design takes into consideration the needs of Datawords. In particular, the advantages for the firm in this experiment fall on the selection of the language pair, the selection of the texts (pretest and posttest), the groups of participants, and the selection of the tools and environment. Considering the profile of the participants, the language used throughout the instruments, documents and references is Spanish. The instruments described in chapter 9 and the data collected through them are reproduced in a publication that functions as appendix to the present research: «Replication Data for: The need for practice in the acquisition of the post-editing skill-set» (Ginovart-Cid 2021). The said appendix is organised as shown in figure 7.5.

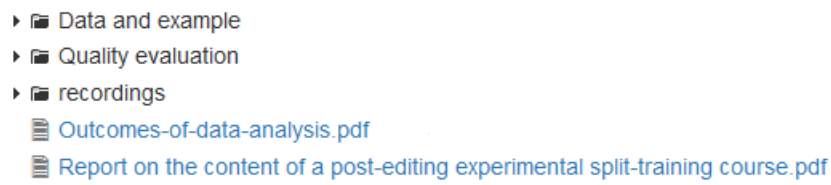

Figure 7.5: Tree of files in dataverse (Appendix)

The "recordings" folder and the `PDF` report titled "Report on the content of a post-editing experimental split-training course" are thoroughly described in chapter 9. In section 9.3, the folder "Data and example" is also introduced.

In chapter 10, we further explore the collected data, and we present the results along with the discussion. This chapter affects the folders "Data and example", "Quality evaluation", and the `PDF` titled "Outcomes-of-data-analysis". Some conclusive remarks are provided in a short chapter at the end of this part (Final thoughts).

[19]Post-editing Practice Application (PEPA), available in Ginovart-Cid (2021). The cited dataset functions as an appendix to the present part.

[20]Definition by Merriam Webster's Dictionary: "a test given to students after completion of an instructional program or segment and often used in conjunction with a pretest to measure their achievement and the effectiveness of the program."

Chapter 8

RESEARCH QUESTIONS

After having defined qualitivity in part I, and having confirmed in part II that current PE practices require, mainly, the skill of spotting errors, the skill of making (right and quick) decisions, and the skill of respecting PE guidelines, our **RQ3** is:

> To which extent does the split-training method on the cited three skills allow for increased qualitivity in the PE performance of French-Spanish semi-professionals?

The **RQ3** takes into consideration a language pair that is relevant for Datawords, as argued in the Introduction. It also represents the Machine Translation Post-editing (MTPE) expertise level discussed in part II, according to which post-editors and firms selling PE services stand currently in a sort of "intermediate" phase of the evolution of the craft. Hence, we consider "semi-professionals" (Master students or young active professionals) as a revelatory audience that could contribute with valuable insight to the experiment described in the next section.

Since it has been demonstrated that specialised basic training enhances the PE performance of novice translators not previously trained at PE in some language pairs, such as English-Russian and English-Spanish (Blagodarna 2019), and considering that, like Muzii (2016) argues, in-house PE training courses should be tailored to a number of factors (engine, client, etc.), three variables are analysed to assess the impact of the split-training proposal on PE qualitivity, as well as the ratio between two of them. In particular:

- ☐ Edits per Hour (EPH) (technical effort)
- ☐ Words per Hour (WPH) (temporal effort)
    - ☐ Ratio between the previous two by calculating the TAUS Efficiency Score (TES) (Ruopp 2015)
- ☐ Translation quality of the product through the Preselected Item Evaluation (PIE) method (see section 2.1.2)

Unlike Stasimioti and Sosoni (2019, p. 129), themselves citing Koponen, Salmi, and Nikulin (2019) and Koponen and Salmi (2017), the type of edit operations[1] are not annotated in the present study. Despite the fact that Edit Distance (ED) and time might not correlate well with cognitive effort (Aranberri and de Gibert 2019, p. 6), these are the metrics used alongside quality, for they are the ones valued within the industry as more cost-effective and accessible. Regarding the TES, it is defined as

> a composite indicator for productivity based on the words per hour measurement and the edit-distance scores: It can offer a rounded assessment of performance; It presents the 'big picture' and can be easier understood than trying to find an answer in the two (or more) other measurements; It can help for the implementation of better analytical methods and better quality data. (Attila Görög 2015b)

According to (Ruopp 2015),

> [t]he number of final edits done in the whole process of producing the translation (and calculated from the character-based edit distance) gives a more reliable productivity score.

However, as discussed in chapter 2, we claim that not only the final edits are relevant, but the totality of the edits done, undone and redone during the process must be considered. In that sense, the limitations and further work regarding TES are considered in Attila Görög (2015a, p. 12), namely: (1) to move from relative to absolute scores —as the 0 score for the less efficient participant of the group is discouraging—, (2) to eliminate outliers, and (3) to add other variables, such as quality assessment.

The language pair French-Spanish is selected for the experiment for several reasons. First, it is a common language pair at Datawords. Second, Machine Translation (MT) (thus, PE) is under-researched in language pairs not involving English. Finally, despite the existence of similar work on the mentioned variables in the field of PE pedagogy, MTPE varies in different language pairs (Koponen 2016a, pp. 134), and, to our best knowledge, this would be the first study for the French-Spanish language pair for Neural Machine Translation (NMT) in an industrial setting. We agree with Kornacki (2018) who, citing Kiraly (2000, p. 122), insists that the process for training translators should be based on real commissions to be representative of translation jobs in the market. Nonetheless, there is no real scenario (outside research) where the same PE job could be needed 34 times, and it is not feasible not to inform the participants of the main characteristics

[1] Also known as Post-editing Action (PEA).

and goals of the project they would be involved in. Hence, our proposal for pretest-treatment-posttest is rather a highly-autonomous simulated project. It has a realistic brief and authentic material (Buysschaert et al. 2018; Canfora 2016, p. 162).

To provide an answer to the enunciated **RQ3**, we depart from four hypotheses that test the variables described above before and after the treatment.

> Hypothesis 1
>
> H0 = The temporal effort, measured in WPH, does not vary significantly before and after the split-training
>
> H1 = The temporal effort, measured in WPH, varies significantly before and after the split-training
>
> Hypothesis 2
>
> H0 = The EPH rate does not vary significantly before and after the split-training
>
> H1 = The EPH rate varies significantly before and after the split-training
>
> Hypothesis 3
>
> H0 = The TES does not vary significantly before and after the split-training
>
> H1 = The TES varies significantly before and after the split-training
>
> Hypothesis 4
>
> H0 = The PIE score does not vary significantly before and after the split-training
>
> H1 = The PIE score before the split-training varies significantly before and after the split-training

Chapter 9

METHODOLOGY

With **RQ3** established, the method to provide an empirical answer to it must be also defined. The methods used so far in PE process research have been summarised by some authors, for instance, Guerberof Arenas (2019, p. 339), citing O'Brien and Simard (2014). They include screen recording, keyboard logging, eye-tracking, Think Aloud Protocol, interviews and questionnaires, etc. Furthermore, some PE tools (such as PET[1], TransLog II[2], TransCenter[3], and PosEdiOn[4])[5] have been developed to specifically address the needs in PE research. In an Industrial Doctorate, the realistic environment is of crucial importance. Hence, the selection of the semi-professional translators (see section 9.1); the design of the materials used for the training provision (see section 9.2); and methodology choices in terms of tools for data collection and analysis (see section 9.3) are performed with the upmost concern of preserving the usual and familiar conditions and scenarios in the everyday PE jobs of the participants.

Considering how the processes in the translation industry sometimes are constrained by the short notice in which a client submits a translation request and the final, non-negotiable, deadline, the split-training methodology described in chapter 4 goes in line with the "learning by tasks methodology" (Orozco and Hurtado Albir 2002, p 390). For a complete description of this methodology the reader can refer to publications by Hurtado (1999) and Hurtado Albir (1996). A similar approach is found in Yang and Wang (2020, p 8), who highlights Pietrzak's (2018) statements on how the acquisition of the proper strategies in PE can be completed "over the course of professional practice".

Considering the PE performance indicators and RQ commented in chap-

[1]https://github.com/wilkeraziz/PET

[2]https://sites.google.com/site/centretranslationinnovation/translog-ii

[3]https://github.com/mjdenkowski/transcenter

[4]https://github.com/aoliverg/PosEdiOn (Oliver et al. 2020).

[5]Some of the cited tools are further described by Denkowski and Lavie (2012), Denkowski, Lavie, et al. (2014), and Vieira (2013).

ter 8, it is considered that the pretest-posttest study design is the most adapted methodology for the present research. Like in Dede (2019), Orozco and Hurtado Albir (2002, p. 389), and Zhang and Torres-Hostench (2019) we apply one treatment (exposure to online PE split-training) to evaluate if the performance before and after the said treatment varies significantly, positively or negatively. Authors like Kornacki (2018, p. 15) also conducted experimental research on how trainee translators acquire technology-related skills —in his case, about Computer-Aided Translation (CAT) tools— with 22 subjects in 2015/2016 and 18 participants in 2016/2017, also with a pretest-posttest dynamic.

The design of this research is considered "pre-experimental" by J. W. Creswell and J. D. Creswell (2017, p. 168), since we do not use a control group, that is, a set of participants performing the pretest and posttest without training provision (treatment) in-between. Other studies, like the one by Orozco and Hurtado Albir (2002), have been performed without control group in the past. For the present research, the choice of not having a control group was made after some peer-reviews suggested that not receiving the treatment intrinsically would lead to lowering the performance of the control group. Moreover, the availability of French-Spanish semi-professionals is also a limiting factor to the creation of a controlled group. Hence, **RQ3** clearly establishes "to what extent" and not "if" PE performance is improved or not.

In terms of duration, as announced in Final thoughts of part II, the training course should represent between 7 to 10 hours to fulfill the practicality requirement of a test (a notion that will be presented in section 9.3.2). The first purpose of the test being to have empirical insights, the announced length should suffice to provide the necessary data. On the other hand, the second purpose of the split-training is to "put up to speed" semi-professional translators in a given PE project. In that sense, it must remain within reasonable duration. On top of that, the setting (Industrial Doctorate) also tends to require a relatively short investment in terms of time, since it is not feasible to stop the production of the whole translation team during too many consecutive hours.

9.1. Participants

The selection of participants is performed by convenience sampling. Three groups are pre-identified on the basis of accessibility to the researcher: Datawords' employees in the Spanish team of project managers and translators; Datawords' freelance database; and Master students at Geneva University. To establish a minimum number of participants, we consider that

statisticians have found that it takes a sample size of about 30 participants to fulfill the assumption that the sample represents a population (Salkind 2017, p. 387). To recruit the participants, a Google form[6] was created and submitted directly to the three above-mentioned groups. The form can be consulted in «Replication Data for: The need for practice in the acquisition of the post-editing skill-set» (Ginovart-Cid 2021)[7].

To make sure the aimed minimum number of 30 participants could be attained, as well as to have reserve if one participant could not complete the study, the recruitment was also extended by probabilistic sampling in platforms such as LinkedIn, and to 3rd and 4th year undergraduate students at Pompeu Fabra University. The form was available in Spanish and it was filled out and submitted between December 2019 and February 2020 by a total of 56 persons. Out of those, 35 were selected. The 21 remaining persons who submitted the form were discarded either on the grounds of not fulfilling the minimal criteria (semi-professional translator[8], French-Spanish language pair, Spanish as native language, not excessive PE experience according to the Advanced and Expert expertise levels commented in part II), either on the grounds of lack of further response to the e-mail subsequently sent.

The final available number of valid participants is 34, for one had two technical errors at the posttest step: one with Qualitivity and one with Flashback Recorder. As a consequence, the qualitivity report recording ED and time, and the screen recording were unavailable. Hence, the contribution of this participant is discarded. We recorded self-reported proficiency in languages but did not ask the participants to take a test. The data gathered about the 34 retained participants show the homogeneity throughout the profiles selected, except for one or two participants in some of the questions. Fifteen are in-house translation project managers and translators at Datawords, eleven are Master students at Geneva University, and the remaining eight are freelance translators. In particular:

- ☐ 11 students, 12 professional translators, and 11 have both profiles.
- ☐ 11 (32%) are 25 years old or younger, 17 (50%) are between 26 and 30 years old, and 6 (18%) are between 31 and 38 years old.
- ☐ 11 (32%) are male, and 23 (68%) are female.
- ☐ 6 have followed some sort of PE training course[9], and the rest of

[6] http://xl8.link/Recruit-gform

[7] Page 6 of file *Report on the content of a post-editing experimental split-training course.pdf*.

[8] Out of the 21, two senior Country Managers from Datawords, with more experience, were asked to perform the pilot study described in section 9.4.

[9] According to their descriptions, these courses were short or specialised, in terms of language pair (English into Spanish) or in terms of audience (for project managers, for in-

participants (28) have never followed any PE course.

- ☐ 28 have never post-edited a Spanish text machine-translated from French; three have post-edited some volume between 1,000 and 6,000 words; two have post-edited more than 7000 words; and one does not know.
- ☐ 30 have less than one year experience accepting MTPE projects; and the rest (4) between 1 and 5.
- ☐ For 30 participants, MTPE projects represent 25% or less of their workload; for 3, it is between 26% and 50%; and for 1 it is more than 76%. Hence, the distribution in terms of the MTPE expertise level introduced in section 7.1[10] consists of: 26 novice participants, 7 intermediate participants, and 1 advanced participant[11].
- ☐ 33 have Spanish as mother tongue; and 1 has Catalan.
- ☐ 31 have a University degree in Translation Studies (TS).
- ☐ 33 selected SDL Trados Studio as one of the CAT tools they are most used to in a multiple-choice question with a maximum of two options. The other selected tools were, in decreasing order, MemoQ, OmegaT[12], Memsource, Wordfast[13], Matecat[14], eLuna[15], and Wezen[16].
- ☐ The 34 participants have the French-Spanish language combination in their professional profile.
- ☐ About the knowledge of French colloquialisms and spoken register, on a Likert scale[17], two participants say it is "1"; two chose the value "2"; ten say it is "3"; eleven chose "4"; and nine chose "5".
- ☐ About their knowledge of the motorcycling domain, on a Likert scale[18], nineteen participants chose "1"; ten chose "2"; four chose "3"; and one, "4".

A well-known problem, intrinsic to translation empirical studies, is the variation between subjects, as highlighted by Koponen (2016a, p 136) and Guer-

---

stance).

[10]More thoroughly analysed in Ginovart-Cid, Colominas, and Oliver (2020).

[11]It should be noted that, when asked, the participant who chose "Más del 76%" as answer to the question "¿Qué porcentaje de tus proyectos de traducción tratas con posedición de traducción automática?" replied that the customer usually enables MT as an interactive resource (the target segments are not filled with the MT output). It is, in that case, MT-aided translation, instead of PE as a service.

[12]https://omegat.org/

[13]https://www.wordfast.net/

[14]https://www.matecat.com/

[15]https://elunaguide.wordpress.com/accessing-eluna/

[16]https://www.wezen.com/translate

[17]From 1 (very low) to 5 (very high).

[18]From 1 (very low) to 5 (very high).

berof Arenas (2019, p. 348). It is hoped that the risks of such variability are reduced to a certain extent thanks to the homogeneity attained in the group described above. The 34 selected participants were sent an e-mail with a 2-page consent statement were the nature and purposes of the study were described in detail. The signed consent form of each participant is collected, at the latest, the day when the warm-up activity is performed. The text is available in Spanish and can be consulted in Ginovart-Cid (2021)[19].

9.2. Material

The material described in the next three sections, including the cited figures and tables, can be consulted in the above-mentioned appendix: «Replication Data for: The need for practice in the acquisition of the post-editing skill-set» (ibid.), in particular, in file *Report on the content of a post-editing experimental split-trainingcourse.pdf*.

9.2.1. Selection of texts

The texts for the pretest and the posttest were purposefully selected from a client of Datawords. This client, Motoblouz, has been briefly presented in the Introduction. Motoblouz requested MTPE as a service in 2014, and since then it remains the client with the longest tradition for the PE service at Datawords. The PE services for Motoblouz often concerned product descriptions from the customer's e-commerce website. In 2018, the client requested the same service for a new text typology. In fact, a new blog was created in their website. Thus, PE was requested on blog articles and entries, which used a more creative style of writing[20].

With the purpose of enabling the MTPE workflow on such text type in the future, as NMT keeps improving, the solution of empowering the post-editors with the right skills is considered of capital importance, and to be accomplished within the project's time constraints. In this experimental setup, the texts used are extracted from Motoblouz's blog. The MT output used at the pretest (see table 1 in pp. 26-31) and the posttest (see table 4 in pp. 109-114) have the characteristics presented in the following tables (9.1 and 9.2), in terms of volumes in words, segments, and characters. In file *Dataset.xlsx* of Ginovart-Cid (ibid.), the volume based on characters is retained for columns F to I.

[19]Page 12 of file *Report on the content of a post-editing experimental split-training course.pdf*.

[20]Including Negative Translatability Indicators (NTI) items such as jokes, metaphors, informal register, colloquialisms, etc.

Table 9.1: DeepL's output (pretest and posttest) in words

Text	# segm	Words	Words segm1	Words last segm	Final # segm	Final # words	Av w/segm
Pretest	78	1250	8	6	76	1236	16.26
Posttest	76	1345	10	12	74	1323	17.88

Table 9.2: DeepL's output (pretest and posttest) in characters

Text	# segm	Char	Char segm1	Char last segm	Final # segm	Final # char	Av c/segm
Pretest	78	7590	55	43	76	7492	98.58
Posttest	76	7995	66	88	74	7841	105.96

The reason behind the length of the selected excerpts is twofold. First, it is considered a medium-length commission for some segments at Datawords, especially for Motoblouz. Second, authors like Huertas Barros and Vine (2018, p. 16) indicate 2500 words as a limit for core translation modules. The sentences used for the exercise on error identification are extracted from the text used at the pretest; they can be consulted in the quiz of figure 22 in the cited appendix. The sentences used for the exercise on decision-making (when to edit or delete and translate from scratch) are also extracted from the text used at the pretest; they are available in the quiz of figure 25 in the cited appendix. The sentences used for the exercise 3.2 (see table 2) are also extracted from the text used at the pretest. The corresponding quiz is found in figure 30.

Finally, for exercise 3.3 a new text of the same blog is selected. It can be consulted in table 3, and its related quiz is found in figure 31. For more context, it is reproduced in its original layout in figure 32. The already mentioned preliminary study (Ginovart-Cid 2018) explored the quality of customised engines versus DeepL, as well as the contrast between two different text typologies: the product description and the blog article. For the present study, the blog article is selected as text type due to two reasons. First, a blog has a certain cultural appeal but in an informational tone, perfect for semi-professionals entering the translation industry, and not yet proficient in a specific domain. Second, the relation it bears to the commercial website and the selling purpose leads to a marketing style of writing that has proved difficult, even for NMT.

In fact, in Ginovart-Cid (ibid.), we argue that the blog article is still too challenging a text type for stock engines, such as DeepL. It is, however, the only option many Language Service Companies (LSCs) and semi-

professionals can have easy or free access to. Hence, the full PE service using DeepL on such text type is a real need for the language industry and future professionals. To conclude, in the same way that Guha and Heger (2014) and other publications did in the past (e.g. Läubli et al. 2013, p. 84), the above-cited proceedings article (Ginovart-Cid 2018) focused on deploying MT. At the light of the results, the present dissertation focuses on how to acquire PE skills for on-boarding collaborators once MT has already been deployed within the listed conditions. That is, a stock MT engine applied to creative-persuasive texts in an "English-free" language pair. In terms of text typology, our choice is similar to the content selected by Läubli et al. (2013, p. 83): "marketing texts from the automobile industry".

9.2.2. Selection of tools and environment

In terms of technology, the selected CAT tool is SDL Trados Studio. In words of Guerberof Arenas (2019, p. 342):

> SDL Trados [...] [is] also used for research, especially if the objective is to have a working environment as close as possible to that of a professional post-editor.

As in O'Curran (2014, p. 113), it is allowed that participants revisit segments like in a real scenario. Indeed, Studio is widely-used both within Datawords and throughout the three groups of participants. Studio allows for a plug-in called Qualitivity[21]. This plug-in has been briefly introduced in section 1.3, and it has the advantage of integrating a widely-used CAT tool quite seamlessly. In practice, it tracks the time (in seconds and milliseconds) each participant spends post-editing each segment, as well as the keystrokes and ED according to the Damerau-Levenshtein metric. Another advantage of this plug-in is that it records every PEA, not only the final output. In other words, it records every editing operation, even the ones that later in the process are undone or modified again[22]. For instance, Attila Görög (2015b) use the Wagner & Fischer algorithm (after the translation process) to calculate the Levenshtein metric in order to produce the TES described above.

The source texts[23] described in the previous section are imported to Studio in a `TXT` format to counter the limitation that the Information and Communications Technology (ICT) skills and other instrumental abilities of the participants are not tested in advance and are assumed to be homogeneous.

[21] Andrew (2018) and Andrew and Filkin (2018).

[22] The fact that all the edits during the process are considered is more accurate in terms of technical effort, as commented in part I when reviewing the literature, especially about the Actual Edit Rate (AER).

[23] Source Text (ST).

The TXT format implies the absence of HTML, XML or formatting tags in the CAT environment, thus reducing the potential variability between participants in terms of technical user perspective (shortcuts, advanced filtering or QA, etc.). Regarding the MT engine, our choice is DeepL. It is considered the top NMT free provider in the web in 2019 (see sections 7.1.2 and 7.1.3). As a Software as a Service (SaaS) solution, this external engine presents a very low total cost of ownership, simple maintenance, limited staffing needs, and a low impact on revenues, as observed by Muzii (2016). Moreover, as seen in section 7.3, for PE training courses at University, it is the one of the mainstream providers too.

Even if eye-tracking is not a technique used in the present thesis, the participants are asked to record their screen. The reason for this is to have qualitative insights (if needed) on some edit operations that are deduced from the Qualitivity report and the quantitative data. Furthermore, it is considered as a back-up to reproduce the exercise of a participant easily, should the plug-in or the CAT tool fail saving at any moment. For this purpose, the selected tool, upon some recommendations of fellow researchers, is the free version of Flashback Express 5 by BlueBerry Studios[24] (unless the participant has another preference), which is also used in the work of Jia et al. (2019, p. 65) and Kornacki (2018, p. 161).

About the training provision, the Google Classroom[25] environment is chosen due to the ease of access and the flexibility it allows for the two kinds of users. Students see the supports and material at the date and moment chosen by the teacher, and the deadline is also clearly displayed. The deliveries of each assignment can be directly performed in the platform, which instantly notifies the trainer. Google Classroom allows for public or private comments and discussion related or independent of each subject matter, and it accepts different formats (such as attachments, files in Google Drive[26], URLs or Youtube videos[27], Google Forms[28], etc.).

At the beginning of this chapter 9 and in the Introduction, it has been highlighted how the nature of the PhD, an Industrial Doctorate, has been a decisive criterion to steer the choices of the tools, environments, and material of this experimental study. Like Läubli et al. (2013, abstract), we are convinced that PE experiments should be carried out on realistic (if not real) translation environments to make them ecologically valid. As a matter of fact, the three tools discussed in previous paragraphs are chosen with the aim of preserving the everyday way of working of participants. Even if,

[24] https://www.flashbackrecorder.com/express/
[25] https://classroom.google.com
[26] https://drive.google.com
[27] https://youtube.com
[28] https://www.google.com/forms/about

for some, Qualitivity, Flashback Recorder, or Google Classroom may have been initially unknown, these tools are deemed to be less intrusive than other software that would require a more thorough learning process.

Since the participants are accustomed to SDL Trados Studio (as per their submitted form seen in subsection 9.1), the Qualitivity plug-in allows us to obtain the required quantitative data without having a direct impact on the usual working methodology of the participants. Nonetheless, to prepare the participants to the few changes in their usual working environment, a warm-up activity is prepared before the study. It is described in next section. Authors like Kornacki (2018, p. 154) consider the stress factor as a relevant criterion for the environment of an experimental setup. We follow the work of Läubli et al. (2013, p. 89) in that no time pressure is applied to participants[29]. Unlike the cited work, we did ask for adherence to a very short client style guide, which is presented in next section.

Kornacki (2018) researched the CAT skills acquisition process and observed that "[a]ny successful training requires [a] correct approach, grounded in the most up-to-date theories and frameworks" (p.23). Moreover, the author (p.59), quoting C. Way (2000, p. 132), highlights that translation courses should be goal-oriented, that is, we should

> train students to reach a level which will allow them to join the professional market as novices and once there, increase their specialisation depending on the field in which they find employment.

The author argues that a digital class can be difficult to organise, since the teacher must incorporate tools such as a Learning Management System[30] (p.67) but concludes that "it is critical to initiate students into the computer-based classroom and allow them to develop all the hardware and software skills that will enable them to enter the market after graduating successfully" (p.71). The next section reports how these two requirements are taken into account during the process of designing the course materials.

9.2.3. Design of documents

It has been commented in previous sections how the split-training methodology is similar to the approaches of "learning by doing". The course materials described in the following paragraphs, are thought to be "scaffolded problem-solving activities where [students] can practice the applica-

[29] It is considered, though, that they all had some degree of pressure or stress derived from the test situation, as well as other obligations scheduled after the pretest and the posttest sessions, whose estimate duration had been communicated in the consent form.

[30] In this case, Google Classroom, as presented before.

tion of the basic skills to realistic situations" (Kiraly and Piotrowska 2014). Google Drive is the application used to upload or link the materials to the relevant subject matter in Google Classroom. The course is named PEPA. The first text that appears when entering PEPA is a short description. The description is displayed in Spanish and it reads as follows (Ginovart-Cid 2020a):

> *En este curso, vas a practicar tres habilidades clave para poseditar resultados de motores de traducción automática (TA):*
>
> - *identificar errores de TA*
> - *decidir qué segmentos poseditar y qué segmentos descartar y traducir desde cero*
> - *respetar directrices de posedición (PE) completa*

The course can be consulted in the cited appendix. Nevertheless, below we provide a summary of its structure and the main goal of the nuclear parts.

- ☐ Foro de preguntas
- ☐ Etapa previa
  - ☐ Tu experiencia previa en PE
    - Goal: To recruit apt participants for the experiment; consider the suitability of the experiment design to their profile.
  - ☐ Consent form[31]
    - Goal: To inform the participants of the different phases and tasks of the study; to inform them of their rights; and to receive their written consent to treat and publish the outcomes.
  - ☐ Warm-up
    - Goal: To present to the participants the environment and the tools, namely Google Classrooms, SDL Trados Studio, Qualitivity, Flashback Express Recorder; and solve any technical issues or answer doubts before the task.
- ☐ Pretest
  - Goal: To collect the data from the resulting Qualitivity report; the screen recordings; and the post-edited text.
- ☐ Introducción al curso
  - Goal: To give a general overview of the goals and content of the online training.

[31] The Google form ("Tu experiencia previa en PE") and the consent form have been described in section 9.1. The form had been sent by e-mail and the consent form was collected the day when the warm-up activity was performed. Hence, these two documents only appeared in PEPA as reference for the participants, since they had already been read and signed.

- ☐ 1. Identificación de errores
  - ☐ Introducción al tema 1
  - ☐ Quiz 1 - identificación de errores
    - Goals: To present the skill of identifying errors in an MT output in the context of post-editing from French into Spanish; and to revisit the text used at the pretest as an exercise to produce explanations about the errors that could be identified.
- ☐ 2. Toma de decisiones
  - ☐ Introducción al tema 2
  - ☐ Quiz 2 - toma de decisiones
    - Goals: To present the skill of decision-making in the context of post-editing, namely when to edit or discard an MT output; and to revisit the pretest extract as an exercise to suggest thinking about which MT output may need to be accepted or rather discarded completely, and translated from scratch.
- ☐ 3. Directrices para poseditar
  - ☐ Introducción a la PE completa
    - Goal: To present the concept of full post-editing with general examples (out of the domain of motorcycling entertainment and marketing).
  - ☐ Quiz 3.2 - ejercicio práctico
  - ☐ Quiz 3.3 - ejercicio práctico
  - Goals: To reflect on which full PE guidelines should or could be applied to a specialised MT output, and produce explanations of the corresponding choices; and to drill the full PE activity to develop the skill of applying guidelines.
- ☐ 4. Revisitamos el pretest
  - ☐ QUIZ - 4. ¿Cómo me fue en el pretest?
    - Goal: To provide for the possibility of reviewing one's own delivery at the pretest, while physical adjustments for lockdown are carried out by all participants due to the Covid-19 pandemic.
- ☐ Postest
  - Goal: To collect the data from the resulting Qualitivity report; the screen recordings; and the post-edited text.
- ☐ Etapa final
  - ☐ Cuestionario retrospectivo
    - Goal: To collect feedback regarding the experience about the pretest-posttest study, and about the split-training course.

The protocol for the warm-up entails reading the consent form, and answering and commenting any doubts and questions the participants may have on the study. The warm-up is organised as a presentation of the Google Classroom as a platform and of PEPA as a course. It concludes by performing an exercise with a small Trados package. It ensured that the participants install Qualitivity and it allowed them to become familiar with the task (later for the pretest and posttest). For instance, they make sure the termbase is displayed and, for the less familiar with Studio, that they remember or know how to create a return package. The instruments used for the warm-up can be consulted in page 15 of the appendix cited at the beginning of the section 9.2. Finally, the participants would deliver to the researcher the said return package along with the Qualitivity report and the screen recording. The warm-up allowed to solve any technical issues (for instance, establish best practices about how to transfer heavy files), and to answer any doubts or questions before the date of the pretest. Upon completion of the warm-up activity, the participants received an e-mail summarising the steps performed together (figure 11 of the appendix).

The PE guidelines are designed by considering TAUS PE guidelines (Massardo et al. 2016), the standard *ISO 18587:2017* (p.8), and previous research discussed in section 2.2.2 in part I. They can be consulted in page 24 of the cited appendix. They include a short style guide and they are used for the pretest, the exercises in lesson 3, and the posttest. A colour code is used to show the activation/deactivation of each guideline according to the requested quality level. Since it is full PE for this experiment, they are all green.

In terms of instructions for the experiment, a part from the PE guidelines, both the pretest (p.24) and the posttest (p.108) are presented with a set of instructions. Furthermore, they are framed within an explicit assignment/brief/commission (p.31 for the pretest and p.114 for the posttest). They are discussed in next section (9.2.3). The three lessons of the course are created according to the three core PE skills identified in the surveys: error identification on NMT output (lesson 1); decision-making about when to edit or discard and translate from scratch (lesson 2); and application of PE guidelines (lesson 3). They are discussed, along with the unplanned lesson 4, in subsection 9.2.3.

Pretest and Posttest

There is a file with instructions[32] created for the pretest and one for the posttest. They are a reminder of what has been presented in the consent form

[32] Available in pages 24 and 108, as indicated before.

and during the warm-up activity. The briefs include contextual information of the MTPE project, as would be given by a client to an any LSCs, or from a Sales department to the Project Manager (PM)[33].

While this contributes to the realistic scenario where different pieces of more or less matching information come from different sources, the PE guidelines and style guided are found in one unique document to ease the lookup of such information by the participants. The style guide is a set of 6 localisation instructions that are given by the requester (customer) in a more or less structured form. Finally, the reference files in `PDF` are provided in case the visual cotext and context of the extracts may be helpful to understand the French ST; they can be consulted in pages 32 and 115, for the pretest and the posttest respectively. It should be noted that, due to Covid-19, the lapse between the pretest and the posttest was longer than expected. This is why the memo in p.106 was created and sent to the participants.

PEPA in Google Classroom

Lesson 1, about error identification, is composed of an introductory YouTube video and a quiz (p.40). The errors have to be copied and pasted from the NMT output given to the participant to the free text entry for each sentence. If there is no error they are instructed to enter number "0" and if the error is an omission the participants should copy and paste the previous token to the omission. As these are complex guidelines for a Google form, it is both written at the top of the quiz and also showed in the above-mentioned video.

Lesson 2, about decision-making, is composed of an introductory YouTube video and a quiz (p.47). The student has to decide whether to edit the output or translate from scratch. To steer the decision, the instruction given to the participants is that they should think of a full PE assignment and that "from scratch" would mean deleting the whole segment (output of DeepL).

Lesson 3 is about applying PE guidelines. It is the longest one, and it contains, as usual, a YouTube video introduction. It also includes a presentation in `PDF` with general examples (not of the motorcycling domain), and two assignments to exercise within Trados Studio (with the relevant references of the ST in `PDF` and the Studio package). The six documents can be consulted in page 56 of the appendix. On this occasion, the type of answer is not a free-text field, nor a radio button, but a checkbox list with the PE guidelines presented above. For this reason, several correct answers cannot be anticipated and the only mean to acknowledge subjectivity is via the

[33] Available in pages 31 and 114, as indicated before.

comments presented to the participant at the end of the quiz (p.134 of the cited appendix).

The feedback that was displayed (for correct, incorrect, and undefined/subjective[34] answers) for lessons 1, 2, and 3 can be consulted in page 134 of the cited appendix. Due to Covid-19 outbreak and the subsequent pandemic, some participants had completed the lesson 3 while others were still relocating to adapt to the lockdown and working remotely. Thus, a new exercise is created to shorten the lapse between ending the split-training in Google Classroom and performing the posttest at a prorogued date: lesson 4. Lesson 4 is a quiz that consists in peer-reviewing some selected translations that fellow participants[35] had delivered at the pretest. The quiz and the document that is subsequently shared as feedback with the 34 participants can be consulted in page 96 of the cited appendix.

Closure of PE experiment

Upon completion of the posttest, the participants receive a certificate, which should encourage their active participation to the experimental course. It can be consulted in page 126 of the cited appendix. The participants are asked to fill out one last questionnaire with their opinion about the pretest and posttest, and the split-training online course. The form can be consulted in page 128 of the appendix. In next section, the methods used during the pretest and posttest to collect the data and its subsequent analysis are described. In section 9.4, the instruments and protocols described in the present chapter are applied with two participants and the lessons learned are commented and used to adapt our instruments and methods accordingly.

9.3. Data collection and analysis

Starting Flashback and opening the Trados package marked the start of the assignment. Hence, as commented in previous section, the quantitative data collection is done by the plug-in Qualitivity and Flashback serves as a back-up source of such recorded data. The first and last segments are excluded at the data analysis stage. For the former one, the participants were still opening the instructions and reference files described in section 9.2 in their working station, and preparing the task. For the latter, some participants either asked the researcher how to stop the Qualitivity, or Flashback, or in what order they should do so; either they asked for guidance regard-

[34]For non-binary errors.

[35]The exercise is anonymous, i.e. the participants do not know whose delivery they evaluate.

ing the creation of a return package in Trados Studio[36], which also added seconds to the last segment that were not spent on the proper PE task.

Regarding the analysis of the data, the basic introductory references are online training courses and tutorials, such as *Data Analysis Tools* by Wesleyan University (2020). A very helpful reference used to select the most appropriate analysis tools is *Statistics for People Who (Think They) Hate Statistics* Salkind (2017). To guarantee a sound data analysis, the Servei d'Estadística Aplicada (SEA) (mentioned in part II) is contacted again to request guidance in the statistical analysis of the collected data. The SEA produces an explanatory report on the pilot study guiding and confirming the data analysis tools described in next section.

9.3.1. WPH (temporal effort) and EPH (technical effort)

As commented in section 9.2, the time and keystrokes are recorded with Qualitivity as it is an unobtrusive tool and it does not require the participants to track and report times, as in Läubli et al. (2013, p. 89) and Plitt and Masselot (2010), or even to click on any other buttons as would be required in tools purposefully designed for PE research. As commented in section 2.1.1, there are a number of publications that found discrepancies in the correlation between ED-based metrics and time (e.g. WPH). For instance, Koponen (2016a, p. 23) notes that

> the correlation between these edit distance metrics and time or human evaluations has been claimed to be good, they do not always accurately reflect these indicators of effort.

In her study of correlation between temporal and technical effort metrics, Tatsumi (2010) found that these two measures do not always correspond well. Nonetheless, ED remains the main measure for technical effort in the field, and also an indirect measure of cognitive effort (Lacruz and Shreve 2014, p. 253). The Qualitivity plug-in allows for an exported report in Ms Excel. An example can be consulted in file *Example-Qualitivity.xlsx* in «Replication Data for: The need for practice in the acquisition of the post-editing skill-set», particularly the tab of the spreadsheet named "Document Activities". In the resulting report for each participant, we work on the following columns:

- Active Seconds (Column V in the cited file)
- Edit Distance (Column Y in the cited file)

The following averages are calculated for the indicators discussed in previous paragraphs, by excluding the first and last segments:

[36] In particular, the group at Geneva University, who required the access to a virtual machine.

- Words per Second (WPS), then WPH: by dividing the total volume of words (see section 9.2) by the total of seconds in column "Active Seconds". To obtain the WPH: $WPS * 3600$.
- Average ED per document: by dividing the aggregated editing distance in column "Edit distance" by the number of last total characters of the MT output. Thus, revisited segments such as segment 1 in the example provided are considered for the ED, while they are excluded of the last total character count.
- EPH: $(WPH * ED)/words\,MT\,output$

For the analysis of the data, we use the Analysis ToolPak complement in Ms Excel[37] and IBM SPSS Statistics[38]. To establish which data analysis method is most appropriate to our sample, we produce Descriptive Statistics for the data. The normality of the data must be checked first. In that sense, IBM SPSS is used to obtain the significance (p-value) of the Shapiro-Wilk test for the WPH, ED and EPH, and quality variables. If the variables are not normally distributed, the non-parametric test of Wilcoxon must be applied. On the contrary, if the variable is normally distributed, the analysis to be applied is the parametric t-Test: Paired Two Sample for Means. The results and discussion are provided in next chapter (10).

9.3.2. Quality: the PIE method

The study in Mariana et al. (2015) evaluates the viability of the Multidimensional Quality Metrics (MQM) rating model to be applied to a well-known certification in translation, as discussed in section 2.1.2. The viability is defined by the authors as the combination of three measures: practicality, reliability and validity. Following Bachman and Palmer (1996, p. 36), Mariana et al. (2015, p. 147) establishes that

> practicality concerns whether the test to determine translation quality can be created and implemented within the constraints of the test designer's given resources.

As the authors, we base it on time cost. In fact, after the study accomplished in Ginovart-Cid (2018), the investment in terms of time to use the MQM framework in Dynamic Quality Framework (DQF) for the quality evaluation is considered too high. In the same vein, the complement that logs keystrokes in Qualitivity is excluded from our analysis.

The PIE (Kockaert and Segers 2014; Kockaert and Segers 2017; Segers and Kockaert 2016), discussed in section 2.1.2 in Part I, seems more cost-effective for the purpose of the present thesis. It is the selected method

[37] Already used for part II.

[38] https://www.ibm.com/products/spss-statistics

to evaluate the quality of the delivered texts, for it is considered to be an adequate procedure for the assessment of translation quality (Van Egdom, Verplaetse, et al. 2019). The PIE method puts the brief in the centre of the assignment (ibid., p. 36), which is capital in a real scenario such as the one at Datawords, considering the Industrial Doctorate framework. In addition, the PIE method may relate test items to learning outcomes, like the PACTE group aimed at relating items to translation competences (ibid., p. 33)[39]. In Van Egdom, Verplaetse, et al. (ibid., p. 47) the main strengths and weaknesses of the PIE method are described.

In particular, the PIE method is considered a variant of Calibration of Dichotomous Items (CDI), which contributes to its reliability. In the present research, the CDI method is also used. Indeed, complementing PIE with CDI is considered an objective way of testing translations and translators, for the selection of the items is not performed by the evaluator (Van Egdom, Verplaetse, et al. 2019, p. 34; Kockaert and Segers 2017, pp. 150–155). The PIE method is presented via an individual conference call to the external raters. The SPSS syntax developed by Lei and Wu (2007)[40] is used to obtain the *p*-value of each item. Thanks to it, 10 items are retained for the evaluation of the pretest and posttest. The *p*-value is indicative of item difficulty and is displayed as a number between 0 and 1. The selected items' *p*-value should lie between 0.27 and 0.79 (Van Egdom, Verplaetse, et al. 2019, p. 42). Indeed, as Lei and Wu (2007, p. 4) puts it:

> Mid range *p*-values (around .50) are desired for norm-referenced tests because they tend to produce larger score variances which in turn lead to higher score reliability estimates.

The above-mentioned SPSS syntax is also applied to obtain the D-index, which reflects the discriminating power of the item.

> The D-index ranges from -1 to 1 with a higher value indicating stronger discriminating power. According to Ebel and Frisbie (1991), an item that has a D-index of .4 or higher is satisfactory, between .30 and .39 requires little or no revision [...]. (ibid., p. 5)

Regarding the pretest, the *p*-value and D-index of the retained 10 items are displayed in page 3 of file *Cronbach-pvalue-dindex.docx* in Ginovart-Cid (2021).

[39]The observation that "[l]earning objectives become far more difficult to specify because they will differ from student to student" by Kiraly (2015) seems important to note here.

[40]The author is contacted because the URL suen.ed.psu.edu/~pwlei/plei.htm is invalid. Dr. Pui-Wa Lei kindly provides the updated link: https://sites.google.com/view/cttitem/homepage

Regarding the posttest, the *p*-value and D-index of the retained 10 items are displayed in page 4 of file *Cronbach-pvalue-dindex.docx* in Ginovart-Cid (2021). As in Eyckmans and Anckaert (2017, p. 49), on the basis of the dataset presented in the following chapter, the discriminatory power of the items is not established using the extreme-group method. Instead, we apply the "more accurate corrected item–total correlations (r_{it} value) were calculated by means of SPSS". Eyckmans and Anckaert (ibid., p. 49) retained items with a discrimination coefficient above .30 following Kockaert and Segers (2014, p. 246). In the pretest, there are two items with very low discriminatory power (D-index = .29), that are nevertheless retained. They are items 1 and 6.

> [T]he Cronbach's Alpha [...] is an indicator of how much the overall internal consistency reliability would increase or decrease if a specific item were deleted from the test. (Chien et al. 2014, p. 43)

The Cronbach Alpha for the pretest is .587. And the Cronbach Alpha for the posttest is .641. Hair (2009) recommends that the Cronbach's alpha cut-off value is set at 0.55, as cited in Samuels (2015, p. 3). A total of 10 points (with 10 items) is the achievable mark per participant in the pretest and the posttest. They are distributed in mistranslations, extra information (additions by the MT system), omissions, grammatical and syntax errors, instances of lack of respect of the termbase or style guide, punctuation errors, and style issues. They are scored once for all participants by the author[41], and they are scored on a second round by two external evaluators[42].

According to Geoffrey S Koby and Melby (2013, p. 176), a test is considered reliable if the student is assigned the same score (with reasonable variation) regardless of who grades the exam. As a consequence, in our study, two external raters are requested to evaluate the delivered post-edited texts for the pretest and posttest. Due to availability reasons, the workload for evaluation is split: each rater assesses 17 participants (pretest and posttest). That is, rater A scores the pretest and posttest of participants 1 to 17, and rater B scores the pretest and posttest of participants 18 to 34.

A template is created to perform the PIE analysis of the quality of the post-edited texts. To ease the work of the raters and make the information easily accessible, the software where the evaluation of quality is performed is Ms Excel. The Ms Excel is structured as follows:

[41] Column J, titled "Nota PIE doctoranda" in tabs "pretest" and "posttest" of file *Delivered-post-edited-texts-and-eval.xlsx* in Ginovart-Cid (2021).

[42] Column H, titled "Nota PIE (0=KO, 1=OK) Evaluadora Externa" in the cited spreadsheet.

- Column A: Participant number (1 to 34)
- Column B: Rater name
- Column C: Item number (1 to 10)
- Column D: ST sentence (in French, with the item highlighted in bold)
- Column E: MT output from DeepL (in Spanish, with the item highlighted in bold)
- Column F: Possible PE guideline and/or instruction from style guide to be applied
- Column G: Delivered post-edited sentence (in Spanish, no highlighting)
- Column H: Score (to be entered by the evaluator: 0 for KO and 1 for OK)
- Column I: Comment (free-text field that the evaluators could use to express doubts or nuances in their decision)

The template has 340 lines (34 participants * 10 items), plus the header, in Spanish, announcing the content of the column as described above. After collecting the raters' filled templates, they are merged into one single file. In it, column B is replaced by the participant ID code, the content of column I is erased to preserve anonymisation, and column J is added with the previously annotated scores from the author. It can be consulted in Ginovart-Cid (2021); in tabs "pretest" and "posttest" in file *Delivered-post-edited-texts-and-eval.xlsx*. Following Eyckmans and Anckaert (2017, pp. 43–44) and Mariana et al. (2015, p. 157), the template is created with a criterion-referenced approach:

> This list is transferred into an Excel matrix with "1" for each time a particular segment of the ST was correctly translated by a particular student and "0" for each time a translation segment was mistakenly translated by a particular student. (Eyckmans and Anckaert 2017, p. 44).

Following Kornacki (2018, p 171), a segment is considered corrected when the "student detected an error and corrected it, or rephrased the segment in a correct way". Admittedly, what can become more complex is defining "in a correct way" precisely.

For the pretest, the 10 selected items concern the following statements to be checked and the corresponding PE guideline:

1. Item 1: The participant finds an alternative to avoid giving an agentive load to the noun phrase "los días" in segment 4. The relevant PE guideline is 5: "Mejora el estilo y la fluidez del texto meta".
2. Item 2: The participant adds the missing definite article before "condiciones" in segment 8. The relevant PE guideline is 1.2: "Asegúrate de que no se ha omitido información".

3. Item 3: The participant replaces “grises” by an understandable image in the target culture in segment 11. The relevant PE guideline is 1: “Asegúrate de la equivalencia semántica entre el texto original y meta”.
4. Item 4: The participant corrects the ellipsis in the DeepL output formed by four dots by deleting one or the whole ellipsis in segment 11. The relevant PE guideline is 4: “Adhiérete a las reglas de puntuación de la lengua meta”.
5. Item 5: The participant replaces the verb “romper” by “batir” to collocate with “récord” in segment 11. The relevant PE guideline is 5: “Mejora el estilo y la fluidez del texto meta”.
6. Item 6: The participant finds a valid translation for the mistranslation “espacio en esta área” in segment 18. The relevant PE guideline is 1: “Asegúrate de la equivalencia semántica entre el texto original y meta”.
7. Item 7: The participant finds an alternative to facilitate the comprehension of the expression “no es superfluo” in segment 19. The relevant PE guideline is 5: “Mejora el estilo y la fluidez del texto meta”. The style guide also applies to this item, namely the first rule: “El estilo es familiar, coloquial”.
8. Item 8: The participant replaces “franceses” by the target culture equivalent in segment 23. The relevant PE guideline is 3: “Respeta la terminología y el manual de estilo del cliente”, and the style guide rule that would apply is: “El blog se debe adaptar para España”.
9. Item 9: The participant uses a valid translation for the expression “bon vieux cuir noir” in segment 25. The relevant PE guidelines are 1 and 5: “Asegúrate de la equivalencia semántica entre el texto original y meta” and “Mejora el estilo y la fluidez del texto meta”.
10. Item 10: The participant adds the missing information in segment 34. The relevant PE guideline is 1.2: “Asegúrate de que no se ha omitido información”.

The 10 items’ ST segment and their corresponding DeepL output are the ones presented in table 9.3, for the pretest.

Table 9.3: 10 items’ ST and MT (pretest)

Item	**Source text**	**MT output**
1	C’est mathématique : les journées les plus courtes de l’année font qu’il est plus fréquent de rouler à moto la nuit.	Es matemático: los días más cortos del año hacen que sea más común conducir una motocicleta por la noche.
2	3. adapter notre pilotage aux conditions de visibilité moindre.	3. adaptar nuestra gestión a condiciones de visibilidad reducida.
Continued on next page		

Table 9.3 – continued from previous page

Item	Source text	MT output
3	La nuit, tous les motards sont gris...	Por la noche, todos los motociclistas son grises....
4	La nuit, tous les motards sont gris...	Por la noche, todos los motociclistas son grises....
5	Déjà peu perceptibles par les autres usagers de la route en temps normal, nous battons des records d'invisibilité dans l'obscurité.	Ya no muy notorios por otros usuarios de la carretera en tiempos normales, rompemos récords de invisibilidad en la oscuridad.
6	En cas de lacune en la matière, n'hésitez pas à y ajouter un brassard réfléchissant.	Si hay un espacio en esta área, siéntase libre de agregar un manguito reflectante.
7	La bagagerie occupe souvent une place de choix dans le champ de vision des automobilistes, la rendre plus visible n'est donc pas superflu.	El equipaje ocupa a menudo un lugar destacado en el campo de visión del conductor, por lo que no es superfluo hacerlo más visible.
8	Les motards du Nord de l'Europe, encore plus concernés par les conditions de luminosité médiocres que nous autres Français, ont pris les devants depuis longtemps :	Los motociclistas del norte de Europa, aún más preocupados por las malas condiciones de iluminación que los franceses, han tomado la iniciativa desde hace tiempo:
9	Vous pouvez aussi enfiler votre gilet fluo par-dessus votre bon vieux cuir noir.	También puedes ponerte el chaleco fluorescente sobre el cuero negro bueno y viejo.
10	Un coup de chiffon pour y voir plus clair	Un trapo para ver más claramente

For the posttest, the 10 selected items concern the following statements to be checked and the corresponding PE guideline:

1. Item 1: The participant edits the MT output as to keep the syntactic referent of the prepositional phrase "à se sentir en sécurité" in segment 48. The relevant PE guideline is 1: "Asegúrate de la equivalencia semántica entre el texto original y meta".
2. Item 2: The participant corrects the mistranslation "tragarse los bolardos" in segment 58. The relevant PE guideline is 1: "Asegúrate de la equivalencia semántica entre el texto original y meta".
3. Item 3: The participant notices and corrects the nuance of the discursive marker "En parlant de" in the MT output. The relevant PE guideline is 1: "Asegúrate de la equivalencia semántica entre el texto original y meta".
4. Item 4: The participant replaces "agarre en ángulo" with a more comprehensible term in segment 68. The relevant PE guideline is 5: "Mejora el estilo y la fluidez del texto meta".

5. Item 5: The participant replaces “tripulación” with a synonym adapted to the context and purpose of the text in segment 68. The relevant PE guideline is 1: “Asegúrate de la equivalencia semántica entre el texto original y meta”.
6. Item 6: The participant replaces the 13 occurrences of “motocicleta” with “moto”. The relevant PE guideline is 3: “Respeta la terminología y el manual de estilo del cliente”, and the rule from the style guide to be applied: “Usar ’moto’ en lugar de ’motocicleta”’. They can be consulted in table 9.5, where the occurrence is marked as “#”.
7. Item 7: The participant has respected the style guide and used the familiar style (addressing the reader as “tú” instead of “usted”) in the 16 occurrences where DeepL’s output had conjugated the verb for the formal style of “usted”. The relevant PE guideline is 3: “Respeta la terminología y el manual de estilo del cliente”, and the rule from the style guide to be applied: “El estilo es familiar, coloquial”. They can be consulted in table 9.6, where the occurrence is marked as “#”.
8. Item 8: The participant has replaced “set” with the translation provided by the termbase in segment 12. The relevant PE guideline is 3: “Respeta la terminología y el manual de estilo del cliente”, and the target term contained in the termbase for the source term “plateau” is “circuito”.
9. Item 9: The participant has added the missing space in segment 68, before the opening quotes for “contrarrestar”. The relevant PE guideline is 4: “Adhiérete a las reglas de puntuación de la lengua meta”.
10. Item 10: The participant detects and modifies the appositional phrase “tensa” as a syntactic calque of French to increase fluency (segment 69). The relevant PE guideline is 5: “Mejora el estilo y la fluidez del texto meta”.

The 10 items’ ST segment and their corresponding DeepL output are the ones presented in table 9.4, for the posttest:

Table 9.4: 10 items’ ST and MT (posttest)

Item	**Source text**	**MT output**
1	a-t-il été effrayé ou impressionné par quelque chose en particulier, a-t-il eu des difficultés à rester en place ou à se sentir en sécurité ?	¿le asustó o impresionó algo en particular, tuvo dificultad para permanecer en su lugar o se sintió seguro/a?
Continued on next page		

Table 9.4 – continued from previous page

Item	Source text	MT output
2	Si vous comptez avaler les bornes, apprenez-lui à se détendre sur la route en soulageant ses bras et ses jambes lors des portions droites à vitesse constante ou lorsque vous êtes à l'arrêt.	Si planea tragarse los bolardos, enséñele a relajarse en el camino aliviando sus brazos y piernas durante las porciones rectas a velocidad constante o cuando esté parado.
3	En parlant de communication, il est indispensable d'établir un « code » entre vous et le passager.	Cuando se habla de comunicación, es esencial establecer un "código" entre usted y el pasajero.
4	Avoir un passager qui cherche à « contrer » la prise d'angle de la moto est une erreur fréquente qui peut mettre l'équipage en difficulté.	Tener un pasajero que intenta"contrarrestar" el agarre en ángulo de la motocicleta es un error común que puede poner en problemas a la tripulación.
5	Avoir un passager qui cherche à « contrer » la prise d'angle de la moto est une erreur fréquente qui peut mettre l'équipage en difficulté.	Tener un pasajero que intenta"contrarrestar" el agarre en ángulo de la motocicleta es un error común que puede poner en problemas a la tripulación.
6	The 13 segments of the ST	The 13 segments with occurrences of "motocicleta"
7	The 16 segments of the ST	The 16 segments with occurrences of a verb or pronoun in the formal style of "usted"
8	Vous vous souvenez de la conduite en duo sur le plateau pour votre permis ?	¿Recuerdas al dúo que conducía en el set para tu licencia?
9	Avoir un passager qui cherche à « contrer » la prise d'angle de la moto est une erreur fréquente qui peut mettre l'équipage en difficulté.	Tener un pasajero que intenta"contrarrestar" el agarre en ángulo de la motocicleta es un error común que puede poner en problemas a la tripulación.
10	N'oubliez pas que vous aurez peu d'informations sur le comportement de la personne derrière vous qui, crispée, peut hésiter à communiquer avec vous.	Recuerda que tendrás poca información sobre el comportamiento de la persona que está detrás de ti que, tensa, puede dudar en comunicarse contigo.

Table 9.5: Item 6

#	Source text	MT output
1	NOS CONSEILS POUR ROULER À 2 EN MOTO	Nuestro consejo para dos personas que viajan juntas en motocicleta

Continued on next page

Table 9.5 – continued from previous page

#	Source text	MT output
2	Vous venez enfin d'acquérir votre nouvelle moto et vous souhaitez emmener un passager moto, ou vous souhaitez gagner en aisance pour rouler en duo ?	¿Acaba de comprar su nueva motocicleta y quiere llevar a un pasajero de motocicleta con usted, o quiere mejorar su comodidad cuando conduce a dúo ?
3	Ce qui change avec un passager moto	Qué cambia con un pasajero de motocicleta
4	Rouler en duo sur une moto n'a rien d'anodin.	Conducir una motocicleta a dúo no es insignificante.
5	À faible vitesse, le problème principal vient du fait que le passager moto ne fait pas « corps » avec la moto de la même façon que le pilote, et son influence sur le centre de gravité (et donc l'équilibre) de la moto peut engendrer plus rapidement un déséquilibre.	A bajas velocidades, el principal problema es que el pasajero de la motocicleta no "encaja" con la motocicleta de la misma manera que el piloto, y su influencia en el centro de gravedad de la motocicleta (y por lo tanto en el equilibrio) puede causar un desequilibrio más rápidamente.
6	Et pourtant, des années plus tard, je pourrais écrire un code de conduite de la passagère moto, tellement on a roulé ensemble.	Y sin embargo, años después, pude escribir un código de conducta para la motocicleta del pasajero, tanto que viajamos juntos.
7	Soyez à l'écoute de votre passager moto	Escuche a su pasajero de motocicleta
8	En fonction de l'expérience du passager moto, ce sera toujours au conducteur de s'adapter à son passager.	Dependiendo de la experiencia del pasajero de la motocicleta, siempre le corresponderá al conductor adaptarse a su pasajero.
9	Encore plus s'il s'agit de la toute première expérience moto pour votre passager.	Más aún si esta es la primera experiencia en motocicleta para su pasajero.
10	Un passager moto capable d'anticiper votre conduite prendra forcément autant de plaisir que vous sur la route.	Un pasajero de motocicleta que es capaz de anticiparse a su conducción tendrá necesariamente tanto placer como usted en la carretera.
11	Cette fatigue sera décuplée si vous roulez avec une moto en position sportive.	Esta fatiga se multiplicará por diez si se conduce con una motocicleta en posición deportiva.
12	Avant toute chose, il convient de briefer le passager moto.	En primer lugar, es importante informar al pasajero de la motocicleta.

Continued on next page

Table 9.5 – continued from previous page

#	Source text	MT output
13	Avoir un passager qui cherche à « contrer » la prise d'angle de la moto est une erreur fréquente qui peut mettre l'équipage en difficulté.	Tener un pasajero que intenta "contrarrestar" el agarre en ángulo de la motocicleta es un error común que puede poner en problemas a la tripulación.

Table 9.6: Item 7

#	Source text	MT output
1	Soyez à l'écoute de votre passager moto	Escuche a su pasajero de motocicleta
2	Restez humble, et bannissez toute envie d'impressionner en roulant au-delà de vos capacités.	Permanezca humilde, y destierre cualquier deseo de impresionar conduciendo más allá de sus habilidades.
3	Encore plus s'il s'agit de la toute première expérience moto pour votre passager.	Más aún si esta es la primera experiencia en motocicleta para su pasajero.
4	Arrêtez-vous fréquemment et échangez avec lui afin de connaître son ressenti :	Deténgase frecuentemente y hable con él/ella para averiguar cómo se siente:
5	Profitez-en pour lui partager vos ressentis en tant que pilote :	Aproveche la oportunidad para compartir con él sus sentimientos como piloto:
6	Un passager moto capable d'anticiper votre conduite prendra forcément autant de plaisir que vous sur la route.	Un pasajero de motocicleta que es capaz de anticiparse a su conducción tendrá necesariamente tanto placer como usted en la carretera.
7	Gardez également en tête que, côté conducteur, la fatigue physique et mentale se fera ressentir aussi plus tôt.	Además, tenga en cuenta que, por parte del conductor, la fatiga física y mental también se sentirá antes.
8	Car vos bras et vos cervicales seront soumis à rude épreuve.	Porque sus brazos y cuello estarán sometidos a una gran tensión.
9	Votre passager aussi se fatiguera plus rapidement.	Su pasajero también se cansará más rápidamente.
10	Si vous comptez avaler les bornes, apprenez-lui à se détendre sur la route en soulageant ses bras et ses jambes lors des portions droites à vitesse constante ou lorsque vous êtes à l'arrêt.	Si planea tragarse los bolardos, enséñele a relajarse en el camino aliviando sus brazos y piernas durante las porciones rectas a velocidad constante o cuando esté parado.
11	Rappelez-vous vos propres débuts…	Recuerde sus propios comienzos...

Continued on next page

Table 9.6 – continued from previous page

#	Source text	MT output
12	Expliquez bien qu'il sera question de rester le plus « neutre » possible sur la moto, dans son axe vertical.	Explique que se trata de permanecer lo más "neutral" posible sobre la bicicleta, en su eje vertical.
13	Mais s'il s'agit d'un « débutant », il est préférable de s'en tenir à un comportement ayant le moins d'incidence possible sur la conduite.	Pero si usted es un "principiante", es mejor atenerse a un comportamiento que tenga el menor impacto posible en la conducción.
14	Soyez compréhensif, et n'oubliez pas que le meilleur moyen d'éviter le stress sera d'adapter une conduite la plus tranquille possible.	Sea comprensivo y recuerde que la mejor manera de evitar el estrés es adaptar la conducción lo más silenciosa posible.
15	En parlant de communication, il est indispensable d'établir un « code » entre vous et le passager.	Cuando se habla de comunicación, es esencial establecer un "código" entre usted y el pasajero.
16	Si vous possédez un intercom, les échanges seront évidemment bien plus simples.	Si usted tiene un intercomunicador, los intercambios serán obviamente mucho más fáciles.

To conclude with the section on methodology for evaluating the quality of the post-edited texts, the construct validity should be defined. As Bachman and Palmer (1996, p. 21) define "construct validity":

> [T]he extent to which we can interpret a given test score as an indicator of the ability(ies), or construct(s), we want to measure.

Still following Mariana et al. (2015, p. 152), validity is considered a non-binary "property of interpreting test results and not a property of the test itself". For this reason, it will be discussed in section 10.3.

9.4. Pilot and study conduct

A pilot study is organised with two participants. The two semi-professional participants to the pilot hold the position of senior Spanish Country Manager at Datawords, but without previous PE experience or knowledge of the domain of the selected texts. Their input and feedback along the different phases of the pilot study was valuable in multiple ways. Piloting the pretest, training provision and posttest sheds light on the suitability of the materials presented in section 9.2. In fact, the feedback from the two participants to the pilot is used to adapt, discard, or accept any parts of the designed full-scale split-training. Similarly, it also guides the planning for the data analysis step.

From October until December 2019, they performed the pretest, the online split-training in Google Classroom, and the posttest. They followed the protocol described in previous sections, that is, they signed the consent form and performed the warm-up as a previous step, and they filled out the retrospective questionnaire as a final step. Moreover, after each activity, there was a physical meeting with the two participants to collect their feedback. In particular, the following observations were made during such sessions:

For the previous step, the participants indicated some typos in the Spanish consent form and in the recruitment form. They also suggested that the warm-up activity included more information on how to see the results from the termbase within SDL Trados Studio, as well as how the resulting reports from Qualitivity would be used by the researcher. Finally, the pilot study also served to point towards a known error message that the Qualitivity plug-in could make pop up in SDL Trados Studio. Instructions to address it are therefore created (file *Report on the content of a post-editing experimental split-training course.pdf* in Ginovart-Cid 2021, p. 21).

About the pretest, the pilot study showed that, like in Sanchez-Torron and Koehn (2016), "[m]ost editing times logged for first segments in our study are in fact spent browsing through the whole document and are therefore unreliable". For this reason, the first segment is excluded from the data to obtain the average PE time (WPH) and edit distance (ED), as has been commented before. About lesson 1, since the error identification does not depend on the assignment or guidelines[43], the answers of the participants can be, to a certain extent, anticipated. In spite of this, the participants to the pilot study noted that one cannot always interpret an error in the same way. For example, in the Google form (Quiz 1), a participant may copy and paste the whole expression, while another participant may copy and paste only the noun with more semantic load.

In conclusion, a participant may act differently in front of a situation not foreseen in the instructions or video, such as having an omission at the beginning of the sentence. To address such risk, several possible answers for one single sentence are considered valid in the quiz, and/or further explained and discuss as comments for all the answers, when displaying the feedback commented in section 9.2.3. Thus, even when the participant answers correctly, a comment will appear explaining why another word could have been selected as part of the error or not. Finally, the purpose of the quiz is to drill the skill of identifying errors in NMT output, but the participant is evaluated in no way about these results, nor are they part of the data analysed. Such information is added to the consent form upon suggestion of the participants

[43] In other words, even in a light PE scenario, an error that would not be edited is still an error.

to the pilot study, as to reduce the stress during the split-training course.

About lesson 2, the answers to the quiz are radio buttons: (1) to post-edit, or (2) to discard and translate from scratch. The participants to the study indicated that it was difficult for them to decide what extent of modification is considered to mean "to discard and translate from scratch". Hence, sometimes they doubt between option (1) and (2). As a result, a reminder is added on the top of the quiz to avoid asking any mathematical calculation from the participants. In the reminder it is expressed that the option (2) should only be chosen when the whole output would be deleted. Like in the previous lesson, either in a correct or in an incorrect answer they receive a comment showing the published translation on the site of the customer, so they can reflect on the amount of editing that they were imagining themselves doing and the HT that was published by the customer.

Regarding lesson 3, the participants to the pilot study indicated that doing the quizzes in the Google forms once, and later post-editing the Trados Studio packages was confusing in terms of chronological order since the content was the same. Hence, the exercises 3.2 and 3.3 are adapted to request that the participant performs the quiz and the actual PE task in parallel. That is, in one screen the Studio package and in another screen the Google form. Since it is not confirmed that all participants have two screens, it is left as a suggestion and there is no further instruction about the order of the task: first the quiz and then post-edit the segment in Studio, or the opposite. It is left as a free choice to the participants. The pilot study did not cover lesson 4 or the Memo for the posttest that have been described in section 9.2.3, because the Covid-19 outbreak could not be predicted. These two instruments were improvised at a later stage.

Regarding the posttest, there were no comments or questions. Everything had been addressed during previous sessions. The feedback collected from the two participants to the pilot let us conclude that the overall length of the experiment corresponded to the estimations made. It also raised awareness on the necessity of explicating certain instructions before the beginning of each phase. Finally, the comments and performance of the two participants guided the preselection of certain items for the translation quality assessment according to the PIE method. Having considered the lessons learned from the piloting and adapted the materials presented in section 9.2 accordingly, and considering how the data is to be analysed, the experimental conduct is ready to be deployed. From a practical perspective, the study is organised and conducted as follows:

- Pretest
    - Datawords participants performed the pretest the 5th and 11th of February 2020, at their two-screen working station in the offices located in Levallois-Perret and Barcelona, respectively. The re-

searcher was present to introduce the test and provide help or answer questions if needed. The warm-up activity had been carried out the previous day (4th and 10th, respectively).

- Freelance participants performed the pretest the 3rd of March of 2020, at home, that is, their usual working station. The researcher launched the pretest remotely via a Google Meet[44] call and the participants could choose to stay in the call (muted and volume off) or to hang up and contact the researcher directly should any question arise. The warm-up activity had been carried out the previous day.
- The participants from Geneva University carried out the pretest the 12th of March of 2020 in the laboratory of the Faculty where the courses that require computer and specific software are conducted. The researcher was present to introduce the test and provide help or answer questions if needed. The warm-up activity had been carried out the previous day.

- Split-training (Gclassroom)
  - The deadline for Datawords participants to complete the split-training was 16th March 2020. For freelance participants, one week later. For the participants from Geneva University the deadline had been set to 3rd April 2020. However, the outbreak of Covid-19 led to lockdown in most European countries. In France, lockdown was set for the 16th March 2020. Most students at the Swiss faculty encountered difficulties when relocating to work and study from home. Their deadline was reset to the 13th April to leave some time to adapt to the situation.
- Posttest
  - Datawords participants performed the posttest the 23rd and 24th March 2020. They conducted it at their working station in the offices located in Levallois-Perret and Barcelona, or at home, depending on the individual possibilities and the lockdown legal requirements of each country (France and Spain, respectively). It was conducted via Google Meet conference call. The participants had the choice to stay in the call (muted and volume off) or to hang up and contact the researcher directly should any question arise during the posttest[45].
  - Freelance participants performed the posttest the 31st March 2020 at home, that is, their usual working station. The re-

[44] https://meet.google.com/.

[45] Admittedly, the lack of physical presence of the lead investigator may have reduced the stress of the participants in comparison to the pretest step.

searcher launched the posttest remotely via a Google Meet call and the participants could choose to stay in the call (muted and volume off) or to hang up and contact the researcher directly should any question arise during the posttest.

- The participants from Geneva University carried out the posttest the 20th April 2020 at home via Google Meet call. The participants could choose to stay in the call (muted and volume off) or to hang up and contact the researcher directly should any question arise during the posttest.

Chapter 10

RESULTS AND ANALYSIS

In the present chapter, the outcomes of the study are examined and reflected upon at the light of the data gathered. The dataset containing the results obtained through the conduct of the study described in the previous chapter is reproduced in file *Dataset.xlsx*, published in Ginovart-Cid (2021). Likewise, the outcomes of the statistical tests applied can be found in file *Outcomes-of-data-analysis.pdf* of the same publication. First, in section 10.1, the findings regarding the temporal effort are discussed. In section 10.2, we comment on the technical effort by means of the EPH rate, and we combine WPH with technical effort, as announced, with TES. In section 10.3, the quality of the post-edited product obtained with the PIE methodology is analysed. Finally, in section 10.4, the participants' opinions regarding the pretest, training course, and posttest are briefly presented.

Before entering into the hypotheses contrast introduced in chapter 8, to select the appropriate statistical method, the Shapiro-Wilk test of normality is performed. This tests analyses the distribution of our data. Normality is declared when the p-value is >0.05. In general, the data for the four studied variables is normally distributed. In other words, the Shapiro-Wilk test results in significance values of p>0.05, except for the time variable at the posttest (p=0.015), as observed in table 2 of file *Outcomes-of-data-analysis.pdf* in Ginovart-Cid (ibid.). In this context, the analysis to be applied is the Student paired-sample t-test[1]. The null hypothesis is not rejected in cases where p-value>0.05, and rejected where p-value<0.05.

10.1. WPH rate

The throughput of the 34 participants is higher at the posttest step than the throughput at the pretest. As can be observed in table 3 of file *Outcomes-of-data-analysis.pdf* in Ginovart-Cid (ibid.). The throughput increases, on

[1] Also named "t-test for dependent samples".

average 143 WPH, and this finding is found to be statistically significant (p=0.009). Thus, the p-value<0.05 makes us reject the null hypothesis:

> H0 = The temporal effort does not vary significantly before and after the split-training course

And we assume the alternative (see chapter 8):

> H1 = The temporal effort varies significantly before and after the split-training course

10.2. EPH rate and Efficiency Score

Unlike the throughput (WPH rate), the EPH rate is lowered at the posttest step. As can be observed in table 4 of file *Outcomes-of-data-analysis.pdf* in Ginovart-Cid (2021). As a first observation, one could deduce that the reason why higher WPH rates are recorded at the posttest step is that the participants apply fewer edit operations on the MT output after the treatment. However, this finding is not statistically significant (p=0.28).

The p-value>0.05 does not allow to reject the null hypothesis.

> H0 = The EPH rate does not vary significantly before and after the split-training

The variation is in negative terms: at the posttest step the participants do less edits per hour. This finding seems to correspond well with the fact that they go faster, as discussed before. It seems fair to assume that the split-training session has had a negative impact, at first glance, on the participants, by reducing their PE performance. One can better grasp this outcome by comparing it to the TES: 45 is the average efficiency score of the 34 participants at the pretest, contrasted to the 38 score obtained at the posttest stage. There could be several reasons to explain the decrease in mechanical efficiency (that is, productivity in its misused sense). Regarding the context of the posttest, it must be reminded that the Covid-19 outbreak had just taken place, thus, the participants may have been in an altered situation, both from a physical and a psychological perspective.

In terms of learning process, it could be hypothesised that the split-training of 7 hours (divided in time for reading/hearing theory and time for autonomous practice) starts to scratch the surface of the process of PE competency acquisition. The awareness about some PE techniques in the participants would lead towards an increased confidence about which edits are necessary, but not yet to the sufficient proceduralisation as to increase the overall efficiency.

Our third hypothesis is tested and the findings go in line with the EPH results. As can be seen in table 5 of file *Outcomes-of-data-analysis.pdf* in Ginovart-Cid (ibid.). The *p*-value=0.052 poses a choice. If we interpret the test strictly, the *p*-value is technically above the threshold of significance, therefore we should not reject the null hypothesis. In a more permissive approach, one could see the *p*-value=0.052 near enough the threshold as to reject the null hypothesis and assume the alternative:

> H1 = The TES varies significantly before and after the split-training

Again, the change is of negative trend: the participants become less efficient, according to TES, after the split-training course.

10.3. Quality of the product

Despite the unsettling situation brought by the global pandemic, the quality of the post-edited texts at the posttest step shows improvement, as can be observed in the data published in file *Dataset.xlsx* and in table 6 of file *Outcomes-of-data-analysis.pdf* in Ginovart-Cid (ibid.). The average score at the pretest phase is 4.56. At the posttest it raises to 6.56. This finding is found to be extremely significant (*p*=2.94E-06[2]).

The inter-rater agreement (also known as inter-rater reliability) (Lommel, Popovic, et al. 2014, p. 33) is a measurement of the degree of coincidence among observers in evaluating specific items. According to Salkind (2017, p. 173), inter-rater reliability can be computed with a very simple formula:

$$IRR = number\,of\,agreements / number\,of\,possible\,agreements$$

As some agreement may be due to chance, the percent of agreement should be presented together with other parameters, for example, kappa statistics (Cohen 1960). The Cohen's kappa can be applied when two raters are used, and also for test-retest reliability evaluation. These are our cases in the present study. Cohen's kappa is

$$\kappa = (Po - Pe)/(1 - Pe)$$

The fact argued in section 2.1.2 that errors can often be analysed in multiple ways, along with the lack of deep training or previous exposure of the external evaluators to the PIE method, are possible sources of disagreement. Having considered this, the inter-rater reliability scores for the pretest and posttest are illustrated in table 10.1.

[2] In other words, or in other numbers, if I may, *p*=0.000002.

Table 10.1: Cohen's Kappa coefficient

Step	Score
Pretest	0.61
Posttest	0.62

Kappa[3] values below 0.4 represent poor agreement; between 0.4 and 0.75, fair to good agreement; and values over 0.75 indicate excellent agreement (Mandrekar 2011, p. 6). By this standards, the agreement with the external evaluators at the two steps (pretest and posttest) is considered good. In absolute terms, that is, dividing the total agreements by the number of evaluations (i.e. participants = 340) as commented above (Salkind 2017, p. 173), the pretest shows a percentage of agreement of 80.6%, and the posttest, 81.2%. The data is reproduced in tab "Summary-IRR" of file *Delivered-post-edited-texts-and-eval.xlsx* in Ginovart-Cid (2021).

To give a couple of examples of disagreement, the external evaluator of participant 33 for item 5[4], at the pretest step, considered the solution of not editing "rompemos" as correct in the following situation:

- ST: Déjà peu perceptibles par les autres usagers de la route en temps normal, nous battons des records d'invisibilité dans l'obscurité.
- MT output: Ya no muy notorios por otros usuarios de la carretera en tiempos normales, rompemos récords de invisibilidad en la oscuridad.
- Relevant PE guideline: "Mejora el estilo y la fluidez del texto meta".
- Post-edited text: Si ya no somos muy visibles para el resto de los conductores, rompemos los récords de invisibilidad por la noche.

The external evaluator assigned a "1" to the above-mentioned delivery of participant 1 to item 5 in the pretest. On the contrary, the lead researcher assigned a "0".

A similar situation occurred at the posttest step, with participant 34 and item 7. The external evaluator assigned a "1" to the delivery of participant 34 to item 7 for the posttest, whereas the lead researcher assigned a "0":

- ST: The 24 occurrences of pronouns and verbs reported in table 9.6, in particular: "Un passager moto capable d'anticiper votre conduite prendra forcément autant de plaisir que vous sur la route.".
- MT output: The 24 occurrences of pronouns and verbs reported in table 9.6. In particular: "Un pasajero de motocicleta que es capaz de anticiparse a su conducción tendrá necesariamente tanto placer como usted en la carretera.".

[3] An easy-to-use calculator online can be found in Scarpellini (2020).

[4] With an item difficulty of .35 *p*-value and a discriminatory power of .57 D-index, as seen in *Cronbach-pvalue-dindex.docx* file in Ginovart-Cid (2021).

- Relevant PE guideline: "El estilo es familiar, coloquial: preferir el tratamiento de *tú*".
- Post-edited text: "Un pasajero capaz de anticiparse a su conducción necesariamente disfrutará de la carretera tanto como tú".

On the contrary, and for item 6[5] in the pretest, the lead researcher assigned a "1" for participant 33, while the external rater esteemed the following delivery inadequate:

- ST: En cas de lacune en la matière, n'hésitez pas à y ajouter un brassard réfléchissant.
- MT output: Si hay un espacio en esta área, siéntase libre de agregar un manguito reflectante.
- Relevant PE guideline: "Asegúrate de la equivalencia semántica entre el texto original y meta".
- Post-edited text: Si no tienes nada, no dudes en agregar un brazalete reflectante.

The solution of the participant may also be considered than the published translation in Motoblouz blog: "Siempre que tengáis carencias en la materia, no dudéis en añadir un brazalete reflectante"[6]. In the published version, the word "reflectante" appeared twice in the previous paragraph, and nine nouns to which the property of "réfléchissant" could potentially be applied: *piezas*, *ribetes*, *chaquetas*, *pantalones*, *equipación*, *zonas*, *material*, *tecnología*, and *chaquetas* again. With all the cited occurrences of the concept being discussed, it is argued that the lexical omission of "matière" or, to put it in other words, the implicitation strategy used by participant 33 to render "espacio en esta área" (as MT output of "lacune en la matière") is an appropriate PEA that does not demand deletion, addition or shift, but rather deleting and rethinking the whole phrase in context.

The p-value<0.05 makes us reject the null hypothesis:

> H0 = The PIE score does not vary significantly before and after the split-training

And we assume the alternative:

> H1 = The PIE score before the split-training varies significantly before and after the split-training

The variation is significant in a positive trend: the participants deliver a higher quality post-edited text after attending the split-training course.

Finally, we measure the effect of one or more independent variables on a dependent variable, including also the interaction between independent

[5]With an item difficulty of .44 p-value and a discriminatory power of .34 D-index, as seen in page 3 of file *Cronbach-pvalue-dindex.docx*.

[6]http://xl8.link/MB-noche

variables, using a linear regression model. The first dependent variable is the TES. As seen in chapter 9, it is the combination of the indicators of temporal WPH and technical EPH effort. According to the proposed regression model, the TES is a variable that can be explained by the other two variables: EPH and WPH. Surprisingly, it is not explained by the interaction between them.

This finding is meaningful to the extent that it confirms the index carries out the function it was created to accomplish: to summarise in a single score the mechanical or instrumental efficiency. Hence, with higher rates of WPH and EPH the TES is expected to increase, by keeping the reference to the studied group of post-editors. However, the split-training also explains the reduction of efficiency for the 34 participants when post-editing MT output in the French-Spanish language pair. Moreover, how the split-training affects EPH, on the one hand, and how it affects WPH, on the other hand, also has a causality relation to our variable. In other words, the split-training lessons on the three selected PE skills leads to faster post-editors, who edit less and, consequently, the participants are less efficient according to the TES index, which measures productivity from the mechanical perspective, without considering the quality of the product.

The second and last studied variable, as dependent of a a Generalised Linear Model, is the quality of the post-edited product. According to figure 4 in file *Outcomes-of-data-analysis.pdf*, the variables that explain the increased quality at the posttest stage are the split-training (treatment), and TES. It can be deduced from the findings exposed in the previous paragraphs that the 34 participants of our sample performed at the stage of the pretest either a higher number of pseudo-editing or a higher number of over-editing operations, or both. Another possibility is that the right PEA was performed, deleted, and redone more often at the pretest, showing insecurity for post-editing.

As a matter of fact, the TES can be relied upon as long as the quality evaluation confirms there have not been a significant number of pseudo-editing operations, or, alternatively, over-editing some parts of the MT output (even repeatedly[7]) while under-editing some others. The split-training on the identification of MT output errors, decision-making, and respect of PE guidelines has had a clear positive effect on the PE competency of the 34 participants of our sample, regarding the three selected skills. Despite the pessimistic global circumstances caused by the Covid-19 pandemic that particularly affected the posttest, after 7 hours of training, they were capable of post-editing more words per hour, while making fewer but more relevant (and possibly more confident) edits, according to the quality scores obtained

[7]That is, executing an edit, undoing it, and redoing it, ever for appropriate PEA.

applying the PIE method.

10.4. Retrospective form

In May 2020, the 34 participants filled out a short form that has been presented in section 9.2.3 to collect retrospective feedback about their perceived impact of the split-training course. The results are described below.

- On a Likert scale from 1 to 5, the usefulness of the exercise Quiz 1 - Identification of errors was graded "4" by almost 50% of the participants.
- On a Likert scale from 1 to 5, the usefulness of the exercise Quiz 2 - Decision-making was graded "5" by 40% of the participants.
- On a Likert scale from 1 to 5, the usefulness of the exercises Quiz 3.2 and Quiz 3.3 - Respect of PE guidelines was graded "5" by over 40% of the participants, and "4" by 42%.
- On a Likert scale from 1 to 5, the usefulness of the improvised feedback of lesson 4: Quiz 4 - "¿Cómo me fue en el pretest?" was graded "5" by 23% of the participants, and "4" by 48%.
- When asked what should be deleted from the split-training course, 32 agreed on "nothing" and two suggested that the answers of "Correct"/"Incorrect" be deleted on exercises that may be more open to subjectivity, such as Quiz 2.
- When asked what should be added to the split-training course, ten participants suggest that more didactic guidance and theory should be included[8]. Two argue that some practice on CAT tools (to learn shortcuts etc.) would also be welcome. Four say they would have liked more personalised feedback on their deliveries. Finally, two suggest technical improvements on the use of the Google Classroom platform.
- On a Likert scale from 1 to 5, 45% of the participants state "4" is their feeling on how much they learned, and 26% would say it is "5".
- 32 say they would like to learn more about PE, and one precises that it is "because I think the extent to which it relates to translation competency is increasing"[9].
- 24 participants proofread the pretest. Only one generated the `TXT` file at that stage.
- 25 participants proofread the posttest. The same participant generated the target `TXT` file at that stage.

[8] For instance, recommended bibliography, more examples and more support material.
[9] Translation by the author.

- When asked about the interrelation between translation and PE competency, over 60% of the participants graded it as very high (5) in a Likert scale from 1 to 5.
- At the end of the experiment, 27 participants knew precisely which MT provider had been used (DeepL).

Finally, in a free-text field the participants could express their opinion on the pretest, the course, and the posttest. The 8 participants that used it, commented positively on the experience. For example, one said (in Spanish):

> *La experiencia me ha parecido muy interesante como primer acercamiento a la posedición. Gracias a la formación, he podido aplicar en el postest conceptos que no había tenido en cuenta en la primera etapa. En cualquier caso, tanto en el pretest como en el postest, los textos me han parecido adecuados para conocer y revisar todo lo aprendido.*

Similarly, at the end of quiz 4, the participants had a free-text field with this question: "¿Te había parecido difícil el pretest? ¿Crees que la formación en línea te ha ayudado a interiorizar las técnicas de la posedición? Puedes explicar aquí tu opinión libremente.". The comments are, in general, positive and in line with the conclusion that further practice is required to lead towards proceduralisation. One extensive opinion from one of the 34 participants is quoted below (in Spanish):

Realizar el pretest sin haber sido formada sobre la posedición no me resultó complicado como tarea en sí, pero durante la realización de dicho test, lo que me resultó difícil fue saber qué tipo de correcciones deben hacerse o no para no caer en la sobrecorrección. Creo que la formación posterior me ha permitido plantearme algunas preguntas a la hora de poseditar. Por ejemplo, vi que hubo ciertos elementos que no corregí durante el pretest por no querer hacer una corrección profunda y pasar mucho tiempo, pero que quizás sí hubiera sido pertinente corregir. Si puedo dar mi opinión sobre la posedición, diría que al empezar no tenía muchas nociones de cómo poseditar y pensaba que se trataba de corregir solamente los errores muy graves, evitando las correcciones de estilo que no fueran graves. Tras la realización del prestest y la formación, tengo la sensación de que habría que profundizar más en las correcciones de la posedición cuando se trate, como en este caso, de una posedición completa y, de que además, no solo hay una opción correcta, lo que dificulta la elección.

Final thoughts

The purpose of the current study was to determine to what extent a short tailored Post-Editing (PE) training course for semi-professionals, designed with the split-technique and based on the three main PE skills, could lead to significant improvement in terms of qualitivity. Returning to the research question posed in chapter 8, four hypotheses are tested to objectively measure and assess the effects of the said treatment to the performance of 34 semi-professional French-Spanish translators.

The first hypothesis investigates temporal effort —measured in Words per Hour (WPH)—; the second, technical effort —measured in Edits per Hour (EPH)—; the third, the TAUS Efficiency Score (TES) (an index resulting from the previous two indicators); and the fourth and last hypothesis explores the influence of the split-training on the quality of the post-edited delivered texts according to the Preselected Item Evaluation (PIE) method, presented in chapter 9.

The results of this investigation show that a short tailored split-training on the top three PE skills can increase the abilities of semi-professional Machine Translation (MT) post-editors, especially in regards of speed and quality. There are, however, two of the four indicators which have not seen significant improvements after the treatment. Those are the technical effort (EPH) and the quantitative or mechanical performance (TES). As argued in chapter 10, these findings could be explained by several reasons. In this particular experiment, the pandemic of Covid-19 and the subsequent lockdown in several countries (it is the case of Spain, France, and Switzerland) meant the posttest step must be performed remotely, which is an important difference from the organisational conditions of the pretest step.

> Post-editing skills are developed gradually, and initial judgments are bound to be reversed. The level of comfort is greatly increased at the end of 100,000 words—the equivalent of a month of full-time post-editing. (Vasconcellos 1986, p. 145)

Following Vasconcellos (ibid.), very often cited in research of this field, it is also argued in chapter 10 that proceduralisation cannot be achieved with such short and autonomous learning setting, without explicit discussion

of meta-techniques and without further hands-on PE activities. In chapter 10, two significant findings suggest that industry-tailored split-training in PE may lead to increasing the quality of the post-edited texts, as well as reducing the temporal effort.

In the mentioned chapter, multiple regression analysis revealed that the the quantitative performance index TES varies according to the time post-editors invest in the task. TES also varies in function of the number of editing operations a post-editor performs. If we consider the definition of TES, explained in chapter 2, one can see how the outcome of such regression model comes as no surprise. Notwithstanding, it is striking that the regression model does not show an impact on the TES index in function of the interaction between the two indicators, that is, temporal and technical effort as per WPH and EPH.

As commented in chapter 10, the split-training (treatment) has a negative impact on the data gathered about EPH and TES. Whereas it could be claimed that the designed course rendered the 34 semi-professionals less efficient, one must take into consideration the quality of the product. The quality of the post-edited product has increased at the posttest step, compared to the pretest, according to the PIE analysis described in the above-mentioned chapter. According to the regression model, the split-training combined with the other three indicators (WPH, EPH, and TES) are the reason behind such qualitative improvements.

Overall, this study strengthens the idea that identifying MT errors, deciding to post-edit or translate from scratch, and post-editing according to a set of guidelines are three key PE skills that can be learned. The present study confirms previous findings (for example: Blagodarna 2019) and contributes additional evidence that suggests how the three main PE skills can be acquired within a real project in a commercial setting. Through a split-training course, the learning process may not be complete. Thus, in the following paragraphs, suggestions for improvement and speculations on future directions of the research in the PE pedagogy field are presented.

Excluding the intrinsic weaknesses of any experiment[10], the most important limitation lies in the fact that Information and Communications Technology (ICT) skills and Translator-relevant Computer Competence (TCC), the Instrumental Input Skill (IIS) discussed in part II, are not evaluated. In fact, this dissertation was not specifically designed to evaluate these factors, which is the reason why further investigation into the learnability of Instrumental input skills is necessary.

More broadly, research is needed to determine the existence of a practically-linguistically-motivated connection or interaction of instrumen-

[10]The threat of testing and the threat of mortality as defined in Blagodarna (2019, p. 140).

tal skills with efficient editing skills.

> [...] [P]articipants have to allocate time to get familiar with the project and to participate. Researchers, on the other hand, have to adapt to the participants' availability. Financial, organisational and infrastructure constraints might turn out as further roadblocks that have to be overcome. (Risku et al. 2019, p. 53)

The observation quoted above is another typical conditioning factor of any human subject research. Similarly, like the work by Ortiz Boix and Matamala (2015) and Fiederer and O'Brien (2009), our experiment is limited in terms of language pair (French into Spanish), text type (blog article with advertising and marketing nuances and purpose), MT engine (DeepL) and sample size (34 participants).

The insights gained from this study may be of assistance to educators in charge of the creation of other French-Spanish PE courses. Likewise, they could be useful to PE courses in other language pairs, either in academic or industrial settings, especially when the goal is to introduce PE to a translation professional for a specialised knowledge domain, or within time constraints. The methods used (three PE efficient editing skills, split-training, PIE...) may be applied or combined with other approaches to enhance our understanding of PE pedagogy. For example, the results presented in the previous chapter and summarised above can also be considered in the light of Yang and Wang (2020, p. 10)'s findings. In essence, it is hoped that such recent progress in the field of PE pedagogy "provide[s] pedagogical suggestions for trainers and instructors to develop targeted interventions aimed at promoting students' academic success [...]". Not only academic success, but also help the industry avoid deficient workforce retention (Fullen 2019).

We would like to conclude the present part with a quotation of Kiraly (2015, p. 24):

> Syllabus design is no longer a task to be accomplished by a teacher alone prior to the start of a course; it becomes a tentative plan that emerges with new challenges and unexpected turns, and one leading to unpredictable goals as a course progresses.

We agree with researchers like Kornacki (2018, p. 83-84) who claim that a major challenge for translation trainers is that the industry is affected by technological developments which have "profound influence on the way translators do their job". In that sense, it is hoped that the compartmentalising or reduction into the three studied PE skills is a first step towards building (a) more holistic educational framework(s) that will prevent, as much

as possible, that future translation professionals end up delivering texts with too high rates of post-editese[11].

[11] Post-editese is a recent object of research, for example in the work by Daems, De Clercq, et al. (2017) and Toral (2019).

Conclusions

> During the last couple of decades, the emergence of new technology has drastically changed the working environment of professional translators.

This is a quote by Anderman and M. Rogers (2000, p. 63). Thus, we have been repeating the same idea for over twenty years now. The following quotation is a comment made by Nico Herbig, whose interview with Samuel Läubli is covered in the Tool Box Journal edition of December 2020.

> Indeed, Post-Editing (PE) requires very different interactions than traditional translation. We see a change from "production", where all text had to be manually entered, to "supervision", where the task changes to capturing and correcting mistakes, as well as manipulating and recombining useful proposals. Naturally, this change already started with TMs[12], but the better MT[13] gets, the more we move away from the production paradigm to supervision and collaboration with the machine. (Zetzsche 2020)

With such acknowledgments made over the past two decades, some research has thoroughly analysed a selection of factors. In the case of this thesis, we investigated how the so-called "capturing" of MT errors is a learnable skill, via a short tailored split-training. Likewise, the designed course in the French-Spanish language pair has attempted to allow for the acquisition of two more skills that come after capturing or spotting the errors: deciding (when to translate from scratch or reuse) and executing (abiding to the project's brief and guidelines).

The following paragraphs provide, firstly, conclusive remarks regarding the main areas of this dissertation; secondly, they seek to suggest recommendations for future work. To this end, the previous literature is brought together with the main outcomes of the survey and experimental parts. Considering the evolution of the translation craft as presented in the literature review provided in part I, the present dissertation has been designed with a mixed methodology. In it, we covered three central RQs. Two, in the first phase, are answered via the combination of three detailed questionnaires with the analysis of PE syllabi and interviews with the relevant educators. The two RQs of part II are summarised below.

> **RQ1** To what extent do the PE stakeholders agree on the definition of PE as an activity and of the post-editor profile?

[12]Translation Memory (TM).
[13]Machine Translation (MT).

> **RQ2** To what extent is the profile of an MT post-editor multi- and transdisciplinary?

The third RQ of the dissertation has been established and addressed in part III. It reads as follows:

> **RQ3** To which extent does the split-training method on the cited three skills allow for increased qualitivity in the PE performance of French-Spanish semi-professionals?

Kornacki (2018, p. 17)[14] noted how surveys of the European Commission indicate that 74% of European translators are freelance, and that Slovakia is the only country to regulate our profession so far. In the localisation industry, the degeneration of the perceived status of professional post-editors discussed in section 2.3, and the fragmented and never-ending-versioned[15] Source Text (ST) are factors that suggest the importance of adequate training, if we want to avoid post-editese (Toral 2019)[16]. With this reality in mind, our study analysed "qualitivity" (quality and productivity) gains in a real(istic) setting thanks to the application of a split-technique training. Like the work by Federico et al. (2012), our research consisted of a field test that measured productivity (temporal and technical efforts) with a commercial Computer-Aided Translation (CAT) tool. In the research presented in this dissertation, the use of a TM was excluded to focus exclusively on Machine Translation Post-editing (MTPE) skills. In that sense, the three RQs devised and the outcomes they lead to constitute valuable insight for a number of stakeholders. In fact, as argued in Ginovart-Cid (2020b), the impact of PE skills in employability is becoming clear in the language market.

As commented in section 7.2, PE modules usually represent a maximum of 12 to 50 contact hours at postgraduate courses, and they often request from the student some work at home, ranging extremely from 8 to 160 hours. According to the reference of 100,000 words provided by Vasconcellos (1986, p. 145), or the equivalent of one month work, it can be concluded that the lowest-end of the studied courses falls 220 hours behind the threshold at which a post-editor acquires comfort and proceduralisation of the activity. The highest-end of the syllabi sample would represent 210 of PE mixing theoretical and practical items, which makes them still fall behind of, at least, 30 hours of hands-on PE to reach the comfort Vasconcellos's established. This is the reason why the contribution of authors like

[14] Page 6 in the preprint.

[15] Texts that are constantly updated through decisions that are less individual (do Carmo 2020a, p. 14), as briefly discussed in section 1.2.

[16] That is, avoid under-editing.

Kiraly and Göpferich are considered a good ingredient to competency-based training. As the former puts it, "learning is meant to be a highly interactive, proactive and transformative process". Proceduralisation is later acquired through work placements, for example, which is in line with the discussion presented in the Final thoughts of II and the Final thoughts of III.

The split-technique in training has proved to be a useful feedforward tool to "make full use of the extended competences that translators [...] are capable of bringing to bear" (Massey 2018, p. 13). In fact, "if the Dreyfuses are right" (Chesterman 2000, p. 80), the concepts are taught/learned through translation studies, and the split-training course has enhanced the decision-making skill and helped the 34 semi-professionals trust their intuition, that is, work faster "but without losing hold of the critical faculty altogether [and it has given them] the experience of *flow*" (ibid., p. 80). Even so, teaching effects are uncertain and hard to assess as recognised in Blagodarna (2019) citing Kiraly (2000). In that sense, it is natural to assume that novice (or even intermediate) translators should "keep on studying [and practising] as active members of emerging lifelong-learning society" (Blagodarna 2019, p. 59). Assuming that 100,000 words (or other referenced volumes of hands-on practice) is not attainable for a student in a stand-alone PE course (O'Brien 2002, p. 104), one could foresee sufficient projects and tasks to reach a similar figure curriculum-wide. That is, MT-aided translation and PE would be **defined** the first year of undergraduate studies. MT-aided translation *could* be **performed** from the first year of undergraduate studies, in several subject matters (audiovisual translation, technical translation, literary translation, scientific and legal translation courses, localization, marketing translation, etc.). As from the second or third undergraduate year, PE could be **introduced** and the difference between both services **reminded**. At that stage, some courses could cover their specialised translation from an MT-aided or PE perspective, or both in different activities. Which begs the question: Is there the necessity or is it relevant at all to have a PE stand-alone course?

Depending on the curriculum of each centre, a course with PE would have more or less technical content, while others would "simply" aim at presenting the possibility of performing feasibility studies in the role of Post-Editor/Project Manager (Sánchez-Gijón 2016). In some of these courses the student would learn about the "notary"[17] role a professional translator may have when approving a given MT output as a valid translated text for the relevant context and scenario. As long as it is clearly established and transparent, the professional-to-be can choose with which Continuing Professional Development (CPD) course or with which work placement their education can be complemented with. With these varied applications of MT, if dis-

[17] Or "arbiter" in terms of Dalla-Zuanna (2020, p. 181).

tributed correctly through translation curricula, one may be able to reduce the leveling effect commented in section 4.2, whereby good students deliver less qualitative translations and, on the contrary, students who used to translate poorly deliver better translations thanks to MT (Schumacher 2020, p. 260), narrowly linked to reducing post-editese too.

Kiraly (2012a, p. 89) talks about the goal of learning in these terms:

> The goal is to contribute to the emergence of independent thinkers, competent heuristic problem-solvers and knowledgeable translators [...] as neo-professionals with a deep knowledge of the panoply of skills and competences they can expect to encounter in the world [...].

Indeed, if educators incorporate Neural Machine Translation (NMT) according to the opportunities and limitations of each subject matter, and keeping Kiraly's (p. 89) goal in mind, then we could avoid or reduce such leveling effect in the long term. All in all, the overall process of skills' acquisition in PE should result in professionals that produce extremely low rates of post-editese in their deliveries, as commented in the Final thoughts of part III. In that sense, initiatives such as the International Network of Simulated Translation Bureaus (INSTB)[18] and supervised work placements that do not fall into "apprenticisation" (Kiraly and Sascha 2019, p. 78) are welcome (EMT Network 2019).

> [T]he curriculum can become a pedagogically sensible and well-structured **introduction**[19] into the professional translators' community of practice. (Kiraly and Sascha 2019, p. 79)

It has been argued in the Final thoughts of parts II and III that different translation courses may cover different PE skills. The three skills selected in this dissertation (identification of MT output errors, decision-making, and respect of PE guidelines) are tentatively considered to pertain to the group of Efficient Editing Skill (EES). These three core PE skills could "be considered a basic set of skills within the translation competence model (like proofreading or information mining)" (Sánchez-Gijón 2014, p. 24), and a variety of courses in translation curricula could address them from different perspectives. The evidence from the exploratory survey-based study of part II suggests that PE happens at a zone of proximal development between several fields, in a so-called systemic interconnectedness (Harding and Cortés 2018, pp. 29–33). Thus, the profile of a post-editor can be considered multi- and transdisciplinary. In other terms, it may be seen as a hybrid profile[20].

[18] https://www.instb.eu/

[19] Emphasis added by the author of this dissertation.

[20] Like the hybrid profile of the jurilinguist (Harding and Cortés 2018, p. 464).

It is therefore capital that stakeholders (in all spheres related to localisation) keep themselves informed about the evolution and changes affecting the market, as it is not just one event, but a constant. As Harding and Cortés (ibid., p. 38) claimed: "the familiar being constantly de-familiarised". In that sense, we must urge ourselves to keep abreast of the progress in the localisation industry. Two examples of recent efforts to be considered in subsequent work in the field are given in the following lines. The first example could be the book *Translation Quality in the Age of Digital Transformation*[21] provides in its Introduction (pp. 28-36) a panorama view of the market that accounts for the pre- and post-Covid situation during the year 2020. It also announces breakthroughs in terms of MT technology in the next three to five years. The second example, the MTPE Training Special Interest Group that started in January 2021[22].

Categorising errors of the post-edited texts into over-, under-, and pseudo-editing was out of the scope at this time. The classification into typology of edits constitutes a piece of further work that could bring more insights to our present conclusions. In other words, annotating the edit operations our 34 participants performed (Koponen and Salmi 2017; Koponen, Salmi, and Nikulin 2019) could shed light on the impact of the findings discussed in the previous paragraphs. As a matter of fact, attributing a weight to the PE mistakes, along with possessing a gold reference to obtain the baseline Translation Error/Edit Rate (TER) of a file, would allow a PE trainer to combine edit efficiency and revision quality in a more comprehensive way than the experiment performed in part III. The metric that responds to the limitations discussed about TAUS Efficiency Score (TES) in chapter 8 is "intervention optimality". It has been proposed recently by Daems and Macken (2020, p. 58):

> Intervention optimality is the weighted harmonic mean of revision quality and edit efficiency, weighting revision quality higher than edit efficiency [...].

The authors (ibid., p. 68) found that professional translators tend to do many preferential changes, either in revision or in PE. Classifying the 10 items studied in our experiment (and increasing the number of items) would help confirm or deny[23] if, as argued in section 10.3, the higher rate of editing

[21] Dalla-Zuanna and Kurz (2020), whose chapter by Wolfram Baur addresses (pp. 100-103) important problems in *ISO 17100:2015*, that must certainly also be addressed for *ISO 18587:2017*.

[22] https://www.gala-global.org/events/events-calendar/mtpe-training-sig-january-2021-call

[23] Even more so if the assessment is performed on the totality of the deliveries, to contrast the results of the Preselected Item Evaluation (PIE) method to those obtained with a holistic assessment.

at the pretest compared to the posttest is due to less over- or pseudo-editing actions at the posttest step (i.e. an enhancement of the PE technique in terms of EES). Moreover, it could also suggest that some Post-editing Action (PEA), even when appropriate, where more often undone, rethought and and redone at the pretest step. This finding would confirm the idea whereby a split-training PE course can render the semi-professional post-editors less hesitant, more confident.

It is hoped that, combined with past and future research (e.g. Massey et al. 2019), the characterisation of the three top PE skills and the experimental work on split-technique training reported in this dissertation can contribute to enrich PE teaching in some tailored settings, academic or professional. In the same vein, the impact of the findings discussed above and in the Final thoughts of part III may be an opportunity to integrate workplace experience within PE courses, as Kiraly and Piotrowska (2014) suggest. We agree with Kiraly (2014) that the work placements combined with appropriate learning techniques (such as speed-training in certain tasks, like the précis writing commented by Bowker 2016, p. 34) are crucial in the translation profession. The design and provision of improved PE and revision training (of any type: speed-, split-, traditional...) could facilitate the acquisition, for novice and semi-professionals, not only of the skills, but of the right sequencing and integration defended by Calvo (2009). This, in turn, may reduce disempowerment of workers (Moorkens 2017, p. 470) and push them to acquire the "role of intercultural, interlingual information brokers and consultants" (Angelone et al. 2019, p. 8). Muzii (2016) claimed that MTPE projects are often difficult to deliver because we ask translators to act as editors, and they consider PE "a minor and diminishing task". While the integration of TM and MT can now pass on to be embedded in any translation course, the PE skill-set and its acquisition —especially Instrumental Input Skill (IIS), and how it interacts with EES— has to be further studied.

To sum up, the academic sphere lacked or lacks some technology in traditional translation courses, considering translation as a highly technologised profession (Schäffner 2019). On the other hand, the industry lacked or lacks some qualitative insights, as it may be underestimating the dangers of post-editese in the long term. This is why the suggestion by Attila Görög (2015b) of adding quality and content difficulty scores in the TES seems a good initiative to the author. In fact, the intervention optimality discussed above comes to fill this gap. As a consequence, more research on the intervention optimality index is required to allow for an easier integration into academic settings, and even-easier applications for the industry.

As argued in the Final thoughts in part III, one possible way of developing further the present research would be to select the top PE skills pertain-

ing to the group of IIS[24] and test how split-training may or may not improve the post-editors technical performance. Thus, a further step would combine the top EES with the top IIS and analyse how their intertwining may impact certain edit operations. In other words, the results from Herbig et al.'s (2019) elicitation study may be further combined with split-technique testing and training to draw conclusions on which PE tasks and which PEA are best supported by which modalities.

In the same way that Massey (2018) suggests that translation students should develop the metacognitive capacity regarding the tools they usually work with, it is our opinion that one avenue of research yet to be explored is the metacognitive PE techniques. In other words, for a given language pair, one can link certain MT errors to certain PE actions (e.g. an omission in the MT output calls for an addition). And one can assess a PEA via the over-, under-, pseudo-, and "possible editing"[25] categories. Yet, there is no research to date regarding the PE techniques that would guide the editing operations, and whose definition would be the justification tool to categorise an over-editing as such, and so on. To give an example: in section 10.3 we analysed the delivery by participant 33 of item 6. Considering the MT output was nonsense, instead of looking for equivalence, modulation, transposition or any other translation strategy, the most efficient course of action, from a linguistic perspective[26] and considering the cotext, was indeed implicitation.

To enrich our PE declarative knowledge, the study of IIS (and their connection to EES), or the correlation of edit operations to translation strategies[27], could be valid paths to gain the non-automatised cognitive resources Beeby, Castillo, et al. (2013) commented on. Later, the automatised and non-automatised resources would be combined in an efficient manner by "activating" the translation routine competence from the TransComp model. do Carmo (2020a) wonders "how growing disappointment and loss of talent across translator communities may affect the sustainability of the business?" and the author draws the reader's attention to the impact that the data-driven approach may have on the localisation industry. If we consider that society, in reality, is data-driven in several aspects, we could claim that the present medium-scale research has brought new insights from a data-driven real scenario. Maybe the way forward for translator training is to incorporate some ingredients from the task-based data-driven learning (Singer 2016) approach?

[24] Automatic Speech Recognition (ASR), interactive, predictive, adaptive MTPE...

[25] Not to be forgotten!

[26] That is, EES.

[27] The borrowing, calque, transposition, modulation, equivalence, etc. by Vinay and Darbelnet (1958), or any other complementary or adapted strategies by translation scholars.

Bibliography

Absolon, J. (n.d.). English-Slovak MTPE Course. http://mtposteditors.com/tests/test_EN_SK_01.html

Absolon, J. (2017). Methodology of MT Post-Editors Training.

Absolon, J. (2018). *The Need for Competency-based Selection and Training of Post-editors*. ASAP-translation.com.

Absolon, J. (2019). *Human Translator 4.0 Manual on the effective use of MT for a modern translator*. ASAP-translation.com.

Akbari, A., & Segers, W. (2017). Translation evaluation methods and the end-product: which one paves the way for a more reliable and objective assessment? *SKASE Journal of translation and interpretation*, *11*(1), 2–24.

Alabau, V., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., Germann, U., González-Rubio, J., Hill, R., Koehn, P., Leiva, L. A. et al. (2014). Casmacat: A computer-assisted translation workbench. *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 25–28.

Albir, P. G. A. H., Galán-Mañas, A., Kuznik, A., Olalla-Soler, C., Rodríguez-Inés, P., & Romero, L. (2020). Translation competence acquisition. Design and results of the PACTE group's experimental research. *The Interpreter and Translator Trainer*, *14*(2), 95–233.

Allen, J. (2003). Post-editing. Amsterdam/Philadelphia, John Benjamins.

Anderman, G., & Rogers, M. (2000). Translator training between academia and profession: A European perspective. *Benjamins Translation Library*, *38*, 63–76.

Andrew, P. (2018). Qualitivity: Explanation of Edit Distance Relative and PEM%. https://tinyurl.com/andrew-distance

Andrew, P., & Filkin, P. (2018). Edit distance and Edit distance relative in Qualitivity. http://xl8.link/Andrew-Filkin-ED

Angelone, E., Ehrensberger-Dow, M., & Massey, G. (2019). *The Bloomsbury Companion to language industry studies*. Bloomsbury Publishing.

Aranberri, N. (2014). Postedició, productivitat i qualitat. *Revista Tradumàtica: tecnologies de la traducció*, (12), 471–477.

Aranberri, N. (2017). What Do Professional Translators Do when Post-Editing for the First Time? First Insight into the Spanish-Basque Language Pair. *HERMES-Journal of Language and Communication in Business*, (56), 89–110.

Aranberri, N., & de Gibert, O. (2019). Estrategia multidimensional para la selección de candidatos de traducción automática para posedición. *Linguamática*, *11*(2), 3–16.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. (1994). Machine translation: An introductory guide. NCC Blackwell.

ASTM International. (Draft Under Development). *New Practice for Analytic Evaluation of Translation Quality* (Standard). ASTM International. USA.

Austermühl, F. (2001). Electronic Tools yor Translators.

Aziz, M., Wilker; Koponen, & Specia, L. (2014). Sub-sentence level analysis of machine translation post-editing effort. *Post-editing of Machine Translation: Processes and Applications* (p. 170). Cambridge Scholars Publishing.

Aziz, W., Castilho, S., & Specia, L. (2012). PET: a Tool for Post-editing and Assessing Machine Translation. *LREC*, 3982–3987.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Balling, L. W., & Carl, M. (2014). *Post-editing of machine translation: Processes and applications*. Cambridge Scholars Publishing.

Bandyopadhyay, S. (2012). *Emerging Applications of Natural Language Processing: Concepts and New Research: Concepts and New Research*. IGI Global.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E. et al. (2009). Statistical approaches to computer-assisted translation. *Computational Linguistics*, *35*(1), 3–28.

Beeby, A., Castillo, L., Fox, O., Galán-Mañas, A., Grauwinkel, G., Kuznik, A., Massana Roselló, G., Neunzig, W., Olalla-Soler, C., Rodríguez-Inés, P. et al. (2013). A retrospective view of PACTE's research into translation competence and its acquisition. *Seminar on Empirical and Experimental Research in Translation (TREC)*.

Beeby, A., Fernández Rodríguez, M., Fox, O., Kuznik, A., Neunzig, W., Rodríguez-Inés, P., Romero Ramos, L., Wimmer, S. et al. (2011). Results of the validation of the PACTE Translation Competence model: Translation project and dynamic translation index.

Belam, J. (2003). Buying up to falling down": A deductive approach to teaching post-editing. *Proceedings of MT Summit IX*, 1–10.

Bernth, A., & Gdaniec, C. (2001). MTranslatability. *Machine Translation*, *16*(3), 175–218.

Bié, L., Cerdà-i-Cucó, A., Degroote, H., Estela, A., García-Martínez, M., Herranz, M., Kohan, A., Melero, M., O'Dowd, T., O'Gorman, S., Pinnis, M., Rozis, R., Superbo, R., & Vasiļevskis, A. (2020). Neural Translation for the European Union (NTEU) Project. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 477–478.

Biel, Ł., Łoboda, K., & Svoboda, T. (2017). *Quality aspects in institutional translation*. Language Science Press.

Black, P. E. (2008). Levenshtein distance, Dictionary of Algorithms and Data Structures [online], US National Institute of Standards and Technology. https://xlinux.nist.gov/dads/HTML/Levenshtein.html

Blagodarna, O. (2019). *Enhancement of post-editing performance: introducing machine translation post-editing in translator training* (Doctoral dissertation).

Blain, F., Senellart, J., Schwenk, H., Plitt, M., & Roturier, J. (2011). Qualitative analysis of post-editing for high quality machine translation. *MT Summit XIII: the Thirteenth Machine Translation Summit [organized by the] Asia-Pacific Association for Machine Translation (AAMT)*, 164–171.

Blanco Claraco, J. L. (2019). Calculadoras de significancia estadística en resultados de encuestas [Calculators of statistical significance in survey results]. *Ciencia explicada.* https://www.ciencia-explicada.com/2012/04/online-calculator-of-statistical.html

Blatt, A., Freigang, K.-H., Schmitz, K.-D., & Thome, G. (1985). Computer und Übersetzen. *Eine Einführung*.

Bond, E. (2018). The stunning variety of job titles in the language industry. *Slator*, *16*. http://xl8.link/Bond2019

Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. University of Ottawa Press.

Bowker, L. (2005). Productivity vs quality. *A Pilot Study on the impact of translation memory systems. Localization Reader*, 133–140.

Bowker, L. (2016). The Need for Speed! Experimenting with "Speed Training" in the Scientific/Technical Translation Classroom. *Meta: Journal des traducteurs/Meta: Translators' Journal*, *61*, 22–36.

Bundgaard, K., & Christensen, T. P. (2019). Is the concordance feature the new black? A workplace study of translators' interaction with translation resources while post-editing TM and MT matches. *The Journal of Specialised Translation*, *31*, 14–37.

Burchardt, A., Harris, K., Melby, A., Uszkoreit, H., Görög, A., Gladkoff, S., & Glazychev, L. (2014). Multidimensional Quality Metrics (MQM) Definition.

Buysschaert, J., Fernández-Parra, M., Kerremans, K., Koponen, M., & van Egdom, G. (2018). Embracing digital disruption in translator training:: technology immersion in simulated translation bureaus. *Revista Tradumàtica. Tecnologies de la Traducció*, *16*, 125–133.

Cadwell, P., O'Brien, S., & Teixeira, C. S. (2018). Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, *26*(3), 301–321.

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*. Sage.

Callison-Burch, C., Koehn, P., Monz, C., Peterson, K., Przybocki, M., & Zaidan, O. F. (2010). Findings of the 2010 joint workshop on statistical machine translation and metrics for machine translation. *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, 17–53.

Calvo, E. (2009). *Análisis curricular de los estudios de Traducción e Interpretación en España. Perspectiva del estudiantado* (Doctoral dissertation). Universidad de Granada.

Calvo, E. (2011). Translation and/or translator skills as organising principles for curriculum development practice. *The Journal of Specialised Translation*, *16*, 5–25.

Canfora, C. (2016). Assessing learning in heterogeneous learning groups in translator training: A role for portfolios. *Towards authentic experiential learning in translator education* (pp. 161–176).

Carl, M., Dragsted, B., Elming, J., Hardt, D., & Jakobsen, A. L. (2011). The process of post-editing: a pilot study. *Copenhagen Studies in Language*, *41*, 131–142.

Carl, M., & Planas, E. (2020). Advances in interactive translation technology. *The Bloomsbury companion to language industry studies*, 361–386.

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

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 109–120.

Chesterman, A. (1997). *Memes of translation* (Vol. 10). Amsterdam: John Benjamins.

Chesterman, A. (2000). Teaching strategies for emancipatory translation. *Benjamins Translation Library*, *38*, 77–90.

Chien, K., Palau, V., Sun, M., Lindhardsen, V., & Seong, Y. (2014). Reading and Writing Achievement Test for Advanced 1 Students in the Community English Program.

Christensen, T. P., & Schjoldager, A. (2016). Computer-aided translation tools–the uptake and use by Danish translation service providers. *The Journal of Specialised Translation*, *25*, 89–105.

Cid-Leal, P., Espín-García, M.-C., & Presas, M. (2019). Traducción automática y posedición: perfiles y competencias en los programas de formación de traductores [http://dx.doi.org/10.6035/MonTI.2019.11.7]. *MonTI*.

Ciobanu, D., Ragni, V., & Secară, A. (2019). Speech Synthesis in the Translation Revision Process: Evidence from Error Analysis, Questionnaire, and Eye-Tracking. *Informatics*, *6*(4), 51.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37–46.

Couto Vale, D. (2017). What does a translator do when not writing? *Empirical modelling of translation and interpreting*, *7*, 209.

Crazover, A. (2010). Un entrepreneur en Chine. https://alexandrecrazover.wordpress.com/2010/01/16/alexandre-crazover-un-entrepreneur-en-chine/

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Daems, J., De Clercq, O., & Macken, L. (2017). Translationese and Post-editese: How comparable is comparable quality? *Linguistica Antverpiensia New Series-Themes in Translation Studies*, *16*, 89–103.

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

Daems, J., & Macken, L. (2020). Post-editing Human Translations and Revising Machine Translations. *Translation Revision and Post-editing: Industry Practices and Cognitive Processes*, 50.

Dalla-Zuanna, J.-M. (2020). Quality evaluation according to the expanded SAE standard J2450 as applied by Volkswagen AG's Language Services. In J.-M. Dalla-Zuanna & C. Kurz (Eds.), *Translation Quality in the Age of Digital Transformation*. BDÜ, Weiterbildungs-und Fachverlagsgesellschaft mbH.

Dalla-Zuanna, J.-M., & Kurz, C. (Eds.). (2020). *Translation Quality in the Age of Digital Transformation*. BDÜ, Weiterbildungs-und Fachverlagsgesellschaft mbH.

De Almeida, G. (2013). *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience*

across two Romance languages (Doctoral dissertation). Dublin City University.

De Almeida, G., & O'Brien, S. (2010). Analysing post-editing performance: correlations with years of translation experience. *Proceedings of the 14th annual conference of the European association for machine translation*, 26–28.

De Cespedes, B. R. (2018). Mind the gap: Language Service Providers' views on the technological training of professional translators. *Optimización de las Competencias del Traductor e Intérprete: Nuevas Tecnologías-Procesos Cognitivos-Estrategias* (pp. 143–162). Frank; Timme.

Dede, V. (2019). Does a formal post-editing training affect the performance of novice post-editors? An experimental study.

de Gibert Bonet, O. (2018). To post-edit or to translate... That is the question: A case study of a recommender system for Quality Estimation of Machine Translation based on linguistic features.

Denkowski, M., & Lavie, A. (2012). TransCenter: Web-based translation research suite. *AMTA 2012 Workshop on Post-Editing Technology and Practice Demo Session*, 2012.

Denkowski, M., Lavie, A., Lacruz, I., & Dyer, C. (2014). Real time adaptive machine translation for post-editing with cdec and TransCenter. *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*, 72–77.

Densmer, L. (2014). Light and Full MT Post-Editing Explained [https://www.rws.com/insights/rws-moravia-blog/bid-353532-Light-and-Full-MT-Post-Editing-Explained/].

do Carmo, F. (2017). *Post-editing: a theoretical and practical challenge for translation studies and machine learning* (Doctoral dissertation). Universidade do Porto.

do Carmo, F. (2020a). 'Time is money'and the value of translation. *Translation Spaces*, *9*(1), 35–57.

do Carmo, F. (2020b). Comparing Post-editing based on Four Editing Actions against Translating with an Auto-Complete Feature. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 421–430.

Dodd, S. C. (1952). Model english for mechanical translation: an example of a national language regularized for electronic translators. *7pp. Repr. as 'Model English'in Locke and Booth (1955)*, 165–173.

Doherty, S. (2016). The Impact of Translation Technologies on the Process and Product of Translation. *International Journal of Communication*, *10*, 23.

Doherty, S., & Gaspari, F. (2013). Effective Post-Editing in Human & Machine Translation Workflows: Critical Knowledge & Techniques. *Centre for Next Generation Localisation*.

Doherty, S., Kenny, D., & Way, A. (2012). Taking statistical machine translation to the student translator.

Doherty, S., & Moorkens, J. (2013). Investigating the experience of translation technology labs: pedagogical implications. *Journal of Specialised Translation*, *19*, 122–136.

Doherty, S., Moorkens, J., Gaspari, F., & Castilho, S. (2018). On Education and Training in Translation Quality Assessment. *Translation Quality Assessment* (pp. 95–106). Springer.

Doherty, S., O'Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. *Machine translation*, *24*(1), 1–13.

Dreyfus, H., Dreyfus, S. E., & Athanasiou, T. (2000). *Mind over machine*. Simon; Schuster.

Dugast, L., Senellart, J., & Koehn, P. (2007). Statistical Post-Editing on SYSTRAN's Rule-Based Translation System. *Proceedings of the Second Workshop on Statistical Machine Translation*, 220–223.

EMT Expert Group and others. (2009). Competences for professional translators, experts in multilingual and multimedia communication. *European Master's in Translation (EMT)*.

EMT Network. (2017). European Master's in Translation Competence Framework. https : / / ec . europa . eu / info / sites / info / files / emt _ competence_fwk_2017_en_web.pdf

EMT Network. (2019). Annual report 2018. https://ec.europa.eu/info/sites/info/files/emt_annual_report_2018_en.pdf

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational research review*, *5*(1), 1–24.

Escartín, C. P., & Arcedillo, M. (2015). A fuzzier approach to machine translation evaluation: A pilot study on post-editing productivity and automated metrics in commercial settings. *Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra)*, 40–45.

Esfandiari, M. R., Shokrpour, N., & Rahimi, F. (2019). An evaluation of the EMT: Compatibility with the professional translator's needs. *Cogent Arts & Humanities*, *6*(1), 1601055.

Eyckmans, J., & Anckaert, P. (2017). Item-based assessment of translation competence: Chimera of objectivity versus prospect of reliable measurement. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16*.

Federico, M., Cattelan, A., & Trombetti, M. (2012). Measuring user productivity in machine translation enhanced computer assisted trans-

lation. *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA)*, 44–56.

Federmann, C. (2012). Appraise: an open-source toolkit for manual evaluation of mt output. *The Prague Bulletin of Mathematical Linguistics*, *98*(2012), 25–35.

Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective? *The journal of Specialised translation*, *11*(11), 52–74.

Fields, P., Hague, D. R., Koby, G. S., Lommel, A., & Melby, A. (2014). What is quality? A management discipline and the translation industry get acquainted. *Tradumàtica*, *12*, 0404–412.

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

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation spaces*, *6*(2), 291–309.

Forcada, M. L., Sánchez-Martínez, F., Esplà-Gomis, M., & Specia, L. (2017). Towards Optimizing MT for Post-Editing Effort: Can BLEU Still Be Useful? *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 183–195.

Fraser, J. (2000). The broader view: How freelance translators define translation competence. *Benjamins Translation Library*, *38*, 51–62.

Fullen, J. E. (2019). *Retaining Millennials in the Workforce* (Doctoral dissertation). Northcentral University.

Gambier, Y. (2006). Pour une socio-traduction. *BENJAMINS TRANSLATION LIBRARY*, *68*, 29.

Gambier, Y. (2010). Translation strategies and tactics. *Handbook of translation studies*, *1*, 412–418.

Gambier, Y., & Van Doorslaer, L. (2009). *The metalanguage of translation* (Vol. 20). John Benjamins Publishing.

Garcia, I. (2011). Translating by post-editing: is it the way forward? *Machine Translation*, *25*(3), 217.

Garcia, I. (2015). Computer-aided translation: systems. *Routledge Encyclopedia of Translation Technology*, 68–87.

Garvin, D. A. (1984). What Does Product Quality Really Mean? *Sloan Management Review*, *26*(1), 25–43.

Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, *23*(3), 333–358.

Gene, V. (2019). Machine Translation: From Translation to Post-editing. *GALA Webinar*. https://www.gala-global.org/events/events-calendar/machine-translation-translation-post-editing

Gene, V. (2020). The Management and Training Challenges of Post-Editing - Parts 1 and 2 [http://xl8.link/Gene2020Part2 and http://xl8.link/Gene2020Part1].

Ginovart-Cid, C. (2018). Statistical & neural MT systems in the motorcycling domain for less frequent language pairs–how do professional post-editors perform? *Translating and the Computer 40: proceedings; 2018 Nov 15-16; London, United Kingdom. Geneva: Editions Tradulex; 2018. p. 66-78.*

Ginovart-Cid, C. (2020a). Post-Editing Practice Application (PEPA). https://classroom.google.com/c/NDM3MzMzMDA4MjFa?cjc=v3xxpuk

Ginovart-Cid, C. (2020b). The Professional Profile of a Post-editor according to LSCs and Linguists: a Survey-Based Research. *Hermes*. https://repositori.upf.edu/handle/10230/46133

Ginovart-Cid, C. (2020c). Report about a survey-based research on machine translation post-editing: common ground and gaps between LSCs, linguists and trainers. http://hdl.handle.net/10230/46218

Ginovart-Cid, C. (2021). Replication Data for: The need for practice in the acquisition of the post-editing skill-set. https://doi.org/10.34810/data36

Ginovart-Cid, C., & Colominas, C. (2020). The MT post-editing skill set: course descriptions and educators' thoughts. Taylor & Francis.

Ginovart-Cid, C., Colominas, C., & Oliver, A. (2020). Language industry views on the profile of the post-editor. *Translation Spaces*.

Ginovart-Cid, C., & Oliver, A. (2020). The post-editor's skill set according to industry, trainers, and linguists. *Maschinelle Übersetzung für Übersetzungsprofis*. BDÜ, Weiterbildungs-und Fachverlagsgesellschaft mbH.

Góis, A., Cho, K., & Martins, A. (2020). Learning Non-Monotonic Automatic Post-Editing of Translations from Human Orderings. *arXiv preprint arXiv:2004.14120*.

Góis, A., & Martins, A. F. (2019). Translator2Vec: Understanding and Representing Human Post-Editors. *arXiv preprint arXiv:1907.10362*.

Göpferich, S. (2009). Towards a model of translation competence and its acquisition: the longitudinal study TransComp. *Behind the mind: Methods, models and results in translation process research*, *4*(4), 11–37.

Göpferich, S. (2013). Translation competence: Explaining development and stagnation from a dynamic systems perspective. *Target. International Journal of Translation Studies*, *25*(1), 61–76.

Görög, A. (2014). Quality evaluation today: the dynamic quality framework. *Proceedings of the 36th Translating and the Computer Conference*.

Görög, A. [Attila]. (2014). Quantifying and benchmarking quality: the TAUS Dynamic Quality Framework. *Tradumàtica*, (12), 0443–454.

Görög, A. [Attila]. (2015a). The TAUS Efficiency Score. Introducing a new score for measuring productivity. https : / / www . slideshare . net / TAUS/taus-efficiency-score-attila-gorog-taus

Görög, A. [Attila]. (2015b). Translation Productivity Revisited. https://blog.taus.net/translation-productivity-revisited

Green, R. (1982). The MT errors which cause most trouble to posteditors. *Lawson (1986)*, 101–104.

Guerberof Arenas, A. (2008). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus The International Journal of Localisation*, *7*(1), 11–21.

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

Guerberof Arenas, A. (2010). Project management and machine translation. *Multilingual computing & technology*, *21*(3), 34.

Guerberof Arenas, A. (2014). Correlations between productivity and quality when post-editing in a professional context. *Machine Translation*, *28*(3-4), 165–186.

Guerberof Arenas, A. (2019). Pre-editing and post-editing. *The Bloomsbury Companion to Language Industry Studies* (p. 333). Bloomsbury Publishing.

Guerberof Arenas, A., Depraetere, H., & O'Brien, S. (2012). What we know and what we would like to know about post-editing. *Revista Tradumàtica*, (10), 211–218.

Guerberof Arenas, A., & Moorkens, J. (2019). Machine translation and post-editing training as part of a master's programme. *Jostrans: The Journal of Specialised Translation*, (31), 217–238.

Guerrero, L. (2017). LSP Perspective: Applying the Human Touch to MT, Qualitative Feedback in MT evaluation (Part 1).

Guerrero, L., & Soloviev, K. (2019). MT Evaluation in Translation Quotations at CPSL with ContentQuo.

Guha, J., & Heger, C. (2014). Machine translation for global e-commerce on eBay. *Proceedings of the AMTA*, *2*, 31–37.

Hair, J. F. (2009). Multivariate data analysis.

Harding, S.-A., & Cortés, O. C. (2018). *The Routledge handbook of translation and culture*. Routledge.

Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M. et al. (2018).

Achieving human parity on automatic chinese to english news translation. *arXiv preprint arXiv:1803.05567*.

Hazbavi, A. A. (2015). Perception and Implementation of Machine Translation Applications by the Iranian English Translators. *International Journal of Cognitive and Language Sciences*, *9*(4), 1115–1120.

Herbig, N., Pal, S., van Genabith, J., & Krüger, A. (2019). Multi-modal approaches for post-editing machine translation. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11.

Hokamp, C., & Liu, Q. (2015). Handycat-an open-source platform for CAT tool research. *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*.

Holmes, J. S. (1988). The name and nature of translation studies. *Translated! Papers on literary translation and translation studies*, *2*, 67–80.

Hönig, H. G. (1991). Holmes'"mapping theory" and the landscape of mental translation processes. *Translation Studies: The state of the art*, 77–89.

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

Huertas Barros, E., & Vine, J. (2018). Current trends on MA translation courses in the UK: changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, *12*(1), 5–24.

Hurtado, A. (1999). Enseñar a traducir. *Metodología en la formación de traductores e intérpretes. Edelsa: Madrid*.

Hurtado Albir, A. (1996). La cuestión del método traductor. Método, estrategia y técnica de traducción. *Sendebar*, *7*(39-57).

Hurtado Albir, A. (2007). Competence-based curriculum design for training translators. *The interpreter and translator trainer*, *1*(2), 163–195.

Hurtado Albir, A. (2008). Compétence en traduction et formation par compétences. *TTR: traduction, terminologie, rédaction*, *21*(1), 17–64.

Hurtado Albir, A. (2015). The acquisition of translation competence. Competences, tasks, and assessment in translator training. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, *60*(2), 256–280.

Hutchins, J. (2003). ALPAC: the infamous report. *Readings in machine translation*, *14*, 131–135.

Hutchins, J. (2007). Machine translation: A concise history. *Computer aided translation: Theory and practice*, *13*(29-70), 11.

Hutchins, J., Hartmann, W., & Ito, E. (2009). Compendium of Translation Software: directory of commercial machine translation systems and computer-aided translation support tools. *Genève: European Association for Machine Translation*.

Hutchins, W. J. (1986). *Machine translation: past, present, future*. Ellis Horwood Chichester.

Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation* (Vol. 362). Academic Press London.

Iizuka, I. (2018). Isn't It Time to Embrace Machine Translation Post-Editing? The Localization Use Case for MT. http : / / xl8 . link / Iizuka2018

ISO Central Secretary. (2008). *Quality Management Systems – Requirements* (Standard ISO 9001:2008). International Organization for Standardization. Geneva, CH.

ISO Central Secretary. (2012). *21500:2012 Guidance on project management* (Standard). International Organization for Standardization.

ISO Central Secretary. (2015). *Translation Services. Requirements for Translation Services* (Standard ISO 17100:2015) [https://doi.org/10.3403/30246895U]. International Organization for Standardization. Geneva, CH.

ISO Central Secretary. (2017). *Translation services. Post-editing of machine translation output. Requirements* (Standard ISO 18587:2017) [https://doi.org/10.3403/30279909U]. International Organization for Standardization. Geneva, CH.

Ive, J. (2017). *Towards a Better Human-Machine Collaboration in Statistical Translation: Example of Systematic Medical Reviews* (Doctoral dissertation). Paris Saclay.

Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, *31*, 61–86.

Johnson, R. L., & Whitelock, P. (1987). Machine Translation as an Expert Task. Cambridge University press.

Joscelyne, A. (2006). Best practices in post-editing. *TAUS*.

Joscelyne, A., & Brace, C. (2010). Postediting in Practice. A TAUS Report.

Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Fikri Aji, A., Bogoychev, N., Martins, A. F. T., & Birch, A. (2018). Marian: Fast Neural Machine Translation in C++. *Proceedings of ACL 2018, System Demonstrations*, 116–121. http://www.aclweb.org/anthology/P18-4020

Jurs, S. G., & Wiersma, W. (1990). Educational measurement and testing. *United States: A Division of Simon & Schuster*.

Kearns, J. (2006). *Curriculum Renewal in Translator Training: Vocational challenges in academic environments with reference to needs and situation analysis and skills transferability from the contemporary*

experience of Polish translator training culture (Doctoral dissertation). Dublin City University.

Kenny, D. (2020). Technology in Translator Training. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 498–515). Routledge.

Killman, J. (2018). Translating the same text twice: An English-Spanish comparative product study of post-edited translations vs. human translations. *The Journal of Internationalization and Localization*, *5*(2), 114–141.

Kiraly, D. (2000). A Social constructivist approach to translator education.

Kiraly, D. (2012a). Growing a project-based translation pedagogy: A fractal perspective. *Meta: journal des traducteurs/Meta: Translators' Journal*, *57*(1), 82–95.

Kiraly, D. (2012b). Skopos theory goes to Paris: Purposeful translation and emergent translation projects. https://goo.gl/6djMMb

Kiraly, D. (2014). *A social constructivist approach to translator education: Empowerment from theory to practice*. Routledge.

Kiraly, D. (2015). Occasioning translator competence: Moving beyond social constructivism toward a postmodern alternative to instructionism. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *10*(1), 8–32.

Kiraly, D. (2016). Beyond the static competence impasse in translator education. *Translation and Meaning [New Series]*, *1*, 129–42.

Kiraly, D., & Massey, G. (2016). *Towards authentic experiential learning in translator education*. V & R Unipress.

Kiraly, D., & Piotrowska, M. (2014). Towards an emergent curriculum development model for the European Graduate Placement Scheme. *Conference proceedings. The future of education*, 368.

Kiraly, D., & Sascha, H. (2019). Towards an emergentist curriculum development model for translator education. *Towards Authentic Experiential Learning in Translator Education*, 59.

Klein, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A. M. (2017). Opennmt: Open-source toolkit for neural machine translation. *arXiv preprint arXiv:1701.02810*.

Klimkowski, K. et al. (2015). Entrepreneurial training in translator and interpreter education. *Roczniki humanistyczne*, *63*(11), 67–83.

Koby, G. S. [Geoffrey S]. (2001). Editor's introduction: Post-editing of machine translation output: Who, what, why, and how (much). *Repairing texts: Empirical investigations of machine translation post-editing processes*, 1–24.

Koby, G. S. [Geoffrey S.], Fields, P., Hague, D., Lommel, A., & Melby, A. (2014). Defining Translation Quality. *Tradumàtica*, *12*, 0413–420.

Koby, G. S. [Geoffrey S], & Melby, A. K. (2013). Certification and job task analysis (JTA): Establishing validity of translator certification examinations. *Translation & Interpreting*, *5*(1), 174–210.

Kockaert, H., & Segers, W. (2014). Evaluation de la traduction: la méthode PIE (Preselected Items Evaluation). *Turjuman*, *23*(2), 232–250.

Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *Journal of Specialised Translation*, (27), 148–163.

Koehn, P., Federico, M., Shen, W., Bertoldi, N., Bojar, O., Callison-Burch, C., Cowan, B., Dyer, C., Hoang, H., Zens, R. et al. (2007). Open source toolkit for statistical machine translation: Factored translation models and confusion network decoding. *Final Report of the Johns Hopkins 2006 Summer Workshop*.

Koehn, P., & Haddow, B. (2009). Interactive assistance to human translators using statistical machine translation methods. *MT Summit XII*.

Koponen, M. (2012). Comparing human perceptions of post-editing effort with post-editing operations. *Proceedings of the seventh workshop on statistical machine translation*, 181–190.

Koponen, M. (2015). How to teach machine translation post-editing? Experiences from a post-editing course. *4th Workshop on Post-Editing Technology and Practice (WPTP4)*, 2–15.

Koponen, M. (2016a). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, *25*, 131–148.

Koponen, M. (2016b). *Machine translation post-editing and effort: Empirical studies on the post-editing process* (Doctoral dissertation). University of Helsinki.

Koponen, M., & Salmi, L. (2015). On the correctness of machine translation: A machine translation post-editing task. *The Journal of Specialised Translation*, *23*(23), 118–136.

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

Koponen, M., Salmi, L., & Nikulin, M. (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation*, *33*(1-2), 61–90.

Kornacki, M. (2018). *Computer-Assisted Translation (CAT) Tools in the Translator Training Process*. Peter Lang. https://www.peterlang.com/view/title/68120

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

Krollmann, F. et al. (1974). Data processing at the translator's service. *Babel*, *20*(3), 121–129.

Lacruz, I., Denkowski, M., & Lavie, A. (2014). Cognitive demand and cognitive effort in post-editing.

Lacruz, I., & Shreve, G. M. (2014). Pauses and Cognitive Effort in Post-editing. *Post-editing of Machine Translation: Processes and Applications* (pp. 246–272). Cambridge Scholars Publishing.

Language Industry Survey. (2019). https://euatc.org/wp-content/uploads/2019/11/2019-Language-Industry-Survey-Report.pdf

Lara, C. P. (2019). SWOT Analysis of the Inclusion of Machine Translation and PostEditing in the Master's Degrees Offered in the EMT Network. *JOURNAL OF SPECIALISED TRANSLATION*, (31), 260–280.

Läubli, S., Fishel, M., Massey, G., Ehrensberger-Dow, M., Volk, M., O'Brien, S., Simard, M., & Specia, L. (2013). Assessing post-editing efficiency in a realistic translation environment.

LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation & Interpreting*, *5*(2), 1–13.

Lee, J., & Liao, P. (2011). A Comparative Study of Human Translation and Machine Translation with Post-editing. *Compilation & Translation Review*, *4*(2).

Lehmann, W. P., Bennett, W. S., Slocum, J., Smith, H., & Pfluger, S. (1981). *The METAL System. Volume I and Volume II. Appendices* (tech. rep.). AUSTIN LINGUISTICS RESEARCH CENTER.

Lei, P.-W., & Wu, Q. (2007). CTTITEM: SAS macro and SPSS syntax for classical item analysis. *Behavior research methods*, *39*(3), 527–530.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, *10*(8), 707–710.

L'Homme, M.-C. (2008). *Initiation à la traductique*. Brossard, Quebec: Linguatech.

Llitjós, A. F., Carbonell, J. G., & Lavie, A. (2005). A framework for interactive and automatic refinement of transfer-based machine translation.

Loffler-Laurian, A.-M. (1984). Machine translation: What type of post-editing on what type of documents for what type of users. *Proceedings of the 10th international conference on Computational Linguistics*, 236–238.

Loffler-Laurian, A.-M. (1986). Post-édition rapide et post-édition conventionelle: l Deux modalités d'une activité spécifique. *Multilingua-Journal of Cross-Cultural and Interlanguage Communication*, *5*(2), 81–88.

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

Lommel, A., & Burchardt, A. (2015). Multidimensional quality metrics (MQM) definition. *URL http://www.qt21.eu/mqm-definition/definition-2015-12-30.html*.

Lommel, A., & Melby, A. K. (2018). Tutorial: MQM-DQF: A Good Marriage (Translation Quality for the 21st Century). *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Papers)*.

Lommel, A., Popovic, M., & Burchardt, A. (2014). Assessing inter-annotator agreement for translation error annotation. *MTE: Workshop on Automatic and Manual Metrics for Operational Translation Evaluation*.

Lörscher, W. (1991). *Translation performance, translation process, and translation strategies: A psycholinguistic investigation* (Vol. 4). G. Narr.

Luong, M.-T., Cho, K., & Manning, C. D. (2016). Neural machine translation. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics: ACL 2016 Tutorial*.

Mandrekar, J. N. (2011). Measures of interrater agreement. *Journal of Thoracic Oncology*, *6*(1), 6–7.

Marg, L. (2018). The Trials and Tribulations of Predicting Machine Translation Post-Editing Productivity. *Language Resources Evaluation Conference(LREC). http://www. lrecconf. org/proceedings/lrec2016/pdf/810_Paper. pdf Accessed*, *26*.

Marheinecke, K. (2016a). Can Quality Metrics Become the Drivers of Machine Translation Uptake? An Industry Perspective. *Translation Evaluation: From Fragmented Tools and Data Sets to an Integrated Ecosystem*, 71.

Marheinecke, K. (2016b). Post-editing Machine Translation. What does it take?

Mariana, V. R., Cox, T., & Melby, A. (2015). The Multidimensional Quality Metric (MQM) framework: A new framework for translation quality assessment.

Marín Juarros, V. (2017). La formación de poseditores mediante el uso de las TIC: estado de la cuestión y una propuesta de estrategia de formación.

Martín-Mor, A., Sánchez-Gijón, P., & i Huerta, R. P. (2016). *Tradumàtica: Tecnologies de la traducció*. Eumo Editioral.

Marzouk, S., & Hansen-Schirra, S. (2019). Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation*, *33*(1-2), 179–203.

Mason, J. B., & Rinsche, A. (1995). *Ovum Evaluates Translation Technology Products*. Ovum.

Massardo, I., Van Der Meer, J., O'brien, S., Hollowood, F., Aranberri, N., & Drescher, K. (2016). TAUS Post-Editing Guidelines [https://www.taus.net/think-tank/reports/postedit-reports/taus-post-editing-guidelines].

Massey, G. (2018). New roles and tasks for the translation profession: educating translators for the digital present and future. *DG TRAD Conference, Luxembourg, 16-17 October 2017*, 36–51.

Massey, G., Kiraly, D., & Ehrensberger-Dow, M. (2019). Training the translator trainers: an introduction.

Matlock-Hetzel, S. (1997). Basic Concepts in Item and Test Analysis.

Maučec, M. S., & Donaj, G. (2019). Machine Translation and the Evaluation of Its Quality. *Natural Language Processing-New Approaches and Recent Applications*. IntechOpen.

Mazur, I. (2009). The metalanguage of localization. Theory and practice. *The Metalanguage of Translation. Amsterdam and Philadelphia: John Benjamins Publishing Company*, 145–165.

Mellinger, C. D. (2017). Translators and machine translation: knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer*, *11*(4), 280–293.

Mesa-Lao, B. (2013). Introduction to post-editing–The CasMaCat GUI. *1*(2015), 305–314.

Mesa-Lao, B. (2014). Gaze behaviour on source texts: An exploratory study comparing translation and post-editing. *Post-editing of Machine Translation: Processes and Applications* (pp. 219–245). Cambridge Scholars Publishing.

Mihalache, I. (2010). Social and economic actors in the evaluation of translation technologies. Creating meaning and value when designing, developing and using translation technologies. *Evaluation of Translation Technology*, *8*, 159.

Moorkens, J. (2017). Under pressure: translation in times of austerity. *Perspectives*, *25*(3), 464–477.

Moorkens, J. (2018). What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, *12*(4), 375–387.

Moorkens, J., Castilho, S., Gaspari, F., & Doherty, S. (2018). Translation Quality Assessment: From Principles to Practice. *Machine Translation: Technologies and Applications*.

Moorkens, J., & O'Brien, S. (2017). Assessing user interface needs of post-editors of machine translation. *Human issues in translation technology* (pp. 127–148). Routledge.

Moorkens, J., O'brien, S., Da Silva, I. A., de Lima Fonseca, N. B., & Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, *29*(3-4), 267–284.

Mossop, B. (2006a). Has computerization changed translation? *Meta: journal des traducteurs/Meta: Translators' Journal*, *51*(4), 787–805.

Mossop, B. (2006b). *Revising and editing for translators*. St Jerome Publishing.

Mossop, B. (2019). *Revising and Editing for Translators*. Routledge.

Muzii, L. (2016). *Post-editing of Machine Translation for Project Managers*. Lulu.com.

Neubauer, P. B. (2015). Unforeseen Consequences: Big Data and the Language Industry. *Translation and Openness* (p. 19). Innsbruck University Press.

Newmark, P. (1988). *A textbook of translation* (Vol. 66). Prentice hall New York.

Nimdzi. (2019). Rise of the machines – the state of machine translation (Report preview for Project Opus). https://www.nimdzi.com/mt-report-preview-opus/

Nimdzi. (2020). Language Technology Landscape v2020.08. https://www.nimdzi.com/nimdzi-language-technology-atlas-2020/

Nitzke, J., Hansen-Schirra, S., & Canfora, C. (2019). Risk management and post-editing competence. *Journal of specialised translation*, *31*, 239–259.

Nitzke, J., Tardel, A., & Hansen-Schirra, S. (2019). Training the modern translator–the acquisition of digital competencies through blended learning. *The Interpreter and Translator Trainer*, *13*(3), 292–306.

Nunziatini, M., & Marg, L. (2020). Machine Translation Post-Editing Levels: Breaking Away from the Tradition and Delivering a Tailored Service. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 309–318.

O'Brien, S. (2002). Teaching post-editing: a proposal for course content. *6th EAMT Workshop Teaching Machine Translation*, 99–106.

O'Brien, S. (2004). Machine translatability and post-editing effort: How do they relate. *Translating and the Computer*, *26*.

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

O'Brien, S. (2006a). *Machine-translatability and post-editing effort: An empirical study using Translog and Choice Network Analysis* (Doctoral dissertation). Dublin City University.

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

O'Brien, S. (2007). An empirical investigation of temporal and technical post-editing effort. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *2*(1), 83–136.

O'Brien, S. (2010). Introduction to post-editing: Who, what, how and where to next [http://www.mtarchive.info/AMTA-2010-OBrien.pdf]. *The Ninth Conference of the Association for Machine Translation in the Americas, Denver, Colorado.*

O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine translation*, *25*(3), 197.

O'Brien, S., & Rossetti, A. (2021). Neural machine translation and the evolution of the localisation sector, Implications for training. *Journal of Internationalization and Localization*, *7*(1-2), 95–121.

O'Brien, S., & Simard, M. (2014). Introduction to special issue on post-editing. *Machine Translation*, *28*(3-4), 159–164.

O'Curran, E. (2014). Translation quality in post-edited versus human-translated segments: A case study. *Third Workshop on Post-Editing Technology and Practice*, 113.

Offersgaard, L., Povlsen, C., Almsten, L., & Maegaard, B. (2008). Domain specific MT in use. *Proceedings of the 12th EAMT conference*, 22–23.

Oliver, A. (2006). La traducció automàtica a internet. *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, (4).

Oliver, A. (2016). *Herramientas tecnológicas para traductores*. Editorial UOC.

Oliver, A., Alvarez, S., & Badia, T. (2020). PosEdiOn: Post-Editing Assessment in PythOn. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 403–410. https://www.aclweb.org/anthology/2020.eamt-1.43

Orozco, M., & Hurtado Albir, A. (2002). Measuring translation competence acquisition. *Meta: journal des traducteurs/Meta: Translators' Journal*, *47*(3), 375–402.

Ortiz Boix, C., & Matamala, A. (2015). Quality assessment of post-edited versus translated wildlife documentary films: a three-level approach. *4th Workshop on Post-editing Technology and Practice (WPTP4)*, (4).

Ovchinnikova, I., & Morozova, D. (2019). Comparative Analysis of Errors in MT Output and Computer-assisted Translation: Effect of the Human Factor. *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, 88–94.

PACTE Group. (2005). Investigating translation competence: Conceptual and methodological issues. *Meta: journal des traducteurs*, *50*(2), 0609–619.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting on association for computational linguistics*, 311–318.

Paradowska, U. (2015). Expert web searching skills for translators â "a multiple-case study. *Constructing translation competence. Frankfurt am Main: Peter Lang*, 227–244.

Pérez Macías, L. (2020). What Do Translators Think About Post-Editing?: a Mixed-Methods Study of Translators' Fears, Worries and Preferences on Machine Translation Post-Editing. *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació*, (18), 11–32.

Pietrzak, P. (2018). The effects of students' self-regulation on translation quality. *Babel*, *64*(5-6), 819–839.

Piotrowska, M., & Tyupa, S. (2014). Translation pedagogy: a new subdiscipline of translation studies. *inTRAlinea*.

Plaza Lara, C. (2014). La enseñanza de la traducción asistida por ordenador: análisis didáctico-curricular y propuesta de integración de la subcompetencia instrumental-profesional en el aula.

Plitt, M., & Masselot, F. (2010). A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague bulletin of mathematical linguistics*, *93*, 7–16.

PMI. (2017). A Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Sixth Edition.

Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, *7*(3), 135–153.

Popović, M. (2018). Error classification and analysis for machine translation quality assessment. *Translation Quality Assessment* (pp. 129–158). Springer.

Popović, M., Avramidis, E., Burchardt, A., Hunsicker, S., Schmeier, S., Tscherwinka, C., Vilar, D., & Uszkoreit, H. (2014). Involving language professionals in the evaluation of machine translation. *Language resources and evaluation*, *48*(4), 541–559.

Popović, M., Lommel, A., Burchardt, A., Avramidis, E., & Uszkoreit, H. (2014). Relations between different types of post-editing operations, cognitive effort and temporal effort. *Proceedings of the 17th annual conference of the european association for machine translation*, 191–198.

Pouliquen, B., & Junczys-Dowmunt, M. (2016). Lecture 7: Neural Machine Translation [https://github.com/mlearningbruno/giancourse/raw/master/lectures/Lecture7.0-NeuralMachineTranslation.pdf].

Pym, A. (2003). Redefining translation competence in an electronic age. In defence of a minimalist approach. *Meta: journal des traducteurs/Meta: Translators' Journal*, *48*(4), 481–497.

Pym, A. (2013). Translation Skill-Sets in a Machine Translation Age. *Meta: Journal des traducteurs/Meta: Translators' Journal*, *58*(3), 487–503.

Pym, A. (2019). How automation through neural machine translation might change the skill sets of translators. *Unpublished manuscript, last modified August*, *29*.

Pym, A., Grin, F., Sfreddo, C., & Chan, A. L. (2013). *The status of the translation profession in the European Union*. Anthem Press.

Quah, C. K. (2006). *Translation and technology*. Springer.

Reifler, E. (1950). Studies in Mechanical Translation, n 1, MT. *Mimeographed*, *10*.

Reifler, E. (1952). Mechanical translation with a pre-editor and writing for mechanical translation. *Conference on Mechanical Translation*, 17–20.

Rico, C. [Celia]. (2017). The ePortfolio: constructing learning in translation technology. *The Interpreter and Translator Trainer*, *11*(1), 79–95.

Rico, C. [Celia], & Díez, P. L. (2013). EDI-TA: Post-editing methodology for Machine Translation. *Multilingualweb-LT Deliverable 4.1. 4. Annex I*.

Rico, C. [Celia], Oliver, A., & Colominas, C. (2020). Jornada sobre l'impacte de la Traducció Automàtica en la docència/aprenentatge de la traducció i segones llengües. https://www.youtube.com/watch?v=atXOM06CWX8&feature=youtu.be

Rico, C. [Celia], & Orzas, P. L. D. (2013). EDI-TA: Training methodology for Machine Translation Post-editing. *Multilingualweb-LT Deliverable 4.1. 4. Annex II*.

Rico Pérez, C., & Torrejón, E. [Enrique]. (2012). Skills and Profile of the New Role of the Translator as MT Post-editor. *Tradumàtica*, *10*, 0166–178.

Risku, H., Rogl, R., & Milošević, J. (2019). Researching workplaces. *The Bloomsbury Companion to Language Industry Studies*, 37.

Robert, A.-M. (2013). Vous avez dit post-éditrice? Quelques éléments d'un parcours personnel. *The Journal of Specialised Translation*, *19*, 29–40.

Robert, I. S., Remael, A., & Ureel, J. J. (2017). Towards a model of translation revision competence. *The Interpreter and Translator Trainer*, *11*(1), 1–19.

Rogers, E. M. (1962). *Diffusion of innovations*. New York, Free Press of Glencoe.

Ruopp, A. (2015). Industry Shared Metrics with the TAUS Dynamic Quality Dashboard and API. http://www.mt-archive.info/15/MTS-2015-Ruopp.pdf

Şahin, M. (2011). Using MT post-editing for translator training.

Saint-André, L. (2011). La formation actuelle en traduction et en postédition : une comparaison des contenus.

Saint-André, L. (2015). *Quelle formation donner aux traducteurs-postéditeurs de demain?* (Doctoral dissertation). Université d'Ottawa/University of Ottawa.

Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. Routledge.

Salkind, N. J. (2017). *Statistics for People Who (Think They) Hate Statistics*. 6th Edition. Sage Publications, Inc., Thousand Oaks, CA.

Samuels, P. (2015). Statistical Methods - Scale reliability analysis with small samples. Birmingham City University, Centre for Academic Success.

Sánchez-Gijón, P. (2014). Developing technical competence for PE. Post-editing modules in translator training programs: Why, What, When, Who, and How.

Sánchez-Gijón, P. (2016). La posedición: hacia una definición competencial del perfil y una descripción multidimensional del fenómeno. *Sendebar*, *27*, 151–162.

Sánchez-Gijón, P., & Torres Hostench, O. (2014). MT post-editing into the mother tongue or into a foreign language? Spanish-to-English MT translation output post-edited by translation trainees. *Workshop on Post-Editing Technology and Practice-AMTA*.

Sanchez-Torron, M., & Koehn, P. (2016). Machine translation quality and post-editor productivity. *Proceedings of AMTA*, *2016*, 16.

Sanz Villa, S. (2015). El traductor ante la traducción asistida por ordenador y la traducción automática con posedición: estudio comparado de actitudes hacia su uso, necesidad y evolución.

Scarpellini, G. (2020). Cohen's kappa free calculator. https://idostatistics.com/cohen-kappa-free-calculator/

Schäffner, C. (2019). Translators' roles and responsibilities. *The Bloomsbury Companion to Language Industry Studies*, 63.

Schumacher, P. (2020). Post-édition et traduction humaine en contexte académique: une étude empirique. *Transletters. International Journal of Translation and Interpreting*, (3), 239–274.

Secretariat, T. M. U. J. (2018). Conclusions and recommendations of JIAMCATT 2018. https://www.iamladp.org/sites/www.iamladp.org/files/jiamcatt-2018-recommendations.pdf

Segers, W., & Kockaert, H. (2016). Can subjectivity be avoided in translation evaluation? *International Maastricht-Lodz Duo Colloquium, Date: 2015/05/21-2015/05/22, Location: Maastricht, The Netherlands*.

Singer, N. (2016). A proposal for language teaching in translator training programmes using data-driven learning in a task-based approach. *International Journal of English Language & Translation Studies*, *4*(2), 155–167.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. *Proceedings of association for machine translation in the Americas*.

Somers, H. (2003). *Computers and translation: a translator's guide* (Vol. 35). John Benjamins Publishing.

Specia, L. (2017, October 11). Translation Quality Webinar: Analyzing Translation Quality. https://www.taus.net/events/webinars/35-translation-quality-webinar-analyzing-translation-quality

Specia, L., Cancedda, N., & Dymetman, M. (2010). A Dataset for Assessing Machine Translation Evaluation Metrics. *LREC*.

Specia, L., & Farzindar, A. (2010). Estimating machine translation post-editing effort with HTER. *Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC 10)*, 33–41.

Stasimioti, M., & Sosoni, V. (2019). Undergraduate Translation Students' Performance and Attitude vis-à-vis Machine Translation and Post-editing: Does Training Play a Role? *Proceedings of Translating and the Computer*, *41*.

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

Stevens, A., & Fuentes Corradi, R. (2016). Best Practices for Enterprise Scale Post-Editing [https://www.sdl.com/download/best-practices-for-enterprise-scale-post-editing/132199/].

Tamchyna, A. (2020). Selection of MT Systems in Translation Workflows. *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 2: User Papers)*, 270–291.

Tatsumi, M. (2010). *Post-editing machine translated text in a commercial setting: Observation and statistical analysis* (Doctoral dissertation). Dublin City University.

TAUS. (2013). TAUS Translation Technology Landscape report. From Luxury to Utility. https://www.taus.net/think-tank/reports/translate-reports/taus-translation-technology-landscape-report

Teixeira, C. (2014). Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. *Third Workshop on Post-Editing Technology and Practice*, 45–59.

Teixeira, C. (2019). Revising computer-mediated translations. *Revising and editing for translators* (pp. 207–224). Routledge.

Teixeira, C., & Moorkens, J. (2017). Creating a tool for multimodal translation and post-editing on touch-screen devices. *Translating and the Computer (TC39) Conference, London, UK*.

Temizöz, Ö. (2013). *Postediting machine translation output and its revision: Subject-Matter Experts versus Professional Translators* (Doctoral dissertation). Universitat Rovira i Virgili.

Temnikova, I. P. (2010). Cognitive Evaluation Approach for a Controlled Language Post–Editing Experiment. *LREC*.

Thelen, M. (2019). Quality and Quality Assessment in Translation: Paradigms in Perspective. *Quality Assurance and Assessment Practices in Translation and Interpreting* (pp. 1–25). IGI Global.

Thicke, L. (2013). Moses in the Mix: A Technology Agnostic Approach to a Winning MT Strategy [2013-06-12].

Tinsley, J. (2014). Beyond Data: Delivering Machine Translation with Subject Matter Expertise. *TAUS MT Showcase*.

Toral, A. (2019). Post-editese: an exacerbated translationese. *arXiv preprint arXiv:1907.00900*.

Toral, A., Castilho, S., Hu, K., & Way, A. (2018). Attaining the unattainable? Reassessing claims of human parity in neural machine translation. *arXiv preprint arXiv:1808.10432*.

Torrejón, E., & Rico, C. (2002). Controlled Translation: A New Teaching Scenario Tailor-Made for the Translation Industry.[online] Available at:< http://www. mt-archive. info. *EAMT-2002-Torrejon. pdf*>.

Toury, G. (2012). *Descriptive translation studies and beyond: Revised edition* (Vol. 100). John Benjamins Publishing.

Tradumàtica Research Group. (n.d.). POST-IT: Equipo de investigación [http://xl8.link/post-it].

Trujillo, A. (1999). *Translation engines: techniques for machine translation*. Springer.

Underwood, N., & Jongejan, B. (2001). Translatability checker: A tool to help decide whether to use MT. *Proceedings of MT Summit VIII: Machine Translation in the Information Age*, 363–368.

Valli, P. (2015). The TAUS Quality Dashboard.

Van den Besselaar, P., Heimeriks, G. et al. (2001). Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. *ISSI*, 705–716.

Van der Meer, J. (2013). Choose your own translation future.

Van Egdom, G.-W., Verplaetse, H., Schrijver, I., Kockaert, H. J., Segers, W., Pauwels, J., Wylin, B., & Bloemen, H. (2019). How to Put the Translation Test to the Test?: On Preselected Items Evaluation and Perturbation. *Quality Assurance and Assessment Practices in Translation and Interpreting* (pp. 26–56). IGI Global.

Van Egdom, G.-W., Vieira, L. N., & Absolon, J. (2018). Towards testing post-editing performance: a futureproof diagnostic. *Tradumàtica*, (16), 0114–124.

Van Ess-Dykema, C. (2011). An Effective Model for Insertion of Translation Technologies into US Government Translation Environments. *Aslib*.

Van Ess-Dykema, C., Phillips, J., Reeder, F., & Gerber, L. (2010). Paralinguist assessment decision factors for machine translation output: a case study. *AMTA*.

Vandepitte, S. (2008). Remapping translation studies: Towards a translation studies ontology. *Meta: journal des traducteurs/Meta: Translators' Journal*, *53*(3), 569–588.

Vasconcellos, M. (1986). Post-editing on-screen: machine translation from Spanish into English. *Proceedings of Translating and the Computer*, *8*, 133–146.

Vasconcellos, M. (1987). A comparison of MT post-editing and traditional revision. *Proceedings of the 28th annual conference of the American Translators Association*, 409–415.

Vasconcellos, M. (1988). *Technology as translation strategy*. John Benjamins Publishing.

Vasconcellos, M., & Bostad, D. A. (1992). Machine translation in a high-volume translation environment. *Computers in translation. A practical appraisal.* (pp. 58–77). Routledge.

Vieira, L. N. (2013). An Evaluation of Tools for Post-Editing Research: The Current Picture and Further Needs. *Proceedings of MT Summit XIV*

Workshop on Post-Editing Technology and Practice, 2 September 2013, Nice, France, 93–101.

Vieira, L. N. (2014). Indices of cognitive effort in machine translation post-editing. *Machine translation*, *28*(3-4), 187–216.

Vieira, L. N., & Alonso, E. (2020). Translating perceptions and managing expectations: an analysis of management and production perspectives on machine translation. *Perspectives*, *28*(2), 163–184.

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

Vinay, J.-P., & Darbelnet, J. (1958). A methodology for translation. *The*.

Vintar, Š., Bago, P., Mikelić Preradović, N., Ciobanu, D., Secara, A., Reiss, C., Žabokrtsky, Z., Nitzke, J., Tardel, A., Vondricka, P. et al. (2019). DigiLing e-Learning Hub: e-Courses for Digital Linguistics.

Wagner, E. [Elizabeth]. (1985). Rapid post-editing of Systran. *Tools for the trade: Translating and the Computer*, *5*, 199–213.

Wagner, E. [Emma]. (1987). Post-editing–Practical Considerations. *ITI Conference I: The Business of Translating and Interpreting*, 71–78.

Wagner, E. [Emma], Bech, S., & Martínez, J. M. (2014). *Translating for the European Union institutions*. Routledge.

Way, A. (2013). Traditional and emerging use-cases for machine translation. *Proceedings of Translating and the Computer*, *35*, 12.

Way, C. (2000). Structuring Specialised Translation Courses: A Hit and Miss Affair? *Developing translation competence*, *38*, 131.

Wendt, C. (2019). The Quantum Leap: Human Parity. https://www.taus.net/events/conferences/71-taus-global-content-summit-amsterdam

Wesleyan University. (2020). Data Analysis Tools. https://www.coursera.org/learn/data-analysis-tools

Wheeler, P. (1995). Offering large-system machine translation as a bureau service. *Machine Translation Archive*.

Wilss, W. (1982). *The science of translation: problems and methods* (Vol. 180). John Benjamins Pub Co.

Winkler, K., Kuhn, T., & Volk, M. (2014). Evaluating the fully automatic multi-language translation of the Swiss avalanche bulletin. *International Workshop on Controlled Natural Language*, 44–54.

Yamagata. (2013). Honda warranty claim project.

Yang, Y., & Wang, X. (2020). Predicting student translators' performance in machine translation post-editing: interplay of self-regulation, critical thinking, and motivation. *Interactive Learning Environments*, 1–15.

Zetzsche, J. (2020). 319th Tool Box Journal. https://internationalwriters.com/toolkit/

Zhang, H., & Torres-Hostench, O. (2019). Cómo enseñar posedición de traducción automática a una segunda lengua: una propuesta didáctica para el aprendizaje de lenguas. *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació*, (17), 153–161.

Zhechev, V. (2014). Analysing the post-editing of machine translation at Autodesk. *Post-Editing of Machine Translation: Processes and Applications* (pp. 2–23). Cambridge Scholars Publishing.

Appendices

Consent form to trainers

Information Notes

1. Industrial PhD: "Lost in Machine Translation, Found in Post-Editing"
2. Supervisor from Universitat Pompeu Fabra: Carme Colominas

Supervisor from Universitat Oberta de Catalunya: Antoni Oliver Supervisor from industry (Datawords): Marina Frattino

1. Doctorats Industrials is the granting body
2. My PhD research, to last between 3 and 4 years and being currently at the 2^{nd} year, is to deepen the knowledge of the post-editor competency model to improve, in particular, FR-ES post-editing (PE) courses, either in the academic or industrial scope.
3. Now, the existing PE courses or syllabus including mention to PE at the postgraduate level are analyzed and compared to understand the currently used contents and methodologies, which will later be contrasted to the results of two exploratory surveys (one to professionals -142 responses- and one to firms -66 responses-). For this, I ask for the contribution of the PE instructors of a list of pre-selected programmes. Such contribution, ideally, is formed by: an online survey (around 10 minutes) & an interview via an online calling software (around 15-20 minutes).
4. The survey does not collect personal data, and its contents are stored by Jotform, the online platform used. All the surveys will be deleted at the end of the PhD research period, which should be in 2021. The interviews shall be recorded. The recordings will be stored in Google Drive, either a non-shared folder or a folder shared with the two academic supervisors (Carme Colominas and Antoni Oliver).

The publications resulting from this research will not mention the name of the University or institution, the name of the programme (Master, etc.) or the name of the syllabus. The findings resulting from this research may be used as an inspiration to (help) design other PE courses, either in the academic or professional sphere. All the data collected via survey or interview is subject to the confidentiality clause in the convention signed between Universitat Pompeu Fabra, Datawords, and Clara Ginovart in the framework of

Doctorats Industrials.

1. Your participation to the survey, to the interview, or to both is voluntary.
2. The contact person is the PhD student: Clara Ginovart Cid. E-mail. clara.ginovart@upf.edu

Express Consent

For the research presented in 'Information Notes', in the framework of Clara Ginovart's PhD, the participant, [___], voluntarily agrees to take part in this research study. I confirm that:

1. I read, understood, and agreed with the "Information Notes"
2. I have had the opportunity to discuss the subject matter or ask questions
3. I have enough information about the project
4. I agree to my interview being audio-recorded
5. I understand that even if I agree to participate now, I can withdraw at any time or refuse to answer any question without any consequences of any kind
6. I understand that under freedom of information legalisation I am entitled to access the information I have provided at any time while it is in storage as specified above
7. I understand that I am free to contact any of the people involved in the research to seek further clarification and information.

Name:
Signature:
Date: